III JORNADAS INTERNACIONALES DE ESTADÍSTICA APLICADA

Casos de jugadas maestras de profesionales de la industria e instituciones y soluciones en contexto de pandemia COVID - 19

Universidad Nacional de Salta
FACULTAD DE INGENIERIA

FACULTAD DE INGENIERIA
UNIVERSIDAD NACIONAL DE JUJUY

INGENIERÍA
UNIVERSIDAD CATÓLICA DE SALTA

Facultad de Ciencias Exactas
UNIVERSIDAD NACIONAL DE SALTA

Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero / UNSE

ESCUELA DE NEGOCIOS
UNIVERSIDAD CATÓLICA DE SALTA

Universidad Nacional de Salta
III Jornadas Internacionales de Estadística Aplicada: Actas de trabajos de investigación de JIEA 2020: Facultad de Ingeniería: Universidad Nacional de Salta /; editado por Angélica Noemí Arenas; Gisella Carla Mautino; compilación de: Gisella Carla Mautino; Angélica Noemí Arenas; Héctor Rubén Tarcaya; ilustrado por María Josefina Méndez; Enrique Ariel Morales del Valle; prefacio de Héctor Iván Rodríguez; prólogo de Francisco Aparicio Izquierdo. - 1a ed. - Salta: Universidad Nacional de Salta, 2021.

Libro digital, DOCX

Archivo Digital: online
ISBN 978-987-633-573-7

Título III Jornadas Internacionales de Estadística Aplicada: Actas de trabajos de investigación JIEA 2020. Facultad de Ingeniería. Universidad Nacional de Salta
ISBN 978-987-633-573-7
1ra. Edición
EUNSa Editorial Universidad Nacional de Salta
Avda. Bolivia 5150 – Salta Capital – CP 4400 – Argentina.
E-mail: seu@unsa.edu.ar
Web: www.unsa.edu.ar
Tel.: +54 387-4258707 – Fax: +54 387-4325745

PRESENTACIÓN DE LAS JORNADAS

Las Jornadas nacieron en el año 2018 en el seno de la Facultad de Ingeniería de la UNSa como respuesta a la necesidad de capacitación del estudiante de ingeniería y carreras afines en la formación por competencias. Parte de ello, es incentivar el uso de la estadística como un instrumento para resolver problemas y también en la toma de decisiones, ambas actividades inherentes a la ingeniería.

Las Jornadas juegan un importante rol en la formación y estímulo, porque el estudiante, no solamente toma vista de las aplicaciones reales de la estadística, a través de los trabajos presentados por los profesionales de distintas empresas y las investigaciones y casos mostrados por docentes e investigadores en una amplia gama de disciplinas, sino que también participan a la par, presentando trabajos realizados en grupos o individualmente como resultado de su aprendizaje durante el cursado de las asignaturas específicas.

> *Las Jornadas componen un ambiente de participación activa de estudiantes, docentes, investigadores de universidad e instituciones, actores del gobierno y profesionales de empresas nacionales e internacionales, con la presentación de trabajos realizados en la práctica y así, dar a conocer la importancia de la estadística como parte de instrumentos en la resolución de problemas y toma de decisiones.*

Estas jornadas crecieron desde el 2018, cuando las cátedras de Probabilidades y Estadística y Diseño Experimental de la Facultad de Ingeniería de la UNSa junto con las Facultades de Ingeniería de la UNJU y UCASAL se propusieron crear las Jornadas con tal objetivo. En el 2019 se realizaron con carácter internacional, donde se sumaron las Facultades de Ciencias Químicas de la Universidad Nacional de Asunción del Paraguay, las Facultades homónimas integrantes del CODINOA, y académicos de las Universidades de Santiago de Chile, Bolivia, Colombia, España y profesionales de México, entre otros.

En esta oportunidad hemos realizado las terceras Jornadas, también de carácter internacional, con la modalidad virtual por la Pandemia de COVID-19. La Conferencia Magistral fue impartida por el Dr. Ing. Francisco Aparicio Izquierdo, prestigioso Profesor de la Universidad Politécnica de Madrid (UPM) y director del Instituto creado por él y que lleva actualmente su nombre: el Instituto Universitario de Investigación del Automóvil (INSIA) de la UPM. Su conferencia versará sobre el valioso aporte de la estadística aplicada, para el equipo de investigadores y expertos coordinada por él, en la investigación científica de accidentes de tránsito, en colaboración con la Dirección General de Tráfico del Gobierno de España.

Las III Jornadas en esta ocasión ha integrado formalmente como organizadores a las Facultades de Ingeniería de la Universidad Nacional de Salta (UNSa), Universidad Católica de Salta (UCASAL), Universidad Nacional de Jujuy (UNJu), Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA), Facultad de Ciencias Exactas y Tecnologías de la Universidad Nacional de Santiago del Estero (UNSE), Facultad de Ciencias Exactas de la UNSa, Facultad de la Escuela de Negocios de la Universidad Católica de Salta (UCASAL) y Facultad de Ciencias Químicas de la Universidad Nacional de Asunción del Paraguay (UNA).

Es *Objetivo* que el año entrante las Jornadas crezcan aún mucho más.

Mg. Ing. Héctor Iván Rodríguez
Profesor Adjunto de las Asignaturas
Probabilidad y Estadística y Estadística Experimental

Secretario de Vinculación y Transferencia
Facultad de Ingeniería UNSA

PRÓLOGO

Debo de iniciar estas líneas, que me han sido pedidas, a modo de prólogo de la publicación que recoge las ponencias presentadas en las III JORNADAS INTERNACIONALES DE ESTADÍSTICA APLICADA*, organizadas por la Facultad de Ingeniería de la Universidad Nacional de Salta (UNSa), la Facultad de Ingeniería de la Universidad Nacional de Jujuy, la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, la Facultad de Ciencias Exactas y Tecnologías de la Universidad Nacional de Santiago del Estero, la Facultad de Ciencias Exactas de la UNSa, la Facultad de Ingeniería y la Escuela de Negocios de la Universidad Católica de Salta y la Facultad de Ciencias Químicas de la Universidad Nacional de Asunción del Paraguay,* y con agradecimiento acompañado de sorpresa porque he tenido valiosas oportunidades de interacción con varias de estas universidades argentinas.

Agradecimiento, por dos motivos fundamentales: el primero, porque es un honor para mí que los organizadores de las Jornadas hayan valorado tan positivamente mi humilde contribución al éxito de las mismas, como para cursarme esta nueva invitación; el segundo, porque desde el año 1991 en que visité por primera vez la UNSa, como conferenciante invitado en otras jornadas destinadas a temas de calidad de las enseñanzas universitarias, y conocí la bella ciudad que le da su nombre y a un buen número de salteños que me brindaron su amistad y su afecto, quedé en deuda perenne con la institución, con sus académicos y con los salteños que tan recordados momentos me brindaron a lo largo de estos treinta años que separan aquel acontecimiento de éste.

Sorpresa, porque yo no soy académico o especialista en área alguna de la Estadística; pienso que la justificación, de esta invitación, debo encontrarla en el hecho de que mi experiencia académica e investigadora me relaciona con la importante temática que es la Estadística Aplicada, como usuario y como coordinador de un cualificado equipo de profesores y expertos en estadística con los que hemos abordado importantes trabajos de investigación científica de accidentes de tráfico, colaborando con la Dirección General de Tráfico del Gobierno de España, con resultados que nos hacen pensar que produjeron aportaciones que ayudaron al “milagro” español en materia de seguridad vial, reduciéndose el número de víctimas mortales, en las carreras y ciudades españolas, de aproximadamente 6000, a menos de 2000, en solo unos cinco años.

Realizadas estas consideraciones previas, que he considerado necesarias para situar mejor a los lectores de estas líneas, acotando lo que pueden esperar de lo que continua, debo expresar mi posicionamiento ante el tema que nos ocupa. Como es bien sabido, los modelos estadísticos tienen como principal finalidad el estudio de fenómenos aleatorios de diferente complejidad y adquieren una mayor importancia a medida de que dicha complejidad se incrementa y las variables de influencia sobre los mismos se hacen más abundantes y aumenta, así mismo, la complejidad de sus relaciones con las variables dependientes del fenómeno, entre las variables de influencia, su variación con el tiempo y otros factores de incidencia directa o indirecta.

Siendo esta la finalidad de la Estadística, especialmente en su orientación aplicada, la aplicación de sus metodologías requiere, como condición imprescindible la existencia de datos, de informaciones con calidad y número suficientes, para asegurar que los resultados de los análisis estadísticos son significativos y también de calidad; por último, una condición complementaria, pero de gran importancia, es contar con herramientas de cálculo apropiadas para implementar algoritmos complejos a un gran número de datos.

Estas tres condiciones fundamentales se dan y se incrementan cada día con el desarrollo de numerosas actividades humanas, científicas, sociales, económicas y otras: la inmensa mayoría de los fenómenos en los diversos campos, son de carácter aleatorio; los cada vez más completos y sofisticados métodos de captación, análisis y clasificación de los valores de las variables asociadas a dichos fenómenos proporcionan gran número de datos con suficiente calidad, el término Big Data refleja cada vez más esta realidad.

Por último, las capacidades de cálculo que ofrecen los ordenadores, desde las grandes máquinas, hasta sus versiones más asequibles a investigadores individuales, completan el círculo virtuoso para facilitar que esta importante rama de la ciencia manifieste plenamente su capacidad para ayudar al progreso de la humanidad en la resolución de los importantes problemas que ofrece el panorama actual ante las grandes necesidades en materias tan vitales como la alimentación de una superpoblación mundial, la salud, la seguridad, la preservación del medio ambiente y otros; sin olvidar ámbitos más concretos como puede ser el incremento de la calidad de los sistemas productivos para incrementar la competitividad de los pueblos; la reducción de los accidentes y víctimas del tráfico rodado y tantos otros.

Es, por tanto, la ciencia Estadística y sus numerosas aplicaciones, un ámbito científico de una enorme importancia actual, que con toda seguridad se incrementará en el futuro, permítanme, antes de terminar este texto, una reflexión final como persona que se beneficia de esta ciencia, desde fuera de ella. Los especialistas en estadística no deben de perder de vista que ésta es una herramienta, poderosa, pero solo una herramienta al servicio de estudios de fenómenos que deben ser bien conocidos y estudiados para valorar los métodos más útiles en cada aplicación, los datos que deben ser utilizados y los niveles de calidad exigibles y para interpretar adecuadamente los resultados obtenidos de la aplicación de los modelos a los datos.

Esta reflexión nos lleva a poner en valor la importancia de trabajar en equipos pluridisciplinares que integren en trabajo “codo con codo” a especialistas conocedores en profundidad del fenómeno a estudiar y expertos en estadística, con la máxima valoración reciproca de la aportación de unos y otros; nuestras experiencias en los estudios de accidentología y seguridad vial confirma esta condición como clave indiscutible del éxito.
Como punto final permítanme una licencia personal, este año mi nieto David ha iniciado los estudios de la carrera de Estadística en la correspondiente Facultad de la Universidad Complutense de Madrid y estoy feliz por ello.

Francisco Aparicio Izquierdo
Dr. Ingeniero Industrial
Catedrático Emérito de la Universidad Politécnica de Madrid (UPM)
Presidente del INSIA (Instituto de Investigación del Automóvil Francisco Aparicio Izquierdo) de la UPM
Presidente de FEIBIM (Federación Iberoamericana de Ingeniería Mecánica)

AGRADECIMIENTOS

Hacemos llegar nuestros agradecimientos a las autoridades de la Universidad, a las autoridades de las Facultades organizadoras, a los Comités de Revisión, Comité Honorario, Comité Ejecutivo y Científico, a los Coordinadores Generales y Colaboradores Estudiantes por las valiosas contribuciones para la realización de las III Jornadas Internacionales de Estadística Aplicada (IIIJIEA).

Se reconoce especialmente al Dr. Ing. Francisco Aparicio Izquierdo por haber impartido generosamente la *Conferencia Magistral de Apertura* de las III Jornadas.

Se agradece la actuación del comité de editores y de compiladores, por plasmar en el Libro de las IIIJIEA las contribuciones recibidas por los autores y la unión de las energías de las distintas disciplinas para hacer de la estadística aplicada un aporte para las ingenierías y carreras afines.

También, se quiere retribuir las actividades de las personas encargadas del diseño y digitalización así como las tareas de soporte en los procesos de edición.

.

Finalmente, los agradecimientos conclusivos son para los protagonistas de las III Jornadas quienes son los autores y revisores, a los primeros por enviar sus publicaciones y apoyar a las (IIIJIEA), y a los segundos por su labor para alcanzar calidad en las publicaciones y ayudar a los autores a mejorar sus trabajos.

COMITÉ DE REVISIÓN

ÁREAS TEMÁTICAS

Six Sigma
Simulación de Procesos
Ingeniería de Confiabilidad
Control Estadístico de Procesos
Muestreo Estadístico
Análisis Multivariado
Estadística Bayesiana
Encuestas
Optimización de Procesos DOE
Data Mining
Pronósticos

COMITÉS ACTUANTES EN LAS III JORNADAS INTERNACIONALES DE ESTADÍSTICA APLICADA

AUTORIDADES - COMITÉ HONORARIO

INSTITUCIONALES

CPN Víctor Hugo Claros - Rector UNSa
Dra. Ing. Graciela Morales - ViceRectora UNSa
Ing. Rodolfo Gallo Cornejo - Rector UCASAL
Ing. Flavio Sergio Fama - Rector UNCA
Dr. Celso Mora - Dir. Ac. UNA
Ing. Héctor Rubén Paz - Rector UNSE

FACULTADES

Ing. Héctor Raúl Casado - Decano FI - UNSa
Ing. Daniel Hoyos - Decano Fac.Cs.Ex. - UNSa
Mg. Ing. Néstor Eugenio Lesser - Decano FI - UCASAL
Ing. Gustavo Alberto Lores - Decano FI - UNJu
Ing. Carlos Humberto Savio - Decano FTyCA - UNCA
Lic. Cynthia Saucedo de Schupmann-Decana- FCQ-UNA
Ing. Pedro Juvenal Basualdo - Decano FCEyT - UNSE
Mg. Lic. Marco Antonio Limarino Cazón - Decano FENEG UCASAL

COORDINACIÓN GENERAL

SECRETARÍAS VINCULACIÓN & COOPERACIÓN

Dra. Delicia Ester Acosta (FI - UNSa)
Mg. Ing. Héctor Iván Rodríguez (FI - UNSa)
Ing. Walter Garzón (Fac.Cs.Ex. - UNSa)
Ing. Juan Francisco Linares (FI - UCASAL)
Ing. Octavio Daniel Coro (FI - UNJu)
Lic. René Alejandro Ramos (FENG -UCASAL)

COORDINADORES OPERATIVOS

Dra. Martha Susana Cañas (FTyCA - UNCA)
Ing. Gisella Mautino (FI - UNSa / UNJu)
Ing. Angélica Arenas (FI - UNSa)
Ing. Manuel Zambrano (FI-UCASAL)
Dr. Daniel Villa (FI-UCASAL)

COMITÉ EJECUTIVO

COORDINACIÓN

Dra. Ivanna Maricruz Lazarte (FTyCA - UNCA)
Lic. José Ignacio Tuero (FI-Fac.Cs.Ex.UNSa)
Ing. Eliana Rizo (FI-UCASAL)
Ing. Elizabeth Elizeche (FCQ-UNA)
Ing. Mariela Ester Rodríguez (Ob.Seg. Min.Gob.Jujuy)

ACTIVIDADES ESPECÍFICAS

Dr. Emilio Almazán (FI-UNSa)
Mg. Ing. Rubén Tarcaya (FI-UNSa)
Ing. Daniel Acuña (FI-UCASAL)
Ing. Amapola Cabrera (FCQ-UNA)
Lic. Nazarena Analía Laureano (Ob.Seg. Min.Gob.Jujuy)

COMITE CIENTIFICO

Dra. Ing. María Soledad Vicente (FI-UNSa)
Dra. Martha Susana Cañas (FTyCA - UNCA)
Dra. Ivanna Maricruz Lazarte (FTyCA - UNCA)
Dra. Lía Orozco (FI-UCASAL/UNSa)
Dra. Ing. Eleonora Erdmann (FI-UNSa)
Dra. Ing. Mercedes Méndez (FI-UNSa)
Dra. Ing. Verónica Beatriz Rajal (FI-UNSa)
Dra. Dolores Gutiérrez Cacciabué (FI-UNSa)
Dr. Ing. Carlos Albarracín (FI-UNSa)
Dr. Ing. Antonio Arcienaga Morales (FI-UNSa/UCASAL)
Dr. Ing. Emilio Almazán (FI-UNSa)
Dr. Ing. Juan Carlos Michalus (FI-UNaM)
Dr. Ing. Raymundo Forradellas (FI-UNCu)
Dr. Ing. Ricardo Palma (FI-UNCu)
Dr. Jorge Viel (FI-UNLaR)
Dr. Ing. Orlando Domínguez (FI-UNSa)
Ing. Jorge Félix Almazán (FI-UNSa)
Mg. Ing. Héctor Iván Rodríguez (FI-UNSa)
Mg. Ing. Pablo Argenti Salguero (FI-UNSa)
Mg. Ing. Roberto Medina (FI-UNSa/UCASAL)
Mg. Héctor Funes (Fac.Cs.Ex.UNSa/UCASAL)
Mg. Ing. Rubén Tarcaya (FI-UNSa)
Mg. Ing. Angélica Arenas (FI-UNSa)
Mg. Ing. Bárbara Villanueva (FI-UNSa)
Esp. Ing. Ricardo Jakúlica (FI-UNSa)
Mg. Ing. Carlos Gabriel R. Herrera (FTyCA - UNCA)
Ing. Gisella Mautino (FI-UNSa)
Ing. Octavio Daniel Coro (FI-UNJu)
Ing. Federico Fabián Quispe (FI-UNSa)
Ing. Juan Francisco Linares (FI-UCASAL)
Lic. María Cristina Ahumada (Cs.Ex.UNSa)
Ing. Noelia Centurión (FCQ-UNA)
Ing. Amapola Cabrera (FCQ-UNA)
Lic. Amanda Marlene Duré (FCQ-UNA)
Lic. Nori Esther Cheein de Auat (FCEyT-UNSE)
Ing. Ricardo Cordero (FCEyT-UNSE)

COLABORADORES ALUMNOS

Sr. Gonzalo Amaya (FI - UNSa)
Srta. Micaela Vargas (FI - UNSa)

DESCRIPCIÓN DE LAS INSTITUCIONES

UNIVERSIDAD NACIONAL DE SALTA

Luego de décadas de postergaciones, en el año 1972 -mediante decreto del 11 de mayo-, cristaliza la creación de la Universidad Nacional de Salta, en un contexto de creación de otras Universidades Nacionales en nuestro país. Puede decirse, que "la UNSa" es una hija neta de la década de los '70 en la Argentina y de los procesos previos, tanto a nivel nacional como en Latinoamérica, por hechos históricos que se comparten. Fue dotada de un claro espíritu participativo, que trascendiera un mero encuadramiento localista y provinciano, para anclar en una realidad que comprendiera la región del noroeste, con toda su tradición y complejidad subregional, étnica y social, e integrada a una Argentina plenamente latinoamericana, sin descuidar nutrirse u coparticipar de los avances científicos a nivel planetario. Su visión estaba influida por los encendidos y riquísimos debates sobre el desarrollo y el subdesarrollo, que hicieran derramar ríos de tinta a notables pensadores del Nuevo y del Viejo Mundo, de modo que apuntara a actualizar críticamente el desarrollo económico y social en su medio, pero irradiándolo a los confines de la Patria y a países hermanos del continente.

Puede decirse sin temor a errar, que en las palabras y en el diseño de su escudo de creación y actual se reflejan los ideales que definirían su vocación. Las palabras de su escudo, atribuidas al poeta salteño Leopoldo J. Castilla son toda una filosofía: "Mi sabiduría viene de esta tierra". Como descripción de este escudo, es oportuno transcribir un párrafo publicado por la Facultad de Ciencias Económicas de la UNSa: "En el dibujo se puede apreciar el sol radiante, identificado con el Inti, que alguna vez pobló estas tierras. Se aprecia una cadena de cerros que es la que colinda hacia el oeste con la Ciudad de Salta. Luego podemos observar una serie de edificios en los cuales se distingue, el cabildo salteño en el centro, la iglesia del Convento de la Orden de Nuestra Señora del Monte Carmelo, popularmente conocida como capilla del Convento San Bernardo, entre otras edificaciones. Se puede apreciar nítidamente la presencia de palmeras, sin lugar a duda una clara referencia a la plaza principal de la
Ciudad de Salta, Plaza 9 de Julio. Luego sigue una guarda con motivos autóctonos... "

Hoy, la tarea completa, de los universitarios y la sociedad en su conjunto, es un esfuerzo importante que debe tener en cuenta fuertes alteraciones generadas a nivel mundial y nacional, una acción permanente y sin tregua para aportar soluciones a la sociedad e identificar las dificultades para ser enfrentadas y superadas...

Para mayor información visite http://www.unsa.edu.ar/web/index.php
Sobre las III Jornadas http://3jiea.ing.unsa.edu.ar/

UNIVERSIDAD NACIONAL DE JUJUY

Surge como tal a comienzos de la década de 1970, aunque sus orígenes se encuentran cuarenta años antes con la proyección hacia Jujuy de otras instituciones de Educación Superior.

En 1930 se instaló la Misión de Estudios de Patología Regional Argentina (MEPRA) dependiente de la Universidad de Buenos Aires y dirigida en sus inicios por el Doctor Salvador Mazza. A mediados de la década de 1940 la Universidad Nacional de Tucumán creó en Jujuy el Instituto de Geología y Minería, el Instituto de Biología de la Altura y la Escuela de Minas “Dr. Horacio Carrillo” un establecimiento de Nivel Técnico. El aporte local fue la habilitación del Instituto Superior de Ciencias Económicas en 1959 que otorgó títulos de nivel superior de Contador Público y Perito Partidor.
En 1973 comienza a funcionar la Universidad Nacional de Jujuy en una provincia con 302.436 habitantes de los cuales 50,6% eran menores de 20 años. Nace con las Facultades de Ciencias Agrarias y de Ingeniería, con las carreras Ingeniería Agronómica, Ingeniería Química, Ingeniería en Minas e Ingeniería Metalúrgica, mostrando el perfil asociado con el desarrollo agroindustrial y minero de sus fundadores.

Actualmente, cerca de cumplir sus primeros 50 años de historia, la Universidad Nacional de Jujuy cuenta con cuatro Facultades, varios institutos de investigación de los cuales algunos dependen también del Consejo Nacional de Investigaciones Técnicas y Científicas de la Nación y otros son cogobernados junto al Estado provincial. Más de 20000 alumnos, casi 1500 docentes, un millar de empleados de apoyo a la gestión y alrededor de 4000 egresados forman parte de la concreción de un sueño que comenzó a cristalizarse cuando en 1972 el Gobernador de Jujuy, Ingeniero Manuel Pérez, tomó la decisión de crear la Universidad en esta provincia.

Para mayor información visite https://www.unju.edu.ar/

UNIVERSIDAD CATOLICA DE SALTA

Próxima a cumplir sus primeros 60 años de vida, la Universidad Católica de Salta es una institución privada, confesional católica, que goza del mayor grado de autonomía. Sus creadores fueron el primer Arzobispo de Salta Monseñor Roberto José Tavella, y el Dr. Robustiano Patrón Costas, fundador del Ingenio y Refinería San Martín del Tabacal.

La Compañía de Jesús tomó en sus manos la organización y dirección académica de la nueva Universidad, estableciendo como condición para su puesta en marcha la existencia de una biblioteca y un campus de cincuenta hectáreas.

Mediante el Decreto Arzobispal de fecha 19 de marzo de 1963, fue creada bajo el lema “Nihil Intentatum”: Nada sin intentar. La Patrona de la misma es Santa Teresa de Jesús, y en el día de su Santoral, se colocó la piedra fundamental en Campo Castañares, de la Ciudad de Salta, el 15 de Octubre de 1966.

Las unidades académicas iniciales fueron la Facultad de Artes y Ciencias, Facultad de Ingeniería, Facultad de Economía y Administración y la Escuela de Servicio Social. Con el pasar de los años se sumaron la Facultad de Arquitectura, la de Ciencias Veterinarias, Ciencias Jurídicas y la Facultad Escuela de Negocios.

En la actualidad, se desempeña como Gran Canciller de la Universidad Católica de Salta el Sr. Arzobispo de Salta Monseñor Mario Antonio Cargnello, siendo su Rector el Ing. Rodolfo Gallo Cornejo.

¡Bienvenidos a los 60 años, Nada sin Intentar!

Para mayor información visite https://www.ucasal.edu.ar/

UNIVERSIDAD NACIONAL DE CATAMARCA

La Universidad Nacional de Catamarca (UNCA) es una institución que conjuga tradición y modernidad, comprometida con el futuro de Catamarca y la Nación, consustanciada con la defensa del patrimonio cultural, histórico y natural desde sus funciones sustantivas de educación, investigación y extensión. La Universidad Nacional de Catamarca (UNCA), de acuerdo a lo que establece su Plan Estratégico, es una universidad pública enraizada en su sociedad, comprometida con su desarrollo, que se erige en creadora de cultura y potenciadora del pensamiento.

Es una entidad comprometida desde su creación con la cultura y la identidad de Catamarca, y con la promoción del desarrollo provincial, por lo que revaloriza el patrimonio socio cultural tangible e intangible y propicia el fortalecimiento de sus vínculos con el sector socio productivo público y privado.

Es un espacio de transformación permanente en el que se impulsa la ciencia y la tecnología, promoviendo un modelo de sociedad basado en el conocimiento, asumiendo de esta forma un rol director y participativo.

La UNCA es una institución que conjuga tradición y modernidad, comprometida con el futuro de Catamarca y la Nación, consustanciada con la defensa del patrimonio cultural, histórico y natural desde sus funciones sustantivas de educación, investigación y extensión.

Para mayor información visite https://www.unca.edu.ar/index.html

UNIVERSIDAD NACIONAL DE ASUNCIÓN

La Universidad Nacional de Asunción (UNA), fundada el 24 de septiembre de 1889, es la primera institución de Educación Superior, la más antigua y con mayor tradición del Paraguay, comprometida con la sociedad desde sus inicios, para impulsar la transparencia, investigación, extensión e innovación.

Cuenta con 14 Facultades, institutos, centros tecnológicos y de investigación que brindan facilidades a la comunidad académica, para la realización de trabajos científicos y el desarrollo de estudios de postgrado, que se traduce en aportes a la sociedad.

La comunidad académica está conformada por más de 55.597 estudiantesy 9.417 docentes, está presente en 22 ciudades y 12 departamentos, con filiales que brindan oportunidad educativa a miles de jóvenes del interior.

La internacionalización es uno de los ejes fundamentales dentro de las estrategias definidas por la UNA, como uno de los mecanismos de respuesta a la globalización para fortalecer la participación de la Universidad en las actividades nacionales e internacionales vinculadas a la Educación Superior.

Para mayor información visite https://www.una.py/

UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO

Con 48 años de historia, la Universidad Nacional de Santiago del Estero tiene como misión la generación, desarrollo, integración y comunicación del conocimiento. Para ello asume con vocación de inclusión y calidad la formación de personas y la contribución al desarrollo en su territorio.

Se constituye como una casa de estudios superiores que gestiona su misión a través de las funciones académicas, de investigación, extensión, vinculación y transferencia del conocimiento y la cultura, como una universidad pública y autónoma.

Creada en 1973, en la actualidad la integran las Facultades de Agronomía y Agroindustrias, Ciencias Exactas y Tecnologías, Ciencias Forestales, Humanidades, Ciencias Sociales y de la Salud y Ciencias Médicas; una Escuela de Nivel Medio de Agricultura, Ganadería y Granja y las Escuelas de Artes y Oficios e Innovación Educativa.

La Universidad Pública Santiagueña camina hacia los 50 años de su creación, y se consolida como referente de excelencia en la gestión académica - científica y en los procesos de internacionalización de la Educación Superior por su alto compromiso ético, responsabilidad y servicio para el desarrollo sustentable del territorio en el que impacta.

Para mayor información visite https://www.unse.edu.ar/

ÍNDICE

Text Mining en publicaciones científicas relacionadas con Covid-19

José Federico Medrano

Facultad de Ingeniería - Visualización y Recuperación Avanzada de Información (VRAIn)
Universidad Nacional de Jujuy
San Salvador de Jujuy, 4600

jfmedrano@fi.unju.edu.ar – 388 4845386

RESUMEN

La Organización Mundial de la Salud calificó al nuevo coronavirus (Covid-19) como una pandemia global en marzo de 2020, significando esto una de las mayores amenazas a nivel mundial y obligando a un cambio en las actividades y rutinas de toda la sociedad, a diferentes escalas. Durante esta crisis, los especialistas en ciencia de la información podrían desempeñar un papel clave para apoyar los esfuerzos de los científicos de la comunidad médica y de la salud para combatir el Covid-19. En este artículo se aplica la técnica de *Text Mining* al conjunto total de publicaciones científicas sobre Covid-19 indexadas en *Microsoft Academic*, para poder describir de forma general la literatura al respecto. Esta descripción se basó en el análisis de frecuencia y coocurrencia de términos y conjunto de términos, y en la similitud entre distintos conjuntos de datos sobre tópicos previamente identificados. Se comprobó que este tipo de enfoque resulta fundamental para procesar y analizar grandes cantidades de información, encontrando relaciones y conocimiento a veces oculto, o filtrando información de manera inteligente. Esta aproximación puede ayudar a la comunidad médica y de salud a extraer información útil e interrelaciones de los estudios relacionados con el coronavirus.

Palabras Claves: Text Mining; Covid-19; Publicaciones científicas; Microsoft Academic.

INTRODUCCION

El brote sin precedentes de la enfermedad por Coronavirus 2019 (Covid-19) (WHO, 2020), causado por un nuevo coronavirus llamado coronavirus 2 (síndrome respiratorio agudo severo) (SARS-CoV-2), que comenzó con los primeros casos en diciembre de 2019, en Wuhan (China), representa uno de los desafíos globales más importantes de este siglo. La pandemia ha tenido y continúa teniendo graves consecuencias para la salud pública, la economía, la política y la sociedad. Desde cierre de ciudades, comercios, pérdidas de puestos de trabajo, cambios en las modalidades a nivel educativo y en el modo de vida de la sociedad a nivel mundial.

En respuesta a esta pandemia, han surgido una gran cantidad de estudios académicos e informes de casos en las principales revistas científicas y médicas internacionales. La mayoría de ellos abordaron cuestiones de investigación relevantes, incluida la evolución y los efectos del virus, así como los posibles factores de riesgo y los hallazgos clínicos, de laboratorio y de imágenes (Cheng, Cao, & Liao, 2020).

Este incremento notable en la producción científico-académica ha pululado las bases de datos bibliográficas con artículos sobre esta nueva temática, generando un reto al momento de buscar material relevante ya sea que se trate de abordar una nueva investigación o investigación relacionada, o responder cuestiones relacionadas a este tema a partir del descubrimiento de nuevo conocimiento.

Como un mecanismo para poder agrupar las publicaciones científicas sobre Covid-19 o facilitar el acceso, existen soluciones como la propuesta por el proyecto *Covid-19 Open Research Dataset* (CORD-19), un enorme dataset de artículos científicos sobre Covid-19 e investigaciones históricas relacionadas con el coronavirus lanzado en Marzo de 2020. CORD-19 está diseñado para facilitar el desarrollo de sistemas de extracción de información y minería de texto sobre su rica colección de metadatos y artículos estructurados de texto completo, como se indica en (Wang, y otros, 2020). En el mismo sentido, el gigante *Elsevier* en enero de 2020, puso a disposición del público el acceso gratuito a más de 19.500 artículos a través de *ScienceDirect* (Elsevier, 2020), en pos de acelerar la lucha contra la pandemia del coronavirus. Así mismo las bases de datos bibliográficas de libre acceso como *Google Scholar* y *Microsoft Academic* también permiten buscar y acceder al material relacionado con esta temática, y han aportado su apoyo al proyecto CORD-19.

Estas iniciativas indican la importancia de contar con investigaciones actualizadas en todo momento y sobre todo poder acceder a estas; ahora bien, a medida que el número de publicaciones relevantes aumenta también aumenta la complejidad para tratarlas y analizarlas de forma manual. Es así que la minería de texto, junto con el procesamiento del lenguaje natural, se puede utilizar para identificar y extraer información o relaciones de datos no estructurados y se ha convertido en un enfoque popular para el análisis de la literatura en una era de investigación que emerge rápidamente.

Este trabajo propone identificar los términos de búsqueda más relevantes y los temas de investigación genéricos de esta enfermedad mediante la realización de un análisis y síntesis de literatura automatizados basados en minería de texto, a partir de la recolección del conjunto de publicaciones relacionadas con Covid-19 indexadas por el motor académico *Microsoft Academic*. Esto permitirá descubrir la literatura representativa sobre cada tema principal de investigación, ayudando en gran medida a encontrar los estudios apropiados sobre los temas objetivos.

METODOLOGIA

CONJUNTO DE DATOS

Para llevar a cabo este trabajo se recolectó el conjunto total de publicaciones relacionadas con

Covid-19 de uno de los motores de búsqueda de material científico más importantes en la actualidad, tal es el caso de *Microsoft Academic*[1]. La recolección se realizó en septiembre de 2020 empleando la *Academic Knowledge API* (AK-API), dicha API permite mediante peticiones http y configuraciones de parámetros y filtros, extraer un conjunto de publicaciones científicas almacenadas en el *Microsoft Academic Graph* (MAG), un enorme grafo de conocimiento de publicaciones académicas que brinda soporte a dicho motor (Wang, y otros, 2020). Como los registros no se encuentran agrupados en un único tema o campo de estudio, fue necesario realizar un relevamiento inicial para conocer todos los distintos nombres de campos de estudio asociados a la temática en cuestión. Así pues, se realizó una consulta filtrando por campo de estudio (campo de metadato *Field Name*) con las siguientes opciones: "Coronavirus", "Coronavirus disease 2019 (COVID-19)", "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)", "2019-20 coronavirus outbreak", "Coronaviridae" y "Coronavirus Infections". Como resultado se obtuvo un total de 95.047 publicaciones, donde el 88% de las mismas poseen año de publicación 2.020.

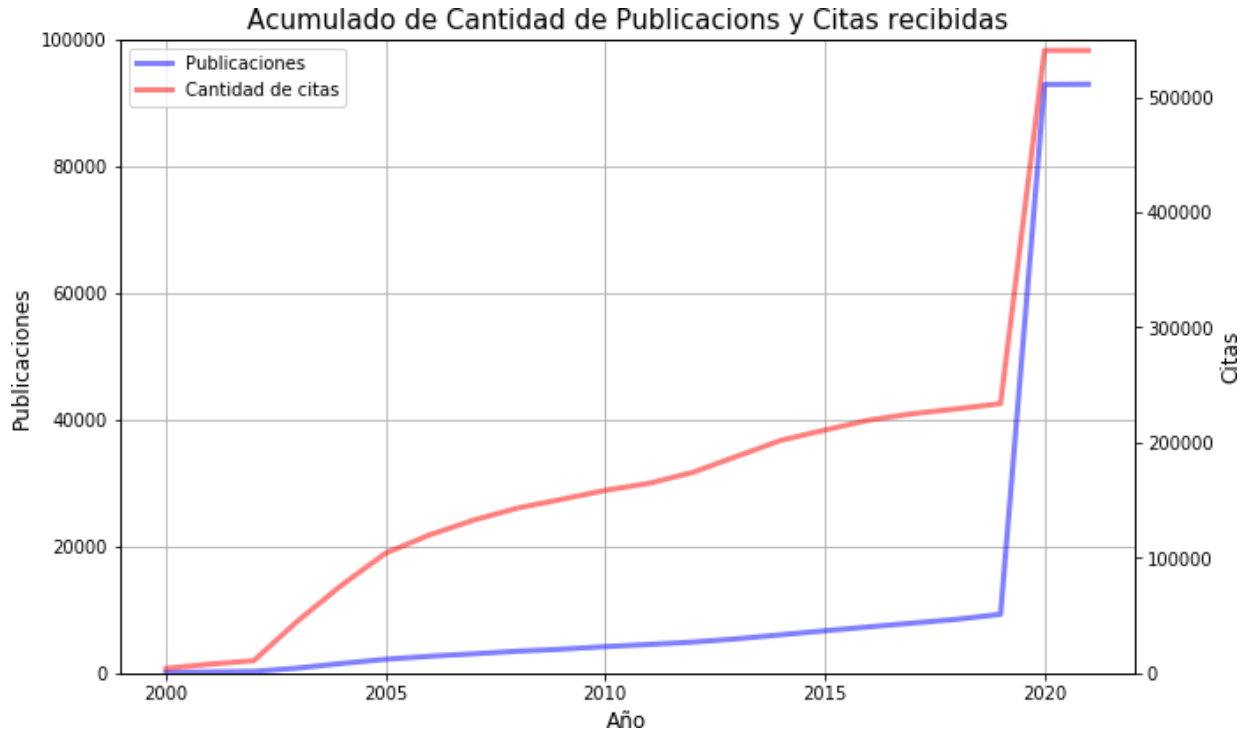


Figura 1: Grafico del total acumulado de publicaciones y citas recibidas sobre coronavirus y enfermedades similares indexadas en Microsoft Academic en los últimos 20 años

En la Figura 1 se observa un resumen de la cantidad de publicaciones y citas acumuladas en los últimos 20 años sobre investigaciones relacionadas con coronavirus. Claro está que la aparición de este nuevo coronavirus (SARS-CoV-2) provocó un aumento increíble en la producción científica, es por ello el gran número de publicaciones-citas en el año 2020. También hay que destacar que existen y existieron otros coronavirus y enfermedades similares, por ello lo interesante de encontrar relaciones no solo con investigaciones pasadas sino también con decisiones y acciones tomadas. Por lo tanto la minería de textos puede dar indicios para encausar investigaciones o centrarse en algún tópico o revisar la bibliografía existente y saber de qué trata

[1] https://academic.microsoft.com/

de manera automática.

TEXT MINING

El *Text Mining* o Minería de Texto, un tipo particular de minería de datos, tiene como objetivo extraer conocimientos útiles como relaciones, patrones y tendencias de datos no estructurados o semiestructurados, por ejemplo, documentos de texto (Feldman & Sanger, 2006). El proceso principal en la minería de textos es transformar el texto en datos numéricos utilizando métodos estadísticos para extraer el contenido textual en una matriz organizada documento-término, que abarca las siguientes dos dimensiones: las palabras (o términos, compuestos por *n* palabras) y los documentos (Moro, Pires, Rita, & Cortez, 2019). Estas técnicas aportan una gran ventaja al momento de analizar corpus textuales de miles de registros, puesto que automatizan parte del proceso de extracción de nuevo conocimiento, facilitando los análisis y la obtención de conclusiones de manera más sencilla. La minería de texto ha sido ampliamente utilizada en diversas áreas del conocimiento como la biomedicina (Kim & Delen, 2018), análisis de sentimiento y opiniones (Liu, 2012), recuperación de información (Meystre, Savova, Kipper-Schuler, & Hurdle, 2008), ciencias sociales (Ignatow & Mihalcea, 2016), entre otros.

PREPROCESAMIENTO

De los registros recolectados solo se trabajó con el título y resumen de estos, puesto que no todos los registros se encuentran a texto completo y porque el resumen y el título aportan información muy relevante y suficiente para una primera aproximación. Los datos se sometieron a un proceso de limpieza y transformación para evitar no solo errores sino sesgos innecesarios. El preprocesado consistió en una serie de pasos, el primero de ellos eliminar los registros que no poseían título o resumen (41.717 registros no poseen resumen, este es un error muy frecuente en las bases de datos bibliográficas de libre acceso debido principalmente a la imposibilidad de indexar correctamente dicho campo). Seguido de esto se eliminaron los registros de idioma distinto al inglés (se emplearon solo los registros en idioma inglés para darle una interpretación más sencilla a las relaciones encontradas, si bien es una buena práctica trabajar sobre un conjunto monolingüe, (Singh, y otros, 2020) ha demostrado buenos resultados sin aplicar filtros de idioma); existían alrededor de 1714 registros en idioma español, 812 en portugués, 705 en francés y 401 en indonesio entre los más frecuentes. Luego se pasó todo el texto a minúsculas, se eliminaron las *stopwords* (palabras vacías y poco descriptivas) (Silva & Ribeiro, 2003) del idioma inglés, se eliminaron los signos de puntuación y se realizó la lematización de los términos de cada documento (Webster & Kit, 1992). Todo este proceso se llevó a cabo en Python empleando la librería NLTK (Bird, 2006) y obteniendo como resultado un total de 49.351 documentos.

DESARROLLO

Para tener una idea del tamaño de cada registro (título + resumen), en la Figura 2 se ofrece un histograma de la longitud de estos. Se puede observar que la distribución del tamaño se asemeja a una distribución normal, con una media de 1058 caracteres. Mientras más descriptivo sea el documento (por lo general la información más detallada e informativa se encuentra en el resumen de la publicación, ya que los títulos suelen ser cortos), mayores y mejores relaciones se podrán hallar.

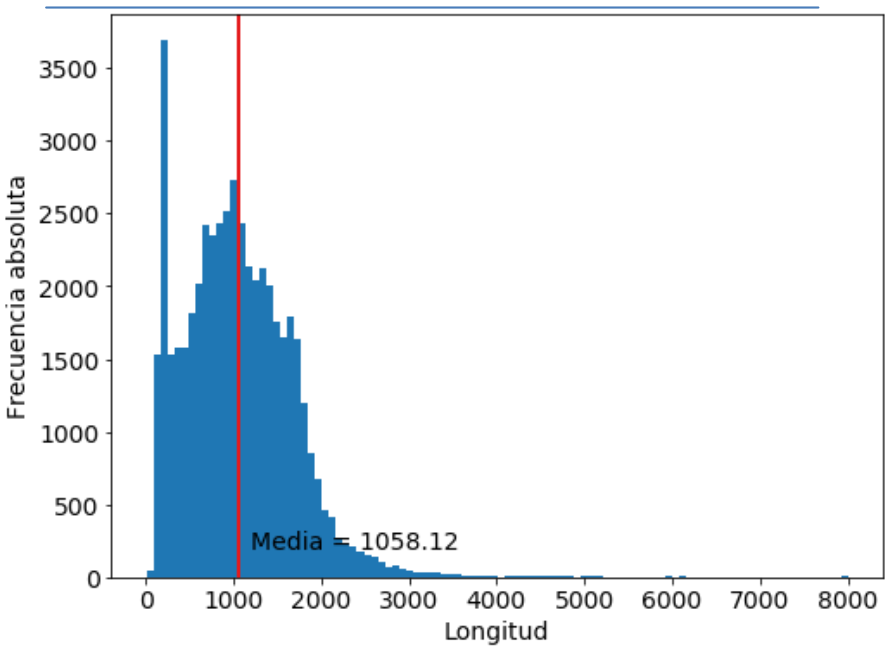


Figura 2: Histograma de la longitud del conjunto textual realizado con *matplotlib.pyplot*

Existen diversas aproximaciones para iniciar este tipo de análisis, uno de ellos es identificar las palabras o términos mayormente empleados dentro del texto como se observa en la Figura 3. En algunos casos los términos más frecuentes son poco descriptivos o discriminativos ya que son comunes a la mayoría de los documentos. Por ello en esquemas como el TF-IDF (del inglés *Term Frequency – Inverse Document Frequency*) son penalizados (Salton & Buckley, 1988). Las cinco palabras más frecuentes son "covid19", "patient", "coronavirus", "disease" y "virus". Esto revela que las investigaciones actuales sobre Covid-19 se han enfocado en el paciente, en la detección de casos tempranos causado por este virus. Revisando un poco más en detalle este listado de frecuencias, se advierte que se trata de una enfermedad respiratoria y severa que afecta gravemente la salud, estas pueden parecer conclusiones obvias pero para una máquina que lo único que hizo fue procesar, limpiar y contabilizar información no lo es.

Otro esquema que ha dado buenos resultados es el empleo de n-gramas, es decir, analizar la frecuencia o coocurrencia de conjuntos de palabras. De este modo no se analizan palabras aisladas sino conjuntos de dos, tres o cuatro términos (bigramas, trigramas o cuatrigramas respectivamente). Para ello, la Figura 4 ofrece una nube de palabras de los 100 bigramas más importantes utilizados en el cuerpo textual de publicaciones analizadas, se puede apreciar que la conjunción más usual es la palabra "covid19" con las palabras "pandemic" (pandemia), "disease" (enfermedad), "epidemic" (epidemia), "patient" (paciente), "outbreak" (brote), "infection" (infección), entre otros. Siguiendo este principio se pueden agrupar palabras de orden superior, aunque en las pruebas realizadas no poseían mucho sentido las agrupaciones encontradas de tres o cuatro términos.

Como se desprende de la última imagen, en la nube de palabras ofrecida (Figura 4), un mayor tamaño de las palabras indica una mayor frecuencia del término empleado, así mismo se puede observar las diversas asociaciones de temas distintos, por ejemplo existen registros que hablan o tratan temas relacionados con las enfermedades respiratorias, con lo síntomas, con el tratamiento, con el cuidado de la salud, con las medidas adoptadas como el distanciamiento social, con el origen del brote, con la relación que existe con enfermedades similares, entre otras cuestiones.

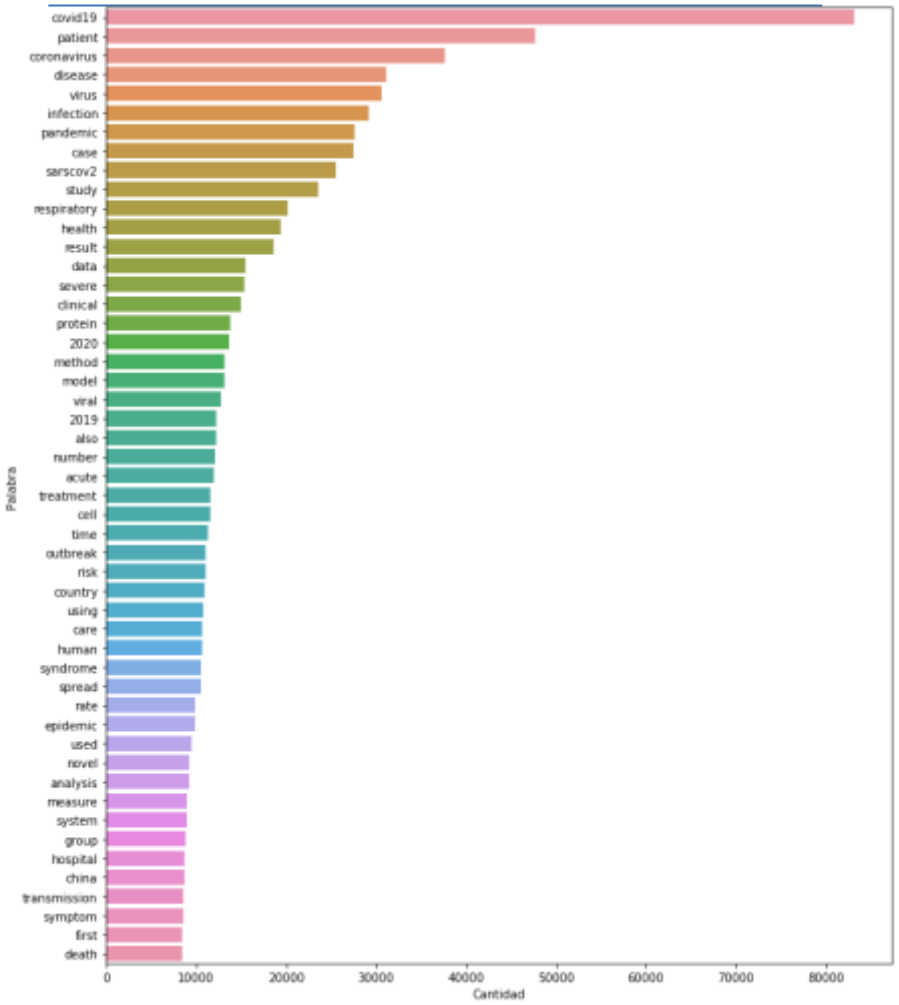


Figura 3: Frecuencia de aparición de las 50 palabras más empleadas en el corpus textual obtenido luego del preprocesamiento, sin distinción de año de publicación.

Es aquí donde se puede vislumbrar la potencia de estos estudios, a partir de esta primera aproximación se pueden desprender nuevas aristas de estudio como la realización de un modelado temático sobre el conjunto textual, la búsqueda de artículos que estén relacionados a un conjunto de términos muy frecuentes, relacionar publicaciones a partir de la coocurrencia de términos. Por ejemplo, el modelado de temas o *topic modeling* es una técnica estadística muy empleada que transforma las palabras relevantes y su frecuencia en una estructura organizada, en la que los documentos se distribuyen en varios temas (Blei, 2012). Esta técnica puede ofrecer un conjunto de agrupaciones de documentos/artículos que estén fuertemente relacionados por el contenido semántico de los mismos, y del mismo modo encontrar documentos que pertenecen a otros temas pero que guardan alguna relación con el tema de origen.

A partir de algunos de los bigramas identificados en la Figura 4, por ejemplo “health care” y “social distancing”, se puede analizar la relación que existe entre las publicaciones que hablan de un tema u otro. En el dataset empleado para este trabajo existen 1614 trabajos que incluyen el bigrama “health care”, mientras que al bigrama “social distancing” lo incluyen 1828 publicaciones.

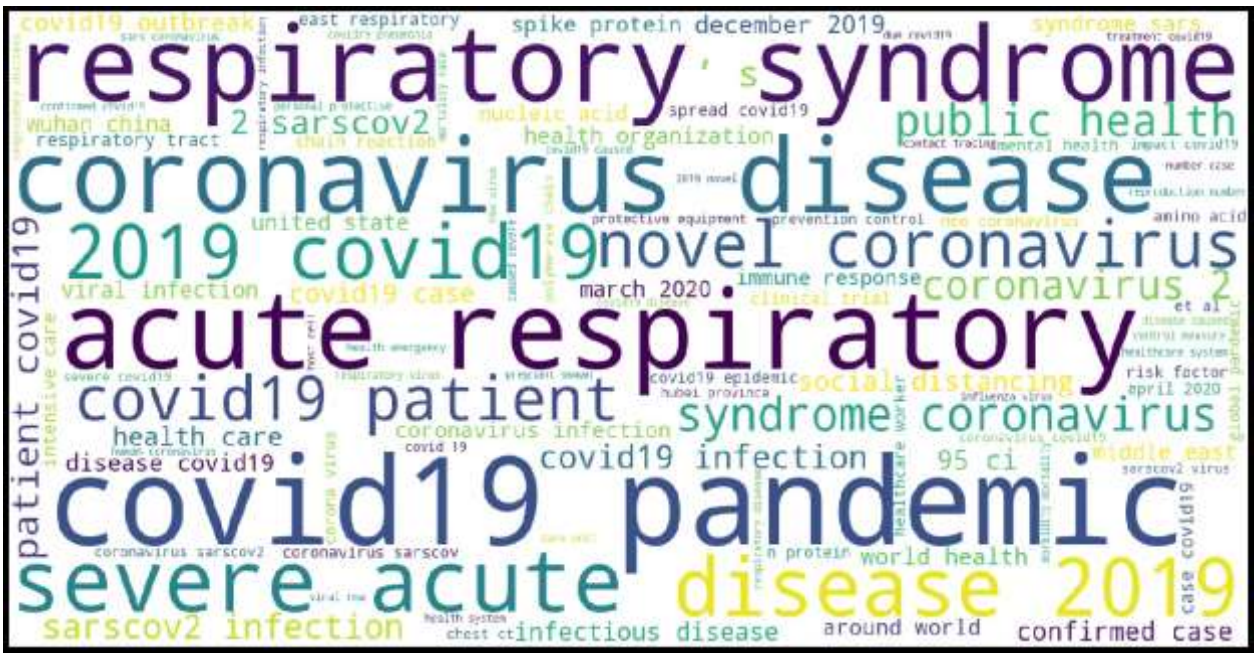


Figura 4: Nube de palabras de los bigramas más frecuentes

Fuera de los términos “health care” (2421), “covid 19” (1339) y “covid19 pandemic” (819) que aparecen en los registros que incluyen al bigrama “health care”, y fuera de los términos “social distancing” (3055), “covid 19” (1525) y “covid19 pandemic” (751) que parecen en los registros que incluyen al bigrama “social distancing”, en la **¡Error! La autoreferencia al marcador no es válida.** se pueden observar los términos (bigramas) más frecuentes y representativos de cada conjunto de registros. Existe una correspondencia entre ambos conjuntos de registros determinada por la existencia no solo de términos en común, sino porque el contenido textual de las publicaciones relacionadas está semánticamente relacionado.

Tabla 1: Términos representativos de la literatura sobre Covid-19 que incluye a "health care" y/o "social distancing"

Bigrama “health care”		**Bigrama “social distancing”**	
Término	Frecuencia	Término	Frecuencia
care worker	552	distancing measure	499
disease 2019	436	public health	333
care system	430	disease 2019	262
acute respiratory	243	covid19 case	215
public health	233	spread covid19	203
respiratory syndrome	225	contact tracing	148
covid19 patient	209	covid19 outbreak	136
severe acute	207	severe acute	133
protective equipment	189	face mask	129
covid19 case	138	health care	128

Este mismo enfoque se puede aplicar a cualesquier conjunto de términos, pudiendo ampliar la cantidad de conjuntos de publicaciones a analizar.

CONCLUSIONES

El brote de Covid-19 está representando uno de los desafíos globales más importantes en este Siglo en todos los ámbitos y a diferentes escalas. Ha supuesto un giro y cambio de paradigmas desde el sistema educativo, económico, social, académico, y claramente afectando el sistema médico y de salud. A la espera de las nuevas olas de contagio, esperando la tan ansiada vacuna que aún no termina de definirse y esperando nuevas soluciones desde el sistema científico-tecnológico, la sociedad sigue en cuarentena con distanciamiento social y la crisis se agudiza cada vez más.

En este trabajo se ha intentado aportar un pequeño granito de arena demostrando la potencia y las posibilidades que ofrece la minería de textos para analizar grandes cantidades de texto no estructurado, en este caso, el conjunto de publicaciones científicas que tratan sobre Covid-19 indexadas por el motor académico *Microsoft Academic*.

Se evidenció cuan factible es procesar miles de publicaciones en busca de nuevo conocimiento, y relaciones a veces poco visibles u ocultas por la enorme cantidad de información. Esta suerte de automatización en el análisis de datos textuales permite obtener conclusiones, responder preguntas, afirmar hipótesis y encauzar nuevas líneas de investigación, a partir de la identificación de temas, tópicos o conjunto de términos, pudiendo evaluar el grado de superposición, similitud y diferencia entre estos temas.

Es así, que del mismo modo que en este trabajo se hallaron e identificaron algunas de las relaciones entre conjuntos de publicaciones, también se puede aplicar la mirada opuesta para demostrar la diferencia que existe entre distintas temáticas o áreas de investigación dentro de un mismo estudio.

BIBLIOGRAFÍA

Bird, S. (2006). NLTK: The Natural Language Toolkit. *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, (págs. 69-72). Obtenido de https://academic.microsoft.com/paper/2143017621

Blei, D. (2012). Probabilistic topic models. *Communications of The ACM, 55*(4), 77-84. Obtenido de https://academic.microsoft.com/paper/2174706414

Cheng, X., Cao, Q., & Liao, S. (2020). An overview of literature on COVID-19, MERS and SARS: Using text mining and latent Dirichlet allocation. *Journal of Information Science*. Obtenido de https://academic.microsoft.com/paper/3082307550

Elsevier. (Marzo de 2020). *Elsevier da acceso completo a su contenido sobre el COVID-19 para acelerar la lucha contra la pandemia*. Obtenido de https://www.elsevier.com/es-es/connect/coronavirus/elsevier-da-acceso-completo-a-su-contenido-sobre-el-covid-19-para-acelerar-la-lucha-contra-el-coronavirus

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Obtenido de https://academic.microsoft.com/paper/2024228866

Ignatow, G., & Mihalcea, R. (2016). *Text Mining: A Guidebook for the Social Sciences.* Obtenido de https://academic.microsoft.com/paper/2748167422

Kim, Y.-M., & Delen, D. (2018). Medical informatics research trend analysis: A text mining

approach. *Health Informatics Journal, 24*(4), 432-452. Obtenido de https://academic.microsoft.com/paper/2772572665

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Obtenido de https://academic.microsoft.com/paper/2108646579

Meystre, S., Savova, G., Kipper-Schuler, K., & Hurdle, J. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform, 17*(1), 128-144. Obtenido de https://academic.microsoft.com/paper/2114388055

Moro, S., Pires, G., Rita, P., & Cortez, P. (2019). A text mining and topic modelling perspective of ethnic marketing research. *Journal of Business Research, 103*, 275-285. Obtenido de https://academic.microsoft.com/paper/2736934374

Salton, G., & Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management, 24*(5), 323-328. Obtenido de https://academic.microsoft.com/paper/1978394996

Silva, C., & Ribeiro, B. (2003). The importance of stop word removal on recall values in text categorization. *Proceedings of the International Joint Conference on Neural Networks, 2003.*, *3*, págs. 1661-1666. Obtenido de https://academic.microsoft.com/paper/1721182246

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., . . . Wang, Y. (2020). A first look at COVID-19 information and misinformation sharing on Twitter. *arXiv preprint arXiv:2003.13907*. Obtenido de https://academic.microsoft.com/paper/3015131281

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies, 1*(1), 396-413. Obtenido de https://academic.microsoft.com/paper/3002924435

Wang, L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., . . . Kohlmeier, S. (2020). CORD-19: The Covid-19 Open Research Dataset. *arXiv preprint arXiv:2004.10706*. Obtenido de https://academic.microsoft.com/paper/3020786614

Webster, J., & Kit, C. (1992). Tokenization as the initial phase in NLP. *COLING '92 Proceedings of the 14th conference on Computational linguistics - Volume 4*, (págs. 1106-1110). Obtenido de https://academic.microsoft.com/paper/2037450062

WHO, W. H. (Febrero de 2020). *WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020*. Obtenido de https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020

10 y 11 de Diciembre de 2020

Identificación de áreas prioritarias de investigación de la estevia en Paraguay mediante el análisis multidimensional de preferencias

Autores: Andres D. Sanabria-Velazquez[1], Guillermo A. Enciso-Maldonado[2], Javier E. Barua-Chamorro[3], H. David Shew[1]

[1]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA.
[2]Centro de Desarrollo e Innovación Tecnológica de Itapúa, Hohenau, Itapúa, Paraguay.
[3]Departamento de Química Biológica, Facultad de Ciencias, Universidad Nacional de Asunción, San Lorenzo, Paraguay.

E-mail address: adsanabr@ncsu.edu

RESUMEN

La estevia (*Stevia rebaudiana*) es un cultivo importante para los agricultores paraguayos, que genera ingresos para más del 30% de las familias rurales. El objetivo de este trabajo fue identificar áreas prioritarias de investigación para factores que limitan la producción de la estevia y conocer la percepción de los investigadores, extensionistas y tomadores de decisiones en Paraguay. Se realizaron encuestas sobre los temas agronómicos con prioridad de investigación para la producción de estevia, en la Facultad de Ciencias Químicas, Universidad Nacional de Asunción, Paraguay y el otro en la Facultad de Agronomía "Universidad Católica Nuestra Señora de la Asunción", María Auxiliadora, Itapúa, Paraguay. Se proporcionó a los encuestados una lista de áreas prioritarias de investigación y se les pidió clasificarlas de 1 (prioridad alta) a 10 (prioridad baja). Participaron un total de 40 sujetos, entre extensionistas, profesores, investigadores públicos, productores de estevia, consultores privados y estudiantes graduados con experiencia en agronomía e investigación de la estevia. La mayoría de los encuestados consideraron enfermedades transmitidas por el suelo como un tópico de alta prioridad de investigación. Enfermedades causadas por virus y nematodos se consideraron de baja prioridad. Para la categoría insectos plaga no se observó un patrón definido de preferencia. Por otro lado, nutrición del cultivo, producción de semillas, manejo orgánico fueron considerados por la mayoría del grupo de encuestados como tópicos más prioritarios que mejoramiento genético.

Palabras Claves: *Stevia rebaudiana, encuesta, fitopatógenos, MDPREF, plagas*

INTRODUCCIÓN

La estevia (*Stevia rebaudiana* [Bertoni] Bertoni) es un cultivo originario del Paraguay y se constituye como una alternativa importante para los pequeños agricultores paraguayos debido a que genera ingresos para más del 30% de las familias rurales y actualmente es el segundo mayor productor mundial (Casaccia et al. 2016). Las propiedades nutracéuticas y como edulcorante natural para su uso en alimentos, así como sus características agronómicas del cultivo de estevia han permitido que este cultivo se extienda por todo el mundo, teniendo un gran un impacto agrícola en países como Japón, China, Taiwán, Corea, México, Estados Unidos, Tailandia, Malasia, Indonesia, Australia, Tanzania, Canadá (Ramesh et al., 2006).

Las investigaciones en estevia se han centrado mayormente en mejorar la calidad del dulzor por medio del mejoramiento genético. Sin embargo, las consideraciones agronómicas deben ser objeto de estudio constante debido a que el inminente cambio climático y la necesidad de hacer mejor uso de los recursos naturales exigen que la producción de cultivos se desarrolle de manera sostenible y eficiente (Ramesh et al. 2006).

Si bien Paraguay tiene buenas condiciones para la producción de estevia, los agricultores tienen dificultades para producir cultivos de manera eficiente y económica. Existen varios factores que limitan la producción, tanto de índole abiótica como biótica. El principal factor abiótico es el estrés hídrico ya que, la estevia es muy exigente en agua, por lo que no se recomienda su cultivo en áreas con precipitaciones menores a los 1400 mm por año sin sistema de riego. Por otro lado, los principales factores bióticos son los daños causados por enfermedades como la cenicilla (*Erysiphe cichoracearum*), el tumbamiento (*Rhizoctonia solani*) o la pudrición del tallo (*Sclerotium dephinii*), entre otras enfermedades causadas por los hongos *Septoria steviae* y *Sclerotinia sclerotiorum* (Casaccia et al. 2016). Mientras que, insectos como pulgones y moscas blancas, cochinillas y ácaros rojos pueden ocasionar daños en el cultivo (Thomas 2000). Otra limitante que se presenta es la poca disponibilidad de variedades comerciales para productores. Actualmente se cuentan con dos variedades clonales “Eireté” y “Katupyry”, mientras que algunos productores utilizan semillas botánicas que resultan en una producción con mezcla de varios genotipos que difieren en sus características morfológicas y fenológicas, dificultando al productor saber el momento oportuno para la cosecha (Casaccia et al. 2016). El objetivo de este trabajo fue identificar áreas prioritarias de investigación agronómica que limitan la producción de estevia en base a la percepción de investigadores, extensionistas y tomadores de decisiones con experiencia en agronomía e investigación de la estevia en Paraguay.

METODOLOGÍA

Se realizaron encuestas sobre las áreas agronómicas con mayor prioridad de investigación para producción de estevia, en la Facultad de Ciencias Químicas, Universidad Nacional de Asunción, Paraguay y el otro en la Facultad de Agronomía “Universidad Católica Nuestra Señora de la Asunción”, María Auxiliadora, Itapúa, Paraguay. En total fueron encuestadas 40 personas entre extensionistas, profesores, investigadores públicos, productores de estevia, consultores privados y estudiantes graduados con experiencia en agronomía y agronomía e investigación de la de estevia, a los que se les proporcionó una lista de áreas prioritarias de investigación y se les pidió clasificarlas de 1 (prioridad alta) a 10 (prioridad baja). Se realizó un análisis de preferencias multidimensional (MDPREF) (Carroll 1972) en SAS para determinar cómo se agrupaban las

prioridades en relación con las preferencias de los encuestados (SAS Institute Inc., Cary, NC). Los datos y el código utilizado para el análisis en SAS se encuentran disponibles en ResearchGate en el siguiente enlace (https://www.researchgate.net/publication/346370502_Encuesta_estevia_py)).

Encuesta
Necesidades de investigación sobre ka'a he'ẽ en Paraguay
Áreas donde opera (marque todas las que correspondan):
_____Investigador Público
_____Docente Investigador
_____Investigador de la industria
_____Extensionista
_____Estudiante universitario
_____Area comercial/producción

Por favor aclarar el nombre de la institución/empresa_____

Por favor, clasifique los temas principales que cree que deberían ser abordados por la investigación sobre ka'a he'ẽ en Paraguay (1 a 10 con 1 "prioridad máxima" y 10 con "prioridad más baja"):

Enfermedades:
____Damping off (*Fusarium* spp.; *Phytium* spp.; *Phytopthora* spp.; *Rhizoctonia solani*; *Sclerotium rolfsii*)
____Podredumbre carbonosa de la raíz (*Macrophomina phaseolina*)
____Pudrición blanca de la raíz (*Sclerotium rolfsii*)
____Mancha foliar de Septoria (*Septoria steviae*)
____Mancha foliar de Alternaria (*Alternaria alternata*)
____Pudrición de la raíz por Sclerotinia (*Sclerotinia sclerotiorum*)
____Enfermedades virales
____Nematodos fitopatógenos
____Enfermedades bacterianas
____ Otros (por favor aclarar): ____

Insectos Plaga:
____Ácaros
____Pulgones
____Trips
____Orugas
____Otros (por favor aclarar): ____

Mejoramiento genético / Manejo cultural:
____Nutrición del cultivo
____Produccion de semillas
____Manejo postcosecha
____Manejo alternativo (orgánico) del cultivo
____Preservación de material genético
____Mejoramiento genético (por favor aclarar característica/as a mejorar): ____

Figura 1. Encuesta sobre de áreas prioritarias de investigación de la estevia en Paraguay.

DESARROLLO

Se utilizó el procedimiento PROC PRINQUAL en SAS para realizar un análisis multidimensional de preferencias (MDPREF) (SAS 1999). El análisis MDPREF es un análisis de componentes principales de una matriz de datos con columnas que corresponden a los encuestados y filas que corresponden a los temas de investigación sobre estevia. Los datos son calificaciones o clasificaciones de la preferencia de cada persona por cada tema de investigación que considera como prioridad. Los datos son la transposición de la matriz de datos de las calificaciones (en otras palabras, las columnas son los encuestados; en la matriz típica, las filas representan a los encuestados). El resultado final de un análisis MDPREF es un biplot de las preferencias (Gabriel 1981). Un biplot muestra a los encuestados y a los temas de investigación en un solo gráfico proyectándolos en el plano de las variables transformadas que explican la mayor variación de los datos. Las calificaciones se realizaron en un 1 (alta prioridad) a 10 (baja prioridad). Los tópicos de investigación se dividieron en tres categorías: enfermedades (principales enfermedades que de limitan la producción y deben ser investigadas), insectos plaga (principales plagas que atacan al cultivo), y manejo cultural (principales técnicas y manejos en la producción de estevia que deben ser investigados) (Figura 1). El siguiente código de SAS permite cargar los datos de los encuestados en las columnas y cada prioridad de investigación de la categoría enfermedades en las filas junto con las calificaciones (Figura 2).

```
title 'Stevia Research Needs in Paraguay Survey';

options validvarname=any;

data steviasurvey;
   input Research  $ 1-20 @40 ('1'n-'40'n) (4.);
   datalines;
Damping off                 1  1  7  1  1  1  1  2  5  5
Charcoal root rot           2  3  1  2  9  5  4  6  7  2
White root rot              4  2  3  6  6  1  7  1  4  1
Septoria leaf spot          8  8  2  4  5  1  5  3  1  3
Alternaria leaf spot        7  7  2  5  4  3  8  4  3  2
Sclerotinia root rot        3  4  6  3  7  1  6  5  2  3
TSWV                        9  9  5  9  8  5 10  8  8  4
Phytonematodes              5  5 10  7  2  5  2  9  9  5
Bacterial diseases          6  6  4  8  3  5  3 10  6  2
;
proc print data=steviasurvey;
run;
```

Figura 2. Código de SAS con percepciones de los encuestados correspondientes a cada prioridad de investigación de la categoría enfermedades.

El siguiente código ejecutan PROC PRINCOMP y crean un gráfico con los valores Eigen del análisis de componentes principales (Figura 3).

```
ods graphics on;
* Principal Component Analysis of the Original Data;
proc princomp data=steviasurvey;
   ods select ScreePlot;
   var 'l'n-'40'n;
run;
```

Figura 3. Código SAS para ejecutar PROC PRINCOMP

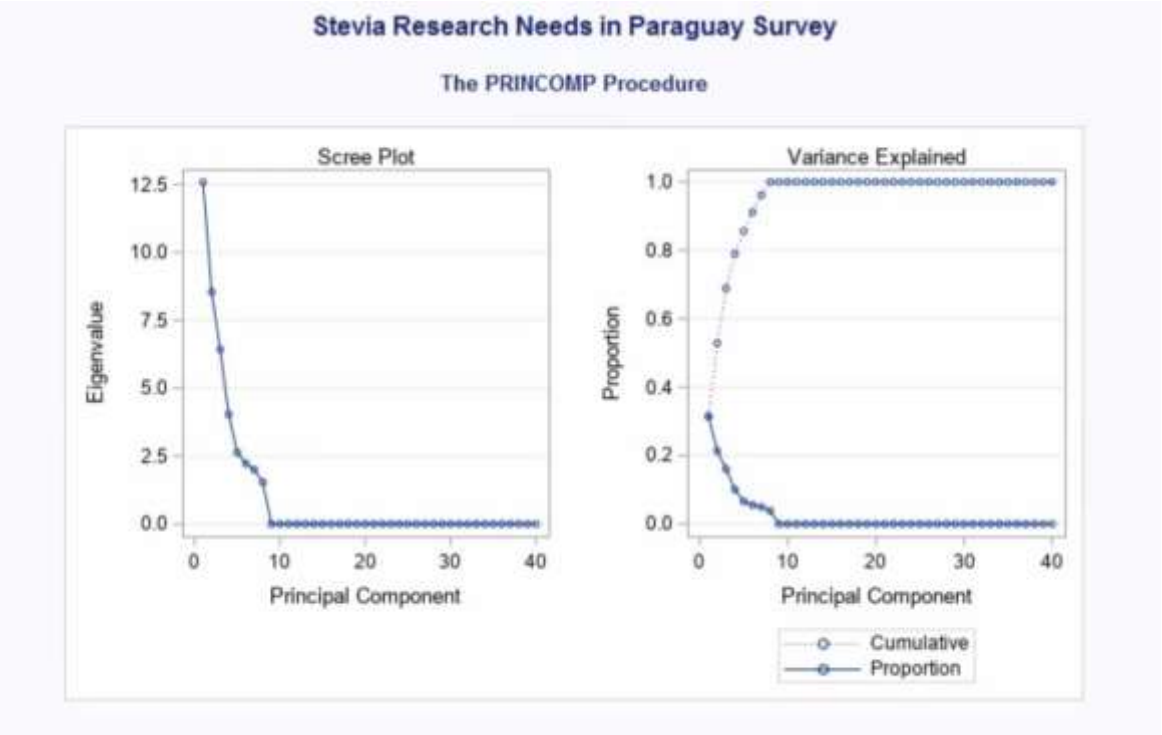


Figura 4. Scree Plot con los valores Eigen del análisis de componentes principales.

PROC PRINQUAL transforma monótonamente las percepciones de los encuestados para maximizar la proporción de varianza explicada por los dos primeros componentes principales. La opción MONOTONE se especifica en la instrucción TRANSFORM para solicitar un análisis MDPREF no métrico; como alternativa, puede especificar la opción IDENTIDAD para un análisis métrico. La opción DATA = especifica el conjunto de datos de entrada, OUT = crea un conjunto de datos de salida, y N = 2 y el método predeterminado METHOD = MTV transforman los datos para que se ajusten mejor a un modelo de dos componentes. La opción REPLACE reemplaza los datos originales con los datos transformados monótonamente en el conjunto de datos OUT =. La opción MDPREF estandariza los puntajes de los componentes a la varianza uno para que la geometría del biplot sea correcta y crea dos variables en el conjunto de datos OUT = llamadas Prin1 y Prin2. La ventaja de crear un biplot basado en componentes principales es que las coordenadas no dependen del tamaño de la muestra. Los siguientes comandos en SAS transforman los datos.

```
* Transform the Data to Better Fit a Two Component Model;
proc prinqual data=steviasurvey out=Results n=2 replace mdpref;
   title2 'Multidimensional Preference (MDPREF) Analysis';
   title3 'Optimal Monotonic Transformation of Preference Data';
   id research;
   transform monotone('1'n-'40'n);
run;
title 'Stevia Research Needs in Paraguay Survey';

options validvarname=any;
```

Figura 5. Código en SAS para ejecutar PROC PRINQUAL y transformar monótonamente las calificaciones las percepciones de los encuestados

El historial de iteraciones que muestra PROC PRINQUAL indica que la proporción de varianza aumenta de un 0.52845 inicial a 0.68198. La proporción de varianza explicada por PROC PRINQUAL en la primera iteración es igual a la proporción acumulada de varianza mostrada por PROC PRINCOMP para los dos primeros componentes principales. La iteración inicial de PROC PRINQUAL realiza un análisis estándar de componentes principales de los datos brutos.

Stevia Research Needs in Paraguay Survey
Multidimensional Preference (MDPREF) Analysis
Optimal Monotonic Transformation of Preference Data

The PRINQUAL Procedure

PRINQUAL MTV Algorithm Iteration History

Iteration Number	Average Change	Maximum Change	Proportion of Variance	Criterion Change	Note
1	0.18097	1.18141	0.52845		
2	0.04456	0.37417	0.66446	0.13600	
3	0.02556	0.26759	0.67463	0.01017	
4	0.01654	0.17240	0.67869	0.00406	
5	0.01028	0.09345	0.68027	0.00158	
6	0.00654	0.07285	0.68092	0.00064	
7	0.00476	0.06829	0.68124	0.00032	
8	0.00362	0.06241	0.68144	0.00021	
9	0.00277	0.04808	0.68159	0.00015	
30	0.00004	0.00247	0.68198	0.00000	Not Converged

WARNING: Failed to converge, however criterion change is less than 0.0001.

Figura 6. Historial de iteraciones del procedimiento PROC PRINQUAL.

Las columnas etiquetadas "*Average Change*" (Cambio Promedio), "*Maximum Change*" (Cambio Máximo) y "*Criterion Change*" (Cambio de Criterio) en la Figura 6, contienen valores que siempre

disminuyen, lo que indica que PROC PRINQUAL está mejorando las transformaciones a una tasa de disminución monótona durante las iteraciones. Esto no siempre sucede, y cuando no sucede, sugiere que el análisis podría estar convergiendo hacia una solución degenerada. El algoritmo no converge en 30 iteraciones. Sin embargo, el cambio de criterio es pequeño, lo que indica que es poco probable que más iteraciones tengan mucho efecto en los resultados.

Interpretación del análisis multidimensional de preferencias

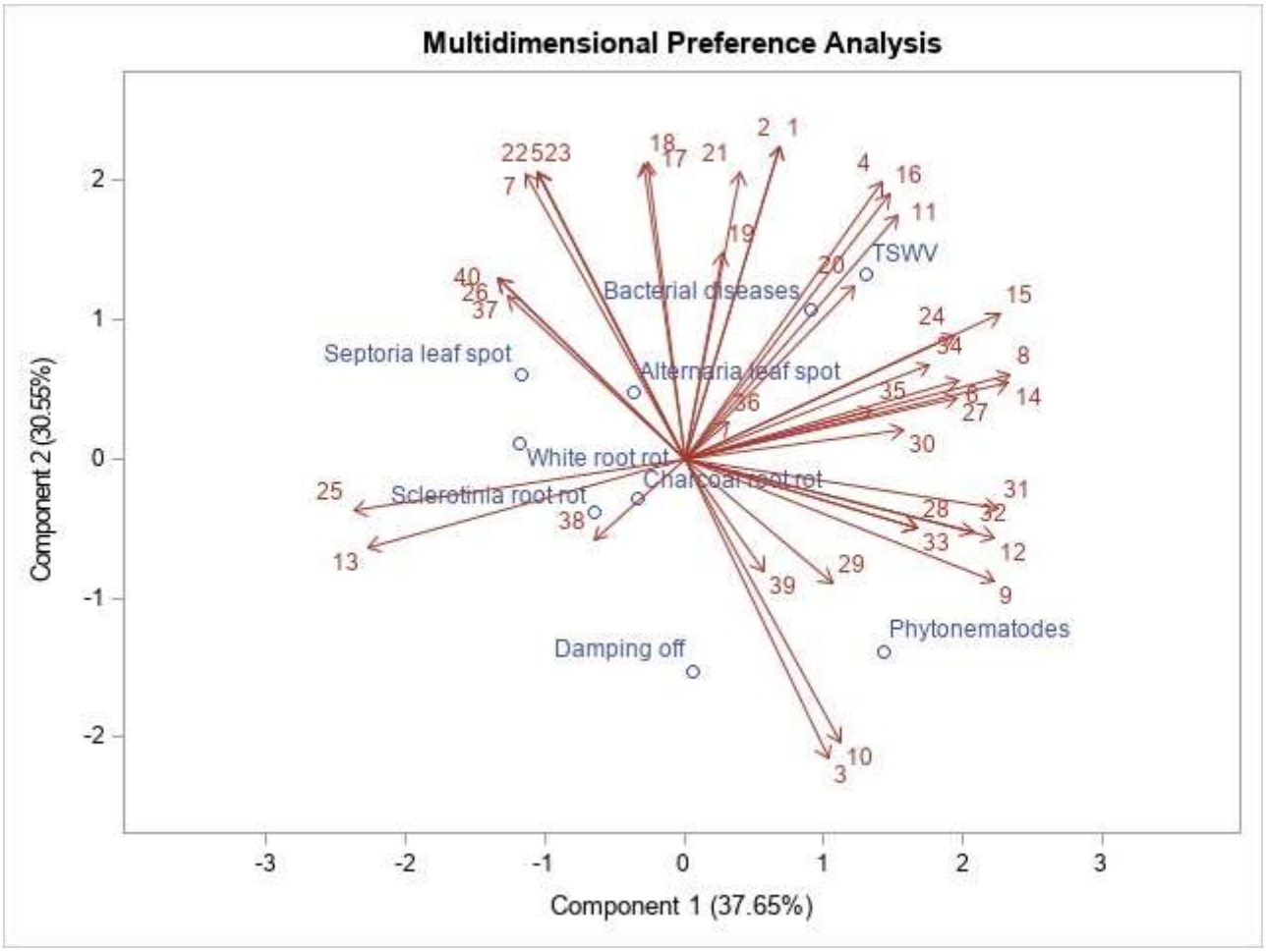


Figura 7. Biplot con las percepciones de los encuestados correspondientes a cada prioridad de investigación de la categoría enfermedades mostrado automáticamente por PROC PRINQUAL cuando ODS Graphics está habilitado y se especifica la opción MDPREF.

Un biplot es un gráfico que muestra la relación entre los valores de las filas y las columnas de una matriz de datos. Dado que el análisis MDPREF se basa en un modelo de componentes principales, las dimensiones del biplot MDPREF corresponden a los dos primeros componentes principales. El primer componente principal es la dimensión más larga a través del biplot MDPREF. El primer componente principal es la preferencia general, que es la dimensión más destacada en los juicios de preferencia. Un extremo apunta en la dirección que, en promedio, es más preferida por los encuestados, y el otro apunta en la dirección menos preferida. El segundo componente principal es ortogonal al primer componente principal, y es la dirección ortogonal que en segundo lugar explica

mejor la variación de los datos. Con un biplot MDPREF, es apropiado representar cada prioridad de investigación (objeto) por un punto y cada encuestado por un vector. Los puntos de cada prioridad de investigación tienen coordenadas que son las puntuaciones de la prioridad de investigación en los dos primeros componentes principales. Los vectores de cada encuestado emanan del origen del espacio y pasan por un punto cuyas coordenadas son los coeficientes del encuestado (variable) sobre los dos primeros componentes principales.
La longitud absoluta de un vector es arbitraria. Sin embargo, las longitudes relativas de los vectores indican ajuste al modelo, siendo las longitudes al cuadrado proporcionales a las comunalidades (que se pueden obtener con el comando PROC FACTOR en SAS). La dirección del vector indica la dirección más preferida por el encuestado, y la preferencia aumenta a medida que el vector se mueve desde el origen.

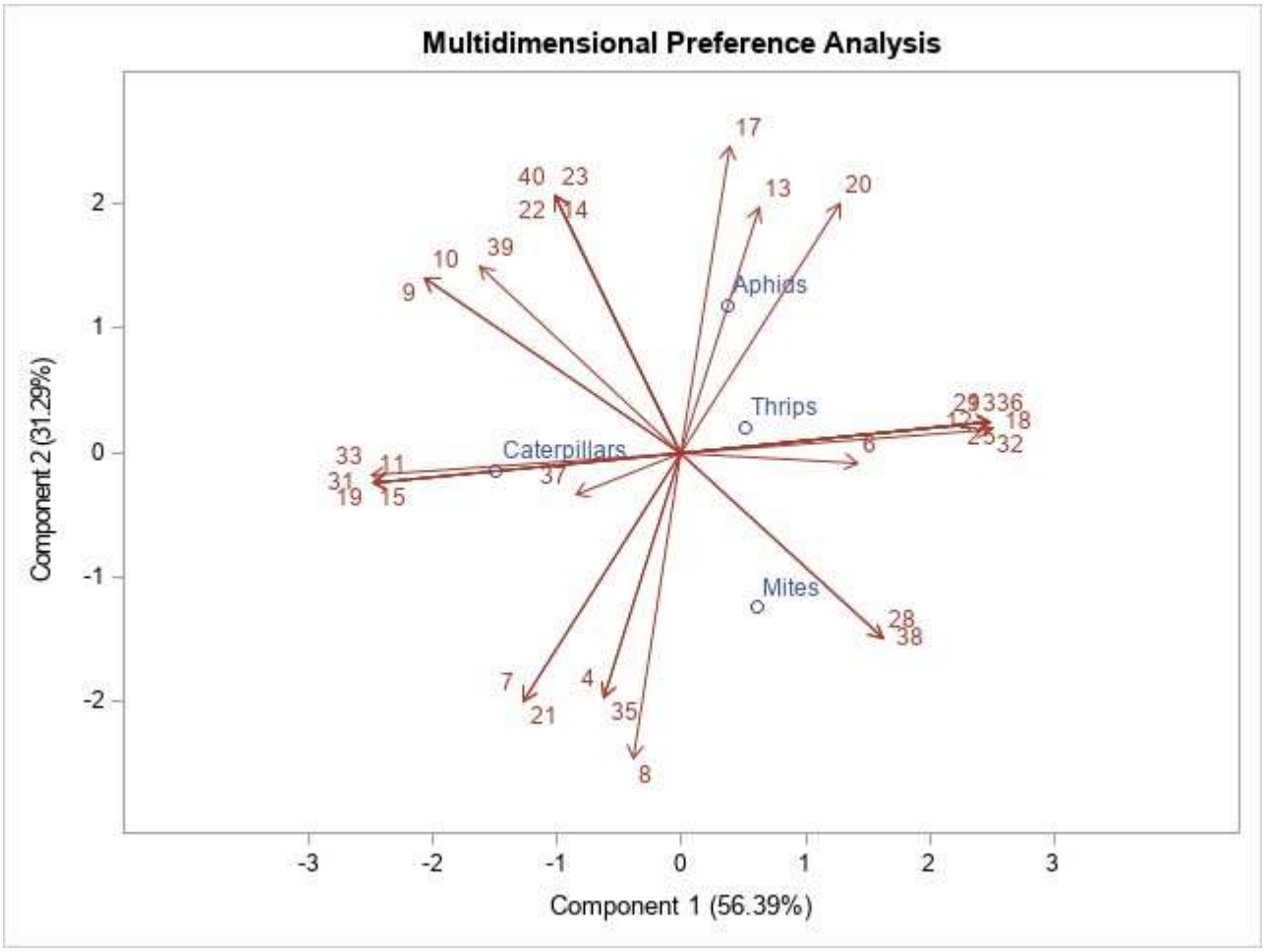


Figura 8. Biplot con las percepciones de los encuestados correspondientes a cada prioridad de investigación de la categoría insectos plagas de la estevia en Paraguay.

Para interpretar el biplot, se deben observar las regiones en el gráfico que contengan grupos de puntos con prioridades de investigación y determinar qué atributos tienen estas prioridades de investigación en común. Los puntos que están estrechamente agrupados en una región de la trama

representan prioridades de investigación que tienen los mismos patrones de preferencia entre los encuestados. Los vectores que apuntan aproximadamente en la misma dirección representan encuestados que tienen patrones de preferencia similares. En el biplot, las prioridades de investigación que se clasificaron por los encuestados como altas se encuentran a la izquierda del espacio, mientras que prioridades de investigación que fueron calificadas como con baja prioridad se ubican a la derecha. La mayoría de los encuestados consideraron enfermedades transmitidas por el suelo como un tópico de alta prioridad de investigación, especialmente *Damping off* y "White root rot" (pudrición blanca de la raíz) seguida por la enfermedad foliar "Septoria leaf spot" (mancha foliar por Septoria. Estos resultados coinciden con lo reportado por otros autores que mencionan que pudrición blanca de la raíz causada por *Sclerotium rolfsii* puede reducir el número de plantas y reducir el rendimiento (Koehler y Shew 2017). De igual manera la mancha foliar de Septoria, se observó en múltiples áreas de producción de estevia alrededor del mundo (Hastoy et al. 2019, Koehler y Shew 2018). Debido a que esta enfermedad causa una defoliación intensa y reduce significativamente los rendimientos, las estrategias de manejo de la mancha foliar de Septoria son de particular importancia en la producción de estevia. En contraste, "Phytonematodes" (nematodos), "Bacterial diseases" (enfermedades bacterianas), "Alternaria leaf spot" (la mancha foliar por Alternaria) y "TSWV" (virus de la marchitez del tomate) se clasificaron como enfermedades de menor prioridad para la estevia. Varios vectores apuntan hacia la esquina superior derecha de la gráfica, hacia una región sin prioridades de investigación. Esta es la región entre enfermedades causadas por virus y nematodos. Esto sugiere que varios investigadores consideran que estas enfermedades con baja prioridad para su investigación.

Modificando el código proveído para el análisis de enfermedades en estevia es posible analizar la percepción de los encuestados en las otras dos categorías restantes: insectos plaga y manejo de la estevia. En la Figura 8 se observan los resultados para la categoría insectos plaga. En este caso no se observa un patrón definido de preferencia, lo cual significa que el grupo encuestado en general no coincide en una plaga en específico como prioritaria y las opiniones se encuentran divididas. Los estudios disponibles sobre plagas de la estevia son escasos ya que los mismos parecen no ser una amenaza para el cultivo (Casaccia et al. 2016).

En cuanto a la categoría correspondiente a manejo de la estevia (Figura 9) las investigaciones menos prioritarias se encuentran a la derecha del plano con varios vectores apuntando a hacia el tópico "Breeding" (mejoramiento genético). Por otro lado, "Crop nutrition" (nutrición del cultivo), "Seed production" (producción de semillas), "Alternative managenement" (manejo orgánico) se encuentran a la izquierda del plano lo cual indica que la mayoría del grupo de encuestados consideran estos tópicos más prioritarios que "Breeding" (mejoramiento genético). En Paraguay existen dos variedades clonales de alto contenido de esteviosidos y rebaudiosidos, la variedad Eirete con mayor contenido en esteviósido y rebaudiósido A, seguida de la variedad Katupyry (Bogado et al. 2020). Es posible que debido a que estas variedades son altamente productivas, los investigadores consideran que no hay una urgencia de obtener nuevos materiales. Varios de los encuestados coincidieron que la producción de semillas es un área de investigación prioritaria, debido a que hasta el momento la producción de estevia se realiza mediante trasplante de plantines clonales lo cual eleva el costo de producción significativamente (Casaccia et al. 2016). También el desarrollo de tecnologías para el manejo orgánico de enfermedades parece ser considerado como tema de prioridad para los encuestados. Algunas alternativas utilizadas en la agricultura orgánica son los agentes de biocontrol, fungicidas a base de cobre y azufre. Sin embargo, estos productos no controlan significativamente enfermedades y plagas en comparación con los controles no tratados (Koehler y Shew 2017). Por lo tanto, es necesario estudiar enfoques

alternativos para el control de enfermedades en sistemas orgánicos.

El análisis multidimensional de preferencias permitió expresar la preferencia de los encuestados en forma visual (Carroll 1972). En lugar de mostrar números sin procesar, el biplot de escala multidimensional (Gabriel 1981) permitió mostrar las relaciones entre las variables; con áreas de investigación de importancia similar agrupadas juntas. Lo obtenido con este análisis ayudará a los investigadores, extensionistas y tomadores de decisiones a enfocar sus esfuerzos en las necesidades prioritarias para la producción de estevia y generar tecnologías que permitan maximizar su producción.

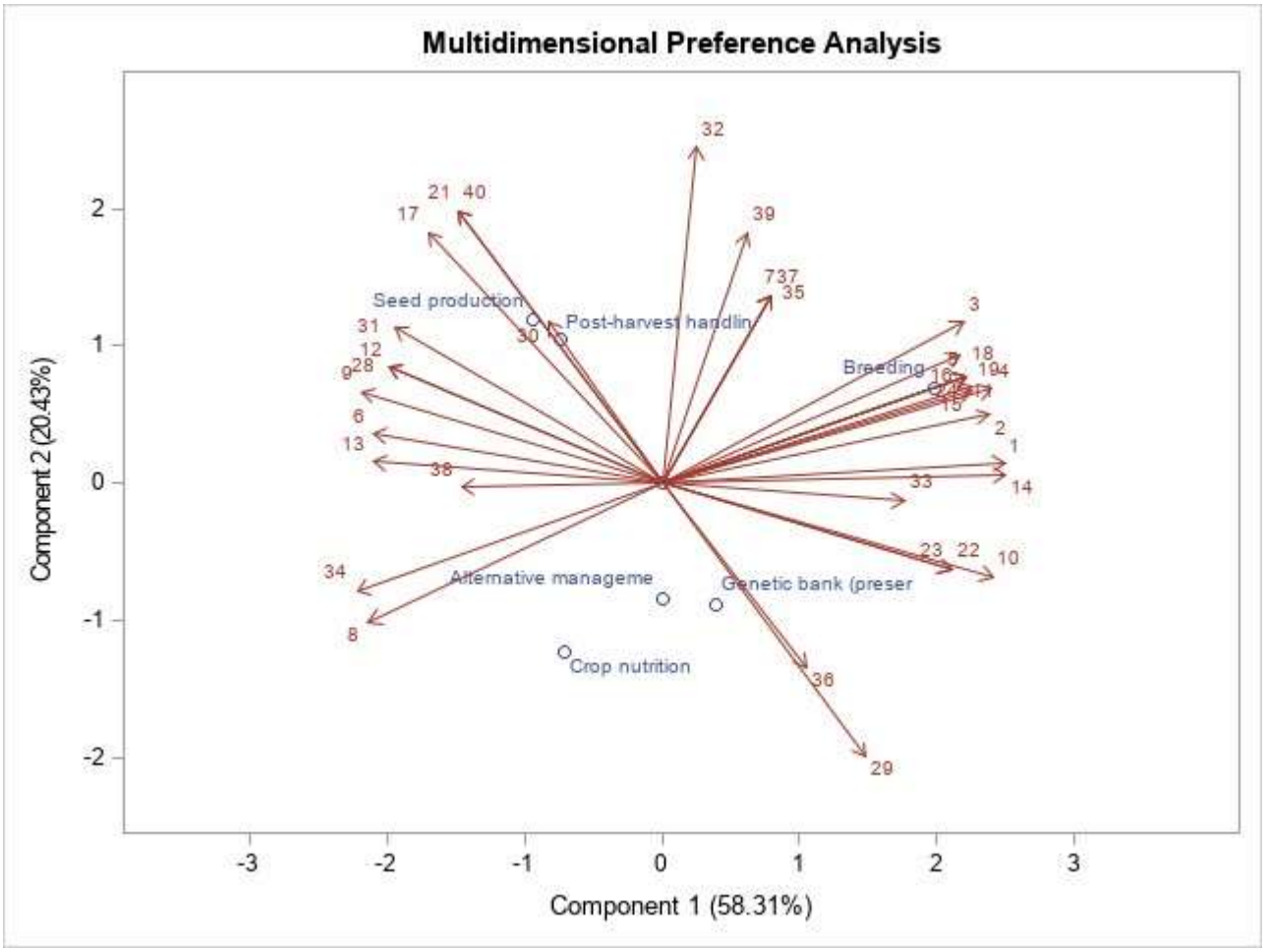


Figura 9. Biplot con las percepciones de los encuestados correspondientes a cada prioridad de investigación de la categoría manejo y mejoramiento de la estevia en Paraguay.

CONCLUSIONES

A través del análisis multidimensional de preferencias se identificó como tópico de alta prioridad de investigación a las enfermedades transmitidas por el suelo, con mayor atención al *Damping off*, "White root rot" (pudrición blanca de la raíz) seguida por la enfermedad foliar "Septoria leaf spot" (mancha foliar por Septoria). Mientras que las enfermedades causadas por virus y nematodos fueron consideradas de baja prioridad para su estudio. Por otro lado, la nutrición del cultivo, la

producción de semillas y el manejo orgánico se encuentran también son tópicos prioritarios de investigación, mientras que el mejoramiento genético actualmente no fue considerado una prioridad. Los insectos plagas fueron consideradas de baja prioridad investigación en el cultivo de estevia.

BIBLIOGRAFÍA

Bogado-Villalba, L., Nakashima, H.N., Britos, R., Iehisa, J.C.M. and Flores Giubi, M.E., 2020. Genotypic characterization and steviol glycoside quantification in a population of *Stevia rebaudiana* Bertoni from Paraguay. *Journal of Crop Science and Biotechnology*, pp.1-8.

Carroll, J. D. 1972. Individual Differences and Multidimensional Scaling. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, vol. 1*, edited by R. N. Shepard, A. K. Romney, and S. B. Nerlove, 105–155. New York: Seminar Press.

Casaccia, J., Rosanna Britos, R., Bozzano, G., Sanabria-Velazquez, A., & Cantero, F. (2016). Ka'a he'ẽ Stevia rebaudiana (Bertoni) Bertoni: La dulce planta de Paraguay para el mundo, alternativa para la diversificación de la finca. IPTA, CIHB, KOPIA, 2016. 116 p.

Hastoy, C; Bihan, Z Le; Gaudin, J; Cosson, P; Rolin, D; Schurdi-Levraud, V. 2019. First report of Septoria sp. infecting Stevia rebaudiana in France and screening of Stevia rebaudiana genotypes for host resistance (En genbank accession numbers for septoria sequences: mh352355-mh352389). :1-39.

Gabriel, K. R. 1981. Biplot Display of Multivariate Matrices for Inspection of Data and Diagnosis. In Interpreting Multivariate Data, edited by V. Barnett, 571-572. Chichester, UK: John Wiley & Sons.

Koehler, AM; Shew, HD. 2017. Enhanced Overwintering Survival of Stevia by QoI Fungicides Used for Management of Sclerotium rolfsii (en línea). Plant Disease 101(8):1417-1421. DOI: https://doi.org/10.1094/PDIS-02-17-0277-RE.

Koehler, AM; Shew, HD. 2018. Field efficacy and baseline sensitivity of *Septoria steviae* to fungicides used for managing Septoria leaf spot of stevia (en línea). Crop Protection 109(March):95-101. DOI: https://doi.org/10.1016/j.cropro.2018.03.006.

Ramesh, K., Singh, V., & Megeji, N. W. (2006). Cultivation of stevia [Stevia rebaudiana (Bert.) Bertoni]: A comprehensive review. *Advances in Agronomy*, *89*, 137-177.

SAS Institute Inc. 1999. SAS/STAT User's Guide, Version 8, SAS Institute Inc., Cary, NC.

Thomas, S. C. (2000). Medicinal plants: Culture, utilization and phytopharmacology. CRC press.

Método Estadístico para Evaluación de Convergencia en Algoritmo de Optimización No Lineal para Call Centers

Ángel Rubén Barberis
Universidad Nacional de Salta
Sede Central
Salta – Argentina
barberis@unsa.edu.ar

Lorena E. Del Moral Sachetti
Universidad Nacional de Salta
Sede Regional Orán
Orán – Salta. Argentina
lorena.dms.7@gmail.com

RESUMEN

La optimización de Call Centers no es un problema fácil de resolver, debido a la complejidad de los modelos matemáticos que derivan de las fórmulas de Erlang. Esta complejidad se traslada a modelos de optimización, que en la mayoría de los casos, se conforman con funciones objetivos no lineales y no derivables. Así como en todas las áreas de la Investigación de Operaciones, resolver estos problemas demanda algoritmos eficientes, rápidos y precisos. La Simulación como herramienta experimental constituye un ambiente esencial para la validación de algoritmos de optimización, sobre todo cuando no se dispone de repositorios de problemas bien definidos con métricas de resultados conocidos con los que se pueda contrastar. En este trabajo se describe una estrategia combinada con simulación estocástica para estudiar la convergencia estadística de algoritmos de optimización no lineal entera en el estudio de los problemas de Call Centers.

PALABRAS CLAVES

Optimización de Call Center, Optimización No Lineal Entera, Convergencia de Algoritmos no lineales.

1. INTRODUCCIÓN

La optimización de recursos operativos en los problemas de Call Centers constituye un proceso difícil de resolver con precisión, debido principalmente a la complejidad de los modelos de Erlang, sobre todo cuando se intenta aproximar éstos a la realidad. Si bien, las formulaciones matemáticas para los modelos de Colas Erlang-A (M/M/*n*+G) están disponibles, éstas son demasiadas complicadas para derivar de ellas soluciones analíticas y algorítmicas para problemas de Call Centers. Por lo que, los investigadores se enfrentan a una tarea difícil al resolver cuestiones como la estimación de la cantidad adicional de operarios cuando el volumen de llamadas entrantes se duplican, o la determinación de la sensibilidad del modelo cuando subyacen errores en la estimación de la paciencia [1], entre otros.

La administración óptima de Call Center persigue, principalmente, dos objetivos contrapuestos: 1) *Dimensionamiento* con el menor costo en la contratación de personal (menor cantidad de personal, menor nivel de servicio); y 2) *Maximización de los niveles de servicio*, que se traduce en encontrar la política de asignación de turnos que conduzca al mayor nivel de la satisfacción del usuario (mayor nivel de servicio, implica mayor número de personal a contratar). Para el primer objetivo, se resuelven problemas de optimización lineal entera. En la literatura hay un amplio desarrollo de investigaciones y algoritmos eficientes sobre la temática [2-5]. Alcanzar el segundo objetivo, implica resolver un problema de optimización no lineal entero, para el cual, no existe un algoritmo general que pueda aplicarse en la búsqueda de la solución, por lo que se

recurre al desarrollo de algoritmos específicos. En la literatura, se registran trabajos en el que se realiza un abordaje conjunto, priorizando la solución al primer objetivo, y en segundo lugar, realizan una aproximación de la solución en la búsqueda del segundo objetivo [6, 7]. Los inconvenientes matemáticos dificultan el desarrollo de procesos analíticos generales que puedan demostrar convergencia de algoritmos de optimización no lineales como consecuencia de la diversidad de enfoques con que se abordan la problemática. La dificultad es aún mayor al no contar, en esta área de investigación, con repositorios de problemas bien definidos, con métricas de resultados conocidos con los que se pueda contrastar. En tanto, otros utilizan la simulación como una herramienta de comparación, apoyados con intervalos de confianza [8].

En los problemas de Call Centers, es muy común el uso de simulación, dado que es una excelente herramienta de soporte en la toma de decisiones [9-11], y muy usado en la comprobación tecnológica para la planificación [12].

Con el objeto de contar en nuestras investigaciones con una herramienta de validación de los procesos de optimización no lineales, se recurre a la estadística para extraer de ella, algunos conceptos como el Análisis de Residuales y la Regla Empírica 68-95-99. Adicionalmente, se incorpora al análisis la evaluación de curtosis para datos muestrales que siguen una distribución normal, que combinados con una herramienta de simulación permiten especificar un método estadístico para evaluar el grado de convergencia de una serie numérica generada por un algoritmo. En el presente trabajo, se describe una estrategia para validar resultados y convergencia obtenidos a partir de algoritmos de optimización, denominada Validación Empírica Residual, que compara estadísticamente los resultados obtenidos por un algoritmo con los obtenidos por simulación, y bajo ciertas premisas de variabilidad estadística se puede concluir si el algoritmo es o no convergente. Este procedimiento se describe en la sección 2. En la sección 3 se presenta un caso práctico de validación de convergencia de un algoritmo de optimización no lineal entera aplicado al problema de Call Center. Finalmente, en la sección 4, se expone las conclusiones.

2. METODOLOGÍA: VALIDACIÓN EMPÍRICA RESIDUAL

2.1 Análisis de Residuales

El *Análisis de Residuales* es la herramienta que se utiliza para evaluar la idoneidad de un modelo de regresión lineal frente a un conjunto de datos experimentales mediante la definición de residuos y la comprobación de supuestos estadísticos a través del análisis de los gráficos de residuales [13-15].

Los principales supuestos estadísticos respecto del modelo de regresión que se comprueban con el gráfico de residuales son [14, 16-18]:

1) La generación de residuos debe tener un comportamiento aleatorio y ser independientes e idénticamente distribuidas.

2) Todos los residuales deben tener igual varianza. Esto quiere decir, que la varianza debe ser constante en todo el rango de concentración dinámica de los residuales. Esta propiedad se conoce como *homocedasticidad*.

3) Todos los residuos son variables aleatorias con distribución (aproximadamente) normal con media 0, por lo que la esperanza residual debe ser 0.

La verificación de las propiedades como la de ser independiente e idénticamente distribuida, conduce al cumplimiento de otros supuestos implícitos [19].

Comprobada la *independencia* residual implica el cumplimiento de supuestos como: ***a***) los residuales no se correlaciona con ninguna otra variable incluida o no en el modelo; ***b***) Los residuos no están agrupados (es decir, las medias muestrales de cualquier conjunto de residuos son todas iguales); ***c***) Los residuos no están auto correlacionados (es decir, no existe autocorrelación temporal o espacial).

Por otro lado, decir que los residuos son *idénticamente distribuidos*, significa que: ***a***) Todos los residuos provienen de la misma distribución. En el caso de una regresión lineal, se asume que todos vienen de una misma distribución normal; ***b***) La varianza residual es homogénea, es decir, se cumple la propiedad de homocedasticidad; ***c***) La media de los residuos es cero en todo el rango de los valores predictores. Cuando los predictores numéricos (covariables) están presentes, implica que la relación entre la variable independiente y dependiente puede describirse adecuadamente mediante una línea recta en el plano.

La diferencia entre los valores observados de la variable dependiente (y) y los valores que predice el modelo ($\hat{y}$) se llama *residual* (e) [20]. También se los conoce como *error*. Es decir:

Residuos (e) = Valores observados (y) - Valores que predice el modelo ($\hat{y}$) $\Rightarrow$ $e = y - \hat{y}$

Donde $\sum e_i = 0$ y $\overline{e} = 0$ siempre que se cumpla los supuestos estadísticos.

En un gráfico residual se muestran los residuos en el eje vertical y la variable independiente en el eje horizontal. Cuando los puntos en el gráfico se dispersan aleatoriamente alrededor del eje horizontal, se considera que el modelo de regresión lineal es apropiado para los datos. Una manera rápida de verificar el supuesto de normalidad es analizar la tendencia del gráfico. Si los residuos se trazan aproximadamente a lo largo o sobre de la línea recta de la tendencia, como en la figura 1, entonces se cumple el supuesto de normalidad [21], lo que implica de inmediato satisfacer los supuestos de implicancia mencionados anteriormente. Desde el punto de vista conceptual, el análisis de residuales conforma una herramienta sencilla que facilita la detección errores y la comprobación del grado de aproximación entre una ecuación empírica o teórica con resultados experimentales [22].

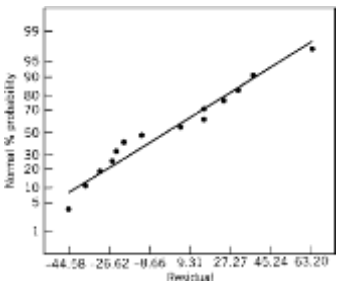


Figura 1: Residuales con distribución normal

2.2 Regla Empírica o Regla 68-95-99

Usar una estimación central como la media o la mediana junto con una medida de variación (como la desviación estándar o el rango intercuartil) en una distribución muestral es una buena manera de describir los valores o el comportamiento de una población. En el caso de que el histograma de frecuencias relativa de los datos tenga la forma de campana (o tiene una aproximación a la distribución normal), la media poblacional y la desviación estándar son la combinación adecuada para estudiar variabilidad o dispersión, y una regla especial los vincula para obtener información bastante detallada sobre la población en general. Esta regla es la llamada *regla empírica*, también conocida como regla 68-95-99.7 [23].

La regla empírica es una regla general que se usa para indicar el porcentaje aproximado de valores muestrales que se encuentran dentro del intervalo dado por las desviaciones estándares respecto de la media muestral, cuando éstos se distribuyen normal [16]. La regla se aplica, generalmente, a una variable aleatoria que sigue una distribución normal, con media μ y desviación estándar σ. Una característica importante expresa que, si la distribución de los datos es más o menos simétrica, unimodal y sigue una ley normal o una aproximación a ella, entonces aproximadamente el 68,27 % de los datos se concentran dentro del rango $\mu \pm \sigma$; el 95,45 % dentro de $\mu \pm 2\sigma$, y 99,73 % dentro de $\mu \pm 3\sigma$. La figura 2 ilustra las tres componentes de la regla 68-95-99. Un punto importante a tener en cuenta, es que la regla empírica no se aplica a conjuntos de datos con distribuciones muy asimétricas.

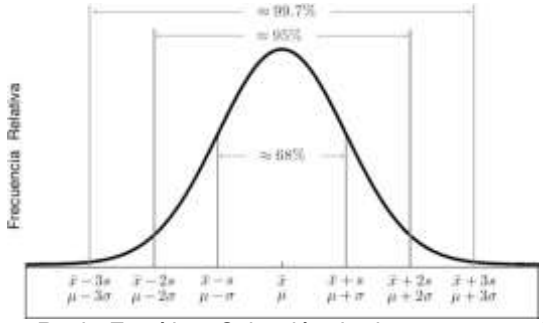


Figura 2: Regla Empírica. Colección de datos porcentuales según la distancia respecto de la media.

2.3 Simulación

La simulación es realizada mediante un simulador de Call Center desarrollado a los efectos de la investigación, cuyo modelo estocástico ha sido debidamente comprobado en [24], cuyo proceso se muestra en tabla 1.

Tabla 1. Esquema del proceso de simulación

Simulación(X, Parámetros, Cant_Jornadas):

X: vector de enteros que contiene la grilla de asignación de agentes para una jornada.
Parámetros: vector de reales que contiene los parámetros preestablecidos requeridos por el modelo de Call Center, como ser NS objetivo, tasa de abandono de llamadas, etc.
Cant_Jornadas: Es la cantidad de jornadas diarias que se quiere simular.

Hacer NS = 0; J = 1
Mientras J ≤ Cant_Jornadas, ***hacer***

- ***Para*** c/intervalo de observación, y hasta completar una jornada diaria, ***hacer***
  - Asignar agentes planificados para el intervalo, si corresponde de acuerdo a X.
  - Determinar cant. de agentes que descansan en el intervalo según planificación y desafectarlos.
  - Generar aleatoriamente llamadas telefónicas entrantes y poner en la cola de espera.
  - Si la cola no está vacía, atender las llamadas con agentes disponibles.
  - Desafectar agentes que hayan cumplido su turno si corresponde según X.
  - Calcular nivel de servicio del intervalo de observación y acumularlo en NS.
- Calcular para la jornada simulada, el promedio del nivel de servicio haciendo NS = (NS/Cant. de intervalos de observación de la jornada).
- Resguardar X en X_old
- Con las llamadas telefónicas entrantes de cada intervalo de observación de la jornada simulada determinar cantidad óptima de agentes usando como punto de partida a X y actualizarlo con el óptimo. Si no se tuvo éxito, restablecer X haciendo X = X_old.
- Estimar nivel de servicio teórico NS_teorico = NS(X).
- Determinar (si es posible) un punto vecino de X del espacio de decisión que mejore NS_teorico y actualizar X. Si no se pudo, hacer X = X_old.
- Acumular NS en NS_parcial.
- Hacer NS = 0; J = J + 1

Devolver: X , (NS_parcial / Cant_Jornadas)

La particularidad de la herramienta es que implementa un dispositivo que permite cumplir dinámicamente la política de asignación de turnos pre-establecida. Este mecanismo no previsto en software de simulación de Call Centers comerciales, habilita y deshabilita cierta cantidad de agentes según la planificación de turnos estipulados, a medida que transcurre la simulación de una jornada laboral. De esta manera, las simulaciones que se realizan son muy cercanas a la realidad operativa de los Call Centers, lo que posibilita formular conclusiones satisfactorias y precisas respecto de la realidad objetiva.

La estocasticidad de las simulaciones por computadora depende de la generación de números pseudo-aleatorios. En ejecuciones sucesivas de simulación, la generación algorítmica de valores pseudo-aleagoritos generan resultados que violan los supuestos estadísticos del análisis de residuales, es decir, se generan estimaciones correlacionados, que, como variables aleatorias, no serían independientes ni idénticamente distribuidas [25]. Para garantizar el proceso estocástico y evitar la problemática generada por la pseudo-aleatoriedad se puede recurrir a dos alternativas para generar muestras que superen test estadísticos de aleatoriedad e independencia:

1) Realizar n ejecuciones distintas de simulación y extraer de ellos un subconjunto de m ($m < n$) resultados de la variable de estudio, preferentemente los de mayor variabilidad. Esto dará lugar a m observaciones. O bien,

2) Realizar n grupos de m ejecuciones distintas de simulación, con n semillas distintas para la generación de pseudo-aleatorios (cada grupo utiliza la misma semilla para las m simulaciones). Se conforma una matriz de tamaño $n \times m$. Luego, se diseña una nueva muestra aleatoria

promediando los valores de cada columna de la matriz, tal como se muestra en la figura 3, generándose así, *m* valores observados de simulación, con características de ser independientes.

De cualquiera de las dos alternativas se obtienen *m* observaciones independientes, idénticamente distribuidas, y aproximadamente normal, muy convenientes para la generación de residuos.

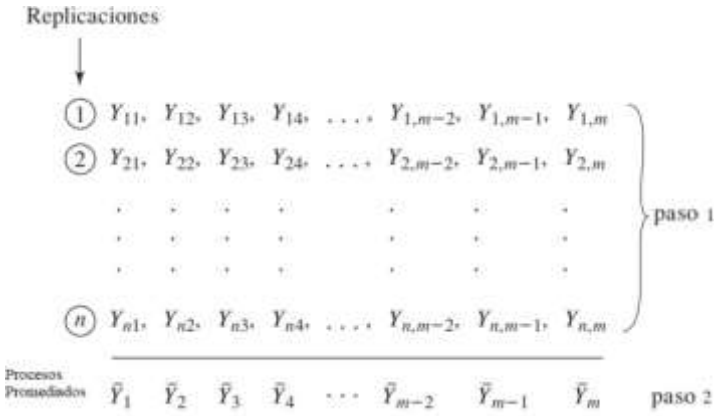


Figura 3: Proceso de selección de muestras aleatorias.

2.4 Proceso de Validación Empírica Residual

Se busca saber si el algoritmo de optimización ejecutado con diferentes puntos iniciales del espacio de decisión converge a un mismo valor, y si éste es próximo o no a la solución del problema en estudio. Las conclusiones se desarrollan a partir de las comparaciones estadísticas entre las salidas de un proceso de simulación y los resultados deterministas obtenidos por el algoritmo objetivo.

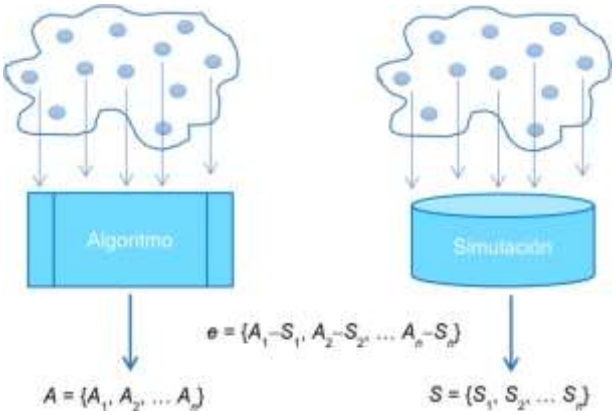


Figura 4: Formación de la muestra de residuales

El procedimiento consiste en generar las observaciones a partir de los resultados del algoritmo que se obtienen de *m* ejecuciones con *m* puntos diferentes de partida del espacio de decisión. Los procesos de simulación se inician también con los *m* puntos diferentes usados por el algoritmo, ver esquema de la figura 4. Si A_i es el valor observado que se obtiene como resultado del algoritmo ejecutado en el *i*-ésimo experimento, y S_i la estimación de la misma variable de estudio obtenido por simulación en un proceso *i*-ésimo, entonces los residuos $r_i = A_i - S_i$ para $i = 1 \ldots m$, conforman la variable aleatoria de residuales *R*. Así, si $A_i \rightarrow \mathrm{E}(S_i)$ entonces $E(R) \rightarrow 0$. Luego de generar la muestra, se realiza el análisis de residuales para asegurar el cumplimiento de los supuestos estadísticos. El incumplimiento de algunos de los supuestos, implica aumentar los tiempos de simulación y cambiar la alternativa de selección de resultados del proceso simulado. Si volviera a no cumplir algunos de los supuestos estadísticos del análisis de residuales, entonces se rechaza la hipótesis de que el algoritmo converge a la solución del problema en estudio. En el caso de verificar todos los supuestos, se considera que el algoritmo tiene una tendencia a converger a un

punto límite. El paso siguiente es analizar el grado de concentración de los residuos alrededor de la medida central. La estimación se obtiene a partir del indicador de Curtosis, que señala el grado de apuntamiento o achatamiento de la distribución muestral [26]. En el análisis, se compara la forma de la curva de la distribución muestral con la distribución normal estándar. Así, si el indicador de curtosis es mayor a 0, se tendrá una distribución muestral Leptocúrtica; si es igual a 0

será mesocúrtica, (indica que se tiene de una distribución normal); en cambio, si el indicador es menor a 0 se tendrá una distribución Platicúrtica [27, 28]. Para una aplicación exitosa de la regla empírica, es deseable un indicador de curtosis mayor o igual a cero, lo que dará lugar no sólo a una validación exitosa de la convergencia del algoritmo, sino también a una convergencia con un grado de precisión aceptable.

Satisfecha la condición de curtosis, se estudiará los resultados derivados del Desvío Estándar (σ) en sus tres valores: σ, 2σ y 3σ , según la *regla empírica 68-95-99*. Es deseable que el 100% de los resultados algorítmicos estén dentro de un rango de aproximación aceptable alrededor de la solución. Al tratarse de un estudio estadístico, se buscará que al menos un cierto porcentaje de los resultados algorítmicos tengan una precisión aceptable, lo que implica fijar una cota superior a 2σ ó 3σ, que dependerá del tipo de problema al que se quiere dar solución. Así por ejemplo, para el caso de estudio que se describe en la sección 3.2, en el que se busca determinar la política de distribución óptima de turnos laborables para los operarios de un Call Center, que maximicen los niveles de servicios (NS), se requiere que 2σ ó 3σ sea menor a 1 – puesto que, el NS se cuantifica en el rango real (0, 1] – para asegurar que al menos el 95% de los resultados algorítmicos estén dentro del rango de precisión.

3. DESARROLLO

3.1 El problema a resolver

En el ámbito de los Call Centers, no hay soluciones algorítmicas que hayan obtenidos resultados óptimos claramente; ni mucho menos, hay consenso en cuanto a la mejor estrategia de implementación. Es por ello, que en la investigación que se desarrolla se trata de diseñar alternativas algorítmicas que buscan optimizar las políticas de planificación y distribución de turnos, desde la perspectiva de la Optimización No Lineal Entera. Así, el problema a resolver es de la forma:

$$\begin{aligned} & máx\ f(x) \\ & s.a. \quad Ax \geq r\ ; \quad x \geq 0;\ x \in \mathbb{Z}^n \\ & \qquad Bx = \Omega; \end{aligned} \qquad (1)$$

Donde $f(x)$ es la función objetivo no lineal con dominio sobre un espacio de decisión discreto, que mide el desempeño del Call Center en término de nivel de servicio (NS). La función es no derivable, no convexa y no configura un problema cuadrático. Las restricciones del problema son lineales convexa y se componen de una matriz $A \in \mathbb{Z}^{m \times n}$, y vectores B, $r \in \mathbb{Z}^m$, y $\Omega \in \mathbb{Z}$.

El algoritmo seleccionado para el experimento es descripto en [29] y tiene la particularidad de ser simple en su diseño y obtener resultados a una velocidad aceptable. En el pseudo-código que se muestra en la tabla 2, $F(x)$ es una función de penalización de la forma $F(x) = f(x)+P(R(x))$ donde $f(x)$ es la función objetivo, $R(x)$ constituye las restricciones de (1) y $P(r)$ es la función que devuelve 0 si x cumple las restricciones del problema, y un valor ponderado negativo en caso de que no las cumpla. También se requiere ajustar los valores de las componentes reales de $x^{(k+1)}$ al vector de enteros más próximo. La tarea de conversión es llevada a cabo por la función *Ajustar en Entero*(v) de la línea 10. Se trata de un proceso complejo que busca el vector integral más próximo al vector real que mejor evalúa a $F(x)$.

Tabla 2. Pseudo-código del algoritmo que resuelve el problema (1).

Algoritmo: Procedimiento básico de optimización con búsqueda direccional

0. Sea $x^{(0)}$ un punto inicial del espacio de decisión.
1. Hacer $x^{(k+1)} = x^{(0)}$
2. Actualizar iteración *k*-ésima.
3. Hacer $x^{(k)} = x^{(k+1)}$
4. **Para** $i = 0, 1, 2, \ldots |D|$
5. | Sea d_i vector direccional de D.
6. | Calcular $\alpha^* = \arg máx\ h(\alpha) = F(x^{(k)} + \alpha \cdot d_i)$
7. | Actualizar $x^{(k+1)} = x^{(k)} + \alpha^* d_i$
8. Si $F(x^{(k+1)}) > F(x^{(k)})$ **volver al paso 2**.
9. Sea $x^a = x^{(k+1)}$
10. Sea $x^{(k+1)} =$ *Ajustar en Entero*(x^a)
11. Si $F(x^{(k+1)}) > F(x^a)$ **volver al paso 2**.
12. Retornar $x^{(k+1)}$

3.2 Resultados Computacionales

La validación del algoritmo, tiene por objetivo dar confiabilidad en la estabilidad y la convergencia hacia un resultado satisfactorio, considerando cierta precisión. El método consta de 7 pasos, de los cuales, los 5 primero corresponde al estudio de la convergencia hacia un punto límite; y los 2 últimos, al estudio de la proximidad del punto límite a la solución del problema de optimización.

Para tener mayor claridad en los experimentos, se muestra en detalle el análisis para un problema de dimensión 6 correspondiente a datos de un Call Center de pequeña envergadura, cuyos parámetros y solución son conocidos. Los parámetros utilizados son iguales tanto para el algoritmo como para el simulador, y se corresponde con: una jornada laboral de 9 hs.; turno de trabajo de 4 hs.; período de observación de 1 hora; AHT de 180 segundos; AWT de 20 segundos; media de abandono igual a 1.5; reintentos de llamadas del 60%, y un nivel de servicio objetivo del 95%. Con estos datos se procede a realizar los experimentos algorítmicos y los de simulación.

1) Asegurar que la muestra de residuales sea independiente e idénticamente distribuida.

Tabla 3. Residuos de 10 experimentos con puntos de partidas diferentes

Experimento	X (de partida)	NS-Algoritmo	NS-Simulación	Residuos
1	5, 1, 2, 2, 6, 3	0,988567835	0,988556579	1,1256E-05
2	5, 2, 1, 2, 6, 3	0,988595998	0,988556676	3,9322E-05
3	5, 3, 0, 5, 3, 3	0,988545779	0,988556877	-1,1097E-05
4	5, 4, 0, 7, 0, 3	0,988569574	0,988556239	1,3334E-05
5	5, 5, 3, 2, 1, 3	0,988555999	0,988557011	-1,0118E-06
6	5, 6, 2, 0, 3, 3	0,988556784	0,988555922	8,623E-07
7	5, 7, 2, 0, 2, 3	0,988549359	0,988556722	-7,3629E-06
8	5, 8, 1, 0, 2, 3	0,988565779	0,988556658	9,1206E-06
9	5, 9, 1, 1, 0, 3	0,988549496	0,988557104	-7,6083E-06
10	5, 5, 0, 6, 0, 3	0,988556779	0,988556848	-6,89E-08
			$\sum e =$	4,6746E-05
			$\overline{e} =$	4,6746E-06
			$\sigma =$	1,404909E-05
			Durbin-Watson=	2,21663062
			Curtosis =	0,445003077

En el experimento de simulación, se realizan 10 grupos de 20 réplicas (ejecuciones) independientes de una simulación de 5.000 jornadas. Se extraen de cada grupo 10 réplicas de mayor variabilidad y se promedian sus resultados. La tabla 3 muestra las 10 observaciones

obtenidas como resultados del algoritmo y del experimento de simulación. La variable de estudio es NS y representa el valor de *f* (*x*) en (1) al ser evaluado en el punto del espacio de decisión. El *vector de planificación* inicial se representa en la tabla con la columna X, y representa el punto factible a partir del cual debe iniciarse la sucesión de aproximación. La comparación de la variable de estudio NS se realiza entre el modelo determinístico del algoritmo y el modelo estocástico de la simulación. Los resultados residuales se muestran en la última columna.

Suponiendo que los resultados dados por el algoritmo son comparables con los obtenidos por simulación, entonces el análisis de residuales debe mostrar que se cumplen los siguientes supuestos:

2) *Los residuales están centradas en torno al eje de abscisas (para verificar el supuesto de normalidad)*.

El supuesto de normalidad se verifica con el gráfico de la figura 5, que muestra la cercanía y la distribución de los residuos muestrales alrededor del eje de abscisa.

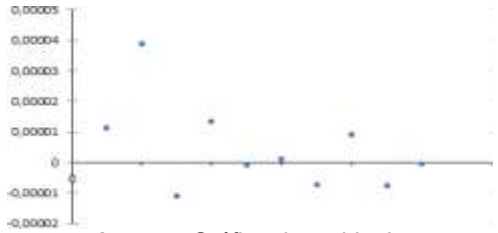


Figura 5: Gráfico de residuales

Del análisis gráfico, se deduce el cumplimiento del supuesto de normalidad.

3) *La distribución de los residuales sigue la ley normal de probabilidad (la sumatoria de los residuales y el promedio muestral deben tender a cero)*.

La tabla 3 muestra que $\sum e_i \to 0$ y $\overline{e} \to 0$, y verificado el supuesto de normalidad con el gráfico de residuales, se acepta que la muestra tiene una distribución aproximadamente normal con media muestral tendiente a cero.

4) *Los residuales no presentan autocorrelación, es decir, la muestra se conforma con residuos independiente entre sí*.

Para verificar que estadísticamente no hay auto-correlación (independencia entre los valores) se usará el indicador de Durbin-Watson [30], cuya expresión se muestra en la figura 6, que toma valor 2 cuando los residuos son completamente independientes, desde el punto de vista teórico.

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}; \quad 0 \le DW \le 4$$

Figura 6: Indicador de Durbin-Watson

Si *DW*< 2 indica auto-correlación positiva, y *DW* >2 se trata de una autocorrelación negativa. En la práctica cuando *DW* se encuentra entre 1.5 y 2.5 se considera que existe independencia. Para ello, se compara el estadístico *DW* con valores críticos según el nivel de significancia α. Los valores se encuentran tabulados para valores límites d_L y d_U [31]. Para determinar el criterio de decisión, se toma en cuenta el estadístico *DW* calculado en la tabla 3. Esto es *DW* = 2,21663062. Para obtener los valores críticos d_L y d_U con un nivel de significancia del 5%, son necesarios el número de muestras *n* = 10 y la cantidad de variables independientes *k* = 1.

Los valores críticos obtenidos de la tabla Durbin-Watson con 5% de significancia son:

$d_L = 0{,}879 \Rightarrow 4 - d_L = 3{,}121$ $\qquad$ $d_U = 1{,}320 \Rightarrow 4 - d_U = 2{,}68$

Se contrasta la hipótesis nula H_0 = *Los residuos no están auto-correlacionados* contra H_a = Los residuos están auto-correlacionados. La figura 7, muestra el área de no rechazo de H_0.

Como $DW \in [d_U ; 4 - d_U] = [1{,}320; 2{,}68]$ no se rechaza la hipótesis nula H_0. Con lo cual, se verifica el supuesto de independencia entre los residuos muestrales.

Figura 7: Criterio de Decisión según [32].

5) ***La dispersión de los residuales es constante (homocedasticidad).***

Para demostrar la homocedasticidad se usa del *Test de Bartlett* [33]. Éste define la hipótesis nula, H_0 que las varianzas de *k* muestras independientes de una población son iguales, frente a la hipótesis alternativa de que al menos dos son diferentes. El estadístico se muestra en la figura 8.

El estadístico de prueba *T* tiene una distribución aproximadamente χ^2_{k-1}. Por lo que, la hipótesis nula se rechaza si $T > \chi^2_{k-1,\alpha}$ con un grado de significancia α.

$$T = \frac{(N-k)\ \ln(S_p^2) - \sum_{i=1}^{k}(n_i-1)\ \ln(S_i^2)}{1+\frac{1}{3(k-1)}\left\{\left(\sum_{i=1}^{k}\frac{1}{n_i-1}\right)-\frac{1}{N-k}\right\}}$$

$$donde\ \ N=\sum_{i=1}^{k} n_i \quad ; \quad S_p^2 = \frac{\sum_{i=1}^{k}(n_i-1)\ S_i^2}{N-k}$$

Figura 8: Expresión del Test de Bartlett

Teniendo en cuenta los valores residuales de la tabla 3, se forman 6 grupos diferentes de 5 elementos residuales cada una, seleccionados al azar. Los grupos se muestran en la tabla 4. Se calculan las varianzas respectivas de cada grupo y se los utiliza para calcular el estimador de Bartlett.

Tabla 4: Selección de grupos diferentes para la evaluación de homocedasticidad.

	M_1	M_2	M_3	M_4	M_5	M_6
1	1,1256E-05	-1,10978E-05	1,1256E-05	3,93222E-05	1,1256E-05	3,93222E-05
2	3,93222E-05	-1,0118E-06	-1,1098E-05	1,33346E-05	-1,1098E-05	-1,0118E-06
3	1,33346E-05	-7,3629E-06	-1,0118E-06	8,623E-07	1,33346E-05	8,623E-07
4	8,623E-07	-7,6083E-06	-7,3629E-06	9,1206E-06	-7,3629E-06	-7,6083E-06
5	9,1206E-06	-6,89E-08	-7,6083E-06	-6,89E-08	-6,89E-08	9,1206E-06
Var(M_i)	**2,1064E-10**	**2,22167E-11**	**7,8224E-11**	**2,5624E-10**	**1,18635E-10**	**3,39505E-10**

El estadístico de Bartlett que se estima es $T = 6{,}8778$, y el valor crítico con significancia del 5% que es lo habitual es $\chi^2_{5;\,0.05} = 11{,}0705$. Dado que $T < \chi^2_{5;\,0.05}$ entonces no se rechaza la H_0, por lo que el supuesto de homocedasticidad queda verificado, y el supuesto de normalidad completamente comprobado.

De esta manera, se verifica el cumplimiento de todos los supuestos del Análisis de Residuales. Por lo que se puede concluir que el algoritmo converge estadísticamente a un punto límite orientado hacia la solución del problema (1).

6) *Análisis de Curtosis*

El paso siguiente consiste en analizar el grado de precisión de la convergencia. Para ello, se procede a medir el grado de apuntamiento o achatamiento de la distribución aproximadamente normal de la muestra de residuales, lo que se logra con el indicador de curtosis. La tabla 3 muestra el cálculo del indicador de curtosis según [34], cuya expresión se muestra en la figura 9.

$$C = \frac{\sum \left(X_i - \bar{X}\right)^4 / n}{\left[\sum \left(X_i - \bar{X}\right)^2 / n\right]^2} - 3 \;\; = 0{,}445003077$$

Figura 9: Indicador de Curtosis (DeCarlo, 1997) calculado según tabla 3.

Al ser C >0 muestra que se trata de una distribución con curtosis positiva del tipo leptocúrtica que es la deseada. Por otro lado, algunos investigadores expresan que los momentos muestrales m_r no son estimaciones insesgadas de los momentos poblacionales μ_r, y proponen otra expresión para el indicador de curtosis que es insesgado para distribuciones normales [35]. La expresión se muestra en la figura 10, donde *S* denota el desvío típico. Teniendo en cuenta la muestra de residuales de la tabla 3, el estimador de curtosis insesgado es *K* = 2,878598261 que es ampliamente mayor a cero (*K* >0).

$$K = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Figura 10: Indicador de Curtosis de Joanes.

Ambos indicadores de curtosis prueban que la distribución muestral es del tipo leptocúrtica. Esto significa que la mayoría de las observaciones residuales se ubica en un entorno cercano a la esperanza *E*(*e*) = 0.

7) Análisis por Regla Empírica → grado de precisión

Teniendo en cuenta el resultado anterior, y por aplicación de la regla empírica *68-95-99*, se tiene que el 99,7 % de las observaciones se concentran en el rango ya sea en [$\bar{x}$ −3σ; $\bar{x}$ +3σ] o bien en [μ −3σ; μ +3σ]. Por lo que, para estimar estadísticamente el grado de proximidad del punto límite a la solución del problema de optimización es necesario fijar una cota a 3σ. Si la muestra residual genera un valor de 3σ mayor que dicha cota entonces se concluiría que el punto límite a la que converge el algoritmo no tiene la proximidad deseada a la solución del problema objetivo. De lo contrario, se acepta al punto límite como una solución aproximada del problema de optimización. Dado que NS − calculado a través de *f* (*x*) en el problema (1) − es un indicador porcentual que se miden en el rango (0, 1], entonces se establece que 3σ < 1. En el caso de estudio que se está desarrollando, se observa en la tabla 3 que σ = 0,00001404909, por lo que el valor de 3σ = 0.00004214727, siendo mucho menor que 1. Así, se acepta que el punto límite a la que converge el algoritmo es próximo a la solución del problema.

Un resultado curioso y puramente experimental se pudo observar que la magnitud del error de aproximación del resultado del algoritmo es del mismo orden de magnitud del desvío estándar (σ). En el experimento σ = 0,00001404909, por lo que, aproximadamente, el resultado final del algoritmo tiene al menos 4 dígitos de precisión, que se considera muy satisfactorio.

4 CONCLUSIONES

El análisis de Residuales es un proceso estadístico que sirve para estudiar características de aproximación en modelos de regresión. En este trabajo, se muestra a dicho proceso como un recurso clave en el estudio de la convergencia de algoritmos de optimización. La combinación del indicador de curtosis con la regla empírica 68-95-99 posibilitó estimar el grado similitud entre los resultados obtenidos por el algoritmo y los de simulación. La técnica también fue aplicada a problemas más grandes, de hasta dimensión 1000, usando el mismo algoritmo mostrado en la sección 3.1, en el que es difícil conocer la solución global, por lo que se confiando en las estimaciones obtenidas por el simulador, la técnica mostró resultados experimentales satisfactorios y alentadores. Posibilitó analizar si la sucesión numérica generada por el algoritmo converge a un punto límite, y si éste último, se encuentra dentro de un rango de aceptación respecto de la solución del problema de optimización. Es arriesgado y prematuro establecer conclusiones contundentes y finales de la aplicabilidad general de la técnica en todas las áreas de optimizaciones en el campo de las Investigaciones Operativas, por lo que es necesario extender el estudio de la aplicabilidad a otros campos del saber.

Como desventaja, la técnica necesita contar con un marco referencial sólido y de resultados reconocidos para realizar las comparaciones estadísticas y obtener resultados que permita de ellos, obtener conclusiones. La técnica de Validación Empírica Residual podría ser una manera de empezar a discutir la convergencia de algoritmos de optimización cuyos resultados están envueltos en un manto de incertidumbre.

REFERENCIAS

[1] Zeltyn S. and Mandelbaum A. (**2005**). Call Centers with Impatient Customers: Many-Server Asymptotics of the M/M/n + G Queue. *Queueing Systems*. Vol. 51 (3):361-402. doi: 10.1007/s11134-005-3699-8

[2] Atlason J., Epelman M. A. and Henderson S. G. (**2007**). Optimizing Call Center Staffing Using Simulation and Analytic Center Cutting-Plane Methods. *Management Science*. Vol. 54 (2):295-309. doi: 10.1287/mnsc.1070.0774

[3] Ingolfsson A., Campello F., Wu X. and Cabral E. (**2010**). Combining integer programming and the randomization method to schedule employees. *European Journal of Operational Research*. Vol. 202 (1):153-163. doi: 10.1016/j.ejor.2009.04.026

[4] Caprara A., Monaci M. and Toth P. (**2003**). Models and algorithms for a staff scheduling problem. *Mathematical Programming*. Vol. 98 (1):445-476. doi: 10.1007/s10107-003-0413-7

[5] Robbins T. R. and Harrison T. P. (**2010**). A stochastic programming model for scheduling call centers with global Service Level Agreements. *European Journal of Operational Research*. Vol. 207 (3):1608-1619. doi: 10.1016/j.ejor.2010.06.013

[6] Koole G. and van der Sluis E. (**2003**). Optimal Shift Scheduling with a Global Service Level Constraint. *IIE Transactions*. Vol. 35 (11):1049-1055. doi: 10.1080/07408170304398

[7] Ingolfsson A., Amanul Haque M. and Umnikov A. (**2002**). Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research*. Vol. 139 (3):585-597. doi: 10.1016/S0377-2217(01)00169-2

[8] Kim S.-M., Nah J.-E. and Kim S.-M. (**2011**). The Staffing Problem at the Call Center by Optimization and Simulation. *IE Interfaces*. Vol. 24 (1):40-50. doi: 10.7232/ieif.2011.24.1.040

[9] Avramidis A. N. and L' Ecuyer P. (**2005**). Modeling and simulation of call centers. *Proceedings of the Winter Simulation Conference.* pp. 144-152. IEEE. Orlando, FL, USA. doi: 10.1109/WSC.2005.1574247

[10] Chokshi R. (**1999**). Decision support for call center management using simulation. *Paper of Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future - Volume 2*. 1999/12/. Vol. 2 pp. 1634-1639. Association for Computing Machinery. doi: 10.1145/324898.325349

[11] Sencer A. and Basarir Ozel B. (**2013**). A simulation-based decision support system for workforce management in call centers. *SIMULATION*. Vol. 89 (4):481-497. doi: 10.1177/0037549712470169

[12] Gulati S. and Malcolm S. A. (**2001**). Call center scheduling technology evaluation using simulation. *Proceedings of the Winter Simulation Conference.* Vol. 2 pp. 1438-1442. IEEE. Arlington, VA, USA. doi: 10.1109/WSC.2001.977467

[13] Asuero A. G. and Gonzalez A. G. (**1989**). Some observations on fitting a straight line to data. *Microchemical Journal*. Vol. 40 (2):216-225. doi: 10.1016/0026-265X(89)90073-8

[14] Cook R. D. and Weisberg S. (**1982**). *Residuals and Influence in Regression*. Monographs on Statistics and Applied Probability. (np. 240) University of Minnesota. School of Statistics, New York: Chapman and Hall. http://hdl.handle.net/11299/37076

[15] Topp R. and Gómez G. (**2004**). Residual analysis in linear regression models with an interval-censored covariate. *Statistics in medicine*. Vol. 23 (21):3377-3391. doi: 10.1002/sim.1731

[16] Black K. (**2010**). *Business Statistics: For Contemporary Decision Making*. 6 Ed. John Wiley & Sons, Inc., USA.

[17] Massart D. L., Vandeginste B. G. M., Deming S. N., Michotte Y. and Kaufman L. (**2003**). *Chemometrics: a Textbook*. Data Handling in Science and Technology. Elsevier Science B.V, Hungary.

[18] Verran J. A. and Ferketich S. L. (**1984**). Residual Analysis for Statistical Assumptions of Regression Equations. *Western Journal of Nursing Research*. Vol. 6 (1):27-40. doi: 10.1177/019394598400600104

[19] Korner-Nievergelt F., Roth T., von Felten S., Guélat J., Almasi B. and Korner-Nievergelt P. (**2015**). Chapter 6 - Assessing Model Assumptions: Residual Analysis. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN*. pp. 75-94. Academic Press. Boston. doi: 10.1016/B978-0-12-801370-0.00006-X

[20] Martin J., Ruiz de Adana D. D. and Asuero A. G. (**2017**). Fitting Models to Data: Residual Analysis, a Primer - Chapter 7. *Uncertainty Quantification and Model Calibration*. pp. 133-173. IntechOpen. Rijeka. doi: 10.5772/68049

[21] Myers R. H., Montgomery D. C. and Anderson-Cook C. (**2009**). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. 3 Ed. John Wiley & Sons, Inc, Hoboken, New Jersey, USA.

[22] Tomàs X., Cuadros J. and González L. (**2006**). *Introducción al Cálculo Numérico*. Institut Químic de Sarrià (IQS). Departament d'Estadística Aplicada, Barcelona, España. https://dialnet.unirioja.es/servlet/libro?codigo=401943

[23] Rumsey D. J. (**2016**). *Statistics For Dummies*. 2 Ed. Wiley Publishing, Inc., Indianapolis, Indiana, USA.

[24] Barberis A. R. and Del Moral Sachetti L. E. (**2011**). Modelización, Simulación y Optimización del personal operativo en la administración de Call/Contact Center. *Paper of Simposio de*

Investigación Operativa 2011 - 40 JAIIO. 29 de Agosto al 02 de Septiembre de 2011. pp. 103-120. SADIO. http://40jaiio.sadio.org.ar/node/160.htm

[25] Law A. M. and Kelton W. D. (**2000**). *Simulation Modelling and Analysis*. Series in Industrial Engineering and Management Science. 3 Ed. McGraw-Hill Higher Education, New York, USA. https://lib.ugent.be/catalog/rug01:001269305

[26] Khurshid A., Hussain E. and Haq M. u. (**2007**). A note on finding peakedness in bivariate normal distribution using Mathematica. *Pakistan Journal of Statistics and Operation Research*. Vol. 3 (2):75-86. doi: 10.18187/pjsor.v3i2.61

[27] Chissom B. S. (**1970**). Interpretation of the Kurtosis Statistic. *The American Statistician*. Vol. 24 (4):19-22. doi: 10.1080/00031305.1970.10477202

[28] Pearson K. (**1905**). Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson. A Rejoinder (Skew Variation, a Rejoinde). *Biometrika*. Vol. 4 (1/2):169-212. doi: 10.2307/2331536

[29] Barberis A. R. and Del Moral Sachetti L. E. (**2019**). Programación No Lineal Entera en la Planificación Óptima de Turnos para un Modelo de Call Center. *Proceedings of VII Congreso de Matemática Aplicada, Computacional e Industrial - MACI 2019. S10-Métodos Numéricos: Algoritmos y Aplicaciones.* Vol. 7 pp. 201-204. ASAMACI. Río Cuarto, Córdoba. Argentina.

[30] Durbin J. and Watson G. S. (**1950**). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*. Vol. 37 (3/4):409-428. doi: 10.2307/2332391

[31] Durbin J. and Watson G. S. (**1951**). Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*. Vol. 38 (1/2):159-177. doi: 10.2307/2332325

[32] Welti L. (2002) *Introducción al Análisis de Regresión Lineal.* [Material de cátedra] [Accedido: Agosto 31 de 2020]. Available from http://imsturex.unex.es/Pagina_TGII/REGRESION.pdf.

[33] Bartlett M. S. (**1937**). Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*. Vol. 160 (901):268-282. http://www.jstor.org/stable/96803

[34] DeCarlo L. T. (**1997**). On the meaning and use of kurtosis. *Psychological Methods*. Vol. 2 (3):292-307. doi: 10.1037/1082-989X.2.3.292

[35] Joanes D. N. and Gill C. A. (**1998**). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*. Vol. 47 (1):183-189. doi: 10.1111/1467-9884.00122

10 y 11 de Diciembre de 2020

Aplicaciones estadísticas en el análisis de procesos de obtención de carbonato de litio en el NOA

Thames Cantolla, Martin; Valdez, Silvana K.; Orce Schwarz, Agustina

Facultad de Ingeniería, INBEMI; Universidad Nacional de Salta, Salta

core.mtc@hotmail.com, +54 9 387 466 9838

RESUMEN

La explotación de salmueras con contenido económico de litio en nuestro país se realiza en distintos salares de la puna argentina. La mayoría de los procesos de obtención de sales de litio involucran etapas de concentración, purificación y precipitación química de sales. El objetivo de este trabajo es determinar la vinculación entre el nivel de concentración por evaporación y la cantidad de producto final obtenido. Para ello se realizó la simulación de dos procesos diferentes para la obtención de carbonato de litio en Aspen Plus v11, empleándose la misma materia prima con igual flujo volumétrico de entrada y variando el grado de concentración. De esta manera se determinó la cantidad de producto final obtenido. Basándose en los resultados obtenidos del análisis de sensibilidad que realiza el software, se propuso una aproximación lineal para vincular las variables seleccionadas. Este tiene un nivel de ajuste del 87% y 64% para el proceso 1 y 2. Considerando que la relación de agua evaporada y cantidad de producto final obtenido, puede emplearse para diseñar y optimizar el proceso en las primeras etapas, los resultados de este trabajo pueden utilizarse para el desarrollo y optimización de los procesos productivos en la industria del litio.

Palabras Claves: litio, procesos, simulación, estadísticas, NOA.

1. INTRODUCCIÓN

Durante los últimos años, el mercado mundial del litio se ha visto incrementado debido a la creciente demanda por parte de los fabricantes de baterías para distintas aplicaciones, en donde la industria automotriz lidera abarcando un 39% del mercado, demanda que de acuerdo a proyecciones para el 2025, alcanzaría a más de dos tercios de la demanda mundial. Desde esta visión, se observa con gran atractivo a los salares sudamericanos enriquecidos en litio, ya que presentan una gran disponibilidad de recursos y costos operativos competitivos [1, 2]. En Sudamérica, la cuarta reserva mundial de litio se localiza en Argentina (país integrante del "triángulo del litio" junto a Bolivia y Chile) representando un desafío local y regional, superar un esquema de extracción minera de litio de alto valor tecnológico [3, 4, 5].

Por su parte, en la zona del Noroeste Argentino (NOA), existen numerosas empresas abocadas a la producción sales del litio, sin embargo, solo 3 de ellas se encuentran en producción (una en Catamarca, otra en Jujuy y una tercera en Salta) con una planta piloto ya en funcionamiento y realizando inversiones para instalar la planta industrial más grande del mundo [5]. De acuerdo a la Secretaria de Minería de Salta, en la provincia existen 50 proyectos de extracción de litio en desarrollo, de los cuales dos se encuentran en fase avanzada para estar en producción para el 2022. Las principales expectativas del gobierno nacional para que el país se convierta en el segundo productor mayoritario de litio a nivel mundial en el año 2022, están puestas en Salta [5].

De esta manera, una empresa que desea iniciar una explotación económica de litio, debe identificar el salar y determinar sus características geológicas, como así también caracterizar la salmuera que este provee. Una vez establecida la factibilidad operativa del salar, se procede al desarrollo o selección del proceso productivo para transformar la salmuera en el producto deseado, carbonato de litio [4, 5, 6]. Este desarrollo o selección del proceso productivo se encuentra íntimamente vinculado con el tipo y calidad de la salmuera a tratar. En un proceso típico de obtención de carbonato de litio, se encuentran presentes las etapas de: evaporación, purificación, encalado, tratamiento con resinas, entre otros. Encontrar una vinculación entre las variables de cada etapa puede resultar beneficioso para la empresa minera, ya que permitiría encontrar el óptimo de funcionamiento del proceso, obteniendo de esta manera un alto rendimiento sin la necesidad de incurrir en un gasto innecesario de recursos [4, 5, 6, 7].

El objetivo de este trabajo es determinar la vinculación entre las variables involucradas en la etapa de concentración por evaporación y la cantidad de producto final obtenido. Para ello se realizó la simulación de dos procesos diferentes para la obtención de carbonato de litio en Aspen Plus v11, empleándose la misma materia prima y variando el grado de concentración. Mediante regresión lineal se determinó la relación de las variables. Los resultados de este trabajo pueden ser empleados por las empresas mineras del sector para el desarrollo y optimización de sus sistemas productivos.

1.1 Aspen Plus

El software llamado Sistema Avanzado para Ingeniería de Procesos o Aspen por sus siglas en inglés es utilizado en diferentes industrias para la simulación de procesos químicos mediante la representación en diagramas de flujo [8, 9]. Este software originado en 1970 puede ser empleado en casi todas las áreas de la ingeniería de procesos, desde la etapa del diseño hasta el análisis de costos y rentabilidad. Cuenta con una biblioteca de elementos y compuestos químicos, como

así también de equipos típicos de la industria como ser: columnas de destilación, intercambiadores de calor, separadores y reactores, entre otros [10, 11].
Mediante la simulación de un proceso es posible representar las transformaciones químicas y físicas que tienen lugar un sistema productivo. Involucrando los modelos matemáticos y sus ecuaciones, balances de masa y energía, puntos de equilibrio, relaciones de equilibrio entre fases, cinética química, etc. [11].

Al utilizar un simulador de procesos, resulta posible:

a) Predecir el comportamiento de un proceso químico.
b) Analizar diferentes escenarios al modificar variables.
c) Optimizar un proceso, ya sea por etapa o en su totalidad.
d) Implementar mejoras o agregar etapas a un proceso.

Cabe destacar que este software es capaz de realizar simulaciones tanto en estado estacionario como en estado dinámico y por ello hacen de este uno de los simuladores más empleados a nivel mundial. Entre las principales funciones que podemos utilizar, tenemos [8, 9, 10, 11]:

a) Generación de gráficos y tablas.
b) Estudio de casos.
c) Dimensionado y evaluación económica de equipos.
d) Estimación de propiedades químicas y termodinámicas.
e) Optimización de procesos.
f) Ajustes de datos experimentales.
g) Determinación de consumos energéticos.
h) Análisis de curvas de funcionamiento.

En la Figura 1 puede observarse la pantalla de simulación de Aspen Plus.

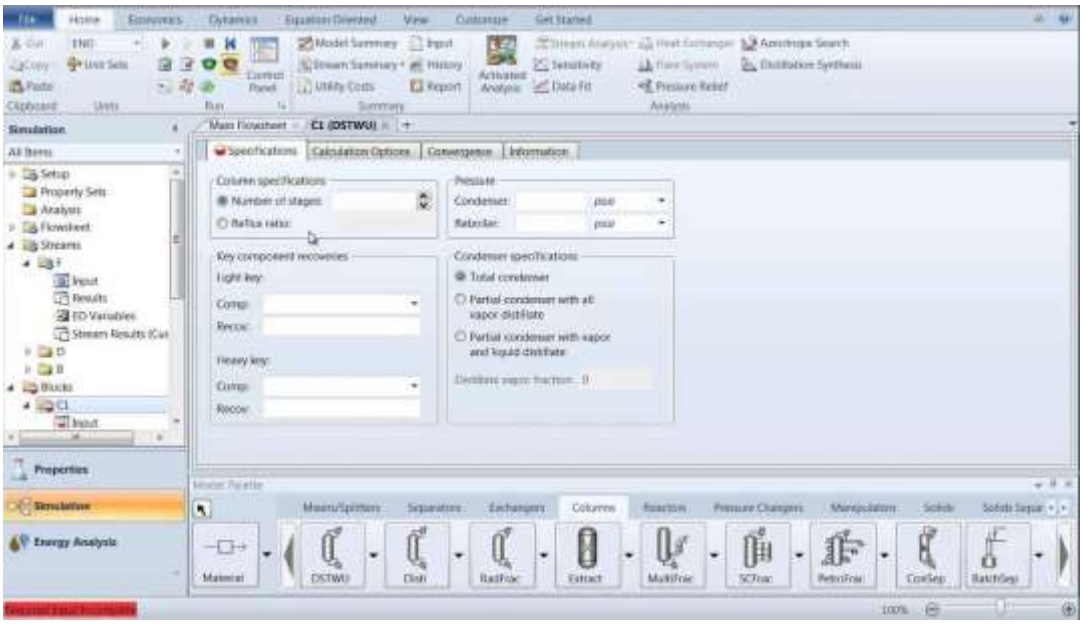


Figura 1 Software de simulación Aspen Plus v11

1.2 Regresión lineal

El objetivo de un modelo de regresión es tratar de explicar la relación que existe entre una variable dependiente, a la que llamamos variable de respuesta, y simbolizamos con "*Y*"; y una o un conjunto de variables independientes o variables explicativas, a las que llamamos $X_1,..., X_n$ [13, 14, 15].
Mediante la aplicación de esta herramienta se puede predecir el comportamiento de la variable de respuesta partir de los valores conocidos de las variables explicativas. Cabe destacar que, cuando se analiza una única variable explicativa, hablamos de regresión lineal simple; mientras que, cuando se analiza un conjunto de variables explicativas, hablamos de regresión lineal múltiple. La Fórmula 1 es la que describe a la regresión lineal [15, 16].

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n \qquad \textbf{Fórmula 1}$$

Donde:

Y = variable de respuesta.
X_{1-n} = variables explicativas.
b_{1-n} = coeficientes de las variables explicativas.
b_0 = constante de la regresión.

De esta manera, cada variable independiente es ponderada, indicando su contribución relativa a la predicción conjunta. El procedimiento del análisis de regresión asegura una predicción máxima a partir del conjunto de variables independientes [13,14].

Debe tenerse en cuenta que el análisis de regresión lineal debe utilizarse sólo cuando las variable dependientes e independientes son métricas. No obstante, ante eventuales condiciones de operación, se podrá utilizar variables no métricas a partir de la transformación de los datos en variables ficticias [13, 14].

Cabe destacarse, que el éxito de la regresión lineal radica en la correcta asignación de las variables dependiente e independientes. La misma debe ser realizada por el investigador o grupo de analistas que trata el problema [14, 15, 16].

2. METODOLOGÍA

2.1 Selección de los procesos a analizar

Para el desarrollo de este trabajo en primer lugar se determinaron los procesos que se analizarían. Para lograr una aplicación práctica, se seleccionaron 2 procesos que son utilizados por empresas minera del medio. Por cuestiones de confidencialidad, llamaremos estos como Proceso 1 y Proceso 2. Los mismos se encuentran representados en las Figuras 2 y 3.

Este proceso requiere algunas modificaciones de acuerdo a las condiciones iniciales de la salmuera, las cuales contienen otros iones como Na^+, K^+, Ca^{2+}, Mg^{2+}, entre otros. A continuación, se describen de manera simplificada las etapas que se llevan a cabo en dicho proceso:

Evaporación natural
El proceso inicia extrayendo la salmuera del salar con bombas, que la transportan por cañería de PVC hacia piletas de evaporación natural en donde la salmuera permanece hasta alcanzar entre 700-2000 ppm de Li^+. El tiempo de permanencia de la salmuera en las piletas dependerá de la concentración inicial, cuanto mayor sea la concentración de Na^+, K^+ y Li^+, menor será el tiempo. En esta etapa se busca reducir el NaCl presente.

Proceso 1 para la obtención de carbonato de litio

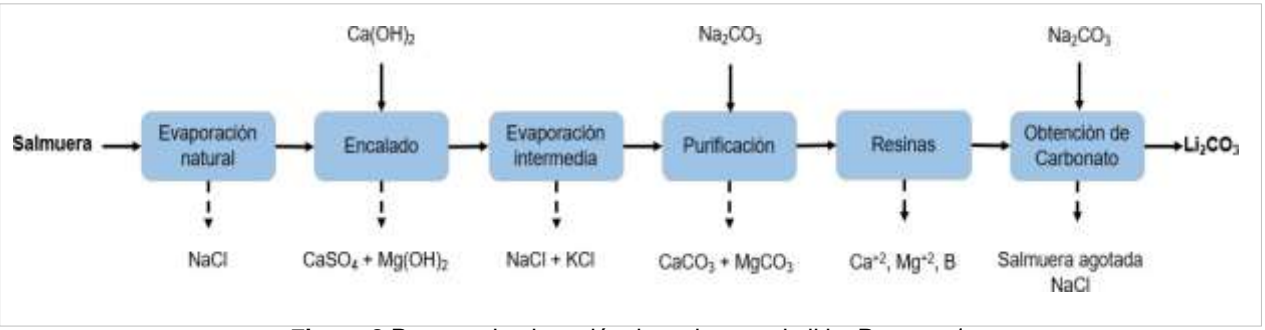


Figura 2 Proceso de obtención de carbonato de litio. Proceso 1

Purificación
Luego de haberse eliminado la mayor cantidad de las impurezas presentes, la salmuera pasa a una etapa de purificación en donde se agrega carbonato de sodio (reactivo de mayor costo) con el propósito de eliminar las impurezas de magnesio y calcio como carbonatos.

Encalado
En esta etapa la salmuera previamente vaporada de la etapa anterior se purifica mediante la reacción con hidróxido de calcio. El agregado de la cal elimina el magnesio y sulfato presentes en la salmuera que se está tratando.

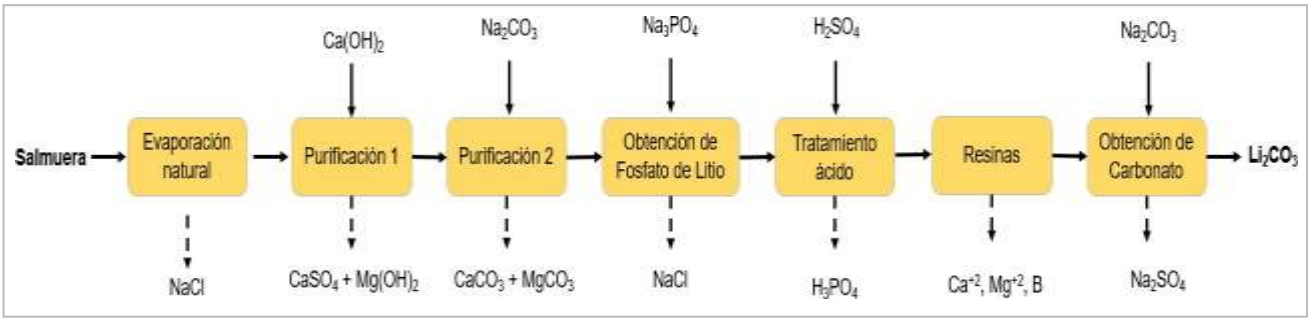
Evaporación intermedia
Una vez eliminadas las impurezas, la salmuera se encuentra lista para una evaporación intermedia en la cual se busca extraer las sales que hayan quedado presentes en la salmuera y aumentar la concentración de litio hasta 30.000 ppm aproximadamente. En esta etapa precipitan los remanentes de NaCl y el KCl.

Paso por resinas
El paso de la salmuera por todas las etapas anteriores nos da como resultado una solución con reducida presencia de iones Ca^{+2}, Mg^{2+} y B pero que continúan representando una impureza para el producto final que se quiere obtener (carbonato de litio). Es por ello que se hace pasar la salmuera por resinas de intercambio iónico, en donde los iones mencionados anteriormente quedan retenidos y se obtiene una solución con una mejor pureza.

Obtención de Carbonato.
Finalmente, la solución es tratada nuevamente con carbonato de sodio en caliente, ahora con el objeto de obtener carbonato de litio. Cabe mencionar que en esta etapa se vuelve a obtener cloruro de sodio, pero en menor cantidad que en las etapas anteriores.

Proceso 2 para la obtención de carbonato de litio

Este proceso presenta algunas modificaciones con respecto al Proceso 1, generando un producto de litio en una etapa intermedia, antes de obtener el carbonato de litio. A continuación, se describen brevemente las etapas que se llevan a cabo en dicho proceso. En Figura 3 se presenta un diagrama simplificado de dicho proceso.



Evaporación natural
Al igual que en el Proceso 1, éste inicia extrayendo la salmuera del salar con bombas, que es transportada por cañería de PVC hacia piletas de evaporación natural en donde la salmuera permanece hasta alcanzar entre 700-2000 ppm de Li^+.

Purificación 1
La salmuera es tratada con hidróxido de calcio, el mismo es agregado a la salmuera para la eliminación de impurezas de calcio y magnesio (al igual que en el Proceso 1).

Purificación 2
Luego la salmuera es tratada con carbonato de sodio con el objeto de continuar eliminando las impurezas de calcio y magnesio que hayan quedado disueltas y no hayan sido eliminadas en la etapa anterior. En este caso, las impurezas se obtienen en forma de carbonatos en ambos casos (carbonato de calcio y carbonato de magnesio).

Obtención fosfato de litio
La solución que ya ha sido purificada en dos etapas anteriores, es tratada con fosfato de sodio, para obtener el fosfato de litio cristalizado, el cual es tratado en la etapa siguiente.
Cabe destacar que en esta etapa se obtiene como desecho cloruro de sodio.

Tratamiento ácido
En esta etapa se trata el fosfato de litio obtenido anteriormente con ácido sulfúrico, esto se realiza para llevar el litio a solución y continuar eliminado cualquier impureza que pueda contener.

Resinas
La solución es acondicionada a pH básico y se la purifica con intercambio iónico. El objetivo de esta etapa es reducir la concentración de boro, calcio y magnesio que afectan la calidad del producto final (carbonato de litio).

Obtención carbonato
Finalmente, a la solución se le agrega nuevamente carbonato de sodio para lograr el precipitado

del carbonato de litio. En esta etapa, se obtiene como desecho el sulfato de sodio.

2.2 Simulación en Aspen Plus

Una vez determinados los procesos, se procedió a realizar la simulación en Aspen Plus siguiendo [9, 10, 11, 12]. Cabe destacar que se emplearon datos provenientes del banco de datos del Instituto de Beneficios de Minerales (INBEMI). Estos datos fueron la composición inicial de la salmuera, como así también las condiciones operativas. De acuerdo a esto, para ambos procesos se cargaron:

a) Composición química de la salmuera a tratar.
b) Unidades de medida.
c) Equipos intervinientes por etapa y su eficiencia.
d) Flujos de materia prima. Para ambos procesos se asignó el mismo flujo volumétrico.

2.3 Selección de variables

Una vez realizada la simulación, se establecieron las variables a analizar de acuerdo a:

- Variable dependiente = Cantidad de producto final (Li_2CO_3), en toneladas/hora.
- Variable independiente = evaporación (grado de concentración)

2.4 Ejecución de la simulación con distintos valores de entrada

Para determinar los puntos a analizar, se ejecutó la simulación en Aspen Plus de forma iterativa, asignando distintos valores a la variable independiente seleccionada. Para realizar esto de forma automática, se empleó la función “Sensitivity Analysis” que trae incorporada el simulador. Esta función permite que se ejecute la simulación, para cada uno de los valores que el usuario establece para una o un conjunto de variables de entrada [9,10,11]. Basándose en datos experimentales y teniendo en cuenta los equilibrios solido-liquido alcanzados y la eficiencia de la filtración, se estableció un intervalo de valores para la variable independiente de [0%, 50%], realizándose un incremento del 5%, cada vez que se ejecutaba la simulación.

3. DESARROLLO

Se realizaron 21 simulaciones para el “Sensitivity Analysis” del Proceso 1 y un total de 10 simulaciones para el Proceso 2. Esta diferencia se debe a las distintas condiciones operativas entre ambos procesos, ya que el Proceso 1 posee dos evaporaciones y el Proceso 2, sólo una. De esta forma, para el Proceso 1, el simulador realiza las iteraciones, asignando el primer valor del intervalo [0%, 50%] a la variable Ev1 y ejecuta la simulación para cada uno de los valores que puede tomar la variable Ev2 (que también puede tomar valores del intervalo [0%,50%]) Una vez completada la simulación, la variable Ev1, toma el segundo valor asignado y se vuelve a repetir el ciclo, esto se repite hasta que la variable Ev1 toma todos los valores posibles del intervalo asignado.
Por otro lado, el Proceso 2, al poseer solo una evaporación (Ev), no requiere realizar la misma cantidad de iteraciones, de esta forma con sólo 10 ejecuciones de la simulación, se cubre el rango de valores asignados.

Para determinar la vinculación entre variables, se analizaron los procesos por separado. De esta manera, con la ayuda de la planilla de cálculo Excel, se ordenaron los resultados de las simulaciones en función de la cantidad de Li_2CO_3 creciente. La variación de la cantidad de producto final obtenido versus las evaporaciones se muestra en la Figura 4, donde se observa saltos debido a las iteraciones que el simulador va realizando a medida que asigna los distintos valores a la variable evaporación. Además, se observa que el mayor grado de evaporación debe realizarse en la etapa 1 para obtener la mayor cantidad de producto final. En las Tabla 1A y 1B se presentan los resultados estadísticos:

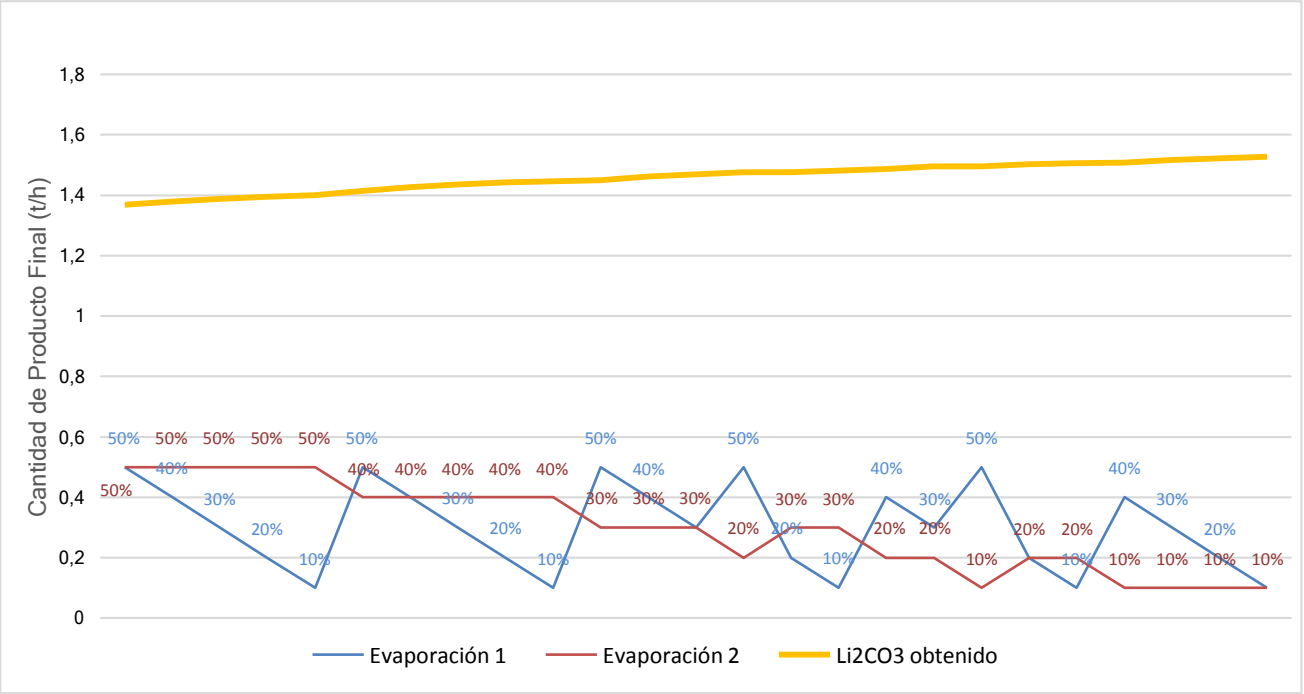


Figura 4 Resultados de la simulación del Proceso 1

Tabla 1*A* Resultados estadísticos para la regresión del Proceso 1

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,931
Coeficiente de determinación R^2	0,867
R^2 ajustado	0,853
Error típico	0,011
Observaciones	21

Tabla 1B Resultados estadísticos para la regresión del Proceso 1

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95,0%*	*Superior 95,0%*
bo	1,5859	0,0129	122,551	0,0000000	1,5587	1,6131	1,5587	1,6131
Ev1	-0,1184	0,0155	-7,656	0,0000005	-0,1508	-0,0859	-0,1508	-0,0859
Ev2	-0,2976	0,0276	-10,776	0,0000000	-0,3557	-0,2396	-0,3557	-0,2396

Luego, al aplicarse el modelo de regresión lineal que trae incorporada la planilla de cálculo, fue posible obtener la Fórmula 2, que corresponde al modelo final para el Proceso 1.

Cantidad de Li_2CO_3 *= 1,586 - 0,118*Evp1 – 0.297*Evp2* **Fórmula 2**

Por otro lado, para el Proceso 2, obtenemos los resultados observados en la Figura 5 y en las Tabla 2A y 2B los resultados estadísticos. En este caso, los mejores resultados se observan para bajos niveles de evaporación.

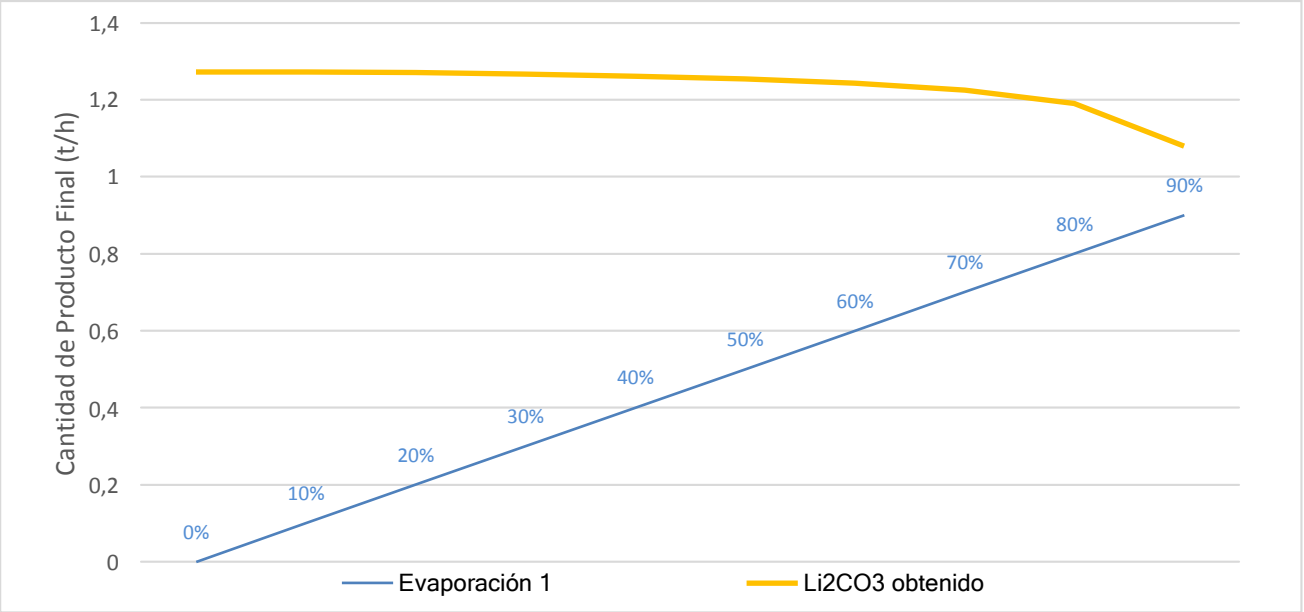


Figura 5 Resultados de la simulación del Proceso 2

Tabla 2A Resultados estadísticos para la regresión del Proceso 2

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,800
Coeficiente de determinación R^2	0,640
R^2 ajustado	0,595
Error típico	0,038
Observaciones	10

Tabla 3B Resultados estadísticos para la regresión del Proceso 2

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95,0%*	*Superior 95,0%*
bo	1,3042	0,0224	58,3191	0,0000000	1,2526	1,3558	1,2526	1,3558
Ev	-0,1579	0,0419	-3,7687	0,0054760	-0,2545	-0,0613	-0,2545	-0,0613

De igual que para el Proceso 1, al aplicarse el modelo de regresión lineal, se obtiene la Fórmula 3:

Cantidad de Li_2CO_3 *= 1,304 -0,158*Evp* **Fórmula 3**

En este caso, puede apreciarse notoriamente que el resultado de las simulaciones para el Proceso 2, resultaron más sencillas al poseer solo una etapa de evaporación. Es así que, a medida que aumenta el grado de evaporación, va disminuyendo levemente la cantidad de Li_2CO_3 obtenido, a partir del 90% de agua evaporada, esta disminución es brusca.

Cabe indicar que, tanto la Fórmula 2 como la Fórmula 3, indican la cantidad de Li_2CO_3, expresado en toneladas/hora, a obtenerse de acuerdo a los valores que adquieran las variables independientes Ev1, Ev2 y Ev, respectivamente para cada modelo.

Finalmente, se señala que los valores para los coeficientes de determinación (R^2) de 0,87 y 0,64 respectivamente para cada proceso, se encuentran dentro de los rangos aceptables para estos sistemas multicomponentes, debido a que su equilibrio químico presenta alta sensibilidad a las condiciones ambientales y de trabajo. De acuerdo a esto, la Fórmula 2 y la Fórmula 3, expresan la cantidad de Li_2CO_3 a obtenerse, para los distintos valores de evaporación que pudiesen aplicarse, con un error del 23% y 36% respectivamente para cada modelo. Debe notarse también que, a pesar de existir una notoria diferencia en la cantidad de simulaciones realizadas para cada proceso, los valores de R^2 no presentan una gran diferencia entre ellos.

4. CONCLUSIONES

Se pudo modelar exitosamente ambos procesos mediante el simulador Aspen Plus. Se observa que, el grado de evaporación (concentración) se relaciona de manera lineal con la cantidad de producto final obtenido. Los valores para los coeficientes de determinación (R^2), se encuentran dentro de los rangos aceptables para este tipo de sistemas. La sencillez matemática para la regresión lineal, hacen de esta metodología una herramienta sencilla que puede aplicarse a sistemas complejos como las salmueras y simular procesos de obtención de carbonato de litio, con una buena aproximación. De los resultados se puede seleccionar el grado de concentración de la salmuera que permita obtener la mayor cantidad de Li_2CO_3 en función del proceso seleccionado. A futuro se plantea la posibilidad de incorporar otras variables independientes para un análisis global del proceso productivo, involucrando de esta manera la obtención de subproductos.

5. BIBILIOGRAFIA

[1] Castello, Andrés; Kloster Marcelo. (2015) “Industrialización del Litio y Agregado de Valor Local: Informe Tecno-Productivo” CIECTI, CABA.

[2] Comisión Chilena de Cobre (COCHILCO). (*2009) “*Antecedentes para una Política Pública en Minerales Estratégicos: Litio*”*. Disponible en: http://ciperchile.cl/pdfs/litio/estudio_cochilco.PDF [Accedido el 5/05/2020].

[3] Conciencia Minera. (2013). “El litio en Argentina”. Disponible en: http://www.concienciaminera.com.ar/2012/05/el-litio-en-argentina/ [Accedido el 2/07/2020].

[4] Manrique, Alejandro. (2014) “Explotación del litio, producción y comercialización de baterías de litio en Argentina” Universidad Nacional de Mar del Plata, Facultad de Ingeniería. E-Book, ISBN 978-987-544-641-0

[5] Comercio y Justicia. (2019). *“*Argentina será el segundo mayor productor global de litio en 2022*”.* Disponible en: https://comercioyjusticia.info/economia/argentina-sera-el-segundo-mayor-productor-global-de-litio-en-2022/ [Accedido el 28/07/2020].

[6] Revista Panorama Minero. (2018) “Litio y tecnología: Cómo obtener más valor de las

salmueras" Disponible en: https://panorama-minero.com/litio/litio-y-tecnologia-como-obtener-mas-valor-de-las-salmueras/ [Accedido el 20/11/2020].

[7] Bravo, Victor. (2019) "Algo sobre el litio". Documento de trabajo. Fundación Bariloche. Departamento de Economía Energética.

[8] Hou, Wei-Feng; Su, Hong-Ye; Hu, Yong-You; Chu, Jian. (2005). "Simulation of industrial catalytic reforming process by developing user's module on ASPEN PLUS platform" Huagong Xuebao/Journal of Chemical Industry and Engineering (China). Vol 56. pp 1714-1720.

[9] Schefflan, Ralph. (2011) "Teach Yourself the Basics of Aspen Plus" Ed. John Wiley & Sons, Inc.

[10] Sandler, Stanley I. (2015) "Using Aspen Plus in Thermodynamics Instruction: A Step-by-Step Guide" Ed. John Wiley & Sons, Inc.

[11] Espínola Lozano, Francisco. (2017) "Tutorial de Aspen Plus, Introducción y modelos simples de operaciones unitarias" Universidad de Jaén. E- Book, ISBN 978-84-9159-039-2

[12] Al-Malah, Kamal I. (2016). "Aspen Plus: Chemical Engineering Applications". Ed. John Wiley & Sons, Inc.

[13] Hair, Joseph F.; Anderson, Rolh Jr.; Tatham, Ronald; Black, William. (1999). "Análisis multivariante". 5º Edición. Prentice Hall Liberia. Madrid.

[14] Ortiz Sánchez, Oswaldo. (2013) "Modelo de regresión de dos etapas aplicado a un cargador de roca en una operación minera superficial" Revista del Instituto de Investigación [en línea]. 2012, vol. 15, no. 29, p. 117-124. ISSN 1561-0888.

[15] Ortiz Sánchez, Oswaldo. (2018). "Modelo analítico para evaluar un yacimiento mineral aplicación en un proyecto de minado superficial". Revista Del Instituto De Investigación De La Facultad De Geología, Minas, Metalurgia Y Ciencias Geográficas, 21(42), 27-34.

[16] Montero Granados, Roberto. (2016) "Modelos de regresión lineal múltiple". Documentos de Trabajo en Ec onomía Aplicada. Universidad de Granada. España.

10 y 11 de Diciembre de 2020

Modelos para el análisis espacial de la tasa de mortalidad por cáncer de próstata en la Provincia de Córdoba

Gonzalez, Mariana Verónica

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba, Argentina.

mariana.gonzalez@unc.edu.ar, 351-4437300

RESUMEN

El cáncer muestra variaciones espaciales y el conocimiento de su patrón de ocurrencia es esencial para identificar grupos de población vulnerables, así como para desarrollar políticas de salud adecuadas para la prevención, el seguimiento y el control (Díaz *et al*., 2010a). En este trabajo se analiza la distribución geográfica de los casos de mortalidad por cáncer de próstata, a nivel de departamentos en la Provincia de Córdoba, Argentina, incluyendo un análisis exploratorio espacial y diferentes enfoques de modelación para la obtención de inferencias respecto de la comprensión y cuantificación del fenómeno. En el marco de los Modelos generalizados mixtos y para variables latentes (*GLLAMM*) se ajustaron diferentes modelos con efectos aleatorios, que permiten considerar la heterogeneidad no observada entre departamentos e incorporan dicha información a la hora de estimar los efectos de covariables de interés.

Los mapas permitieron detectar que las tasas de mortalidad por cáncer de próstata en la Provincia de Córdoba siguen un patrón no aleatorio en su distribución espacial, existiendo una concentración de valores elevados del Cociente de Mortalidad Estandarizado (*CME*) hacia el centro-este provincial.

Las pruebas bondad de ajuste de los modelos revelaron la superioridad del modelo Poisson con ordenada aleatoria respecto del modelo Poisson clásico.

Palabras Claves: cáncer, distribución espacial, Poisson, *GLLAMM*

INTRODUCCION

El monitoreo de la variación geográfica en la distribución de enfermedades y la investigación para comprender las razones subyacentes a dicha variación son, habitualmente, un punto de partida importante en Epidemiología. Se han identificado muchos factores de riesgo significativos como resultado de los hallazgos en el análisis de patrones geográficos de las enfermedades.

Desde la perspectiva de la epidemiología espacial es posible estudiar la ocurrencia de eventos de salud-enfermedad (enfermedades, defunciones) en localizaciones específicas y sus factores condicionantes, incluyendo la producción de mapas de enfermedad o exposición y al análisis estadístico de estos datos. Una característica de este tipo de estudios es, además, que los datos son frecuentemente discretos y asumen la forma de conteos de casos de enfermedad o muerte dentro de regiones (Lawson, 2001).

Mientras el mapeo de enfermedades infecciosas es una práctica bien establecida, la creación de mapas de enfermedades no transmisibles, como el cáncer, está menos desarrollada. En efecto, el cáncer muestra variaciones espaciales y el conocimiento de su patrón de ocurrencia es esencial para identificar grupos de población vulnerables, así como para desarrollar políticas de salud adecuadas para la prevención, el seguimiento y el control (Díaz *et al*., 2010a).

En Argentina, sin embargo, existen pocos estudios de naturaleza epidemiológica vinculados a la distribución espacial de mortalidad por cáncer (Díaz *et al*., 2009; Díaz et al., 2010a). En la Provincia de Córdoba funciona, desde el año 2003, el Registro Provincial de Tumores (RPT), el cual abarca toda la provincia, cubriendo de este modo el 9% de la población argentina. La mayor parte de los estudios basados en datos del RPT, consideran la dimensión temporal para los tipos de cáncer más comunes, pero son pocos los antecedentes de análisis del patrón espacial de los datos.

El cáncer de próstata (CP) es el segundo más común en hombres (GLOBOCAN, 2018) y es, a su vez, la quinta causa de muerte por cáncer en los hombres en Argentina (Instituto Nacional de Cáncer. Ministerio de Salud de la Nación. Argentina, 2017). La supervivencia promedio para este tipo de cáncer es relativamente alta, aunque es notablemente mayor en los países de altos ingresos (Niclis, 2011b). El riesgo de CP aumenta con la edad y se diagnostica en muy pocas personas de 50 años o menos (Grönberg, 2003).

En este trabajo se analiza la distribución geográfica de los casos de mortalidad por CP, a nivel de departamentos en la Provincia de Córdoba, Argentina, incluyendo un análisis exploratorio espacial. En el marco de los Modelos generalizados mixtos y para variables latentes (*GLLAMM*) se ajustaron modelos con efectos aleatorios, que permitieron considerar la heterogeneidad no observada entre departamentos. Los resultados obtenidos permitieron detectar zonas de riesgo, verificando que las tasas de mortalidad por cáncer de próstata en la Provincia de Córdoba siguen un patrón no aleatorio en su distribución espacial.

METODOLOGIA

Se utilizaron los datos de defunciones obtenidos de la Dirección de Estadísticas e Información de Salud de la Nación Argentina y de la Dirección de Estadísticas y Censos de la Provincia de Córdoba. Asimismo, se obtuvieron estimaciones poblacionales, para cada año intercensal, por interpolación exponencial a partir de la información censal 1980, 1991, 2001 y 2010 de Instituto Nacional de Estadística y Censos (INDEC).

A partir de la información de defunciones y población, se calcularon las tasas de mortalidad (por 100.000 habitantes) específicas por sexo y estandarizadas por edad, para la provincia de Córdoba, sus 26 departamentos y para la serie temporal 1986-2011.

Posteriormente, se trabajó en la construcción de la estructura de vecindades, una de las primeras cuestiones a resolver en los análisis que involucran el espacio. Se determinó una lista de vecinos basada en regiones con límites contiguos, que comparten uno o más puntos límites (criterio de contigüidad tipo reina) y se construyó el mapa de contactos (Figura 1).

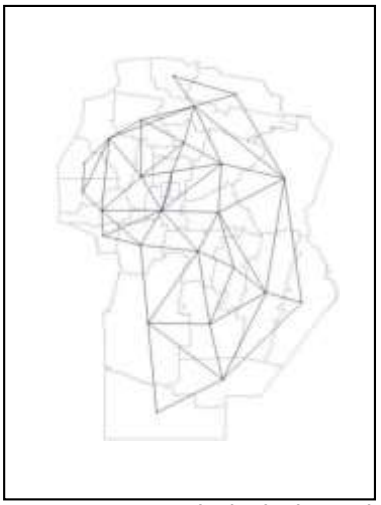


Figura 1. Mapa de contactos para el criterio de contigüidad tipo reina.
Provincia de Córdoba.

Como estimación del riesgo relativo de cada área se calculó el cociente de mortalidad estandarizado (CME), a partir de los casos observados (o_i) y los casos esperados (e_i) para cada departamento de la Provincia de Córdoba, así como sus intervalos del 95% de confianza. A los fines de visualizar la distribución geográfica, por departamentos, de los cocientes de mortalidad, se representaron sus valores en el mapa. Su utilizaron escalas monocromáticas (Pickle *et al*., 1999), para inducir a una menor variabilidad y los límites de clases se definieron empleando cinco cuantiles determinados a partir de la distribución observada.

El riesgo relativo θ_i, específico para cada área, es una variable aleatoria. Condicional a θ_i, los conteos observados son variables independientes e idénticamente distribuidas (*i.i.d*) Poisson con media $\mu_i = e_i . \theta_i$, esto es:

$$o_i|\theta_i, e_i \sim \text{Poisson}\,(e_i\theta_i)$$

Empleando el comando *gllamm* del software *Stata* se ajustó un modelo Poisson con ordenada aleatoria, para la Provincia de Córdoba, período 1986-2011.

$$\ln(\mu_i) = \ln(e_i) + \beta_0 + \zeta_i\,,$$

donde las ordenadas aleatorias representan la heterogeneidad no observada entre áreas $\zeta_i \sim i.i.d$, $\zeta_i \sim (0, \psi)$ y $\ln(e_i)$ es una variable *offset* (covariable con coeficiente 1). El propósito de ésta variable *offset* es asegurar un modelo que incluya el *CME*:

$$\ln(\mu_i) - \ln(e_i) = \ln(\mu_i/e_i) = \beta_0 + \zeta_i$$

ya que $CME_i = \mu_i/e_i$.

Habiendo obtenido estimaciones por Máxima Verosimilitud de los parámetros del modelo fue deseable asignar valores a los efectos aleatorios ζ_i para áreas particulares. El método de predicción sugerido por Rabe-Hesketh y Skrondal (2005) para la predicción de los *CMEs* es el Empírico Bayesiano (*Empirical Bayes Prediction*) que provee valores más estables para áreas con poblaciones pequeñas. Así, la estimación Bayesiana de ζ_i usa no sólo las respuestas μ_i, sino también la distribución a priori de ζ_i (normal con media cero y varianza estimada) para predecir valores de los efectos aleatorios individuales:

$$Posterior(\zeta_i|\,\mu_i) \propto \text{Prior}\,(\zeta_i).Likelihood(\mu_i|\,\mathbf{X}, \zeta_i)$$

Específicamente, la estimación del valor esperado a posteriori del *CME* para un área *i* será:

$$\hat{CME}_i = \text{E}\left[\exp(\hat{\beta}_0 + \zeta_i)\middle|\, o_i, e_i\right] = \int \exp(\hat{\beta}_0 + \zeta_i)\, Posterior(\zeta_i|\, o_i, e_i)\, d\zeta_i .$$

Así, el modelo Poisson con ordenada aleatoria se combinó con la predicción de los *CMEs* a través del método empírico Bayesiano, empleando el comando *gllapred* de *Stata*.
Dado que el modelo Poisson estándar se encuentra anidado en el modelo Poisson con intercepto aleatorio, se utilizó el test de cociente de verosimilitud para realizar comparaciones, empleando el comando *lrtest* de *Stata.* El *AIC* y el *BIC*, calculados empleando el comando *estimates stats* de *Stata*, se emplearon también para evaluar la bondad de ajuste de los modelos.

DESARROLLO

La distribución espacial (Figura 2), para el cáncer de próstata, exhibe una concentración de valores elevados del *CME* hacia el centro-este provincial. Se trataría de una zona en riesgo, teniendo en cuenta que los casos observados superan a los esperados, en función de la estructura poblacional. Dentro de ésta zona, los departamentos Minas y Sobremonte presentaron valores elevados del *CME*, además de amplios intervalos de confianza, en comparación con el resto de los departamentos (figura 3). Sin embargo, hay que tener en cuenta que son departamentos con baja densidad

poblacional, lo que condiciona las interpretaciones. Los *CMEs* más bajos se presentaron en los departamentos Pocho, Río Seco y Río Cuarto, los dos primeros con un reducido tamaño poblacional.

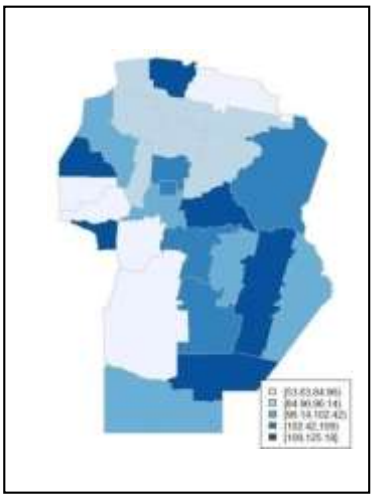


Figura 2. Mapeo de los Cocientes de mortalidad estandarizados en porcentaje.
Provincia de Córdoba. Período 1986-2011.

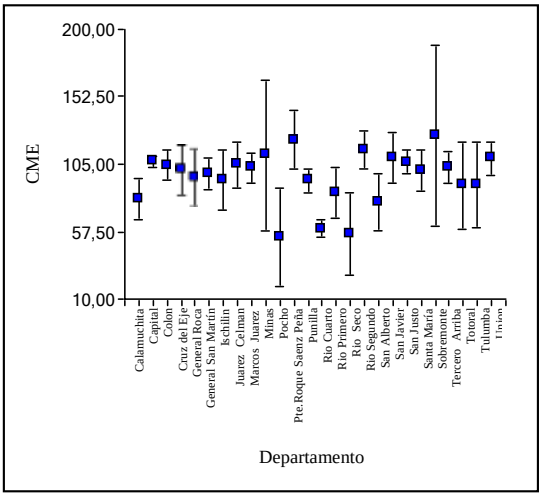


Figura 3. Cocientes de mortalidad estandarizados e intervalo de confianza.
Provincia de Córdoba. Período 1986-2011.

La Tabla 1 muestra las estimaciones por máxima verosimilitud para el modelo Poisson con ordenada aleatoria, período 1986-2011. Se observa una varianza del intercepto de 0,022, mayor que cero, evidenciando la existencia de superdispersión

	Estimaciones	Desvío estándar	Valor *p*
β_0 (const)	-0,040	0,035	0,254
ψ	0,022		

Tabla 1. Estimaciones para el modelo Poisson con intercepto aleatorio.
Provincia de Córdoba. Período 1986-2011.

Dado que el modelo Poisson estándar se encuentra anidado en el modelo Poisson con intercepto aleatorio, se aplicó el test de cociente de verosimilitud (Tabla 2). Como puede observarse, el modelo Poisson debe ser rechazado en favor del modelo Poisson con ordenada aleatoria. Este último modelo resulta ser el que mejor ajusta a los datos, en función de los indicadores de Bondad de ajuste (Tabla 3)

	Valores
-2 Log-Likelihood	117,35
p-valor	0,000

Tabla 2. Resultados del Test de cociente de verosimilitud para los modelos Poisson y PIA.
Provincia de Córdoba. Período 1986-2011.

	Poisson	Poisson con ordenada aleatoria
AIC	360,77	245,43
BIC	362,03	247,94

Tabla 3. Indicadores *AIC* y *BIC* para los diferentes modelos.
Provincia de Córdoba. Período 1986-2011.

Habiendo obtenido las estimaciones de los parámetros del modelo, se efectuaron las correspondientes predicciones de los riesgos relativos, que se representaron en los mapas, a fin de visualizar su distribución espacial (Figura 4). La distribución geográfica de los riesgos estimados para el cáncer de próstata muestra que los mayores riesgos se registran en la zona este de la Provincia, involucrando los Departamentos San Justo, Río Segundo, Unión y Presidente Roque Sáenz Peña.

También presentan riesgos superiores al 100% los Departamentos Capital y San Javier. Este gradiente coincide, en términos generales, con la zona de riesgo detectada al analizar la distribución espacial de los *CMEs* (Figura 2).

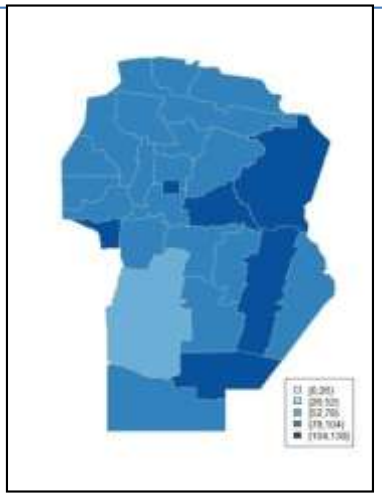


Figura 4. Mapeo de los riesgos relativos estimados con el modelo.
Provincia de Córdoba. Período 1986-2011.

CONCLUSIONES

La representación y análisis de mapas de mortalidad por enfermedad se han convertido actualmente en una herramienta básica para el análisis de la salud pública regional. Proporcionan un rápido resumen visual de información geográfica compleja y permiten identificar patrones en los datos que de otro modo podrían pasar inadvertidos en las presentaciones tabulares (Elliott y Wartenberg, 2004). No obstante, como afirma Lawson (2001), si bien los mapas proporcionan información visual importante sobre la distribución espacial, deben interpretarse en conjunto con la información estadística. El uso de escalas monocromáticas en su construcción reduce la variabilidad, aunque existe una cierta arbitrariedad en la elección del número de clases para la paleta de colores, que puede sesgar las interpretaciones. Las observaciones correspondientes a los casos brutos (datos crudos, tal y como se observan) pueden ser representadas directamente sobre la estructura geográfica de una zona obteniendo un mapa, por ejemplo, de los casos de muerte por una determinada enfermedad. Sin embargo, esta sencilla representación gráfica no siempre proporciona información de interés. Cualquier interpretación de la estructura de los casos incidentes está limitada por la falta de información sobre la distribución espacial de la población que podría estar “en riesgo” de padecer la enfermedad y que, consecuentemente, ha generado esos casos incidentes. A la representación de los casos brutos, se preferirá, en general, la representación de razones que permiten incorporar el efecto de la población a riesgo, tal como se hizo en este trabajo.

En esta investigación se emplearon mapas de mortalidad en conjunto con modelos específicos, lo

que permitió identificar que la distribución espacial de la tasa de mortalidad por cáncer de próstata, en la Provincia de Córdoba, sigue un patrón no aleatorio. Específicamente, el análisis de los mapas correspondientes a los cocientes de mortalidad estandarizados permitió identificar áreas en riesgo, con un número de casos observados superior al esperado, teniendo en cuenta la distribución poblacional. Una concentración de valores elevados del CME fue detectada hacia el centro-este

provincial, donde los departamentos de Presidente Roque Sáenz Peña, Unión y Río Segundo registran los valores más altos. Hacia el oeste de la Provincia, los departamentos Minas, Sobremonte y San Javier también presentaron valores correspondientes al quintil superior de la distribución, siendo los dos primeros departamentos con una baja densidad poblacional.

En relación a la modelación, si bien es lógico suponer que los conteos de casos se distribuyen Poisson con una esperanza diferente dentro de cada área, generalmente se asume que la esperanza es función de un parámetro de riesgo relativo constante θ_i. Este modelo implica, además, que todo exceso de riesgo debe ser captado por los e_i, de manera que en su cálculo debe considerarse la información de aquellas variables que, estando disponibles, pueden afectar la distribución del riesgo subyacente. En este sentido, los casos esperados en este trabajo se calcularon aplicando el método de estandarización indirecta que toma las tasas específicas por estrato de edad de la población estándar, las que se promedian utilizando como ponderaciones los tamaños de los estratos de la población estudiada. Esta estandarización se justifica en el hecho de que, en Epidemiología, la mayoría de las tasas (incidencia, prevalencia y mortalidad) son fuertemente dependientes del grupo etario. Sin embargo, podría ser interesante, tal como sugieren Lawson, Brown y Rodeiro (2003) incorporar ajustes por variables del Censo Nacional, particularmente relativas a mediciones socioeconómicas o indicadores de privación. En este sentido se sugiere avanzar en la aplicación de modelos que permitan, además de modelar la heterogeneidad no observada entre áreas, incorporar el efecto de covariables socieconómicas, disponibles para el año 2001 y para toda la provincia, a partir del Censo Nacional de Población, Hogares y Viviendas.

Los avances en la disponibilidad de datos y en los métodos analíticos para tratarlos han proporcionado nuevas oportunidades para extender las investigaciones epidemiológicas tradicionales, a escala nacional o regional, hasta el estudio de las variaciones en las enfermedades a nivel local, o de área pequeña (Walter, 2000). Sería deseable que tales investigaciones contemplaran además, factores de riesgo para la salud con relevancia a nivel local, tales como exposiciones ambientales, distribución local de condiciones socioeconómicas y los hábitos y estilos de vida (Elliott y Wartenberg, 2004).

BIBLIOGRAFIA

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Anselin L., Bao S. (1997). *Exploratory Spatial Data Analysis*. En Recent developments in spatial analysis (Eds. Fischer y Getis), Springer-Verlag, Berlín; pp.35-59.
- Breslow N. and Clayton D. (1993). *Approximate inference in generalized linear mixed models.* Journal of the American Statistical Association 88:9-25.
- Cliff A., Haggett P. (1992). *Atlas of disease distributions: analytic approaches to epidemiological data.* Oxford: Blackwell..
- Díaz M. P., Osella A., Aballay L., Muñoz S., Lantieri M., Butinof M., Meyer Paz R., Pou S., Eynard A., La Vecchia C. (2009). *Cancer incidence pattern in Cordoba, Argentina*. European Journal of Cancer Prevention. 18:259-266.
- Díaz M. P., Corrente J., Osella, A., Muñoz S., Aballay L. (2010[a]). *Modeling Spatial Distribution of Cancer Incidence in Cordoba, Argentina*. Applied Cancer Research 30(2)245-252.
- Díaz M., García F., Caro P., Díaz M.P. (2010[b]). *Modelos Mixtos Generalizados para el estudio de la asociación entre algunas variables socioeconómicas y las tasas de incidencia de cáncer en localidades de Córdoba, Argentina*. Instituto Interamericano de Estadística. 62, 178, pp. 99-117.

- Dirección de Estadísticas e Información de Salud (2011). *Agrupamiento de causas de mortalidad por división político territorial de residencia, edad y sexo. República Argentina*. Año 2009. Boletín N° 131. Buenos Aires: Ministerio de Salud, Presidencia de la Nación.
- Elliott P, Wartenberg D. (2004). *Spatial epidemiology: current approaches and future challenges*. Environmental Health Perspectives; 112(9): 998-1006.
- Getis A., Ord J. (1992). *The Analysis of Spatial Association by Use of Distance Statistics*. Geogmphicd Analysis 24 (July), 189-206.
- Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.
- Grönberg H. (2003). *Prostate cancer epidemiology*. Lancet; 361:859-64.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press Cambridge.
- Hall S., Kaufman J., Millikan R., Ricketts T., Herman D., Savitz D.(2005). *Urbanization and breast cancer incidence in North Carolina, 1995–1999.* Ann Epidemiol 15:796–803.
- Lawson A. (2001). *Statistical Methods in Spatial Epidemiology*. John Wiley & Sons, Ltd.
- Lawson A., Browne W., Rodeiro C. (2003). *Disease Mapping with WinBUGS and MLwiN.* Wiley Ed.
- Niclis C., Díaz M.P., La Vecchia C. (2010). *Breast cancer mortality trends and patterns in Córdoba, Argentina in the period 1986–2006.* European Journal of Cancer Prevention 2010, 19:94–99.
- Niclis C., Pou S., Bengió R., Osella A., Díaz M. (2011[b]). *Tendencias en la mortalidad por cáncer de próstata en Argentina 1986-2006: análisis joinpoint y de edad-período-cohorte.* Cad. Saúde Pública, Rio de Janeiro, 27(1):123-130.
- PAHO; Health Surveillance and Disease Management Area (2007). *Health Statistics and Analysis Unit*. PAHO Regional Mortality Database. Direct adjusted mortality rate using the World Population Prospects 2006 Revision.
- Pickle L., Mungiole M., Jones G., White A. (1999). *Exploring spatial patterns of mortality: the new atlas of United States mortality.* Statistics in Medicine. 18. 3211-3220.
- Pou S., Osella A., Eynard A., Niclis C., Diaz M. (2009). *Colorectal cancer mortality trends in Córdoba, Argentina.* Cancer Epidemiology 33 (2009) 406–412.
- Pou S. (2012). *Estudio comparativo de la relación cáncer- dieta en regiones sanitarias de la Provincia de Córdoba empleando la estrategia de modelos multinivel.* Trabajo de Tesis para optar al Título de Doctor en Ciencias de la Salud. Facultad de Ciencias Médicas. Universidad Nacional de Córdoba.
- Rabe-Hesketh S., Toulopoulou T., Murray R. (2001). *Multilevel modeling of congnitive function in schizophrenic patiens and their first degree relatives.* Multivariate Bahavioural Research 36, 279-298.
- Rabe-Hesketh S., Skrondal A., Pickles A. (2002). *Reliable estimation of generalized linear mixed models using adaptative quadrature.* . Stata Journal 2.
- Rabe-Hesketh S., Skrondal A., Pickles A. (2005). *Maximun likelihood estimation of limited and discrete dependent variable models with nested random effects.* . Journal of Econometrics 128.
- Rabe-Hesketh S. and Skrondal A. (2005). *Multilevel and longitudinal modelling using Stata*. College Station, TX: Stata Press.
- Rabe-Hesketh S., Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata.* Third Edition. Stata Corp LP. College Station, Texas.
- Walter S. (2000). *Disease mapping: a historical perspective*. En: Elliott P, Wakefield J, Best N, Briggs DJ (eds). Spatial epidemiology: methods and applications. Oxford: Oxford University Press, 223-252.

10 y 11 de diciembre de 2020

NECESIDADES DOCENTES EN CLASES VIRTUALES DERIVADAS DE LA PANDEMIA COVID-19

(FACULTAD DE INGENIERIA DE LA UNIVERSIDAD NACIONAL DE SALTA)

Ing. Gisella Mautino - Mg. Ing. Héctor Iván Rodríguez
Facultad de Ingeniería - Universidad Nacional de Salta – Salta Argentina
gmautino@ing.unsa.edu.ar - Cel +5493884440566
ivan@ing.unsa.edu.ar - Cel +5493874129731

RESUMEN
Debido a la Pandemia originada por el COVID19, se suspendieron las actividades presenciales por la situación de aislamiento, esto llevó a que las universidades dispongan sus actividades de manera virtual. Surge la necesidad de conocer la opinión de los docentes de la Facultad de Ingeniería de la Universidad Nacional de Salta a efectos de tomar decisiones orientadas a la gestión de la planta docente. Se realizó una encuesta a todo el personal docente de la Facultad de Ingeniería de la Universidad Nacional de Salta, sobre las principales necesidades, tanto técnicas como formativas, para el dictado virtual, también se consultó sobre barreras que obstaculizan la enseñanza a través de la plataforma Moodle y la virtualidad, tanto desde el lado docente como del estudiantil. Otro aspecto consultado fue el grado de acuerdo o desacuerdo respecto a modalidad y el nivel de avance de las actividades virtuales que los docentes están realizando. El objetivo del presente estudio es dar a conocer la experiencia de la Facultad en el dictado virtual mediante la opinión docente.

Palabras Claves: Clases Virtuales; Pandemia, Covid19, Docentes

INTRODUCCIÓN

Los docentes respondieron positivamente a la iniciativa de la encuesta. Del resultado surgieron dos talleres de capacitación que se diseñados con la temática demandada por la mayoría y fue una fuente de información para la gestión de resoluciones necesarias por parte del Decanato en referencia a actividades del dictado, tiempos y evaluaciones.

METODOLOGIA

Fecha de consulta: 7 y 8 de abril del 2020
Universo: Docentes activos de la Facultad de Ingeniería de la Universidad Nacional de Salta que dictan clases en el primer cuatrimestre y que no son ayudantes de segunda categoría.
Tipo de relevamiento: Censo, consulta realizada por mail a toda la población docente definida en el Universo.
Instrumento de recolección: Cuestionario estructurado para auto llenado.
Planta docente activa total: 313 personas
Planta docente del Universo: 254 personas
Cantidad de respuestas obtenidas: 186

COMPOSICION DE LA POBLACION DOCENTE RESPECTO A LA ENCUESTA

Ilustración 1

Se planificó realizar la consulta a toda la población docente de la Facultad de Ingeniería que dicta clases en el primer cuatrimestre, exceptuando auxiliares de 2da categoría. El objetivo es contar con un censo que permita conocer, no solo la opinión, sino también el volumen de respuesta, a efectos de tomar decisiones acordes a la situación y también evaluar el funcionamiento de la estructura comunicacional de la Facultad. Se comunicó por mail, a través del departamento de cómputos de la Facultad, durante dos días (7 y 8 de abril) obteniéndose 186 respuestas de un total de 254 docentes a los que se les envió la consulta por mail. Este resultado constituye un nivel de respuesta del 73,2% de docentes que respondieron, lo que constituye afirmar que los datos que refleja la encuesta representan al 73,2% de la población docente de la Facultad de Ingeniería que dicta clase en el primer cuatrimestre, exceptuando a los auxiliares de 2da. Aún si se calcula el nivel de respuesta sobre la base de la planta docente activa de 313 personas, se tendría una representatividad del 60%. Por lo tanto, los resultados del presente informe se constituyen en la opinión de la mayoría.

Es necesario aclarar que el 34,5% de la población que no respondió la encuesta, podría estar constituida por:

- Docentes con voluntad de responder pero que no leyeron el mail o lo hicieron de manera tardía.
- Docentes que leyeron el mail, pero decidieron no contestar.
- Docentes a los que no les llegó el mail.

En los puntos que refieren a la falta de respuesta por razones que hacen a la estructura comunicativa, es donde se hace necesario trabajar para asegurar que la comunicación llegue en tiempo y forma.

COMPOSICION DE LA POBLACION QUE RESPONDIO LA ENCUESTA
A continuación, los datos reflejan la opinión del 65,5% de la población docente de la Facultad de ingeniería.

El 52% respondió desempeñarse con rango de profesor. 60% afirmó tener tareas de responsabilidad de catedra, lo que significa que hay auxiliares que declaran responsabilidades de cátedra. Esto puede deberse a una posible confusión de esos docentes auxiliares, o desconocimiento del significado formal de responsabilidad de cátedra, o bien que sean efectivamente responsables en otra materia.

Ilustración 2

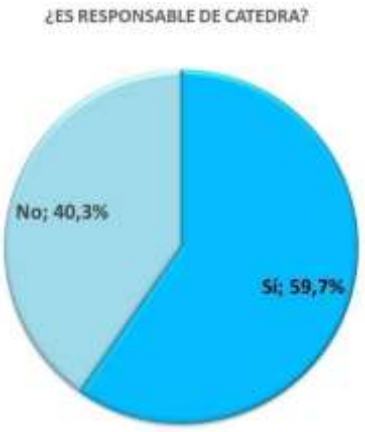


Ilustración 3

		CARGO			
		Profesor titular o Asociado	Profesor Adjunto	JTP o auxiliar de primera	Total
		% del N de tabla	% del N de tabla	% del N de tabla	% del N de tabla
¿ES RESPONSABLE DE CATEDRA?	Sí	10,2%	34,9%	14,5%	59,7%
	No	1,1%	5,9%	33,3%	40,3%
	Total	11,3%	40,9%	47,8%	100,0%

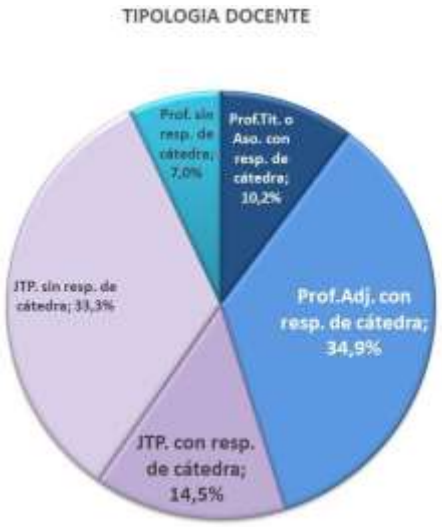
Tabla 1

		CARGO		
		Profesor titular o Asociado	Profesor Adjunto	JTP o auxiliar de primera
		% del N de columna	% del N de columna	% del N de columna
¿ES RESPONSABLE DE CATEDRA?	Sí	90,5%	85,5%	30,3%
	No	9,5%	14,5%	69,7%
	Total	100,0%	100,0%	100,0%

Tabla 2

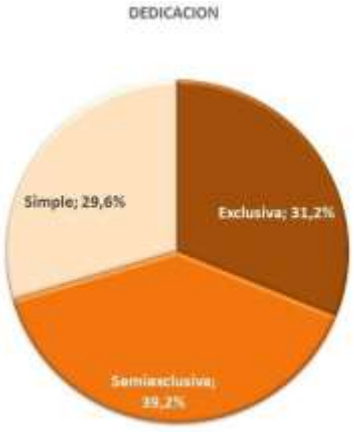
CARGO, RESPONSABILIDAD DE CATEDRA Y DEDICACION

A efectos de analizar de manera conjunta cargo y responsabilidad se construye una tipología docente que refleja 45% de profesores con responsabilidad de cátedra y 14,5% de auxiliares que declaran responsabilidad de cátedra. Respecto a la dedicación puede observarse una distribución bastante pareja, prácticamente un tercio de cada una.



		% del N de columna
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	10,2%
	Prof.Adj. con resp. de cátedra	34,9%
	JTP. con resp. de cátedra	14,5%
	JTP. sin resp. de cátedra	33,3%
	Prof. sin resp. de cátedra	7,0%
	Total	100,0%

Ilustración 4



		% del N de columna
DEDICACION	Exclusiva	31,2%
	Semiexclusiva	39,2%
	Simple	29,6%
	Total	100,0%

Ilustración 5

		TIPOLOGIA DOCENTE					
		Prof.Tit. o Aso. con resp. de cátedra	Prof.Adj. con resp. de cátedra	JTP. con resp. de cátedra	JTP. sin resp. de cátedra	Prof. sin resp. de cátedra	Total
		% del N de tabla	% del N de tabla	% del N de tabla	% del N de tabla	% del N de tabla	% del N de tabla
DEDICACION	Exclusiva	5,9%	11,8%	1,1%	8,1%	4,3%	31,2%
	Semiexclusiva	3,8%	15,1%	6,5%	11,8%	2,2%	39,2%
	Simple	0,5%	8,1%	7,0%	13,4%	0,5%	29,6%
	Total	10,2%	34,9%	**14,5%**	33,3%	7,0%	100,0%

Tabla 3

El dato llamativo es el 7% de docentes auxiliares que declaran dedicación simple con responsabilidad de cátedra. Ese 7% representa el 50% del total de auxiliares que declaran responsabilidad de catedra, por lo que podría tratarse de alguna confusión.

CARRERA DONDE DICTA CLASE

Puede observarse que todas las carreras están representadas en esta encuesta, y que muchos docentes dictan clases en más de una carrera, por ello las suma excede el 100%.

Ilustración 6

		CARRERA					
		ING. CIVIL	ING. ELECTROMECANICA	ING. INDUSTRIAL	ING. QUIMICA	TUTA SEDE CENTRAL	TUTA SEDE SUR
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
CARGO	Profesor titular o Asociado	9,9%	5,6%	7,6%	11,7%	4,5%	0,0%
	Profesor Adjunto	37,0%	30,6%	41,8%	33,8%	50,0%	44,4%
	JTP o auxiliar de primera	53,1%	63,9%	50,6%	54,5%	45,5%	55,6%

Tabla 4

EFECTIVIDAD DE LA CLASE VIRTUAL RESPECTO A LA PRESENCIAL

Que el 43% tenga la percepción de alto grado de efectividad y 6,5% piense es bajo o nulo, siendo el tiempo transcurrido muy corto para lograr experiencia en el tema, significa una valoración positiva importante del grado de efectividad que se puede tener en transmitir el contenido de una materia de manera virtual. Valoración ésta, dada por personas idóneas para opinar en el tema como lo son los propios docentes.

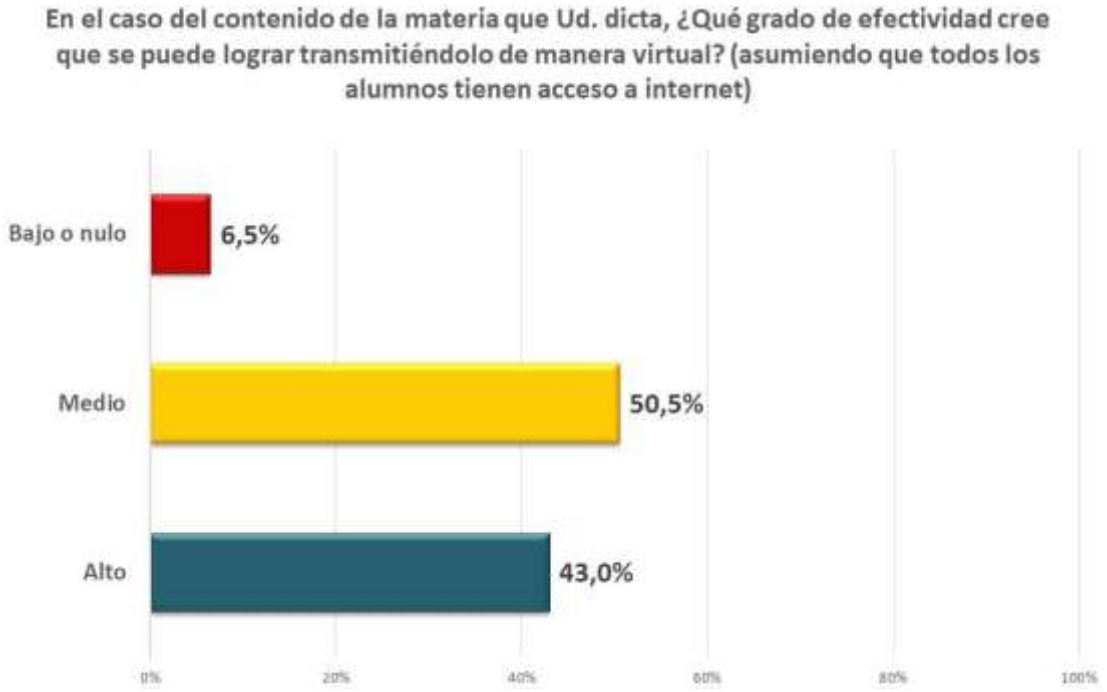


Ilustración 7

		TIPOLOGIA DOCENTE				
		Prof.Tit. o Aso. con resp. de cátedra	Prof.Adj. con resp. de cátedra	JTP. con resp. de cátedra	JTP. sin resp. de cátedra	Prof. sin resp. de cátedra
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
En el caso del contenido de la materia que Ud. dicta, ¿Qué grado de efectividad cree que se puede lograr transmitiéndolo de manera virtual? (asumiendo que todos los alumnos tienen acceso a internet)	Alto	47,4%	49,2%	33,3%	41,9%	30,8%
	Medio	52,6%	43,1%	59,3%	51,6%	61,5%
	Bajo o nulo	0,0%	7,7%	7,4%	6,5%	7,7%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 5

		DEDICACION		
		Exclusiva	Semiexclusiva	Simple
		% del N de columna	% del N de columna	% del N de columna
En el caso del contenido de la materia que Ud. dicta, ¿Qué grado de efectividad cree que se puede lograr transmitiéndolo de manera virtual? (asumiendo que todos los alumnos tienen acceso a internet)	Alto	41,4%	43,8%	43,6%
	Medio	53,4%	50,7%	47,3%
	Bajo o nulo	5,2%	5,5%	9,1%
	Total	100,0%	100,0%	100,0%

Tabla 6

DIFICULTADES PARA TRANSMITIR LA CLASE VIRTUAL CON ALTA EFECTIVIDAD

Opiniones expuestas por el 6,5% de docentes que afirman que la efectividad es nula o baja

Básicamente se observan tres componentes que integran la dificultad para transmitir clases virtuales, a saber:

1. Factor Técnico
2. Factor Psicológico
3. Factor de destreza

Factor de Técnico: Mas de la mitad de las dificultades expuestas son referidas a problemas de conectividad, por lo que la solución al problema es de orden técnico.

Factor Psicológico: La mitad del 6,5% de docentes que opinan que no se puede transmitir la clase con alta efectividad, afirman que lo presencial es superior a lo virtual, esta opinión requiere el esfuerzo de persuasión en demostrar lo contrario con hechos concretos, si es que ello realmente es así.

Factor Destreza: un tercio de las razones expuestas son referidas a la toma de evaluaciones y la realización de prácticas en laboratorio o en campo. Estos problemas en gran parte se pueden mitigar con la capacitación en herramientas informáticas, por ejemplo, los programas de realidad aumentada están comenzando a aplicarse para la realización de prácticas en todas las disciplinas. Hay herramientas para la toma de evaluaciones de manera virtual.

Ilustración 8

Opiniones expuestas por el 50,5% de docentes que afirman que la efectividad es mediana

Puede observarse que los factores enunciados anteriormente están vigentes en este grupo de docentes, en los cuales aumenta la preocupación por la toma de exámenes y actividades de laboratorio y campo. El 46,8% de los docentes que, entre sus opiniones, expresaron que la clase presencial es superior a la virtual representa el **23,6%** de la planta docente que respondió la encuesta.

El orden de las preocupaciones enunciadas, son generalmente compartidas por docentes con independencia del cargo y dedicación. Se observan algunas diferencias por carrera, Ingeniería Química y TUTA, son las más preocupadas por las prácticas de laboratorio y de campo. A los docentes de la TUTA también les preocupan significativamente la conectividad técnica. El problema de la toma de evaluaciones preocupa más significativamente a ingeniería Civil, Electromecánica y Química.

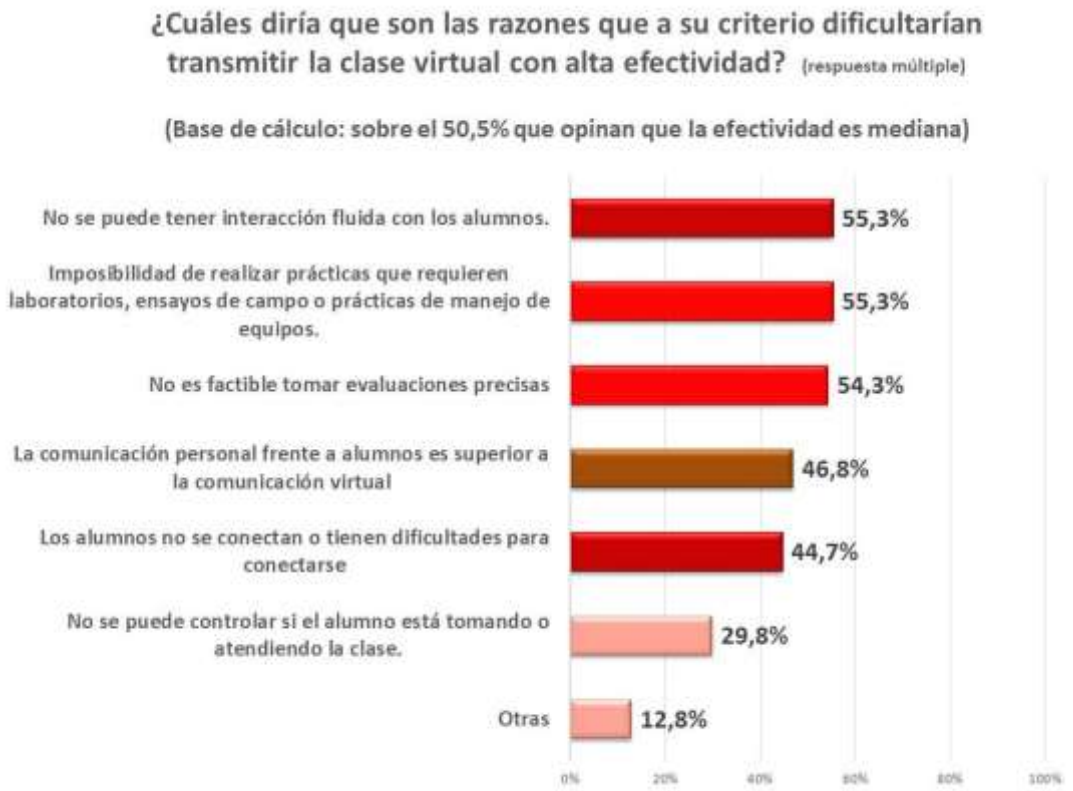


Ilustración 9

		TIPOLOGIA DOCENTE				
		Prof.Tit. o Aso. con resp. de cátedra	Prof.Adj. con resp. de cátedra	JTP. con resp. de cátedra	JTP. sin resp. de cátedra	Prof. sin resp. de cátedra
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Cuáles diría que son las razones que a su criterio dificultarían transmitir la clase virtual con alta efectividad? (respuesta múltiple))	Imposibilidad de realizar prácticas que requieren laboratorios, ensayos de campo o prácticas de manejo de equipos.	60,0%	67,9%	37,5%	53,1%	50,0%
	No se puede tener interacción fluida con los alumnos.	70,0%	53,6%	50,0%	56,3%	50,0%
	No se puede controlar si el alumno está tomando o atendiendo la clase.	30,0%	14,3%	37,5%	40,6%	25,0%
	La comunicación personal frente a alumnos es superior a la comunicación virtual	40,0%	42,9%	43,8%	53,1%	50,0%
	No es factible tomar evaluaciones precisas	70,0%	46,4%	43,8%	71,9%	12,5%
	Los alumnos no se conectan o tienen dificultades para conectarse	50,0%	50,0%	37,5%	43,8%	37,5%
	Otras	20,0%	10,7%	6,3%	12,5%	25,0%

Base de cálculo: sobre el 50,5% de docentes que afirman que la efectividad es mediana

		DEDICACION		
		Exclusiva	Semiexclusiva	Simple
		% del N de columna	% del N de columna	% del N de columna
¿Cuáles diría que son las razones que a su criterio dificultarían transmitir la clase virtual con alta efectividad? (respuesta múltiple)	Imposibilidad de realizar prácticas que requieren laboratorios, ensayos de campo o prácticas de manejo de equipos.	61,3%	54,1%	50,0%
	No se puede tener interacción fluida con los alumnos.	51,6%	59,5%	53,8%
	No se puede controlar si el alumno está tomando o atendiendo la clase.	29,0%	27,0%	34,6%
	La comunicación personal frente a alumnos es superior a la comunicación virtual	45,2%	40,5%	57,7%
	No es factible tomar evaluaciones precisas	51,6%	54,1%	57,7%
	Los alumnos no se conectan o tienen dificultades para conectarse	48,4%	43,2%	42,3%
	Otras	16,1%	13,5%	7,7%

Base de cálculo: sobre el 50,5% de docentes que afirman que la efectividad es mediana

		CARRERA					
		ING. CIVIL	ING. ELECTROMECANICA	ING. INDUSTRIAL	ING. QUIMICA	TUTA SEDE CENTRAL	TUTA SEDE SUR
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Cuáles diría que son las razones que a su criterio dificultarían transmitir la clase virtual con alta efectividad? (respuesta múltiple))	Imposibilidad de realizar prácticas que requieren laboratorios, ensayos de campo o prácticas de manejo de equipos.	37,2%	44,4%	47,6%	57,9%	50,0%	70,0%
	No se puede tener interacción fluida con los alumnos.	53,5%	58,3%	57,1%	47,4%	57,1%	40,0%
	No se puede controlar si el alumno está tomando o atendiendo la clase.	25,6%	36,1%	23,8%	42,1%	35,7%	30,0%
	La comunicación personal frente a alumnos es superior a la comunicación virtual	53,5%	55,6%	47,6%	42,1%	42,9%	30,0%
	No es factible tomar evaluaciones precisas	58,1%	61,1%	50,0%	65,8%	42,9%	50,0%
	Los alumnos no se conectan o tienen dificultades para conectarse	48,8%	47,2%	42,9%	50,0%	64,3%	60,0%
	Otras	14,0%	16,7%	14,3%	5,3%	7,1%	30,0%

Base de cálculo: sobre el 50,5% de docentes que afirman que la efectividad es mediana

DEMANDA DE ASISTENCIA TECNICA PARA EL DICTADO VIRTUAL DE CLASES

El 36,6% de los docentes afirman no requerir asistencia técnica para el dictado virtual de clases, el resto necesita asistencia parcial o total. Este resultado es interesante analizar desde el punto de vista de las expectativas y percepciones.

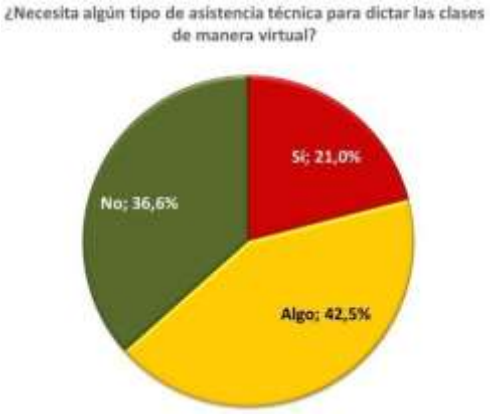


Ilustración *10*

		¿Necesita algún tipo de asistencia técnica para dictar las clases de manera virtual?			
		Sí	Algo	No	Total
		% del N de tabla	% del N de tabla	% del N de tabla	% del N de tabla
En el caso del contenido de la materia que Ud. dicta, ¿Qué grado de efectividad cree que se puede lograr transmitiéndolo de manera virtual? (asumiendo que todos los alumnos tienen acceso a internet)	Alto	5,9%	19,9%	17,2%	**43,0%**
	Medio	12,9%	20,4%	17,2%	**50,5%**
	Bajo o nulo	2,2%	2,2%	2,2%	**6,5%**
	Total	**21,0%**	**42,5%**	**36,6%**	100,0%

Tabla 7

El 36,6% que dice no necesitar asistencia técnica para el dictado virtual, está compuesto por tres segmentos:

1) una mínima fracción (2,2%) que piensa en la imposibilidad de lograr eficacia con el dictado virtual, estos constituyen el grupo de docentes que no aprueba al sistema virtual y por ende no demandan asistencia técnica para incurrir el mismo, ya que es lógico pensar que lo que se asume innecesario o inútil no se requiere ni se demanda. Este segmento representa el **2,2%** del total de la población docente que respondió la encuesta.
2) Los que no requieren asistencia técnica y opinan que la efectividad es **alta.** La lógica indica que los que no requieren asistencia técnica y opinan que la eficacia es alta, deben estar constituidos por el segmento de docentes que ya tendrían solucionado el dictado virtual y estarían teniendo experiencias favorables. Este segmento representa el **17,2%** del total de la población docente que respondió la encuesta.
3) Los que no requieren asistencia técnica y opinan que la efectividad es **mediana**. Representan también el **17,2%** del total de la población docente que respondió la encuesta. Es lógico analizar que este segmento puede estar constituido de la siguiente manera:

- Una parte que se asuma capacitado en las herramientas virtuales y que la experiencia les hace percibir que no se logra ni logrará total eficacia en transmitir conocimiento.
- Una parte que se asuma capacitado en las herramientas virtuales y que aún no estén logrando total eficacia, pero **no** están cerrados en la posibilidad de mejora en el tiempo.
- Una parte que reconoce alguna eficacia, pero sin predisposición a recibir capacitación en herramientas virtuales.

Analizando el restante 63,4% de docentes que requieren asistencia técnica total o parcial puede distinguirse otros tres segmentos más, a saber:

4) Los que requieren asistencia técnica parcial o total y opinan que la efectividad es **alta**. Constituyen el segmento de los que creen en el sistema virtual y reconocen la necesidad de capacitación en el tema. Este segmento representa el **25,8%** del total de la población docente que respondió la encuesta.
5) Los que requieren asistencia técnica parcial o total y opinan que la efectividad es **mediana**. Constituyen el segmento de los expectantes, los que tienen dudas en el sistema virtual, pero reconocen que algún beneficio otorga y es necesaria la capacitación en el tema. Este segmento es potencialmente persuasible cuando se tengan resultados concretos sobre la eficacia del sistema virtual. Representan el **33,2%** del total de la población docente que respondió la encuesta.
6) Los que requieren asistencia técnica parcial o total y opinan que la efectividad es **nula o baja**, no creen en el sistema virtual, pero se diferencian por reconocer la necesidad de capacitación en el tema, otorgarían el beneficio de la duda. Representan el **4,4%** del total de la población docente que respondió la encuesta.

		¿Necesita algún tipo de asistencia técnica para dictar las clases de manera virtual?			
		Sí	Algo	No	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	21,1%	57,9%	21,1%	100,0%
	Prof.Adj. con resp. de cátedra	24,6%	35,4%	40,0%	100,0%
	JTP. con resp. de cátedra	29,6%	37,0%	33,3%	100,0%
	JTP. sin resp. de cátedra	14,5%	45,2%	40,3%	100,0%
	Prof. sin resp. de cátedra	15,4%	53,8%	30,8%	100,0%
	Total	21,0%	42,5%	36,6%	100,0%
DEDICACION	Exclusiva	20,7%	43,1%	36,2%	100,0%
	Semiexclusiva	21,9%	45,2%	32,9%	100,0%
	Simple	20,0%	38,2%	41,8%	100,0%
	Total	21,0%	42,5%	36,6%	100,0%
CARRERA	ING. CIVIL	23,5%	39,5%	37,0%	100,0%
	ING. ELECTROMECANICA	22,2%	36,1%	41,7%	100,0%
	ING. INDUSTRIAL	26,6%	35,4%	38,0%	100,0%
	ING. QUIMICA	20,8%	33,8%	45,5%	100,0%
	TUTA SEDE CENTRAL	27,3%	31,8%	40,9%	100,0%
	TUTA SEDE SUR	27,8%	38,9%	33,3%	100,0%
	Total	21,0%	42,5%	36,6%	100,0%

Tabla 8

NECESIDADES TECNICAS PARA EL DICTADO VIRTUAL DE CLASES
Opina el 63,4% de docentes que necesitan asistencia técnica total o parcial.

La principal necesidad es de capacitación en toma de evaluaciones, le siguen las de seguimiento a alumnos y el dictado a través de reuniones virtuales.

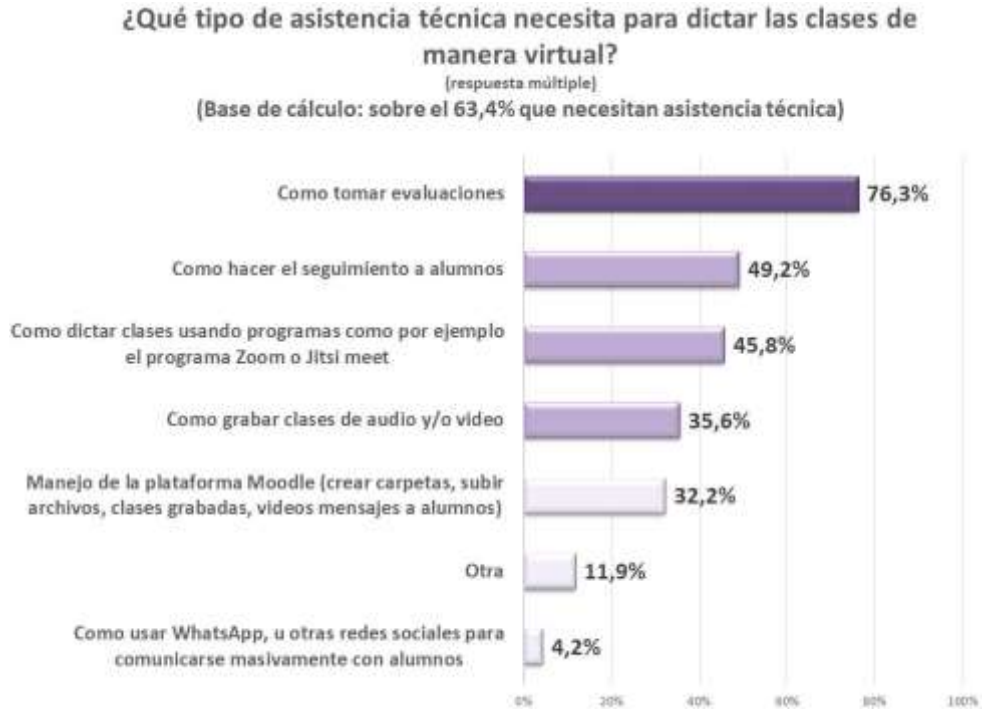


Ilustración 11

		¿Qué tipo de asistencia técnica necesita para dictar las clases de manera virtual? (respuesta múltiple)						
		Manejo de la plataforma Moodle (crear carpetas, subir archivos, clases grabadas, videos mensajes a alumnos)	Como usar WhatsApp, u otras redes sociales para comunicarse masivamente con alumnos	Como dictar clases usando programas como por ejemplo el programa Zoom o Jitsi meet	Como grabar clases de audio y/o video	Como hacer el seguimiento a alumnos	Como tomar evaluaciones	Otra
		% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	33,3%	0,0%	26,7%	20,0%	26,7%	86,7%	6,7%
	Prof.Adj. con resp. de cátedra	17,9%	5,1%	46,2%	38,5%	56,4%	82,1%	12,8%
	JTP. con resp. de cátedra	44,4%	0,0%	55,6%	38,9%	33,3%	55,6%	11,1%
	JTP. sin resp. de cátedra	40,5%	5,4%	48,6%	37,8%	59,5%	78,4%	10,8%
	Prof. sin resp. de cátedra	33,3%	11,1%	44,4%	33,3%	44,4%	66,7%	22,2%
DEDICACION	Exclusiva	27,0%	2,7%	45,9%	32,4%	51,4%	83,8%	10,8%
	Semiexclusiva	36,7%	6,1%	42,9%	26,5%	46,9%	69,4%	14,3%
	Simple	31,3%	3,1%	50,0%	53,1%	50,0%	78,1%	9,4%
CARRERA	ING. CIVIL	41,2%	7,8%	47,1%	35,3%	51,0%	78,4%	7,8%
	ING. ELECTROMECANICA	38,1%	4,8%	50,0%	40,5%	31,0%	64,3%	11,9%
	ING. INDUSTRIAL	36,7%	6,1%	44,9%	36,7%	46,9%	75,5%	8,2%
	ING. QUIMICA	35,7%	4,8%	45,2%	33,3%	45,2%	76,2%	11,9%
	TUTA SEDE CENTRAL	38,5%	15,4%	46,2%	53,8%	53,8%	69,2%	0,0%
	TUTA SEDE SUR	25,0%	8,3%	50,0%	33,3%	66,7%	83,3%	25,0%

Base de cálculo: sobre el 63,4% de docentes que necesitan asistencia técnica total o parcial.

ESTADO DE SITUACION ACTUAL

La mayoría de la población docente de la Facultad de Ingeniería de la Universidad Nacional de Salta, en todas sus carreras, afirma tener resuelto en gran medida el dictado virtual de clases. En promedio representan el 73,1% de la población docente que respondió la encuesta. Los que mayormente tienen resuelto el problema son los docentes de Ingeniería Química (79,2%) y los que están por debajo del promedio general son los de TUTA CENTRAL con el 59,1%.

Solo un 3,2% de la población docente afirma dificultes en comenzar el dictado de clases de manera virtual.

Ilustración 12

El resto se constituye en el 23,7% de docentes que tienen dificultades, pero ya han iniciado el dictado, por lo menos parcialmente, de las clases. Este segmento podría rápidamente integrarse al de los que afirman tener resuelto el problema, si se les brinda capacitación en el uso de herramientas de video conferencias, como los son el Zoom o Jitsi meet, los cuales también permiten la grabación, esto solucionaría gran parte del problema de este segmento.

El problema de la evaluación es transversal a todos los segmentos.

(ver los tablas a continuación).

		¿En cuál de las siguientes situaciones diría que se encuentra la cátedra a la que Ud. pertenece?		
		Tenemos resuelto en gran medida, el dictado de las clases teóricas y prácticas de manera virtual	Tenemos dificultades, pero ya empezamos a dictar algunas clases	Tenemos dificultades que nos impiden empezar a dictar las clases de manera virtual
		% del N de columna	% del N de columna	% del N de columna
¿Qué tipo de asistencia técnica necesita para dictar las clases de manera virtual? (respuesta múltiple)	Manejo de la plataforma Moodle (crear carpetas, subir archivos, clases grabadas, videos mensajes a alumnos)	28,8%	35,3%	75,0%
	Como usar WhatsApp, u otras redes sociales para comunicarse masivamente con alumnos	3,8%	2,9%	25,0%
	Como dictar clases usando programas como por ejemplo el programa Zoom o Jitsi meet	36,3%	**67,6%**	50,0%
	Como grabar clases de audio y/o video	26,3%	52,9%	75,0%
	Como hacer el seguimiento a alumnos	46,3%	55,9%	50,0%
	Como tomar evaluaciones	**78,8%**	**70,6%**	**75,0%**
	Otra	11,3%	11,8%	25,0%

Tabla 9

		¿En cuál de las siguientes situaciones diría que se encuentra la cátedra a la que Ud. pertenece?			
		Tenemos resuelto en gran medida, el dictado de las clases teóricas y prácticas de manera virtual	Tenemos dificultades, pero ya empezamos a dictar algunas clases	Tenemos dificultades que nos impiden empezar a dictar las clases de manera virtual	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	84,2%	10,5%	5,3%	100,0%
	Prof.Adj. con resp. de cátedra	75,4%	21,5%	3,1%	100,0%
	JTP. con resp. de cátedra	77,8%	22,2%	0,0%	100,0%
	JTP. sin resp. de cátedra	62,9%	33,9%	3,2%	100,0%
	Prof. sin resp. de cátedra	84,6%	7,7%	7,7%	100,0%
	Total	73,1%	23,7%	3,2%	100,0%
DEDICACION	Exclusiva	79,3%	19,0%	1,7%	100,0%
	Semiexclusiva	74,0%	24,7%	1,4%	100,0%
	Simple	65,5%	27,3%	7,3%	100,0%
	Total	73,1%	23,7%	3,2%	100,0%
CARRERA	ING. CIVIL	75,3%	23,5%	1,2%	100,0%
	ING. ELECTROMECANICA	75,0%	23,6%	1,4%	100,0%
	ING. INDUSTRIAL	70,9%	27,8%	1,3%	100,0%
	ING. QUIMICA	79,2%	18,2%	2,6%	100,0%
	TUTA SEDE CENTRAL	59,1%	31,8%	9,1%	100,0%
	TUTA SEDE SUR	66,7%	33,3%	0,0%	100,0%
	Total	73,1%	23,7%	3,2%	100,0%

Tabla 10

DIFICULTADES EXPRESADAS PARA EL DICTADO DE CLASES VIRTUALES
(opina el 26,9% de los docentes)

Ante la consulta al 26,9% de los docentes que manifestaron algún tipo de dificultad en el inicio de clases virtuales, puede observarse que la conectividad a internet resulta ser uno de los problemas para el 40% de esta fracción de docentes. Otra observación, es que la mitad tendría que atender situaciones particulares de cátedra. El 22% afirmó no disponer de tiempo para la preparación de clases virtuales, lo que representa sobre la población total de docentes el **5,92% sin tiempo disponible**.

Ilustración 13

		¿Qué tipo de dificultades tiene o tienen para dictar clases de manera virtual? (respuesta múltiple)					
		Mal servicio de Internet o no dispone del mismo	Difícil acceso a las herramientas virtuales necesarias para dictar clases	Cuestiones que tienen que ver con la particularidad de la cátedra	Cantidad insuficiente de personal docente con relación al nro. de alumnos	No dispone del tiempo necesario para preparar clases adaptadas al dictado virtual	Otra
		% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	66,7%	33,3%	0,0%	33,3%	0,0%	66,7%
	Prof.Adj. con resp. de cátedra	18,8%	25,0%	62,5%	25,0%	43,8%	31,3%
	JTP. con resp. de cátedra	66,7%	33,3%	66,7%	0,0%	0,0%	0,0%
	JTP. sin resp. de cátedra	43,5%	17,4%	39,1%	13,0%	26,1%	34,8%
	Prof. sin resp. de cátedra	50,0%	0,0%	100,0%	0,0%	50,0%	0,0%
DEDICACION	Exclusiva	33,3%	16,7%	66,7%	25,0%	25,0%	25,0%
	Semiexclusiva	47,4%	21,1%	57,9%	15,8%	21,1%	36,8%
	Simple	36,8%	26,3%	31,6%	10,5%	36,8%	26,3%
CARRERA	ING. CIVIL	35,0%	25,0%	50,0%	15,0%	5,0%	30,0%
	ING. ELECTROMECANICA	50,0%	5,6%	50,0%	16,7%	16,7%	16,7%
	ING. INDUSTRIAL	39,1%	8,7%	52,2%	30,4%	39,1%	30,4%
	ING. QUIMICA	50,0%	6,3%	50,0%	25,0%	37,5%	18,8%
	TUTA SEDE CENTRAL	11,1%	33,3%	44,4%	22,2%	33,3%	11,1%
	TUTA SEDE SUR	0,0%	33,3%	50,0%	33,3%	16,7%	16,7%

Base de cálculo: sobre el 26,9% que expresa alguna dificultad.

OPINION SOBRE LA PARTICIPACIONDE ALUMNOS

Es importante destacar que existe correlatividad entre la percepción de grado de eficacia de la clase virtual con la de conectividad de los alumnos.

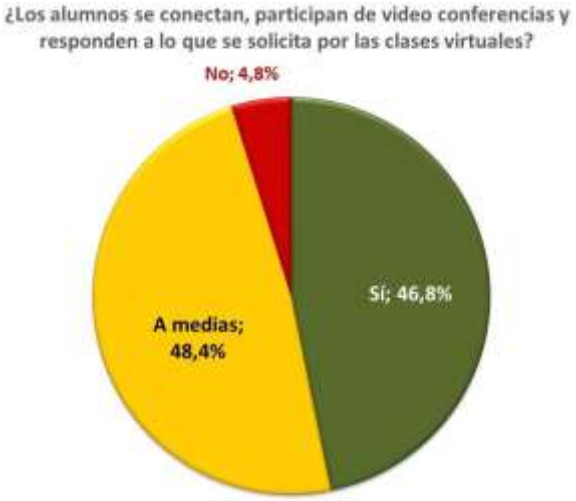


Ilustración 14

		¿Los alumnos se conectan, participan de video conferencias y responden a lo que se solicita por las clases virtuales?			
		Sí	A medias	No	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	52,6%	42,1%	5,3%	100,0%
	Prof.Adj. con resp. de cátedra	53,8%	41,5%	4,6%	100,0%
	JTP. con resp. de cátedra	40,7%	48,1%	11,1%	100,0%
	JTP. sin resp. de cátedra	41,9%	54,8%	3,2%	100,0%
	Prof. sin resp. de cátedra	38,5%	61,5%	0,0%	100,0%
	Total	46,8%	48,4%	4,8%	100,0%
DEDICACION	Exclusiva	53,4%	41,4%	5,2%	100,0%
	Semiexclusiva	43,8%	52,1%	4,1%	100,0%
	Simple	43,6%	50,9%	5,5%	100,0%
	Total	46,8%	48,4%	4,8%	100,0%
CARRERA	ING. CIVIL	45,7%	48,1%	6,2%	100,0%
	ING. ELECTROMECANICA	45,8%	50,0%	4,2%	100,0%
	ING. INDUSTRIAL	45,6%	50,6%	3,8%	100,0%
	ING. QUIMICA	50,6%	42,9%	6,5%	100,0%
	TUTA SEDE CENTRAL	22,7%	54,5%	22,7%	100,0%
	TUTA SEDE SUR	33,3%	61,1%	5,6%	100,0%
	Total	46,8%	48,4%	4,8%	100,0%

Tabla 11

		En el caso del contenido de la materia que Ud. dicta, ¿Qué grado de efectividad cree que se puede lograr transmitiéndolo de manera virtual? (asumiendo que todos los alumnos tienen acceso a internet)		
		Alto	Medio	Bajo o nulo
		% del N de columna	% del N de columna	% del N de columna
¿Los alumnos se conectan, participan de video conferencias y responden a lo que se solicita por las clases virtuales?	Sí	75,0%	27,7%	8,3%
	A medias	22,5%	68,1%	66,7%
	No	2,5%	4,3%	25,0%
	Total	100,0%	100,0%	100,0%

Tabla 12

OPINION SOBRE LA SOLUCION PARA LA TOMA DE EVALUACIONES

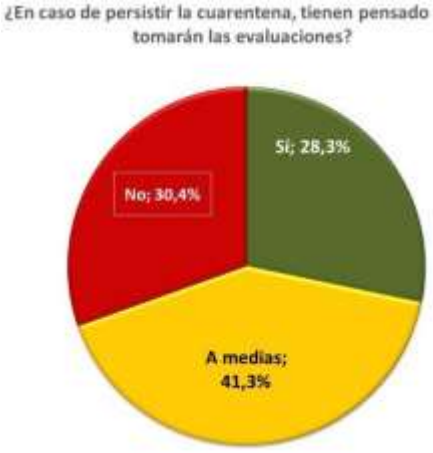


Ilustración 15

		¿En caso de persistir la cuarentena, tienen pensado como tomarán las evaluaciones?			
		Si	A medias	No	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	31,6%	31,6%	36,8%	100,0%
	Prof.Adj. con resp. de cátedra	31,3%	43,8%	25,0%	100,0%
	JTP. con resp. de cátedra	38,5%	34,6%	26,9%	100,0%
	JTP. sin resp. de cátedra	21,0%	43,5%	35,5%	100,0%
	Prof. sin resp. de cátedra	23,1%	46,2%	30,8%	100,0%
	Total	28,3%	41,3%	30,4%	100,0%
DEDICACION	Exclusiva	29,8%	36,8%	33,3%	100,0%
	Semiexclusiva	23,3%	47,9%	28,8%	100,0%
	Simple	33,3%	37,0%	29,6%	100,0%
	Total	28,3%	41,3%	30,4%	100,0%
CARRERA	ING. CIVIL	27,5%	37,5%	35,0%	100,0%
	ING. ELECTROMECANICA	35,7%	35,7%	28,6%	100,0%
	ING. INDUSTRIAL	29,5%	43,6%	26,9%	100,0%
	ING. QUIMICA	27,6%	43,4%	28,9%	100,0%
	TUTA SEDE CENTRAL	22,7%	31,8%	45,5%	100,0%
	TUTA SEDE SUR	23,5%	29,4%	47,1%	100,0%
	Total	28,3%	41,3%	30,4%	100,0%

Tabla 13

COMPARACION ENTRE CLASES PRESENCIALES Y VIRTUALES

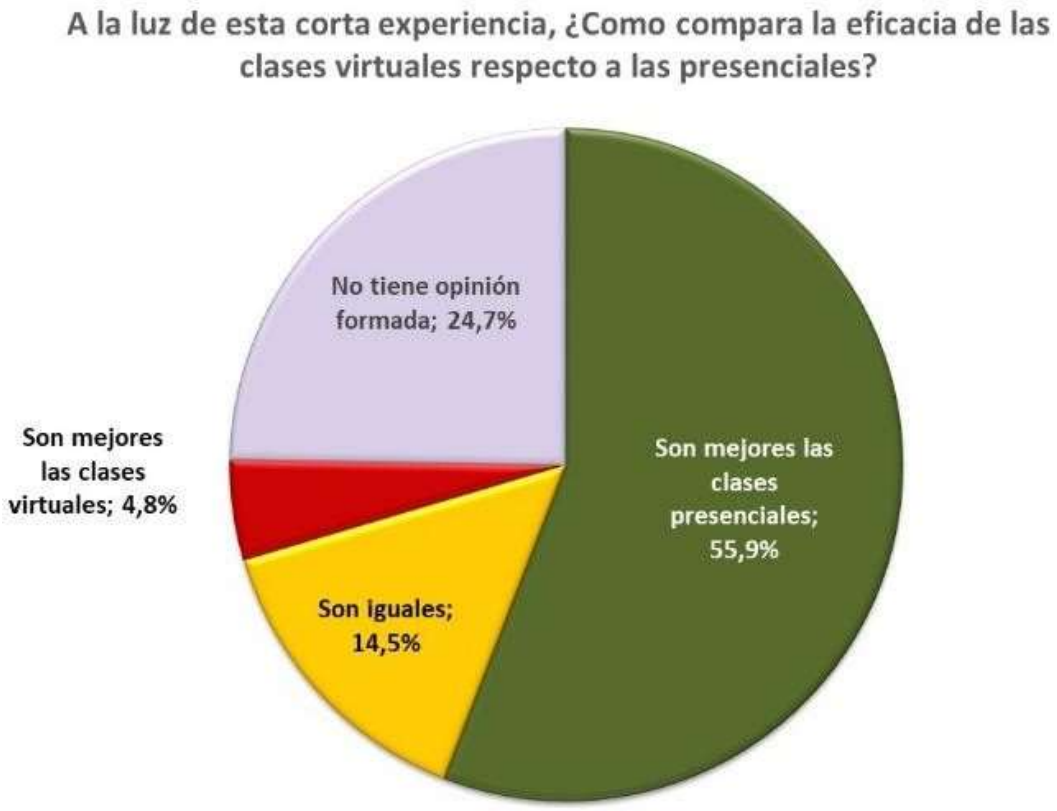


Ilustración 16

El resultado es esperable, la mayoría piensa que son mejores las clases presenciales, es natural oponerse a nuevos cambios, requiere tiempo, máximo en esta situación impuesta de repente por la pandemia. En este marco es significativo el resultado del 14,5% de docentes que opinan que las clases virtuales y presenciales son iguales, a éstos se les suman el 4,8% de docentes que valoran como mejor a las clases virtuales. Se observa un 24,7% de quienes no emiten opinión, es decir esperan ver resultados para evaluar, esto significa que no tienen preconceptos negativos respecto a la clase virtual.

Puede observarse que en la medida que el docente piensa que es alto el grado de efectividad de la clase virtual, aumenta la proporción que opina que es mejor a la presencial, o por lo menos igual.

		A la luz de esta corta experiencia, ¿Como compara la eficacia de las clases virtuales respecto a las presenciales?				
		Son mejores las clases presenciales	Son iguales	Son mejores las clases virtuales	No tiene opinión formada	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	42,1%	15,8%	5,3%	36,8%	100,0%
	Prof.Adj. con resp. de cátedra	52,3%	15,4%	6,2%	26,2%	100,0%
	JTP. con resp. de cátedra	70,4%	18,5%	0,0%	11,1%	100,0%
	JTP. sin resp. de cátedra	59,7%	9,7%	6,5%	24,2%	100,0%
	Prof. sin resp. de cátedra	46,2%	23,1%	0,0%	30,8%	100,0%
	Total	55,9%	14,5%	4,8%	24,7%	100,0%
DEDICACION	Exclusiva	41,4%	19,0%	3,4%	36,2%	100,0%
	Semiexclusiva	58,9%	16,4%	4,1%	20,5%	100,0%
	Simple	67,3%	7,3%	7,3%	18,2%	100,0%
	Total	55,9%	14,5%	4,8%	24,7%	100,0%
CARRERA	ING. CIVIL	60,5%	11,1%	4,9%	23,5%	100,0%
	ING. ELECTROMECANICA	62,5%	11,1%	4,2%	22,2%	100,0%
	ING. INDUSTRIAL	51,9%	12,7%	6,3%	29,1%	100,0%
	ING. QUIMICA	53,2%	18,2%	3,9%	24,7%	100,0%
	TUTA SEDE CENTRAL	63,6%	9,1%	0,0%	27,3%	100,0%
	TUTA SEDE SUR	61,1%	11,1%	0,0%	27,8%	100,0%
	Total	55,9%	14,5%	4,8%	24,7%	100,0%

Tabla 14

		En el caso del contenido de la materia que Ud. dicta, ¿Qué grado de efectividad cree que se puede lograr transmitiéndolo de manera virtual? (asumiendo que todos los alumnos tienen acceso a internet)		
		Alto	Medio	Bajo o nulo
		% del N de columna	% del N de columna	% del N de columna
A la luz de esta corta experiencia, ¿Como compara la eficacia de las clases virtuales respecto a las presenciales?	Son mejores las clases presenciales	30,0%	73,4%	91,7%
	Son iguales	26,3%	6,4%	0,0%
	Son mejores las clases virtuales	11,3%	0,0%	0,0%
	No tiene opinión formada	32,5%	20,2%	8,3%
	Total	100,0%	100,0%	100,0%

Tabla 15

<

PERCEPCIÓN DE CONSECUENCIAS

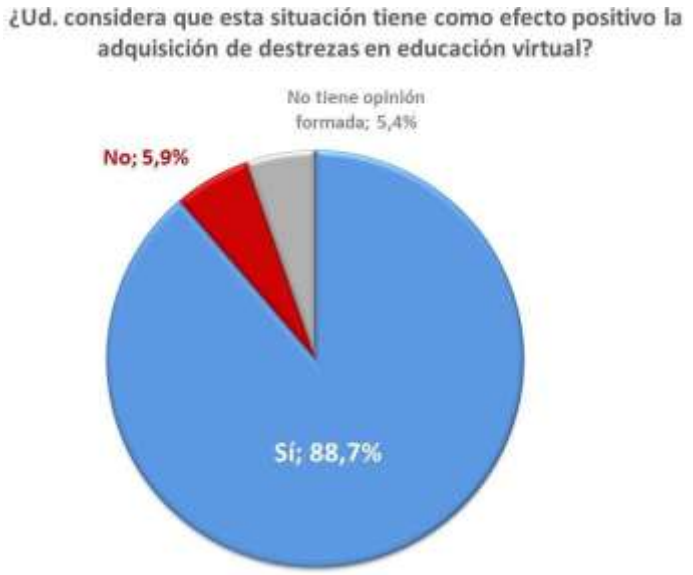


Ilustración 17

		¿Ud. considera que esta situación tiene como efecto positivo la adquisición de destrezas en educación virtual?			
		Sí	No	No tiene opinión formada	Total
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
TIPOLOGIA DOCENTE	Prof.Tit. o Aso. con resp. de cátedra	89,5%	0,0%	10,5%	100,0%
	Prof.Adj. con resp. de cátedra	92,3%	4,6%	3,1%	100,0%
	JTP. con resp. de cátedra	92,6%	7,4%	0,0%	100,0%
	JTP. sin resp. de cátedra	82,3%	9,7%	8,1%	100,0%
	Prof. sin resp. de cátedra	92,3%	0,0%	7,7%	100,0%
	Total	88,7%	5,9%	5,4%	100,0%
DEDICACION	Exclusiva	89,7%	5,2%	5,2%	100,0%
	Semiexclusiva	90,4%	5,5%	4,1%	100,0%
	Simple	85,5%	7,3%	7,3%	100,0%
	Total	88,7%	5,9%	5,4%	100,0%
CARRERA	ING. CIVIL	85,2%	8,6%	6,2%	100,0%
	ING. ELECTROMECANICA	86,1%	8,3%	5,6%	100,0%
	ING. INDUSTRIAL	89,9%	5,1%	5,1%	100,0%
	ING. QUIMICA	84,4%	7,8%	7,8%	100,0%
	TUTA SEDE CENTRAL	81,8%	13,6%	4,5%	100,0%
	TUTA SEDE SUR	88,9%	5,6%	5,6%	100,0%
	Total	88,7%	5,9%	5,4%	100,0%

Tabla 16

PRINCIPALES CONCLUSIONES

Se puede dar solución a gran parte del problema, mejorando la conectividad; ampliando la capacidad de la plataforma Moodle para alojar videos; continuar con la capacitación continua del uso de la plataforma Moodle; iniciar capacitación en el uso de programas para reuniones virtuales que se puedan aplicar para a impartir clases; realización de grabaciones de clases y manejo de herramientas para la toma de evaluaciones.

Si bien aún la mayoría tiene la percepción que las clases presenciales son mejores a las virtuales, no se constituyen en una población pasiva ante el problema impuesto por la Pandemia del Covid-19. El 73,1% afirman tener resuelto el dictado virtual de clases, mientras que otro 23,7% expresa que están realizando el esfuerzo en resolver el dictado a través de la tecnología virtual, lo que demuestra una población altamente proactiva.

La situación docente se presenta favorable y puede resumirse en los siguientes seis sectores:

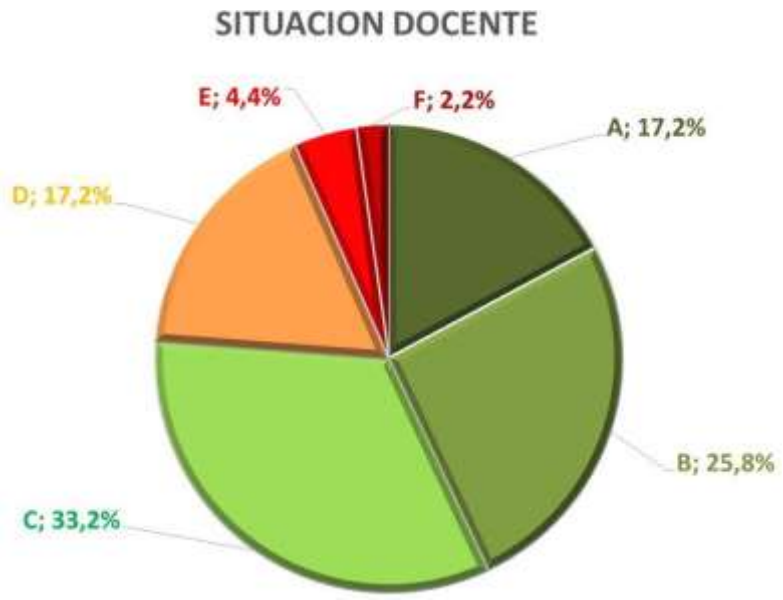


Ilustración 18

A) 17,2% No requieren asistencia técnica y opinan que la efectividad de las clases virtuales es alta. Tendrían solucionado el dictado virtual. Sector altamente favorable a la virtualidad.

B) 25,8% Si requieren algún grado de asistencia técnica y opinan que la efectividad de las clases virtuales es alta. Creen en el sistema virtual y reconocen la necesidad de capacitación. Sector a favor de la virtualidad.

C) 33,2% Si requieren algún grado de asistencia técnica y opinan que la efectividad de las clases virtuales es mediana. Dudan en el sistema virtual, pero reconocen que algún

beneficio otorga y es necesaria la capacitación. Sector con alta potencialidad a ser persuasible en estar a favor de la virtualidad.

D) 17,2% No requieren asistencia técnica y opinan que la efectividad es mediana. Son expectantes y constituyen el Sector de mediana potencialidad a ser persuasible en estar a favor de la virtualidad.

E) 4,4% Si requieren algún grado de asistencia técnica, pero opinan que la efectividad de las clases virtuales es nula o baja. No creen en el sistema virtual, pero le otorgan el beneficio de la duda.

F) 2,2% Perciben imposibilidad de lograr eficacia con el dictado virtual, no lo aprueban y no demandan asistencia técnica.

BIBLIOGRAFÍA

[1] Greimas, A. J., *Semántica estructural. Investigaciones metodológicas* [1966], Madrid, 1971; "Elementos para una teoría de la interpretación del relato mítico", en *Communications*, nº 8 (trad.: *Análisis estructural del relato*, pp. 45-86, y luego recogido en *Du Sens. Essais sémiotiques*, París, 1970, pp. 185-230 (también en *En torno al sentido*, Madrid, 1973, pp. 219-269, con el título de "Contribución a la teoría de la interpretación del relato mítico"); *La semiótica del texto: ejercicios prácticos. Análisis de un cuento de Maupassant*, Barcelona, Buenos Aires, 1983.
[2] Rodríguez Gómez G., Gil Flores J. y García Jiménez E. (1996) *Metodología de la Investigación Cualitativa. España.* Aljibe.
[3] Garrat, D. (2003) My qualitative research journey. Cresskill, New Jersey: Hampton Press Inc.
Briones G. “Métodos y Técnicas de Investigación”. Trillas 1995.
[4] Hernández, Fernández Baptista. “Metodología de la Investigación”. McGraw Hill 1994. Colombia.
[5] Padua J. “Técnicas de Investigación” FCE-Colegio de México 1982, México.
Sabino, Carlos A. El Proceso de Investigación. Buenos Aires: Edit. Lumen.1996
Salkind, Neil J. Métodos de Investigación. México: Prentice Hall. 1999.
[6] Sierra Bravo R. Tecnicas de investigación Social Teoría y ejercicios, Décima edición, Editorial Paraninfo 1995 Madrid.
[7] Taylor, S.J. y R. Bogdan. Introducción a los métodos cualitativos de investigación. Barcelona: Paidós. 1987.

10 y 11 de Diciembre de 2020

Análisis Inteligente de la población carcelaria de la Provincia de Jujuy

Guillermo S. Ávila, José H. Farfán, Mariela Rodriguez

Facultad de Ingeniería – Universidad Nacional de Jujuy - Jujuy[1]

avilags@gmail.com[1] jhfarfan@fi.unju.edu.ar[1],
mariela.rodriguez@fi.unju.edu.ar[1]

RESUMEN

El presente trabajo procesa los datos recogidos el 31 de diciembre de 2019 por el censo de la población carcelaria realizado por el Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP), estos datos son de dominio público y son provisto en la página web del Ministerio de Justicia y Derechos Humanos de la República Argentina. Filtra el DataSET con el objetivo de describir estadísticamente las características de la población carcelaria de los establecimientos penitenciarios existentes en la Provincia de Jujuy. Tomando la calificación conductual de los detenidos como objetivo de estudio, busca identificar los atributos más relevantes, y con ellos desarrollar reglas de clasificación inteligente. Para esto muestra los resultados de la experimentación utilizando diferentes algoritmos de clasificación y comparando sus grados de certezas para los modelos desarrollados por los mismos. Este análisis inteligente de la población carcelaria de la provincia de Jujuy, servirá para explicar estadísticamente las problemáticas que aquejan no solo a los detenidos sino a toda la sociedad jujeña.

Palabras Claves: SNEEP, Minería de Datos, establecimientos penitenciarios, población carcelaria, patrones.

INTRODUCCIÓN

El 16 de julio de 2020 dos presos fallecieron en la provincia de Jujuy, luego de tres horas de protesta y quema de colchones en el Establecimiento Penitenciario Nº1 de Internos Varones

Mayores del barrio Alto Gorriti de la Provincia de Jujuy, donde el reclamo se ocasiona por supuestos casos de coronavirus y diversos pedidos vinculados al régimen de visita, provisión de alimentos y aceleración de en la resolución de causa penales.
Para la realización del presente trabajo los autores se han basado en el DataSet provisto por el Ministerio de Justicia y Derechos Humanos, el cual brinda datos de dominio público el Sistema Nacional de Estadísticas sobre Ejecución de la Pena o SNEEP.
En este conjunto de datos se detalla la información recolectada en el censo realizado sobre el total de la población detenida al día 31 de diciembre de cada año, en cada establecimiento de la República Argentina. Los datos relevados corresponden a la Provincia de Jujuy del año 2019 y sobre ellos se intenta encontrar los patrones más relevantes para la obtención de las calificaciones de conducta, aplicando para ello técnicas de Minería de Datos y la herramienta para el análisis de datos Rapidminer.
Para llevar a cabo el cumplimiento de los objetivos propuestos se procede a utilizar la metodología KDD junto con la herramienta Rapidminer, un software Open-Source para el análisis y minería de datos, en su versión 9.7 con Licencia Estudiantil.
La unidad de análisis son las personas alojadas en los establecimientos. El censo recaba la siguiente información sobre cada interno: edad, sexo, nacionalidad, estado civil, nivel de instrucción, situación laboral, lugar de residencia, jurisdicción judicial, situación legal, fecha de detención, fecha de condena, establecimiento de procedencia, tipo de delitos imputado, participación en trabajo remunerado, en actividades de capacitación laboral, en actividades recreativas, asistencia médica, vistas, alteraciones al orden, sanciones disciplinarias, **calificaciones de conducta**, tentativas de fugas o evasiones, tentativa de suicidios, lesiones recibidas, duración de la condena, medidas de seguridad, reincidencia, régimen de progresividad, salidas transitorias, régimen de semilibertad, programa de prelibertad, prisión discontinua, semidetención, reducción de pena, mujeres alojadas con sus hijos.

METODOLOGÍA

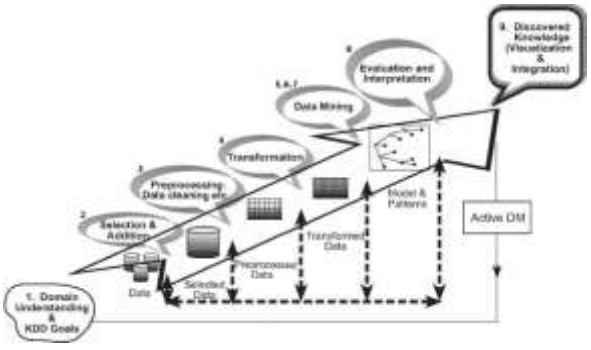


Figura 1. Esquema del Descubrimiento de Conocimiento en Base de Datos [1]

La Minería de Datos es el núcleo de todo un proceso llamado Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery in Databases – KDD) (ver fig. 1), el cual es un proceso metodológico para encontrar un “modelo” válido, útil y entendible que describa patrones de acuerdo a la información, y el modelo es la representación que intenta explicar ese patrón en los datos. KDD abarca los siguientes 9 pasos: 1) abstracción del escenario, 2) selección de datos, 3) limpieza y pre-procesamiento, 4) transformación de los datos, 5) elección de tareas de Minería de Datos, 6) elección del algoritmo, 7) aplicación del algoritmo, 8) evaluación e interpretación y 9) entendimiento del conocimiento.

Se parte de un conjunto de datos, el cual es una colección de información, ya sea cuantitativa o cualitativa y que está compuesto por variables o atributos (columnas) que representan las propiedades de un fenómeno o suceso, y casos (filas) que significan los diferentes sucesos que se presentaron en el escenario. Esto constituye es la materia prima del KDD. Ahora, aquí sus fases:

1 – Abstracción del escenario

Busca entender la problemática y el contexto para proponer soluciones viables y reales. Conociendo las propiedades, limitaciones y reglas del escenario en estudio, para posteriormente definir las metas a alcanzar.

2 – Selección de los datos

Del conjunto de datos recolectados y ya definidos los objetivos por alcanzar, se deben elegir datos disponibles para realizar el estudio e integrarlos en uno solo que puedan favorecer a llegar a alcanzar a los objetivos del análisis.

3 – Limpieza y pre-procesamiento

En esta etapa se determina la confiabilidad de la información, es decir, realizar tareas que garanticen la utilidad de los datos. Para esto se hace la limpieza de datos (tratamiento de datos perdidos o remover valores atípicos). Esto implica eliminar variables o atributos con datos faltantes o eliminar información no útil para este tipo de tareas.

4 – Transformación de los datos

En esta etapa se mejora la calidad de los datos con transformaciones que involucran ya sea reducción de dimensionalidad o bien transformaciones como por ejemplo convertir los valores que son números a categóricos (discretización).

5 – Selección de la apropiada tarea de Minería de Datos

Fase en la que se elige el paradigma apropiado de Minería de Datos, ya sea la clasificación, regresión o agrupación, según los objetivos que se haya planteado para la investigación (predicción o descripción), la primera ocupada para encontrar un modelo que sea utilizada para casos futuros y desconocidos; mientras que la segunda solo para observar su comportamiento.

6 – Elección del algoritmo de Minería de Datos

Posteriormente se procede a seleccionar la técnica o algoritmo, o incluso más de uno para la búsqueda del patrón y obtener conocimiento. El meta-aprendizaje se enfoca en explicar la razón por la que un algoritmo funciona mejor en determinadas problemáticas, y para cada técnica existen diferentes posibilidades de cómo seleccionarlas.

7 – Aplicación del algoritmo

Una vez seleccionado las técnicas el paso siguiente es aplicarlo a los datos ya seleccionados, limpiados y procesados. Es posible que la ejecución de los algoritmos sean varias intentando ajustar los parámetros que optimicen los resultados.

8 – Evaluación

Una vez aplicado los algoritmos al conjunto de datos, procedemos a evaluar los patrones que se generaron y el rendimiento que se obtuvo para verificar que cumpla con las metas planteadas en las primeras fases. Para realizar esta evaluación existe una técnica que se llama Validación Cruzada , el cual realiza una partición de los datos dividiéndose en entrenamiento y prueba

9 – Aplicación

Si todos los pasos se siguen correctamente y los resultados de la evaluación se satisfacen, la última etapa es simplemente aplicar el conocimiento encontrado al contexto y comenzar a resolver sus problemáticas. Si de lo contrario, los resultados no son satisfactorios entonces es necesario regresar a las anteriores etapas a realizar algún ajuste, analizando desde la selección de los datos hasta en la etapa de evaluación.

DESARROLLO

Para la etapa de desarrollo se tiene en cuenta los procesos realizados según cada uno de los objetivos propuestos.

Compresión del dominio y establecimiento de los objetivos

Para esto se ha realizado una investigación sobre la temática abordada a un problema de dominio público, con respecto al censo realizado en Diciembre del año 2019. La fuente se obtiene del Ministerio de Justicia y Derechos de la Nación. El dataSet cuenta con un total de 100.634 encuestas a privados de la libertad de Argentina, el objetivo de este trabajo es analizar la población carcelaria de Jujuy y se en este caso se cuenta con 1157 registros de Jujuy.

Objetivos: Para la realización del presente trabajo se persiguen tres objetivos:

- Describir estadísticamente la población carcelaria de la Provincia de Jujuy.
- Filtrar los 4 factores más influyentes en la calificación conductual de los presos.
- Determinar un modelo de clasificación para la evaluación-conductual y los parámetros que más influyen en ella.

En todo aprendizaje de Minería de Datos es importante definir un label u objetivo del conjunto de atributos del DataSet de trabajo. En este trabajo se ha definido al atributo label como "clasificación_conducta", en base al análisis de todos los atributos del DataSet original.

Selección

En esta etapa se realiza una selección de los datos excluyendo datos redundantes, faltantes o datos erróneos que no se ajustan a los objetivos a los que se pretende llegar, contribuyendo al rendimiento del modelo generado, a través de la reducción de las dimensiones de los datos y al nivel de procesamiento generado en el equipo informático.

Para la visualización de los datos se procede a utilizar la codificación de archivo UTF-8, tal como se muestra en el gráfico Nº1:

Gráfico Nº1: codificación del archivo de datos.

En la etapa de limpieza de datos se filtra el DataSet dejando solo los datos correspondientes únicamente a la provincia de Jujuy contabilizando, contabilizando un total de 1157 personas, tal como se observa en el gráfico Nº2.

Gráfico Nº2: filtrado del DataSet para la provincia de Jujuy.

Descripción gráfica estadística de la población carcelaria de la Provincia de Jujuy

Se considera necesario que, para el cumplimiento de este objetivo, se debe tener en cuenta el uso del análisis visual de los principales atributos descriptivos de los 87 que posee el DataSet de estudio..

Distribución por sexo de la población

El primer análisis trabajado es el de la distribución del Sexo de la población carcelaria, para ello se obtiene el gráfico Nº3 que se muestra a continuación:

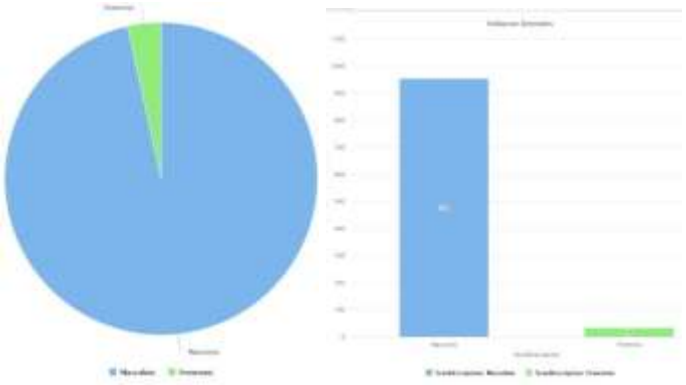


Gráfico Nº3: distribución por sexo de la población carcelaria.

Se puede observar que gran parte de la población son de sexo masculino y una ínfima proporción pertenece al sexo femenino. Esto puede deberse, según en opinión de los autores del actual trabajo a muchas razones, por ejemplo que la población carcelaria femenina muchas veces no tienen condenas firmes por dudas o falta de pruebas y cuando la poseen pueden tener el beneficio de la prisión domiciliaria.

Distribución por edades

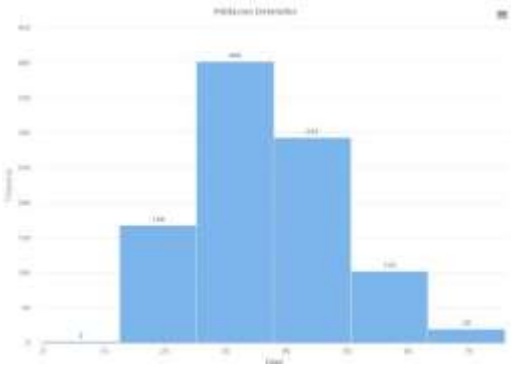


Gráfico Nº4: distribución por edades.

Otro análisis muy interesante a realizar es el de la distribución por edades de la población carcelaria tal como se muestra en el gráfico Nº4.

Se puede observar que el rango etario de la población de reclusos está concentrado mayormente entre los 25 y 38 años aproximadamente. Aunque también es importante la cantidad de detenidos menores de edad. Tomando como referencia a una eminencia como lo es el Dr. Abel Albino[1] en la república argentina, esto puede deberse a la falta de educación debido a que estas personas no pudieron desarrollar adecuadamente su cerebro durante los primeros años de vida, teniendo como consecuencias problemas de salud graves (como por ejemplo la desnutrición) y que consecuentemente afectan sus impulsos (más violencia, irritabilidad, uso de armas, drogas, etc) que desembocan en actos de delincuencia.

□ Distribución por Delito

Tal como se observa en el Gráfico Nº5 llama la atención que el principal delito causante de la pena de privación de la libertad sea Violaciones. Luego le siguen Robos, Homicidios Dolosos y por causas de Estupefacientes (ley 23.727). Esto puede ser debido a que las violaciones al ser un delito tan grave y horroroso para la víctima la Justicia lo castiga con severidad, no así los delitos contra la propiedad como robos y hurtos que para tal fin deben acumularse las causas y pesan aquellos con los que se actuó de gravedad, también es importante destacar que para los delitos contra la propiedad la víctima por descreimiento de la justicia no radica la denuncia porque se piensa que la justicia no va a actuar o el delincuente saldrá rápidamente en libertad.

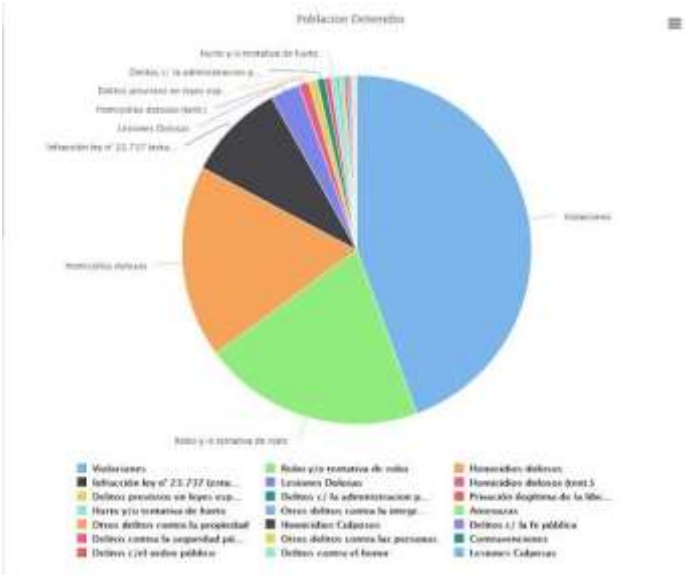


Gráfico Nº5: distribución por delitos.

Como se observa en el gráfico Nº6, 3 de cada 10 privados de la libertad aún no tiene condena. Debido a que muchas veces estos casos están siendo judicializados. Por ejemplo, el caso de Alexis Mamani, el niño que fue asesinado por su madre que aún no tiene condena firme.

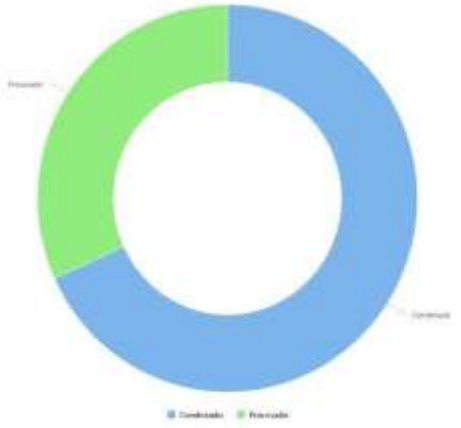


Gráfico Nº6: distribución de población carcelaria con condena firme.

□ Distribución por establecimiento y Jurisdicción

En el gráfico Nº 7, el color celeste muestra la población de los Establecimientos de Jurisdicción Provincial y el color verde los de Jurisdicción Federal. Como se puede ver, el establecimiento con más población es la penitenciaria N°1 de Gorriti.

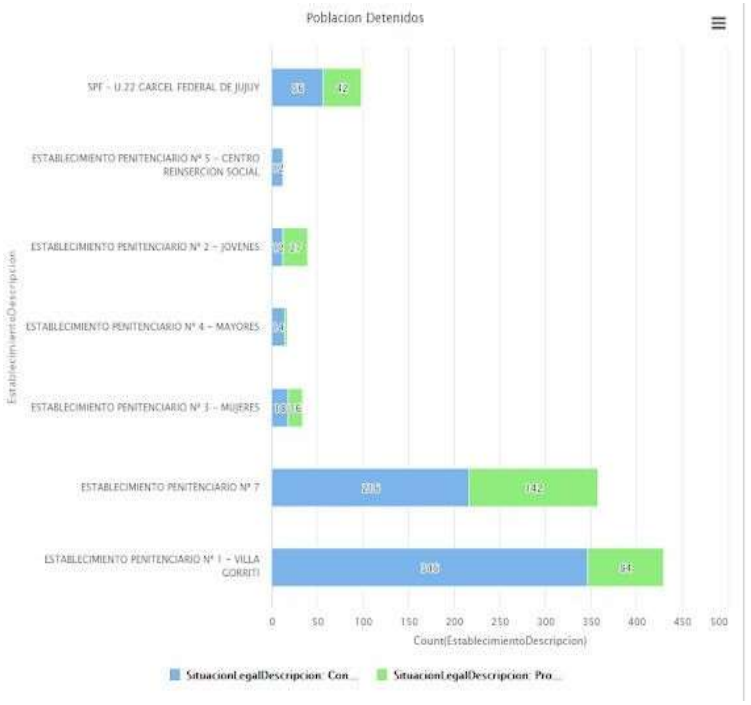


Gráfico Nº7: distribución de población carcelaria con condena firme.

En el gráfico Nº 8 se muestra la población de los Establecimientos pero se distingue en verde los que presentan procesos y en celeste los que ya poseen condena.

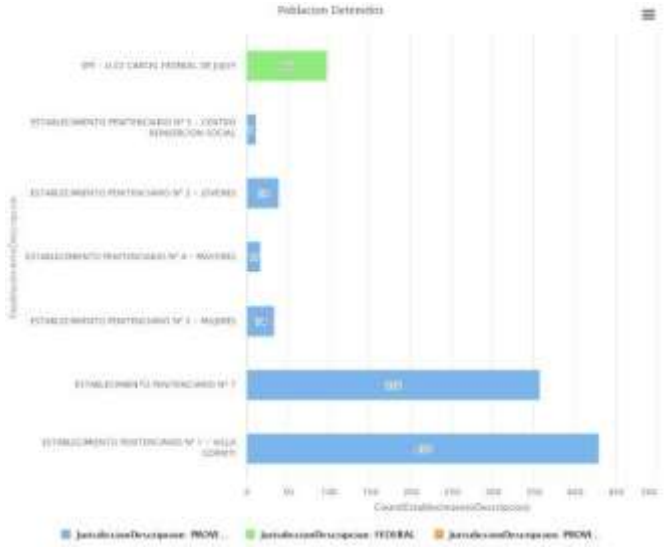


Gráfico Nº8: distribución de población carcelaria con condena firme.

Distribución de la Calificación de Conducta

En el gráfico Nº 9 en relación a la calificación de conducta predomina la conductual Buena, Regular y Muy buena (con más del 75% de la población sumada estas tres categorías). Se podría inferir que esto sucede porque los reclusos quieren ver reducida su condena mediante actos de buenas conductas o tener la oportunidad de recibir otros beneficios.

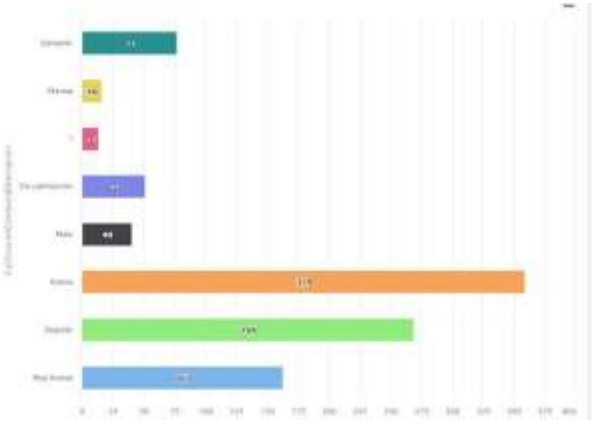


Gráfico Nº9: distribución de la calificación de la conducta.

- **Filtrar los 4 factores más influyentes en la calificación conductual de los presos**

El DataSet cuenta con atributos nominales y sus correspondientes id, el cual clasifica en forma numérica dichos atributos, como en este DataSet no todos los atributos numéricos cuentan con su descripción, entre los atributos numéricos los candidatos en donde se establece el atributo CalificacionConductual como label y se buscan entre el resto de parámetros numéricos los parámetros más influyentes.

Se seleccionan los atributos candidatos como se ve en el gráfico Nº10:

Gráfico Nº10: atributos seleccionados.

Y se selecciona el label u objetivo de clasificación, como se ve en el gráfico Nº11:

Gráfico Nº11: atributos seleccionados.

Se usa el operador Weight by Correlation para ponderar los atributos con respecto al label objetivo:

Gráfico Nº12: atributos seleccionados.

El el gráfico Nº13 se observa que los patrones con más peso en la Calificación de Conductas son:

1. Participación en Programa Laboral.
2. Atención Médica en el Último Año.
3. Participación en el Programa Educativo.
4. Sanciones Aplicadas.

Y esto son los que se utilizaran en los la búsqueda de un modelo de clasificación.

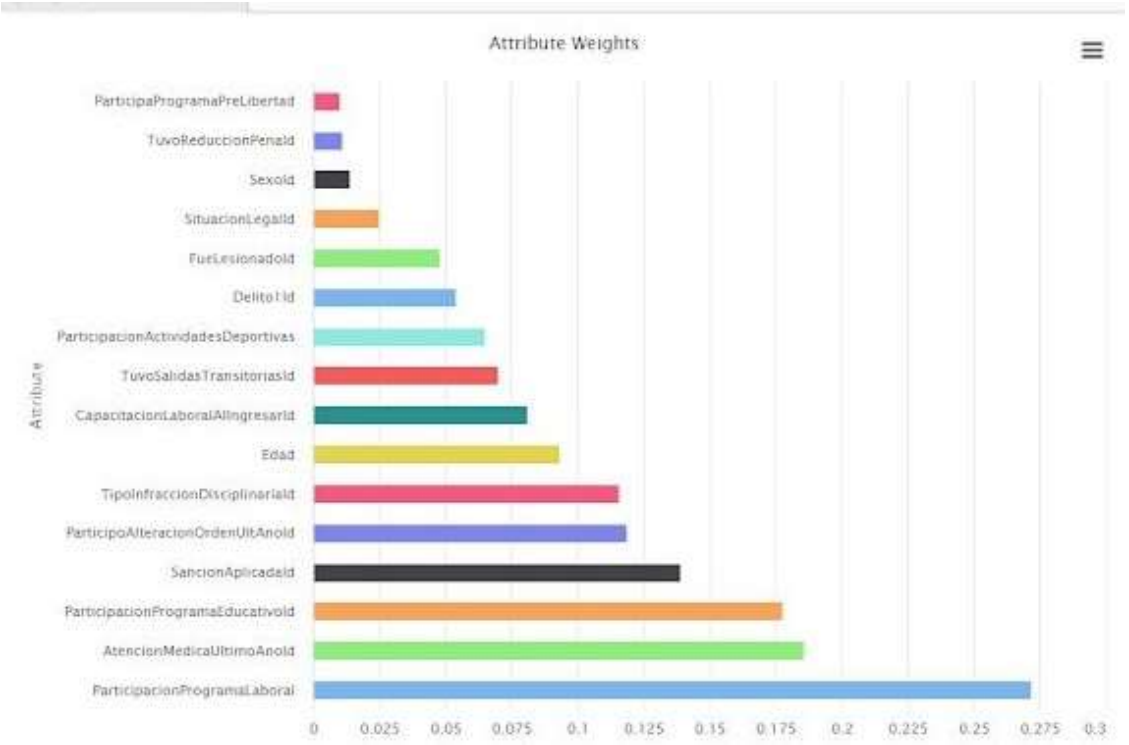


Gráfico Nº13: atributos seleccionados.

El el cuadro Nº1 se observa que el Identificador (ID) y la descripción de los valores que pueden tomar los atributos con los que se trabaja

ATRIBUTO	ID	DESCRIPCION
1. Participación en Programa Laboral.	0	NO PARTICIPA
	1	PARTICIPA
2. Atención Médica en el Último Año.	1	SI
	2	NO
	11	SI - ASMA O EPOC
	12	SI - CHAGAS
	13	SI - DIABETES
	14	SI - ENFERMEDADES MENTALES
	15	SI - HEPATITIS
	16	SI - HIPERTENSIÓN
	17	SI - HIV
	18	SI - TUBERCULOSIS
	19	SI - OTRAS ENFERMEDADES
3. Participación en el Programa Educativo.	1	Si - educación formal - PRIMARIA (EGB)
	2	Si - educación formal - SECUNDARIA (Polimodal)
	3	Si - educación formal - TERCIARIO
	4	Si - educación formal - UNIVERSITARIA
	5	Si - educación no formal (Cursos)
	6	No participa de ningún programa educativo
4. Sanciones Aplicadas.	1	Traslado a otro estab. régimen mas severo
	2	Traslado a otra sección régimen mas severo
	3	Perm. aloj. indiv. o celda hasta 7 f/de Sem sus/al
	4	Perm. aloj. indiv. o celda hasta 15 días inint.
	5	Susp/restric tot/parc. derec. regl. hasta 15 días
	6	Exclus. de activ. común hasta 15 días
	7	Exclus. de activ. recreat/dep. hasta 10 días
	8	Amonestación
	9	Otra Sanción
Calificacion Conducta	1	Ejemplar
	2	Muy buena
	3	Buena
	4	Regular
	5	Mala
	6	Pésima
	7	Sin calificación

Cuadro Nº1: Atributos. Identificadores y Descripción.

- **Determinar un modelo de clasificación para la evaluación-conductual y los factores que más influyen en ella.**

Minería de Datos

Mediante los atributos seleccionados en el paso anterior se procede a buscar las características a través de los operadores correspondientes a Random Forest, para llevar a cabo una comparación de los resultados obtenidos por cada uno de ellos, a través del operador Random Forest.

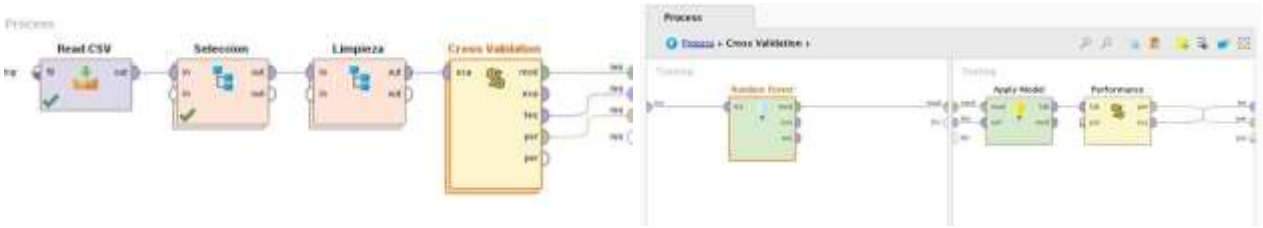


Gráfico Nº14: operadores para Random Forest.

Se obtiene un performance del 42,76%, tal como se ve en el gráfico Nº15.

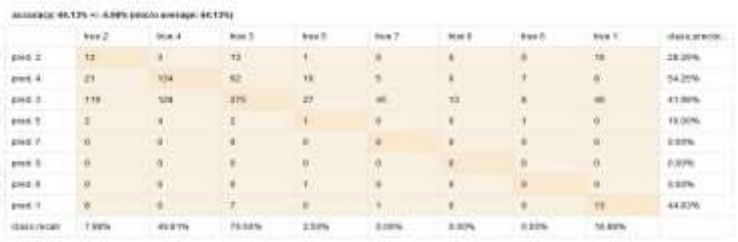


Gráfico Nº15: performance para el operador Random Forest.

A continuación se muestran las características obtenidas por uno de los árboles de Random Forest en el gráfico Nº16:

Gráfico Nº16: árbol del operador Random Forest.

También se procede a utilizar el operador de la técnica J48 (gráfico Nº17):

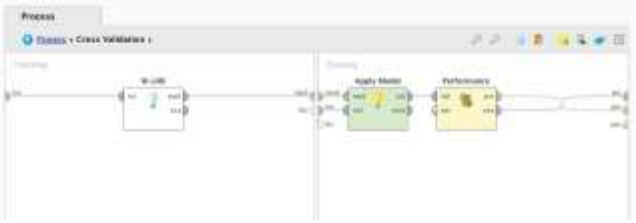


Gráfico Nº17: operadores para el operador J48.

Y se obtiene un performance del 44,23% como se muestra en el gráfico Nº18:

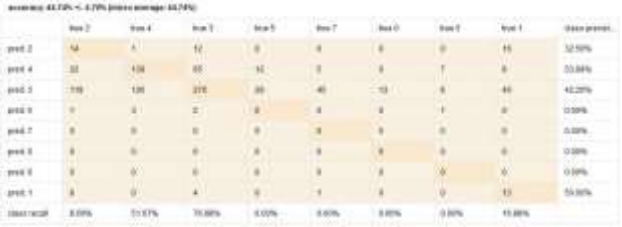


Gráfico Nº18: performance para el operador Random Forest.

Se muestra también árbol de decisión formado (gráfico Nº19):

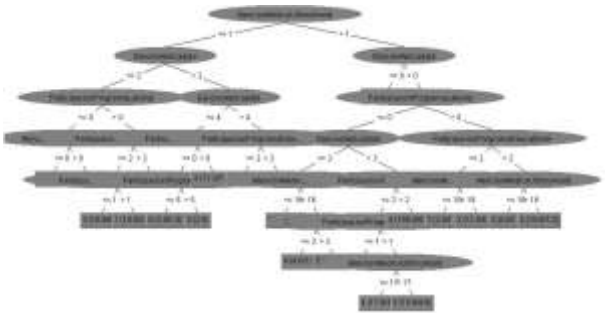


Gráfico Nº19: árbol de decisión para el operador J48.

A continuación se muestra el gráfico Nº20 con las características obtenidas para J48 (la W es por la implementación con Weka):

Gráfico Nº20: árbol obtenido con el operador J48.

Teniendo en cuenta las características obtenidas se puede decir que:

Los internos que tuvieron atención médica

-Los detenidos que obtuvieron la mayor calificación de conducta, igual a 7, son los que tuvieron:
- Una atención médica en el último año con un id mayor a 1 y menor o igual a 18.
- Una sanción aplicada con un id mayor a 0.
- Participación en programa laboral con un id mayor a 0.

- Participación en programa educativo con un id menor o igual a 2.

-Los detenidos que obtuvieron la menor calificación de conducta, igual a 1, son los que tuvieron:
- Una atención médica en el último año con un id menor o igual a 1.
- Una sanción aplicada con un id menor o igual a 2.
- Participación en programa laboral con un id mayor a 0.
- Participación en programa educativo con un id mayor a 1 y menor o igual a 2.

CONCLUSIONES

Guillermo Sanhueza, doctor en Trabajo Social y Sociología de la Universidad de Michigan, criticó la política carcelaria y de reinserción social afirmando que *"Las cárceles son el reflejo de la sociedad que estamos construyendo y es el fracaso de nosotros como sociedad"* [2].

Analizando los factores que se denunciaron como causa del motín realizado durante el año 2020, se observa que los reclamos de los reclusos presentan un respaldo estadísticos sólo en algunos casos. Como la sobrepoblación carcelaria y la cantidad de detenidos con procesos sin condenas.

Además si se toma la crítica de Sanhueza y se interpola con las principales causas de delitos de los detenidos, y sumado a esto la gran cantidad de noticias y casos de los últimos tiempo, se concluye que la educación ética y ciudadana para evitar los casos de violación y abusos en la provincia de Jujuy tiene un déficit enorme. Y esto se trasluce con el dato que la mayoría de los delitos de los privados de la libertad responden a violación o violencia sexual.

El objetivo fue propuesto por un integrante del Departamento de Trabajo Social del Servicio Penitenciario de Jujuy, el Lic. Villa Arredondo, quien además brindó asesoría en el dominio de estudio, y recalcó que la baja performance de los atributos utilizados para llegar a la calificación de los reclusos se debe a que existen factores no censados para la elaboración de dichas ponderaciones. Además señaló que dichas calificaciones surgen de un equipo interdisciplinario que trabaja en los establecimientos penitenciarios de Jujuy, y su interés estaba dado en saber si las metodologías aplicadas para la calificación estaban reflejadas estadísticamente en el censo anual que lleva a cabo el Ministerio de Justicia y Derechos Humanos de la Nación.

Algunas de las variables que han resultado significativas en su relación con la calificación de conducta que obtienen los reclusos en las cárceles de la provincia de Jujuy son: participación en programa laboral, participación en programa educativo y sanciones aplicadas, lo cual parece tener sentido para que la mayoría de los reclusos presenten buena conducta.

Resulta de interés el descubrir que las mujeres tengan muchísima menor cantidad de personas de su propio sexo dentro de las cárceles en comparación a los hombres o que casi la mitad de los delitos se tratan de violaciones, lo cual puede dar lugar a futuras investigaciones para buscar las causas de este hallazgo.

Además se podría extrapolar la metodología del presente trabajo a un dominio más amplio como una región como el NOA o a nivel nacional. Además de procesar los censos anteriores para mostrar la evolución de diferentes atributos y armar indicadores de evolución. Todo este análisis de

datos ayuda a una mejor toma de decisión del personal que trabaja en los establecimientos penitenciarios de Jujuy.

REFERENCIA BIBLIOGRÁFICAS

[1] [Dr. Abel Albino - Diario Infobae] ""Un chico que desarrolla el cerebro en un 20, 30 o 40% aprenderá a sumar o a restar, pero jamás irá a la Universidad". Recuperado de: https://www.infobae.com/salud/2018/01/06/un-chico-que-desarrolla-el-cerebro-en-un-20-30-o-40-aprendera-a-sumar-o-a-restar-pero-jamas-ira-a-la-universidad/. Fecha de consulta: 20/11/2020.

[2] [Gonzalo Castillo - Diario U de Chile], "Doctor en Trabajo Social: La cárcel es el reflejo de la sociedad que construimos". Recuperado de: https://radio.uchile.cl/2016/05/23/doctor-en-trabajo-social-la-carcel-es-el-reflejo-de-la-sociedad-que-construimos/#:~:text=%E2%80%9CLas%20c%C3%A1rceles%20siempre%20hablan%20de,justa%20y%20digna%E2%80%9D%2C%20se%C3%B1ala . Fecha de consulta: 20/11/2020.

BIBLIOGRAFÍA

[Documentación – Rapidminer, 2020], Recuperado de: https://docs.rapidminer.com/ Fecha de consulta: 25/11/2020.

[Datos del Sistema Nacional de Estadísticas sobre Ejecución de la Pena - SNEEP, 2020], Recuperado de: http://datos.jus.gob.ar/dataset/sneep/. Fecha de consulta: 25/11/2020.

[Dirección Nacional de Política Criminal - Ministerio de Justicia y Derechos Humanos Argentina], "Una gestión penitenciaria integral". Recuperado de: http://www.jus.gob.ar/media/1126013/Una_Gestion_Penitenciaria_Integral_SNEEP.pdf. Editorial Infojus, 2011.

[Dirección Nacional de Política Criminal - Ministerio de Justicia y Derechos Humanos Argentina], Estadísticas de Política Criminal - FILTRADO INTERACTIVO SNEEP. Recuperado de: https://www2.jus.gov.ar/dnpc/. Fecha de consulta: 20/11/2020.

[Metadatos del Sistema Nacional de Estadísticas sobre Ejecución de la Pena - SNEEP, 2020], Recuperado de: https://github.com/datos-justicia-argentina/Sistema-Nacional-de-Estadisticas-sobre-Ejecucion-de-la-Pena-SNEEP/blob/master/Sistema-Nacional-de-Estadisticas-sobre-Ejecucion-de-la-Pena-SNEEP-metadata.md. Fecha de consulta: 25/11/2020.

[Ministerio de Justicia y Derechos Humanos Argentina], Resolución Portal de Datos Abiertos de la Justicia Argentina. Recuperado de: http://datos.jus.gob.ar/resoluciones/RESOL-2016-986-E-APN-MJ.pdf. Fecha de consulta: 20/11/2020.

10 y 11 de Diciembre de 2020

Machine learning aplicado a la detección explícita de plagio

Autores: Ramos, Pablo Nicolás; Perez, Ricardo Daniel; Valdiviezo, Melisa Rocío; Farfán, José Humberto; Rodríguez, Mariela Ester; Vega, Ariel Alejandro; Sánchez Rivero, Víctor David

Facultad de Ingeniería, Universidad Nacional de Jujuy. San Salvador de Jujuy.

Datos de contacto: email: pablonicolasrr777@gmail.com; sanexto@gmail.com.

ABSTRACT

En los últimos años, el plagio de textos, sobre todo en pleno 2020 con el inicio de la pandemia y la educación a distancia, ha aumentado notablemente debido al gran volumen de información y documentos de dominio público a los que se puede acceder desde la web. Entonces, resulta importante proveer de una herramienta informática capaz de detectar plagio textual. En el presente trabajo se propone un nuevo método de segmentación de textos, que denominamos "Gramas de palabras de parada n", el cual pretende reducir la cantidad de falsos positivos en la detección explícita de plagio, utilizando técnicas de aprendizaje automático supervisado, más precisamente, Support Vector Machine. Este clasificador fue evaluado utilizando el corpus Webis Crowd Paraphrase Corpus 2011 (Webis-CPC-11).

Palabras Claves: Machine Learning, Plagio Explícito, Gramas de palabras de parada n, Segmentación

1. INTRODUCCIÓN

El plagio es una acción que se ha incrementado de manera exponencial en los últimos años. Esto se debe al gran volumen de información y documentos de dominio público a los que se puede acceder desde la web. La Real Academia de la Lengua Española presenta una definición para el acto de plagiar (Real Academia de la Lengua Española, 2019):

"Copiar en lo sustancial obras ajenas, dándolas como propias."

Existe un amplio conjunto de objetos susceptibles de ser plagiados: las imágenes, la música, las

ideas y los documentos. Para este trabajo es de interés aquel plagio que se presenta en la realización de un documento escrito (plagio de textos). En este ámbito el acto de plagiar, significa incorporar fragmentos de un documento escrito por otro autor, sin darle el crédito correspondiente (Pérez Afonso, 2013).

Actualmente, los problemas concernientes al plagio, y en particular al plagio de textos, se han visto incrementados debido al fácil acceso a grandes fuentes de información a través de la web, más precisamente en medios electrónicos tales como bibliotecas electrónicas u otros sitios que almacenen información de interés. Por ello, es importante aplicar técnicas de minería de texto que permitan abordar la tarea de detección automática de plagio.

Es ampliamente reconocida la importancia de suministrar herramientas informáticas que permitan a los usuarios la detección automática de plagio. Esta detección se dificulta cuando, a diferencia de la copia exacta de algún fragmento original, se modifica el mismo cambiando el orden de las palabras, reemplazando las palabras por sinónimos, sustituyendo oraciones largas por cortas, etc. En este contexto, la detección de plagio se puede clasificar en términos generales como explícita o intrínseca. En la primera es requisito, además del documento a analizar, disponer de una colección de documentos de referencia con la cual realizar una búsqueda y comparación; y se provee como resultado las secciones del documento en donde se produjo el plagio y las secciones que se utilizan de los documentos de referencia. Por el contrario, en la detección intrínseca no se requiere disponer de esta colección y únicamente se realiza un análisis de estilo en el documento para detectar variaciones de escritura. La salida del detector, en este caso, son los fragmentos de texto que estilísticamente se desvían del modelo observado a lo largo del documento.

Este trabajo aborda el estudio, comprensión y aplicación de técnicas de aprendizaje automático (machine learning) supervisado en la detección explícita de plagio.

2. MARCO TEÓRICO

2.1. DETECCIÓN AUTOMÁTICA DE PLAGIO

El plagio, el reuso de información no autorizada, y sin referencia de texto, es un fenómeno que ha cobrado relevancia debido a la gran cantidad de documentos e información que se ha generado en la web. Debido a la magnitud del problema, la revisión manual de los documentos científicos en busca de plagio es prácticamente inviable (Barrón-Cedeño et al., 2010). Por ello, surgen los detectores automáticos de plagio como una medida de prevención y corrección para asistir al humano en la detección de plagio de textos. Estas herramientas tienen como objetivo asistir al humano en la detección, proveyéndoles de pruebas posibles de un potencial caso de plagio. La decisión final, la tiene el experto, en este caso, el par evaluador de artículos científicos de ciencias de la computación.

En este contexto, la detección automática de plagio se puede clasificar, en términos generales, como explícita o intrínseca (Funez y Errecalde, 2011). En la primera, se debe proporcionar una colección de documentos de referencia, $\boldsymbol{D}$, además del documento sospechoso a analizar; donde se provee como resultado las secciones en donde se produjo el plagio y la sección que se utilizó en el archivo de referencia. La detección intrínseca, en cambio, no utiliza un corpus de referencia y realiza un análisis de estilo en el texto para detectar las variaciones en el estilo de escritura; la salida del detector en este caso, son los fragmentos de texto que estilísticamente se desvían del modelo observado a lo largo del texto. Existe una tercera variante, denominada detección automática de plagio translingüe que consiste en detectar plagio entre documentos de diferentes idiomas. El presente trabajo de investigación aborda el primer enfoque.

2.2. CARÁCTERÍSTICAS INDICADORAS DE PLAGIO

La tarea de detección de plagio necesita que se seleccionen un conjunto de características de un documento que permitan clasificar los documentos sospechosos de plagio, de los originales. Barrón Cedeño (2008) y Pérez Afonso (2013) mencionan un conjunto de características que se pueden utilizar para detectar posibles casos de plagio. Ellas son:

- Vocabulario utilizado. Analizar el léxico utilizado en algún escrito, con respecto a documentos escritos previamente por el mismo investigador. La existencia de una alta cantidad de vocabulario nuevo, puede ayudar a detectar si un académico realmente es quién escribió documento.
- Cantidad de texto común entre documentos. Es poco habitual que dos documentos escritos de manera independiente, tengan grandes cantidades de texto en común.
- Distribución de palabras. Es poco usual que la distribución en el uso de las palabras a través de textos escritos independientemente sea la misma.
- Estructura sintáctica del texto. Un indicador de plagio es que dos textos compartan una estructura sintáctica común.
- Largas secuencias de texto en común. Es poco probable que dos textos independientes (incluso cuando traten el mismo tema), compartan largas secuencias de caracteres o palabras consecutivas.
- Orden de similitud entre textos. Si existe un conjunto significativo de palabras o frases comunes en dos textos, y si el orden de ocurrencia de las coincidencias de los textos es el mismo, se está ante un posible caso de plagio.
- Frecuencia de palabras. Es poco común que las palabras halladas en dos textos independientes sean usadas con la misma frecuencia.
- Legibilidad del texto. Resulta poco probable que dos autores, $\mathfrak{A}_i$ y $\mathfrak{A}_j$, compartan las mismas métricas de legibilidad.

Estas características nos permiten comprender la esencia intrínseca de los documentos de texto, y nos introducirnos en los conceptos primitivos de la detección automática de plagio.

2.3. DETECCIÓN EXPLÍCITA DE PLAGIO

La detección explícita de plagio consiste en comparar un documento sospechoso $\boldsymbol{s}$, con un corpus de referencia $\boldsymbol{D}$, que se componen de documentos de textos que se suponen originales. Barrón Cedeño (2008) define formalmente este enfoque de detección de plagio de texto como:

"Dado un corpus $\boldsymbol{D}$, conformado por un conjunto de documentos que se suponen originales, y un documento sospechoso $\boldsymbol{s}$, la tarea de detección de plagio puede reducirse a realizar una comparación exhaustiva del texto en $\boldsymbol{s}$ sobre el corpus $\boldsymbol{D}$ para responder a la pregunta: ¿Existe algún fragmento $\boldsymbol{s_k} \in \boldsymbol{s}$ que esté incluido en algún documento de $\boldsymbol{D}$?

En las próximas subsecciones se muestran distintos métodos presentados por distintos autores en este enfoque de detección de plagio.

2.3.1. REPRESENTACIÓN COMPUTACIONAL DE DOCUMENTOS TEXTUALES

2.3.1.1. N-GRAMS

Un n-gram es una subsecuencia de $\boldsymbol{n}$ elementos de una secuencia dada. En el ámbito del procesamiento del lenguaje natural, dichos elementos son palabras o caracteres.

Uno de los métodos más populares en el estado del arte es aquel que se basa en la comparación de n-grams. Diversos autores tales como Barrón Cedeño (2008, 2009), Basile et al. (2009),

Stamatatos (2011), entre otros, han aportado a la detección automática de plagio, introduciendo el concepto de n-grams. Los métodos tienen la misma estructura: se toman n-grams del documento ($\boldsymbol{n}$ caracteres o palabras consecutivas), se calcula en algunos casos la frecuencia (Basile et al., 2009), en otros sólo el conjunto de n-grams (Nawab et al., 2010), y se computa la distancia con alguna función.

La justificación para utilizar este enfoque es que dos textos independientes tienen un nivel muy bajo de n-grams en común, siempre y cuando se considere un valor $\boldsymbol{n} > \mathbf{1}$. De hecho, la frecuencia de aparición de n-grams en un mismo documento suele ser muy baja. El uso de n-grams varía en forma y tamaño según el autor, algunos suelen utilizar n-grams de caracteres, otros de palabras y en algunos casos un híbrido de los anteriores.

En el trabajo realizado por Lyon et al. (2001, 2004) señalan que los mejores resultados se obtienen considerando trigramas. Así, tanto el documento sospechoso $\boldsymbol{s}$ como cada uno de los documentos $\boldsymbol{d_i} \in \boldsymbol{D}$ son codificados en forma de n-grams para luego compararlos. Para determinar si el documento $\boldsymbol{s}$ puede ser plagio del documento $\boldsymbol{d_i}$ se han propuesto dos medidas principales: semejanza (R) y contención (C) (Lyon et al., 2001).

La medida de similitud es útil cuando los conjuntos de n-grams a comparar provienen de textos de longitud semejante (comparaciones documento a documento, sentencia a sentencia, etc.). Considerando un documento de referencia $\boldsymbol{d_i}$ y, uno sospechoso $\boldsymbol{s}$, la semejanza se define por medio de la ecuación 1. La $\boldsymbol{R}$ hace referencia a Resemblance (semejanza).

$$\boldsymbol{R(s|d)} = \frac{|\boldsymbol{N(d_i)} \cap \boldsymbol{N(s)}|}{|\boldsymbol{N(d_i)} \cup \boldsymbol{N(s)}|} \qquad (\mathbf{1})$$

donde $N(\cdot)$ es el conjunto de n-grams en $\cdot$. Siempre toma valores entre 0 y 1, correspondiente este último a la igualdad total entre ambos conjuntos. Esta ecuación representa al conocido Coeficiente de Jaccard (Jaccard, 1901) entre los conjuntos de n-grams de ambos textos.

Si los documentos que se comparan no poseen una longitud similar (comparaciones sentencia a documento, por ejemplo), se utiliza le medida de contención, la cual se define mediante la ecuación 2.

$$\boldsymbol{C(s_k|d_i)} = \frac{|\boldsymbol{N(s_k)} \cap \boldsymbol{N(d_i)}|}{|\boldsymbol{N(s_k)}|} \qquad (\mathbf{2})$$

donde $\boldsymbol{s_k}$ es alguno de los fragmentos en el documento sospechoso $\boldsymbol{s}$. Esta medida toma valores dentro del intervalo [0, 1]. Se hace necesario definir un umbral que, al ser superado, considere que el texto sospechoso $\boldsymbol{s}$ sea candidato potencial de plagio del documento $\boldsymbol{d_i} \in \boldsymbol{D}$.

2.3.2. ASPECTO LÉXICO

El aspecto léxico hace referencia a las palabras usadas en los documentos de texto. Las palabras son los elementos mínimos del habla que expresan una idea concreta como una cosa, un atributo o una relación (Harley, 2006). Las ideas en los documentos son expresadas a través de las palabras, y, por tanto, una forma de identificar si las ideas han sido plagiadas es comparar las palabras que son empleadas en dos textos.

Formalmente un documento de texto $\boldsymbol{d_i}$ conformado por la secuencia $\langle \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n} \rangle$ de $\boldsymbol{n}$ palabras $\boldsymbol{w_i}$ es representado mediante sus vocablos, sin importar la secuencia en la que aparecen en $\boldsymbol{d_i}$, al usar sólo el aspecto léxico del texto.

Generalmente, los métodos léxicos que se emplean en la detección de plagio son: el Modelo del Espacio Vectorial (VSM, del inglés Vector Space Model) y el modelo de la bolsa de palabras (BOW,

del inglés Bag of Words).

2.3.2.1. MODELO DE LA BOLSA DE PALABRAS

Este modelo transforma el texto $\boldsymbol{d_i}$, en un conjunto de términos $\boldsymbol{t_j}$ (3). El modelo de bolsa de palabras sólo captura aquellas palabras que están dentro del documento de texto, sin tener en cuenta los términos ausentes, como el Modelo del Espacio Vectorial. Al modelar los textos como un conjunto de palabras, el cálculo de la similitud de los documentos se realiza con medidas de comparación de conjuntos (Sánchez-Vega, 2016). Las dos medidas de comparación que se usan con mayor frecuencia son el Coeficiente de Dice (4) y Coeficiente de Jaccard (5).

$$\boldsymbol{BOW(t_j) = \{t_1, t_2, \dots, t_{n-1}, t_n\}} \quad (\mathbf{3})$$

$$\boldsymbol{Dice(t_1, t_2) = \frac{2|t_1 \cap t_2|}{|t_1| + |t_2|}} \quad (\mathbf{4})$$

$$\boldsymbol{Jaccard(t_1, t_2) = \frac{|t_1 \cap t_2|}{|t_1 \cup t_2|}} \quad (\mathbf{5})$$

Luego de aplicar una medida de similitud utilizando alguno de los coeficientes recién descriptos, al igual que utilizar el coseno de similitud en la representación del Modelo del Espacio Vectorial, se emplea una función umbral y se determina la clase predicha por el algoritmo de machine learning.

2.4. APRENDIZAJE AUTOMÁTICO

El aprendizaje automático (AA, o Machine Learning, por sus siglas en inglés) es la rama de la Inteligencia Artificial que tiene como objetivo desarrollar técnicas que permitan a las computadoras aprender. Es decir, se trata de crear algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de una información suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento, es decir, un método que permite obtener por generalización un enunciado general a partir de enunciados que describen casos particulares (Gámez Granados, 2017).

2.5. APRENDIZAJE SUPERVISADO

Los algoritmos de aprendizaje automático se pueden clasificar en algoritmos no supervisados, y algoritmos supervisados. En el primer grupo, la respuesta es conocida; y en el segundo, también llamado clúster, no se tienen las posibles respuestas (Calvo Torres, 2017). En el presente trabajo se proporciona una propuesta metodológica utilizando un método de machine learning supervisado, Support Vector Machine.

2.5.1. ALGORITMOS SUPERVISADOS

Los algoritmos de aprendizaje supervisado son aquellos donde el objetivo es crear, a partir de un conjunto de ejemplos de entrenamiento, para los cuales se conoce de antemano la salida correcta (datos etiquetados), un modelo numérico capaz de realizar predicciones precisas para nuevos ejemplos no contemplados durante el entrenamiento (Arjona y Díaz, 2017).

De manera formal:

- Dado un conjunto de datos etiquetados $\mathcal{D}$, donde $\mathcal{D} = (x, y)/x \in Y$ A $y \in f$, o sea tenemos un conjunto de valores de salidas f,
- Una función objetivo desconocida $f: Y \to f$,
- Se busca calcular una función $f: Y \to f$ usando $\mathcal{D}$ tal que $f(x) \cong f(x) \forall\, x \in Y$

Los problemas de clasificación están ligados con el Aprendizaje Supervisado, de hecho, es una de

las tareas más destacadas en el AA. En la siguiente sección se describe este proceso. La siguiente figura (1) ilustra el esquema genérico del Aprendizaje Supervisado.

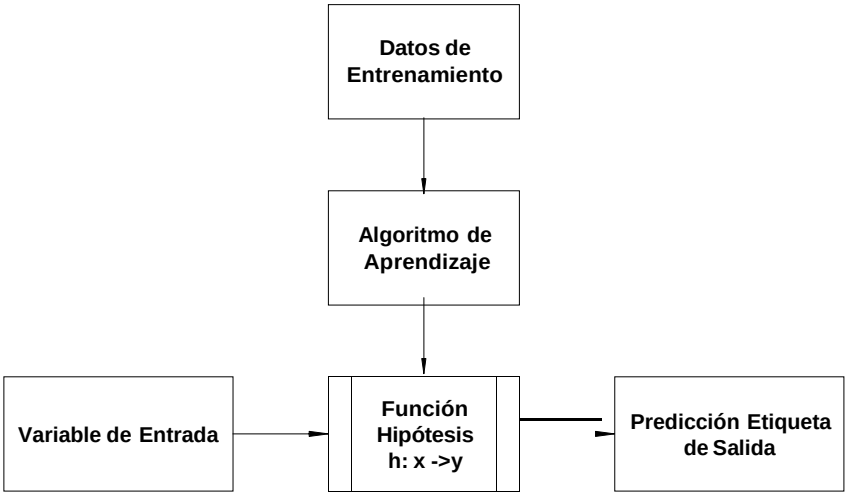


Figura 1. Esquema genérico de Aprendizaje Supervisado.

2.5.1.1. SUPPORT VECTOR MACHINE (SVM)

SVM, o bien, las llamadas Máquinas de Soporte Vectorial o SVM (del inglés Support Vector Machine), son métodos de clasificación capaces de trabajar en espacios de alta dimensionalidad. Estos métodos surgen de los trabajos sobre la teoría del aprendizaje estadístico y fueron introducidos en los años noventa por Cortes y Vapnik (1995).

Para entender la esencia de estas técnicas, se deben tener en cuenta los siguientes supuestos: 1) se tienen dos dimensiones y, 2) se tienen dos categorías.

Sea un conjunto de datos como el que se puede observar en la figura (2). Estos datos están representados como un vector *n*-dimensional. El caso más simple del SVM, consiste en tratar de encontrar un hiperplano lineal que separe las dos categorías, aquí representadas como cuadrados y círculos. En este caso el hiperplano de separación es una recta y la dimensión $n = 2$.

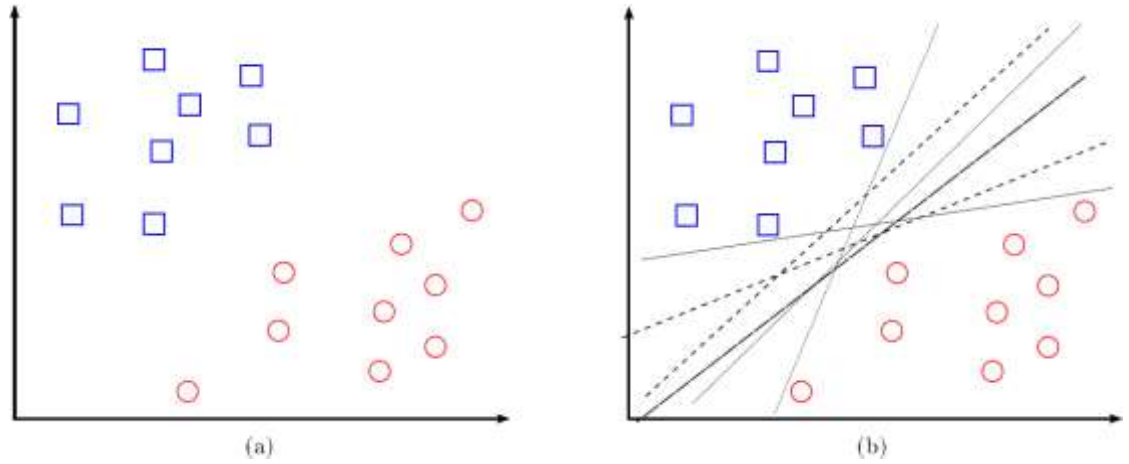


Figura 2: (a) Conjunto de datos del ejemplo. (b) Posibles hiperplanos de separación de las dos categorías.

Existe un número infinito de posibles hiperplanos (líneas) que realicen la clasificación, pero, ¿cuál es la mejor y cómo la definimos? Para responder está cuestión, lo que se busca es seleccionar un hiperplano de separación óptima, es decir, aquella solución que permita un margen lo más amplio posible entre las categorías, debido a esto, también se denomina clasificador de máximo margen. El hiperplano que cumple esa condición es el equidistante a los ejemplos más cercanos de cada clase. Estos puntos determinan el margen y el hiperplano, y se denominan vectores de soporte. En la figura 3, son los que tienen el color compacto.

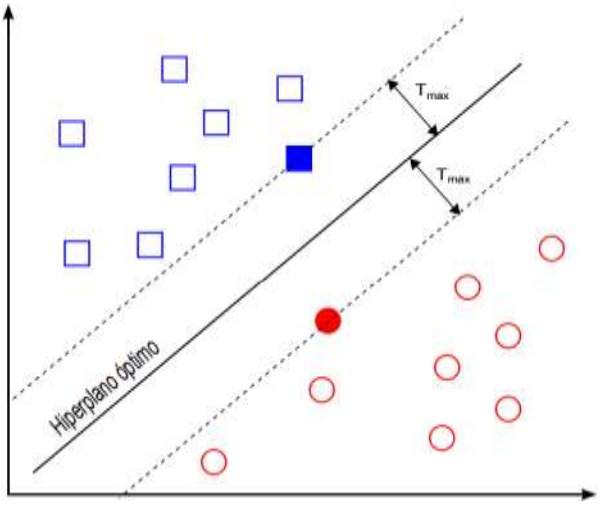


Figura 3: Imagen donde se encuentra el hiperplano óptimo y los vectores de soporte para el ejemplo dado.

En este caso, se clasifica como cuadrado cualquier dato que estuviese por encima del hiperplano, y como círculo, cualquiera que estuviese por debajo.

2.5.1.1.1. SVM para la clasificación binaria de ejemplos linealmente separables

Se considera una muestra de aprendizaje D dada por:

$$\boldsymbol{D} = \{(\boldsymbol{x_1}, \boldsymbol{y_1}), (\boldsymbol{x_2}, \boldsymbol{y_2}), \dots, (\boldsymbol{x_n}, \boldsymbol{y_n})\}$$

donde $\boldsymbol{x_i} = (\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \dots, \boldsymbol{x_{im}})$ con $\boldsymbol{x_i} \in \boldsymbol{R^m}$ $\boldsymbol{e}$ $\boldsymbol{y_i} \in \{-\mathbf{1}, +\mathbf{1}\}$ para $\boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{n}$

Un hiperplano de separación se define como una función capaz de separar los elementos del conjunto $\boldsymbol{D}$ sin error, tal como se ilustra en la Figura 2. La función del hiperplano de separación está dada por:

$$\boldsymbol{h(x)} = (m_1\boldsymbol{x_1} + m_2\boldsymbol{x_2} + \cdots + m_n\boldsymbol{x_n}) + \boldsymbol{b} = m \cdot \boldsymbol{x} + \boldsymbol{b}$$

Donde $m = (m_1, m_2, \dots, m_n)$ es el vector ortogonal al hiperplano, $\boldsymbol{b} \in \boldsymbol{R}$.

El hiperplano de separación $h(x)$ cumple las siguientes restricciones para todo elemento x_i, con $i = 1, \dots, n$ del corpus de referencia:

$$\boldsymbol{h(x_i)} = m \cdot \boldsymbol{x_i} + \boldsymbol{b} > \mathbf{0} \qquad \boldsymbol{si} \qquad \boldsymbol{y_i} = +\mathbf{1}$$

$$\boldsymbol{h(x_i)} = m \cdot \boldsymbol{x_i} + \boldsymbol{b} < \mathbf{0} \qquad \boldsymbol{si} \qquad \boldsymbol{y_i} = -\mathbf{1}$$

o, de forma más concreta:

$$y_i \cdot h(x_i) > 0 \qquad i = 1, \dots, n$$

Lo que se conoce como margen de hiperplano de separación, que no es más que la distancia entre dicho hiperplano y el ejemplo más cercano de cualquiera de las dos clases, se denota por r. Teniendo en cuenta esta definición, se considera como hiperplano de separación óptimo aquel que consiga el máximo margen de separación.

3. PROPUESTA METODOLÓGICA

En esta sección se presenta el enfoque propuesto. Primeramente, se presenta un modelo para la detección explícita de plagio mediante el uso de paráfrasis, el cual consiste en identificar plagio textual en una base de datos mediante el uso de dos modelos de aprendizaje automático.

3.1. APROXIMACIÓN A LA DETECCIÓN EXPLÍCITA DE PLAGIO

En la tarea de detección explícita de plagio se proponen dos modelos de aprendizaje automático supervisado, un clasificador Support Vector Machine (SVM) y un clasificador Naive Bayes.

3.1.1. ELECCIÓN Y PROCESAMIENTO DEL CORPUS DE ENTRENAMIENTO

Para el entrenamiento de ambos clasificadores se utiliza el corpus Webis Crowd Paraphrase Corpus 2011 (Webis-CPC-11). Se trata de un corpus publicado en el año 2011 para tareas de detección explicita de plagio. El mismo contiene 7.859 textos en idioma inglés, donde cada texto tiene su respectivo par plagiado parafraseado. Entre otros corpus publicados en diversos repositorios públicos, es el que mayor cantidad de textos contiene, de ahí su elección. Además, es importante mencionar que en dichos repositorios no se disponen de corpus de textos del idioma español.
En el corpus, cada texto, de ahora en adelante muestra (caso de plagio), está representado por tres (3) archivos (ver figura 4):

- *original.txt:* contiene el texto original, sin plagiar.
- *paraphrase.txt*: contiene el correspondiente texto plagiado parafraseado.
- *metadata.txt*: contiene información acerca del autor del plagio y el tiempo empleado en la elaboración del plagio.

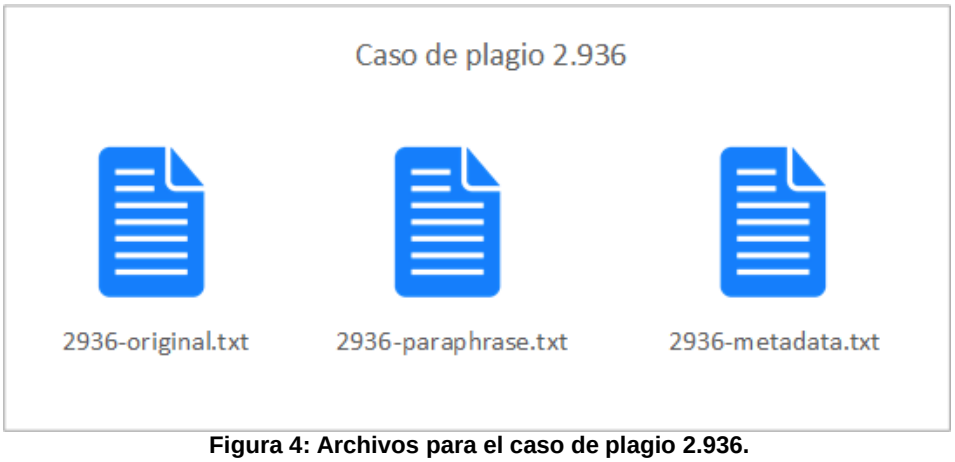


Figura 4: Archivos para el caso de plagio 2.936.

En una primera etapa de procesamiento y filtrado de los textos del corpus, con el objetivo de lograr un corpus de entrenamiento lo más limpio posible, se programó un script en Python para llevar a cabo dicha tarea. Se trata del script *corpus_purging.py*, que elimina los pares de archivos *original*

y *paraphrase* sin ningún contenido y también elimina todos los archivos *metadata* (ver figura 5).

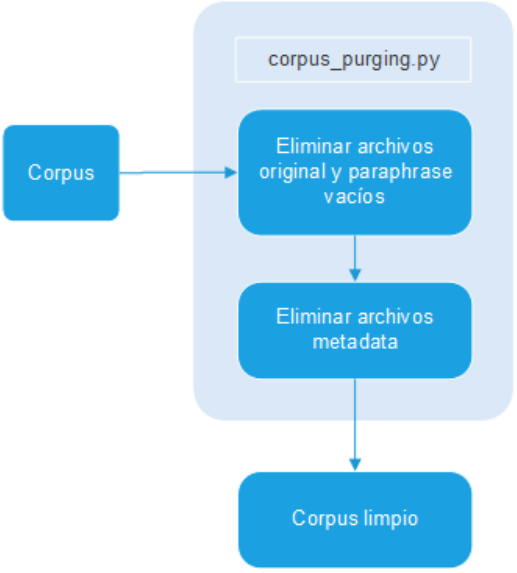


Figura 5: Etapas en el procesamiento del corpus.

Eliminar los archivos de texto sin ningún texto o contenido es clave, dado que su inclusión en la etapa de entrenamiento podría sesgar el criterio de detección de plagio en ambos clasificadores. Los archivos *metadata* se eliminan simplemente porque no resultan útiles para el presente trabajo.

3.1.2. GENERACIÓN DEL CONJUNTO DE DATOS DE ENTRENAMIENTO

En la etapa de entrenamiento, de cualquier modelo predictivo supervisado, es requisito contar con datos etiquetados, los clasificadores planteados para la detección explicita de plagio no son la excepción. Por ello es importante contar con un conjunto de datos de entrenamiento, donde algunas variables del conjunto actúan como variables independientes (también llamadas variables predictoras, de características o, de entrada) y otra variable actúa como variable dependiente (también llamada variable a predecir, objetivo o de salida), como se puede apreciar en la figura 6.

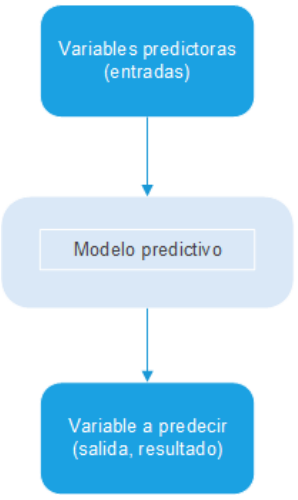


Figura 6: Esquema de un modelo predictivo supervisado.

Sin embargo, el corpus solo contiene archivos de textos aislados, es decir, no está en un formato valido para el entrenamiento de los modelos predictivos propuestos. Por lo tanto, hay que adoptar alguna estrategia para generar un conjunto de datos valido a partir de dichos textos.
En un primer paso se definen las variables predictoras y la variable objetivo. Se proponen como variables predictoras las métricas de similitud de texto *Jaccard Similarity*, *Dice Similarity*, *Cosine Similarity* y *Bag Similarity*. Y como variable objetivo una variable binaria que adopta un valor u otro dependiendo si se trata de un caso de plagio o no.

Para cada par de texto *original* y *paraphrase* del corpus se calculan las correspondientes métricas de similitud y se normalizan sus valores en el intervalo cerrado $[0; 1]$. El cálculo de las métricas no se realiza directamente sobre los textos sino sobre los tokens de sus gramas. Para ello se realiza lo siguiente:

1. Se elimina de *original* y *paraphrase* cualquier contenido irrelevante, se eliminan saltos de línea, espacios adicionales (dos o más caracteres espacio), números y caracteres especiales.
2. Se obtienen sus correspondientes gramas, aplicando el método de *gramas de palabras de parada* n, con $n = 100$. El valor de n se determinó experimentalmente.
3. Se eliminan las palabras de parada de todos los gramas obtenidos.
4. Se tokeniza cada grama, así cada grama está representado por una lista o bolsa de tokens (palabras).
5. Se obtienen las raíces (radicalización) de cada una de las palabras (tokens).
6. Se calculan las métricas de similitud de cada lista i-ésima de *original* con cada lista j-ésima de *paraphrase*.
7. Se obtiene un promedio de las métricas calculadas en el paso 6, para así obtener métricas globales a nivel de documento o texto. En el cálculo del promedio se sigue el siguiente criterio:
    a. Si existen listas que tienen al menos una de las métricas mayor o igual que $0{,}5$, entonces el promedio de las métricas se calcula únicamente sobre estas listas.
    b. Si existen listas que tienen al menos una de las métricas mayor o igual que cero, entonces el promedio de las métricas se calcula únicamente sobre estas listas.
    c. Si ninguna de las condiciones anteriores se cumple, entonces el promedio es cero para todas las métricas.

La variable objetivo adopta valor 1 (uno) si se trata de un caso de plagio y valor 0 (cero) en caso contrario. Es un caso de plagio si alguna de las métricas globales supera el umbral de $0{,}5$. Alcanzar un umbral de $0{,}5$ implica un plagio del 50% por lo que se podría considerar un caso de plagio intensional, de ahí la elección de este valor, menores valores podrían implicar clasificaciones erróneas decantando en falsos positivos.
Para extender aún más la cantidad de muestras de casos de no plagio, se calcula también la similitud entre cada texto *original* y algún texto *paraphrase* aleatorio distinto del vinculado al texto *original* (ver figura 7).

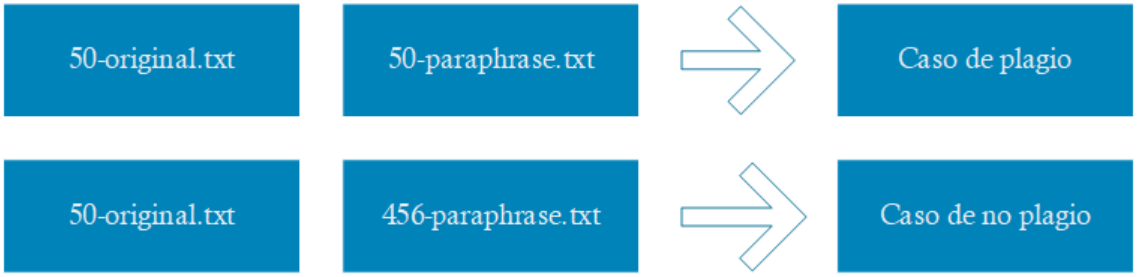


Figura 7: Generación de casos de no plagio.

La generación del conjunto de datos se realiza con el script *corpus_csv_generator.py*, el mismo genera un conjunto de datos en formato CSV (del inglés comma-separated values, valores separados por coma), un formato valido para el entrenamiento de los clasificadores propuestos. En la figura 8 se puede apreciar la estructura de éste CSV.

index	source	suspicious	jaccard_similarity	dice_similarity	cosine_similarity	bag_similarity	plagiarism
0	1-original.txt	1-paraphrase.txt	0.3	0.46	0.46	0.43	0
1	1-original.txt	5478-paraphrase.txt	0	0	0	0	0
2	10-original.txt	10-paraphrase.txt	0.15	0.26	0.26	0.26	0
3	10-original.txt	5899-paraphrase.txt	0.02	0.03	0.04	0.02	0
4	100-original.txt	100-paraphrase.txt	0.42	0.59	0.59	0.56	1
5	100-original.txt	2761-paraphrase.txt	0	0	0	0	0
6	1000-original.txt	1000-paraphrase.txt	0.08	0.15	0.17	0.1	0
7	1000-original.txt	6645-paraphrase.txt	0.01	0.02	0.03	0.02	0
8	1001-original.txt	1001-paraphrase.txt	0.52	0.68	0.68	0.68	1
9	1001-original.txt	5577-paraphrase.txt	0.01	0.01	0.02	0.01	0
10	1002-original.txt	1002-paraphrase.txt	0.88	0.93	0.93	0.93	1
11	1002-original.txt	7276-paraphrase.txt	0.01	0.02	0.03	0.02	0
12	1003-original.txt	1003-paraphrase.txt	0.1	0.18	0.2	0.13	0
13	1003-original.txt	7423-paraphrase.txt	0	0	0	0	0
14	1004-original.txt	1004-paraphrase.txt	0.3	0.46	0.47	0.42	0
15	1004-original.txt	1228-paraphrase.txt	0.02	0.04	0.04	0.04	0
16	1005-original.txt	1005-paraphrase.txt	0.38	0.56	0.56	0.48	1
17	1005-original.txt	328-paraphrase.txt	0	0	0	0	0
18	1006-original.txt	1006-paraphrase.txt	0	0	0	0	0
19	1006-original.txt	1553-paraphrase.txt	0.01	0.03	0.03	0.03	0
20	1007-original.txt	1007-paraphrase.txt	0.25	0.4	0.4	0.38	0
21	1007-original.txt	5144-paraphrase.txt	0.02	0.05	0.05	0.03	0
22	1008-original.txt	1008-paraphrase.txt	0.42	0.59	0.59	0.59	1
23	1008-original.txt	6020-paraphrase.txt	0.02	0.03	0.03	0.02	0

Figura 8: Estructura del conjunto de datos de entrenamiento.

3.1.3. GRAMAS DE PALABRAS DE PARADA N

La segmentación de los textos es muy importante cuando se calculan las métricas de similitud entre los mismos, ya que cada segmento de un texto es comparado con los segmentos del otro texto. Hay varias estrategias de segmentación, las más sencillas suelen ser las más rápidas de implementar, pero frecuentemente no obtienen un buen rendimiento. La estrategia más sencilla es dividir los textos en segmentos del mismo tamaño y compararlos unos con otros. Esta estrategia

de segmentación es conocida en la literatura como gramas de tamaño n, donde n es la longitud de los segmentos o gramas.

Una mejor opción es obtener segmentos o gramas de longitudes dinámicas. En el presente trabajo, se propone un nuevo método de segmentación dinámico denominado *Gramas de palabras de parada* n, donde n es un parámetro ajustable que se determina experimentalmente.
El método establece que, dado un texto t, que se compone de m palabras de parada (stopwords, en inglés), descomponer t en segmentos (gramas) que contengan n palabras de parada, con $n \ll m$.

Por ejemplo, al aplicar el método en el siguiente texto con un valor de $n = 5$.

> *El oeste de Texas divide la frontera entre México y Nuevo México. Es muy bella, pero áspera, llena de cactus, en esta región se encuentran las Davis Mountains. Todo el terreno está lleno de piedra caliza, torcidos arboles de mezquite y espinosos nopales. Para admirar la verdadera belleza desértica, visite el Parque Nacional de Big Bend, cerca de Brownsville. Es el lugar favorito para los excursionistas, acampadores y entusiastas de las rocas. Pequeños pueblos y ranchos se encuentran a lo largo de las planicies y cañones de esta región. El área solo tiene dos estaciones, tibia y realmente caliente. La mejor época para visitarla es de diciembre a marzo cuando los días son tibios, las noches son frescas y florecen las plantas del desierto con la humedad en el aire.*

Se obtendría el siguiente conjunto de segmentos $\boldsymbol{S}$:
$\boldsymbol{S} = \{$
"El oeste de Texas divide la frontera entre México y Nuevo México.",
"Es muy bella, pero áspera, llena de cactus, en",
"esta región se encuentran las Davis Mountains. Todo el terreno",
"está lleno de piedra caliza, torcidos arboles de mezquite y espinosos nopales. Para admirar",
"la verdadera belleza desértica, visite el Parque Nacional de Big Bend, cerca de Brownsville. Es",
"el lugar favorito para los excursionistas, acampadores y entusiastas de",
" las rocas. Pequeños pueblos y ranchos se encuentran a lo largo",
"de las planicies y cañones de esta región.",
"El área solo tiene dos estaciones, tibia y realmente caliente. La mejor época para visitarla",
"es de diciembre a marzo cuando los días",
"son tibios, las noches son frescas y florecen las plantas ",
"del desierto con la humedad en el aire."
$\}$

Como se puede observar, en cada segmento o grama hay $n = 5$ palabras de parada. Para este texto justamente coincide que el último segmento tiene exactamente 5 palabras de parada, pero podría haber menos o incluso no haberlas ya que se trata del final del texto.

3.1.4. SELECCIÓN DE CARACTERÍSTICAS RELEVANTES

Luego del tratamiento de los valores atípicos el siguiente paso es la selección de aquellas variables

predictoras o de características más relevantes del conjunto de datos, es decir, seleccionar aquellas que aporten mayor información en la tarea de categorización o clasificación de plagio.
En primer lugar, se realiza una puntuación (scoring) de cada una de las variables de características, es decir, de cada una de las métricas de similitud. Para el cálculo de las puntuaciones se propone realizar un análisis de la varianza del valor estadístico F de cada una de las características del conjunto de datos. Esta estrategia de puntuación es normalmente conocida en la literatura bajo el seudónimo *ANOVA F-value*, ANOVA por sus siglas en inglés de *ANalysis Of VAriance* (Análisis de la Varianza). Este método es adecuado cuando las variables de características $x_1, \dots, x_n$ son numéricas y la variable objetivo y es una variable categórica (plagio / no plagio).
La librería de Python, scikit-learn, ya proporciona una implementación para este método de puntación a través de la llamada a la función $f\_()$.
A partir de las puntuaciones se realiza un ranking de las características (ver Figura 9), así la primera del ranking es la métrica *Cosine Similarity* y la última es la métrica *Jaccard Similarity*.

	Feature	Score
2	cosine_similarity	142868.529789
1	dice_similarity	140038.196188
3	bag_similarity	103025.538845
0	jaccard_similarity	36528.721491

Figura 9: Ranking de las características.

La selección de las características se realiza de manera independiente sobre cada modelo predictivo. En dicha selección se evalúa el error del modelo a medida que se van considerando y agregando al modelo más características del ranking, empezando por la primera y sucesivamente. El error es un promedio de los errores obtenidos por el modelo en 10 (diez) iteraciones. En cada iteración el modelo es entrenado y evaluado con el 80% y el 20% del conjunto de datos, respectivamente. En cada iteración el muestreo (partición) es realizado de manera aleatoria y estratificada, es decir, en cada partición se conserva la proporción de casos de plagio y no plagio con respecto al conjunto de datos original sin particionar.
A continuación, se describe la configuración del modelo Support vector machine a partir de los cuales se procede a realizar la selección de las características relevantes.

3.1.5. CONFIGURACIÓN DEL MODELO SVM

Se trabaja sobre un modelo Support Vector Machine con kernel lineal, con un valor de 1 para el hiperparámetro de ajuste C, es el valor por defecto que la librería de Python, *sklearn,* da al mismo. La elección de un kernel lineal se debe a que es adecuado y tiene un muy buen rendimiento en clasificaciones binarias, es el caso de la variable de salida (plagiarism), un variable categórica binaria que toma dos valores posibles, 0 (no plagio) y 1 (plagio).
En la figura 10 se muestra el error obtenido por el modelo con cada nueva característica que se considera.

	Número de características	Error	ΔError
0	1	0.0	0.0
1	2	0.0	0.0
2	3	0.0	0.0
3	4	0.0	0.0

Figura 10: Tabla con los errores del modelo *Support Vector Machine* a medida que se consideran más características.

Se concluye que a medida que se agregan nuevas características el error no disminuye, por lo tanto, es suficiente considerar únicamente la característica (métrica) *Cosine Similarity*, la primera del ranking. Consiguiendo así reducir considerablemente la dimensión del vector de características $\mathbf{X}$, ahora $\mathbf{X}$ es un vector unidimensional, sin penalizar el rendimiento de los modelos. Esta reducción en la dimensión decantará en un menor tiempo de cómputo en el entrenamiento de los modelos finales.

3.1.6. AJUSTE DE HIPERPARÁMETROS

El ajuste de los hiperparámetros de los modelos predictivos es importante ya que permite ajustar las predicciones (clasificaciones) para evitar problemas de sobreajuste (en inglés overfitting) y subajuste (en inglés underfitting). Los hiperparámetros son usualmente coeficientes que acompañan a las ecuaciones respectivas de los modelos y que no se entrenan con los datos, y que deben ser establecidos experimentalmente.
En las siguientes secciones se describe el procedimiento para ajustar los hiperparámetros de cada modelo predictivo propuesto.

3.1.6.1. HIPERPARÁMETRO DE REGULARIZACIÓN C, DEL MODELO SVM

Al ser el kernel lineal un kernel simple en su formulación matemática, el único hiperparámetro de ajuste para el modelo Support Vector Machine es el hiperparámetro de regularización C. En otros kernel, más complejos, se tienen más hiperparámetros de ajuste, pero no es el caso en cuestión.
La estrategia para encontrar el valor óptimo del hiperparámetro es exactamente la misma que la descripta anteriormente para el modelo bayesiano, en donde se considera el error del modelo como criterio de selección del valor óptimo.

En la generación de los diferentes valores del hiperparámetro se parte de un valor inicial $C_0 = 1$, cómo se puede observar en la figura 11. El valor inicial C_0 es valor por defecto que asigna la librería de Python, sklearn, a dicho hiperparámetro.

En referencia a la figura 11, se puede observar que mayores valores del hiperparámetro C no implican una disminución en el error del modelo, y teniendo en cuenta que mayores valores del mismo implican un mayor tiempo de cómputo en la etapa de entrenamiento del modelo, se concluye que un valor de $C = 1$ es un valor óptimo.

	C	Error	ΔError
0	1	0.0	0.0
1	10	0.0	0.0
2	100	0.0	0.0
3	1000	0.0	0.0
4	10000	0.0	0.0
5	100000	0.0	0.0
6	1000000	0.0	0.0
7	10000000	0.0	0.0
8	100000000	0.0	0.0
9	1000000000	0.0	0.0
10	10000000000	0.0	0.0

Figura 11. Tabla con los errores del modelo Support Vector Machine para diferentes valores del hiperparámetro *C*.

3.1.7. VALIDACIÓN DEL MODELO

En la construcción de ambos modelos predictivos, el conjunto de datos se particiona en dos subconjuntos, un subconjunto de entrenamiento (80% del conjunto de datos) y un subconjunto de prueba (20% del conjunto de datos), es decir se aplica el principio de Pareto 80/20. Previo a la partición se mezcla aleatoriamente el conjunto de datos. Además, la partición se realiza de manera estratificada, es decir, se mantiene la proporción de cada extracto (plagio y no plagio) con respecto al conjunto de datos original.

La validación de cada modelo se realiza aplicando la técnica de validación cruzada de K iteraciones (k-fold cross-validation). Esta técnica permite comprobar si el modelo sufre de overfitting (sobreajuste) o underfitting (subajuste). En la aplicación de la técnica se utiliza un valor de $k = 10$ (valor recomendado), por lo que se realizan 10 iteraciones. En cada iteración se divide el subconjunto de datos de entrenamiento, en dos subconjuntos adicionales: un subconjunto de entrenamiento más pequeño (80% del subconjunto de entrenamiento) y un subconjunto de validación (20% del subconjunto de entrenamiento). La partición se realiza nuevamente de manera aleatoria y estratificada, conservando las proporciones.

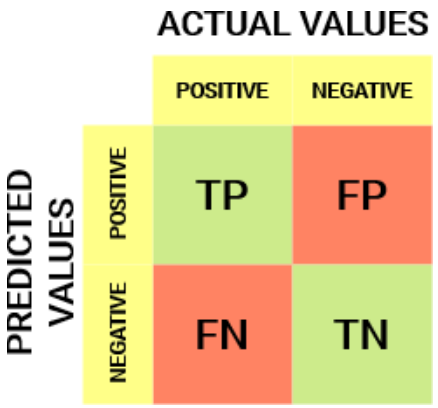
En cada iteración k_i se realiza el entrenamiento y validación del modelo con un subconjunto de entrenamiento y validación diferente (aleatorio) y se calculan las correspondientes métricas de rendimiento.

4. DESARROLLO

4.1. MÉTRICAS DE EVALUACIÓN

La matriz de confusión es una matriz $\boldsymbol{N\ x\ N}$ que se utiliza para evaluar el rendimiento de un modelo de clasificación donde $\boldsymbol{N}$ es el número de clases objetivo (labels). La matriz compara los valores objetivo reales con los predichos por el modelo de aprendizaje automático. Esta técnica indica que

tan bien se desempeña un modelo de clasificación y que tipo de errores está cometiendo.
Para un problema de clasificación binaria, se tiene una matriz de $\mathbf{2\ x\ 2}$ como se muestra en la figura 12 con 4 (cuatro) valores:



12. Matriz de confusión binaria (con dos clases objetivo).

La variable de destino tiene dos valores: positivo o negativo. Las columnas representan los valores reales de la variable objetivo y las filas los valores predichos por el modelo predictivo.
Verdadero positivo (TP, por sus siglas en inglés), representa la cantidad de veces que el valor predicho es positivo, siendo el real positivo. Verdadero negativo (TN, por sus siglas en inglés), representa la cantidad de veces que el valor real negativo, coincide con el valor negativo predicho por el modelo. Falso positivo (FP, por sus siglas en inglés) o error de tipo 1, representa la cantidad de veces que el valor real es negativo pero el modelo lo predice positivo. Falso negativo (FN, por sus siglas en inglés) o error de tipo 2, representa la cantidad de veces que el valor real es positivo pero el modelo lo predice negativo.
Entonces se definen las métricas de rendimiento *accuracy*, *precision*, *recall*, *f1_score* y *error* de la siguiente manera:

$$\boldsymbol{accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{TN} + \boldsymbol{FN}} \quad (\mathbf{6})$$

$$\boldsymbol{precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \quad (\mathbf{7})$$

$$\boldsymbol{recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \quad (\mathbf{8})$$

$$\boldsymbol{f1_{score}} = \frac{\mathbf{2}}{\frac{\mathbf{1}}{\boldsymbol{recall}} + \frac{\mathbf{1}}{\boldsymbol{precision}}} \quad (\mathbf{9})$$

$$\boldsymbol{error} = \frac{\boldsymbol{FP} + \boldsymbol{FN}}{\boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{TN} + \boldsymbol{FN}} \quad (\mathbf{10})$$

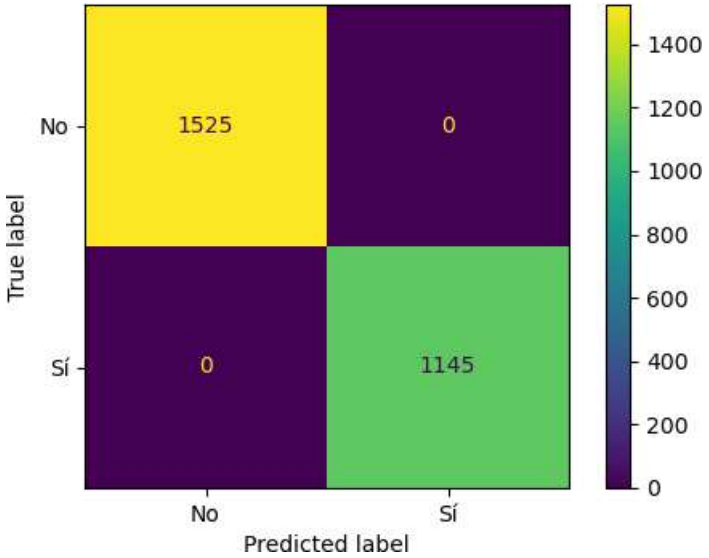
Una curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una curva que muestra el rendimiento de un modelo de clasificación en todos los

umbrales de clasificación. Esta curva representa dos parámetros:

- Recall
- Precision

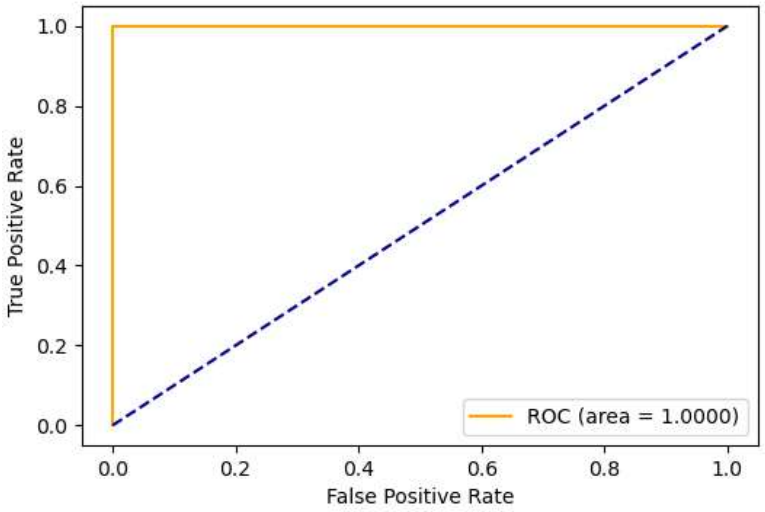
4.2. EXPERIMENTOS Y RESULTADOS

En la evaluación del rendimiento del modelo Support Vector Machine, se realiza el entrenamiento y prueba del mismo con el subconjunto de entrenamiento y prueba obtenido en un principio. La configuración del modelo es la descripta en la sección **3.1.7.**
Para evaluar el modelo se utiliza la matriz de confusión sobre el conjunto de prueba, la misma se visualiza en la figura 13.



13. Matriz de confusión del modelo Support Vector Machine.

Se observa que el modelo clasifica correctamente todas las instancias que se le presentan, es decir, obtiene una performance óptima. $Recall = Accuracy = Precision = f1_{score} = 1$ y $error = 0$.
Luego se muestra la curva ROC, para ello se procede a realizar las predicciones, dichas predicciones se obtienen como probabilidad, es decir valores en el intervalo $[0, 1]$. Para cada clase se obtiene una probabilidad, por lo tanto, habrá una probabilidad para no plagio (p) y otra probabilidad para plagio (1 - p). En este caso nos interesa la probabilidad de plagio para mostrar en la curva ROC la proporción de falsos positivos y verdaderos positivos. A continuación, en la figura 14 se muestra la curva ROC perteneciente al clasificador SVM lineal.



14. Curva ROC del modelo Support Vector Machine.

La curva ROC nos indica que tan balanceadas se encuentran las métricas recall y precision. Cuanto más parecidas sean, mejor es la performance del clasificador. Al igual que en la matriz de confusión, el rendimiento es óptimo.

4.3. ENTORNO EXPERIMENTAL

Aquí se presenta una breve descripción de cómo se configuró la parte experimental. Estos pasos deberían ser suficientes para reproducir, de forma general, los resultados obtenidos en la detección de plagio explícito. El entorno de desarrollo y prueba se configura de la siguiente manera:

- Computadora Portátil Core i5 con Procesador Intel y 6 GB de Memoria RAM.
- Windows 10, 64-bits
- Anaconda 3, Python 3.8 (+ librerías)
- Jupyter Notebook 6.1.5 (Análisis exploratorio)
- Spyder IDE (Scripts)

Se utilizó un ordenador portátil de 6 núcleos con 6 GB de memoria RAM para el desarrollo de los algoritmos y ejecución de los experimentos. El lenguaje de programación elegido fue Python, en su versión 3.8, que se ejecuta en el sistema operativo Windows 10 de 64 bits.

5. CONCLUSIONES

En la detección explícita de plagio, el método de segmentación propuesto elimina los falsos positivos que se generaban cuando se utilizaba la técnica de segmentación por sentencias. En los primeros experimentos, cuando se utilizaba la segmentación por sentencias, el modelo Support Vector Machine no alcanzaba un rendimiento perfecto y en la práctica se obtenían falsos positivos. En contraposición, cuando se experimentó con el método “Gramas de palabras de parada n” (con $\boldsymbol{n = 100}$), se eliminó esa limitación y se alcanzó una puntuación perfecta, con $\boldsymbol{f1_{score} = 1.}$ En este enfoque, el paso previo a la realización del modelo de inteligencia artificial, jugó un papel preponderante, ya que la limpieza o no, de los textos influye notablemente en el rendimiento global de los algoritmos de machine learning. Para ellos se estudiaron varias técnicas de procesamiento del lenguaje natural, más precisamente, aquellas técnicas que eliminan las palabras o caracteres que carecen de poca importancia para el análisis en cuestión. Para reducir el vocabulario de cada

sentencia se utilizó un algoritmo de stemming, cuya aplicación consiste en reducir una palabra dada, a su raíz morfológica, porque la persona que comete plagio suele agregar sufijos o prefijos a las palabras copiadas para "camuflar" el delito que está cometiendo. Los experimentos principales se realizaron utilizando documentos escritos en inglés, pero ambos métodos pueden ser utilizados para trabajar con textos en español, o cualquier otro idioma, ya que no se utilizan características dependientes del idioma, proporcionando un punto de partida para explorar con otras lenguas. Esta es una observación importante, ya que muchos enfoques lingüísticos utilizan características que dependen del idioma y son más complejos de extender a otras lenguas.

6. REFERENCIAS

1. Arjona, N. F. y Díaz, A. G. (2017). "Detección de puntos clave en rostros aplicando random forest" (tesis de grado). Universidad Nacional de Jujuy, Argentina.
2. Barrón Cedeño, L. A. (2008). "Detección automática de plagio en texto" (tesis de maestría). Universidad Politécnica de Valencia, España. Obtenido de: https://riunet.upv.es/handle/10251/12186
3. Barrón Cedeño, A., Rosso, P. (2009). "On Automatic Plagiarism Detection Based on n-grams Comparison". English. En: Advances in Information Retrieval. Ed. por Mohand Boughanem y col. Vol. 5478. Lecture Notes in Computer Science. Springer Berlin Heidelberg, págs. 696-700. ISBN: 978-3-642-00957-0. Obtenido de: http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=4ED3D701892331F6EF7819FE969EC7B3?doi=10.1.1.204.3915
4. Barrón Cedeño, L. A., Vila, M. y Rosso, P. (2010). "Detección automática de plagio: de la copia exacta a la paráfrasis". Jornadas informativas de lingüística forense, Madrid, España. Obtenido en: http://personales.upv.es/prosso/resources/BarronEtAl_JLF10.pdf
5. Basile, C., Benedetto, D., Caglioti, E., Cristadoro, G., y Degli Esposti, M. (2009). "A plagiarism detection procedure in three steps: selection, matches and 'squares". En: SEPLN 2009 Workshop on Uncovering Plagiarism, Authorshipand Social Software Misuse (PAN 09. 2009, págs. 1-9.). Obtenido de: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.204.3915&rep=rep1&type=pdf
6. Calvo Torres, M. (2017). "Text Analytics para Procesado Semántico" (tesis de maestría). Universidade da Coruña, España. Obtenido de: http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1475.pdf
7. Cortes, C. y Vapnik, V.. (2009). Support-vector networks. Mach Learn 20, 273–297 (1995) doi:10.1007/BF00994018. Obtenido de: http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf
8. Funez, D. G., Errecalde, M.L. (2011). "Detección de plagio intrínseco usando la segmentación de texto". XVIII Congreso Argentino de Ciencias de la Computación. San Luis, Argentina. Obtenido de: http://sedici.unlp.edu.ar/handle/10915/18580
9. Gámez Granados, J. C. (2017). "Uso de técnicas de aprendizaje para clasificación ordinal y regresión" (tesis doctoral). Universidad de Granada, España. Obtenido de: http://decsai.ugr.es/Documentos/tesis_dpto/235.pdf
10. Harley, H. (2006). "English words: A Linguistic Introduction". Michigan, Estados Unidos: Wiley. ISBN: 9780631230311.
11. Jaccard, Paul. (1901). "Etude de la distribution florale dans une portion des Alpes et du Jura". Bulletin de la Societe Vaudoise des Sciences Naturelles. 37. 547-579. 10.5169/seals-266450.

12. Lyon, C., Malcolm, J., y Dickerson, B. (2001). “Detecting short passages of similar text in large document collections”. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 118–125, Pennsylvania.
13. Lyon, C., Barrett, R. y Malcolm, J. (2004). “A theoretical basis to the automated detection of copying between texts, and its practical implementation in the Ferret plagiarism and collusion detector”. In Proceedings of Plagiarism: Prevention, Practice and Policies Conference, Newcastle, UK.
14. Pérez Afonso, J. (2013). “Detección Intrínseca de Plagio” (tesis de maestría). Universidad Politécnica de Valencia, España. Obtenido de: https://riunet.upv.es/handle/10251/43831
15. Real Academia de la Lengua Española. (2019). “plagiar | Definición de plagiar - Diccionario de la lengua española - Edición del Tricentenario“. Madrid. España. Obtenido de https://dle.rae.es/?id=TIZy4Xb
16. Sánchez-Vega, J. F. (2016). “Identificación de plagio parafraseado incorporando estructura, sentido y estilo de los textos” (Tesis Doctoral). INAOE, Puebla, México. Obtenido de: http://personales.upv.es/prosso/resources/SanchezPhD.pdf
17. Stamatatos, E. (2011). “Plagiarism detection using stopword n-grams”. En: Journal of the American Society for Information Science and Technology 62.12, págs. 2512-2527. ISSN: 1532-2890. DOI: 10.1002/asi.21630. URL: http://dx.doi.org/10.1002/asi.21630.

10 y 11 de diciembre de 2020

Análisis comparativo del modelo de *Flipped Learning* virtual y presencial contra clases tradicionales en INFORMÁTICA - Facultad de Ingeniería - UNSa.

Autores: José Ignacio TUERO; Néstor Javier HURTADO; Héctor Iván RODRÍGUEZ; Gisella MAUTINO; Rubén TARCAYA

Institución: Facultad de Ingeniería, Universidad Nacional de Salta. Argentina.

jituero@gmail.com +54 9 3876 858 432 njhurtado@hotmail.com +54 9 3875 104 116

RESUMEN

Flipped Learning (FL-*Aprendizaje Invertido*) es un modelo pedagógico, de tendencia global que aceleró su penetración durante la pandemia, caracterizado por un hecho saliente: el estudiante gestiona su aprendizaje guiado por material confeccionado previamente por los docentes a través de medios TICs. Propone invertir el modelo de la clase tradicional expositiva por parte del docente, donde se imparten contenidos de forma pasiva para el estudiante y que implicaban -cada vez más- un olvidado posterior estudio del material y desarrollo de tareas "en casa" lejos de los expertos y compañeros; por un modelo donde el alumno pre-estudie conceptos introductorios vía TICs y desarrollen actividades cognitivas superiores en los momentos de encuentro sincrónico.

El pre-estudio requiere que la cátedra desarrolle material con su visión, dosificado en pequeñas píldoras cognitivas ("cercanas etaria y culturalmente, vía Youtube", a las nuevas generaciones) que son evaluadas (muy asiduamente -con tendencia diaria: cada hora o fracción prevista; para que esto sea operativamente factible, debe ser automatizado e integrado al LMS-Learning Management Systems). En el momento del encuentro sincrónico, el docente, como facilitador, complementa los saberes hacia niveles más complejos expuestos en la taxonomía de Bloom[1].

Este trabajo presenta principal y estadísticamente los detalles salientes de las últimas dos experiencias de dictado totalmente virtuales durante 2020; compara así, tres modalidades didácticas implementadas en idénticas períodos previos y dictados: BL-Blended Learning (2019, 1° y 2° dictado), presencial clásico (2018, 1° y 2° dictado) contra la virtual antes mencionada.

[1] *Si se quiere ahondar en detalles específico (desde el plano didáctico-pedagógico, procesos de evaluaciones, o verosimilitud de identidades del auditorio) de la versión propia de Flipped Learning implementada se recomiendo comenzar por las presentaciones con evaluación que se publicaron en paralelo en: las 1° Jornadas de Virtualidad e Innovación Educativa de la UNSa (paralelas a las presentes) y en las símiles previas con jornada final del 20/11/2020 del CONFEDI. Asimismo, se invita a solicitar nuevas instanciaciones de un curso taller de cómo implementar la metodología citada en cada cátedras que se dictó en la Facultad de Ingeniería de la UNSa a fines de octubres del corriente, cuya duración sería 20 hs reloj aproximadas.*

El objeto de esta presentación es mostrar la solidez que la estadística le confiere a los resultados obtenidos empíricamente sobre escenarios operativos "de producción" reales. Las experiencias abarcaron a toda la población y todos los temas desarrollados en la cátedra para seis cohortes (doble dictado anual) desde 2018 hasta el presente 2020 (como se dijo), impartida para las carreras de ingenierías "tradicionales": Química, Civil, Industrial, Electromecánica de la Facultad de Ingeniería de la Universidad Nacional de Salta.

Palabras Claves: Flipped-Learning, Aprendizaje-Invertido, Clase-Invertida, Enseñanza-Virtual, Blended-Learning, Estadística-inferencial, Análisis de varianzas, Comparación de medias, ...

INTRODUCCIÓN

Contexto de las experiencias pedagógicas de implementación de Aprendizaje Invertido:

Se reestructuró e impartió todo el contenido de *Informática*[2] basado en FL con clases teórico-prácticas: presenciales durante dos cohortes de 2019 y, totalmente virtual, para las correspondientes del 2020. Estos cuatro dictados (agrupados de a dos) se contrastan estadísticamente contra igual cantidad del 2018, impartidas bajo el enfoque "clásico presencial tradicional" basado en el post-estudio. Siendo los dictados del 2018 representativos y similares al histórico de los cinco años previos.

Así, se analiza formalmente según un diseño experimental por observación, consistente en tres tratamientos, con dos réplicas cada uno.

Objetivos pedagógicos-operativos perseguidos al sostener FL-totalmente virtual, durante la eventualidad pandémica:

- Para el alumnado: que encuentren motivación, a través de FL y medios que le son cercanos, para aprehender contenidos y gestar competencias, transformando pasividad en proactividad.
- Para la materia: al menos, sostener los indicadores de rendimiento alcanzados[3], específicamente, en los últimos años y, dada la realidad existente de virtualidad plena. Si fuera posible, cimentar y potenciar las bases de cambios propuestas por el paradigma educativo FL, en la modalidad. Todo propendiendo a un aprendizaje más activo, participativo y que el alumno tome conciencia que él es el centro y su principal artífice. Proyectar qué actividades podrían mejorar el rendimiento y la efectividad en la materia, tendiendo a medir el "valor agregado" que dichas actividades o vertientes le aportan al enfoque.

[2] *Informática, materia curricular de las cuatro carreras de Ingeniería (Química, Civil, Industrial y Electromecánica). En todos los Planes de Estudio se encuentra situada en el 1° año, 2° cuatrimestre; pero se re-dicta a "contra-cuatrimestre" para evitar el abandono y desgranación, al igual todas las materias del CCA (Ciclo Común Articulado, convenio firmado entre todas las Facultades de Ingeniería de Universidades Nacionales del NOA). Al igual que todas las materias de las ingenierías, el régimen es promocional. El staff docente autor del presente trabajo, se encuentra a cargo de la asignatura desde sus inicios como tal en el nuevo plan de estudio, teniendo 15 años de experiencia acumulada desde entonces con dos dictados anuales (o sea acumula experiencia durante 30 dictados).*

[3] *Se consigna que el primer autor, profesor adjunto responsable de cátedra, como el segundo autor -JTP regular- se desempeñan desde el año 2015 impartiendo la materia desde su primera vez como parte de la currícula del Plan de Estudio de todas la ingenierías y acorde al convenio del consorcio de Facultades homólogas de Universidades Nacionales del NOA. Con doble dictado anual (se imparte en el segundo cuatrimestre del primer año y se "re-dicta" en primer cuatrimestre del año siguiente al del ingreso de cursantes de las carreras de grado).*

Aspectos salientes de la experiencia pedagógica implementada:

El FL implementado tiene innovaciones y particularidades producto de experiencias de BL, durante los dos dictados en 2019; los que facilitaron una acelerada adaptación. Además, permitieron estructurar un compendio de buenas prácticas y recomendaciones para guiar una "reconversión" que podrían extrapolarse a otras asignaturas/realidades análogas.

Las originalidades están basadas en la articulación de tres momentos diferenciados para cada temática abordada. Estos tres momentos son: pre-estudio (guiado, pero autónomo y asíncrono por parte del estudiante), momento sincrónico (orientado a la participación, actividad práctica aplicativa, impulsando la gestación de saberes superiores expuestos en la revisión de Bloom (Anderson-Krathwohl, 2001) y un momento que incluimos y denominamos: post-sincrónico.

Principales momentos que caracterizan el FL implementado: cada temática, es articulada con las posteriores (en la secuencia del desarrollo de contenidos). Cada tema se distribuye bi-semanalmente. En la primera, de lunes a jueves, se facilita (en forma guiada) las actividades del pre-estudio; el viernes está reservado para el encuentro sincrónico.

La segunda semana, se destina al post-estudio: tareas de refuerzo cognitivo, "remediales" (individuales como colectivos) y cierres evaluativos de la temática. Esta segunda semana está inter-solapada con la del pre-estudio de la próxima temática articulada. De esta forma, se optimizan tiempos y rendimiento.

1. Pre-ESTUDIO: totalmente asíncrono, introductorio y autónomo a cargo del alumno (individual o grupalmente, a su voluntad), propone reorganizar la gestión de aprendizaje diaria con píldoras cognitivas mediadas por TICs generadas por los docentes de la cátedra (y vinculadas con otras materias de las ingenierías[4]). Se prevén actividades de aplicación práctica de conceptos auto-aprendidos y evaluaciones segmentadas e integradas también asíncronas e individualizables a través del LMS.

2. Clase Sincrónica-VIRTUAL: momento de encuentro común destinado al despliegue de actividades cognitivas para alcanzar saberes superiores, fomentando la participación activa del alumnado. La construcción cognitiva tiende a ser facilitada por el docente (experto) con actividades lúdicas mediadas por TICs. Se realizan evaluaciones cortas masivas on-line integradas al LMS.

3. POST-Sincrónico: momento de maduración de saberes (individual/grupal, a voluntad) asíncrono de temáticas desarrolladas previamente. Existen consultas guiadas por docentes, con remediales y recuperaciones (por inconvenientes técnicos o cognitivos) producidos durante la semana previa (primera en el desarrollo de esa temática). Esta segunda semana, dentro de cada temática, está solapada con el "*Pre-Estudio"* del próximo saber.

[4] *Todos los temas impartidos semanalmente de la materia tienen incorporados, al menos un video de materias superiores donde se usan o requieren herramientas informáticas. En este video, los responsables o docentes de las otras materias, explican y desarrollan una temática específica -que requiere, o puede requerir- de la informática; luego integramos el desarrollo de ejercicios de programación. A modo ilustrativo, la actual vice-rectora es la partícipe del 1° video de introducción a la programación y el 3° autor de la quinta clase.*

Lo positivo de este tipo de cooperación inter-cátedras es que los alumnos reciben temáticas específicas de expertos en ellos. Para las materias superiores, los videos también le son de utilidad y pueden re-utilizarse internamente. El impacto en la motivación y que el alumno vislumbra conceptos en hechos concretos y prácticos fue realmente valiosísimo; constituyeron asimismo una piedra basal para el replanteo y estructuración de TPs sobre temáticas específicas. Integración y optimización en tiempo y esfuerzo son practicables.

Instrumentos que facilitan logros operativos: para la evaluación de logros operativos de aprendizaje se utilizaron cuestionarios de tipo múltiple-choice en los tres momentos de FL. Cuestionarios asincrónicos de pre-estudio y actividades basadas en los TPs, cuestionarios cortos sincrónicos y un cuestionario asincrónico como cierre de post-estudio. También se evaluó la participación en foros de intercambios y consultas. El LMS de la Facultad fue una herramienta TIC fundamental para la publicación de los instrumentos, recepción de las actividades solicitadas y evaluación por parte de los docentes.

Complementaciones formativas:

El segundo autor se capacitó en la enseñanza y aprendizaje con el enfoque FL logrando la certificación internacional en español emitida por la Universidad de La Rioja, España.

Actualmente el mismo está en la etapa final de defensa de tesis de la *"Maestría en procesos educativos mediados por tecnologías"*, relativa a esta temática y respecto a este ámbito aplicativo; dicha maestría a distancia es con titulación de la Facultad de Ciencias Sociales de la Universidad Nacional de Córdoba, Argentina.

Previo a los desarrollos e implementaciones "en terreno" de los enfoques de FL en 2019, se realizaron indagaciones bibliográficas para identificar los momentos y actividades planteadas por el modelo FL e incluso entre cada una de las modalidades impartidas cada año: 2019 y 2020, se introdujeron variantes y aportes de mejora desde lo didáctico / pedagógico basadas en presunciones apriorísticas del plano citado que se quieren convalidar con este estudio formal inferencial. Asimismo, se compartió internamente dentro de la Facultad de Ingeniería un taller para la trasmisión de la experiencia y las expectativas sobre que otras asignaturas puedan experimentar logros similares.

METODOLOGÍA ESTADÍSTICA

El análisis inferencial formal desarrollado en esta presentación, se enfoca en *prueba paramétrica de análisis de varianzas de más de dos muestras para comparar medias*; consistente en tres tratamientos, con dos réplicas cada uno. En ellos se impartieron todos los contenidos de la currícula de la asignatura INFORMÁTICA bajo estas tres modalidades:

1. El modelo FL en la modalidad totalmente virtual (durante los dos dictados del 2020),
2. El modelo FL con encuentros presenciales (durante dos idénticos dictados del 2019),
3. El modelo clásico tradicional de enseñanza centrado en el post-estudio, precedidos por clases expositivas de docentes (también abordado durante las dos cohortes del 2018).

El principal objetivo consiste en comparar la efectividad de las modalidades y el rendimiento de poblaciones de alumnos que, dependiendo de en qué cohorte cursó, hicieron sus primeras experiencias en introducción a la programación estructurada y orientada a objetos, como en herramientas complementarias informáticas bajo tres enfoques distintos de aprendizaje de todos los contenidos de la asignatura.

En el desarrollo del estudio se utilizó una metodología de investigación con un enfoque mayormente cuantitativo, que incluye aspectos de tipo cualitativo. Algunos de estos objetivos cualitativos considerados en este trabajo fueron:

- Interpretar y entender cómo el modelo de enseñanza y aprendizaje FL puede potenciar el aprendizaje activo por parte del estudiante.

- El desarrollo de competencias básicas dentro de lo propulsado por la Facultad / CONFEDI.
- La motivación en los alumnos de Informática de la Facultad de Ingeniería de la UNSa respecto al rol de materias y su interrelación con otras.

Recolección de Datos

Los datos sobre los cuales se basó el presente estudio, son los que se recabaron formalmente de la asignatura, durante los seis dictados (dos por año) para los tres tratamientos considerados: 2018 ("enseñanza clásica"), 2019 (FL-presencial) y 2020 (FL-virtual).

Se consideraron los alumnos inscriptos al principio de cada cuatrimestre, excluyendo a aquellos que "Nunca Concurrieron" (NC) y los "Abandonos al Inicio" o tempranos (AI). En consecuencia, los guarismos fueron analizados siempre sobre los Alumnos Cursantes (AC). Sobre ellos se distinguió: entre los que Promocionaron (sea en 1°etapa de dictado = P1 ó en 2°etapa recuperatoria = P2); de aquellos que quedaron Libres (en un 1°parcial/recuperación = L1, en un 2°parcial/recuperación = L2 ó en el Global integrador = LG). Asimismo, cabe consignar que también se discriminó a los Abandonos Tardíos (AT) que corresponden a alumnos que optan por no presentarse a las instancias evaluativas o hasta el momento de cierre del dictado, (estando en capacidad reglamentaria de cursado de hacerlo; imposibilitando cualquier acción de la cátedra).

	2°c=dictado 2020	1°c=redictado **2020**	**2°c dictado 2019**	1°c=redictado **2019**	**2°c=dictado 2018**	1°c=redictado **2018**
Cantidad **Inscriptos**	142	143	159	119	233	142
No concurrieron o Abandono temprano (NC ó AB_1°mes)	23	19	36	28	31	34
% Porcentaje de NC / inscriptos	16,20%	13,29%	22,64%	23,53%	13,30%	23,94%
Cursaron el dictado	**119**	**124**	**123**	**91**	**202**	**108**
Promocionaron (P1+P2) sobre los que asistieron	109	114	105	79	175	91
% Porcentaje Promocionados / Cursaron	91,60%	91,94%	85,37%	86,81%	86,63%	84,26%
Nota promedio Promoción (media P1yP2)	7,67	8,31	9,15	8,59	8,37	8,29
Abandonos Tardíos (2°+3°mes o NO asisten al parcial/recuperat./glob.)	6	10	6	3	14	10
% Porcentaje Abandonos Tardíos/Cursaron	5,04%	8,06%	4,88%	3,30%	6,93%	9,26%
Libres x reprobar parcial/recuperación (L1+L2+LG)	4	0	12	9	13	7
% Porcentaje Libres (reprobaron) / Cursaron	3,36%	0,00%	9,76%	9,89%	6,44%	6,48%
Desgrane Tardío (AT + L1, L2, LG)	**10**	**10**	**18**	**12**	**27**	**17**
% Porcentaje desgrane tardío / Cursaron	8,40%	8,06%	14,63%	13,19%	13,37%	15,74%

Tabla 1: Datos Base considerados en el análisis: tres tratamientos, con dos réplicas cada uno.

Respecto a la **información cualitativa**, proviene de instrumentos como cuestionarios y encuestas, proporcionados por el LMS (Moodle) que cuenta la Facultad, ya sea para recolectar información cuali/cuantitativa y específica sobre las opiniones y percepciones de los alumnos.

La encuesta de opinión final sobre FL y la cátedra, se realizó de forma anónima al total de los alumnos al final del cursado de cada modalidad FL implementada y estuvo accesible desde el aula virtual Moodle de la materia sobre la que se realizó la experiencia. Esta incluyó preguntas de respuestas cerradas y abiertas. Los objetivos de la misma fueron conocer:

- las percepciones de los alumnos sobre su propio proceso de aprendizaje.
- las valoraciones de los alumnos sobre el modelo FL específico.
- la evaluación de los alumnos respecto al modelo enseñanza/aprendizaje.

Sobre el sistema evaluativo: los cuestionarios sobre temas técnicos de contenido, se desarrollaron en la plataforma Moodle durante todo el cursado y periódicamente en cada unidad temática. Estos incluyeron preguntas de tipo verdadero-falso, *multiple-choice*, *filling the blanks* y matemáticas.

Técnicas de Procesamiento de Datos

Existen varias pruebas estadísticas que permiten comparar las medias de una variable entre dos o más grupos. Estas pruebas se pueden aplicar cuando se cumplen una serie de supuestos necesarios, bajo diferentes condiciones de aplicación.

Prácticamente todas las hipótesis planteadas se pueden analizar bajo una base paramétrica o una base no paramétrica. La elección de esta base depende básicamente de las características inherentes a la variable a analizar. Las pruebas paramétricas son más potentes que las pruebas no paramétricas, pero exigen que se cumplan una serie de supuestos como:

1. La normalidad en la distribución de la variable.
2. La homocedasticidad (homogeneidad de varianzas).
3. Independencia de las observaciones.

Sin el cumplimiento de los supuestos, estas pruebas pierden todo su potencial y resulta imprescindible recurrir a sus homólogas no paramétricas. En este trabajo se optó por las pruebas paramétricas para realizar una comparación de medias mediante un análisis de varianzas.

1.- Prueba de Normalidad

Dado que la cantidad de datos de cada una de las dos poblaciones, de los tres tratamientos abarcados, es mayor que 50 se realizó la prueba de Kolmogorov-Smirnov; (se usó *SPSS Statistics)*.

Planteo de hipótesis

H_o: Los datos de todas las poblaciones tienen una distribución normal.

H_1: Los datos de al menos una población no tiene una distribución normal.

Nivel de significancia	Criterio de decisión
Alfa = 5% (0.05)	Si p-valor >= 0.05 se acepta la H_0 y se rechaza la H_a.
	Si p-valor < 0.05 se rechaza la H_0 y se acepta la H_a.

Decisión y conclusión

Prueba de Kolmogorov-Smirnov para 1 muestra

		RESIDUOS
N		18
Parámetros normales[a,b]	Media	0,0000
	Desviación estándar	1,96359
Máximas diferencias extremas	Absoluta	,190
	Positivo	,190
	Negativo	-,190
Estadístico de prueba		,190
Sig. asintótica (bilateral)		,084[c]

a. La distribución de prueba es normal.

b. Se calcula a partir de datos.

c. Corrección de significación de Lilliefors.

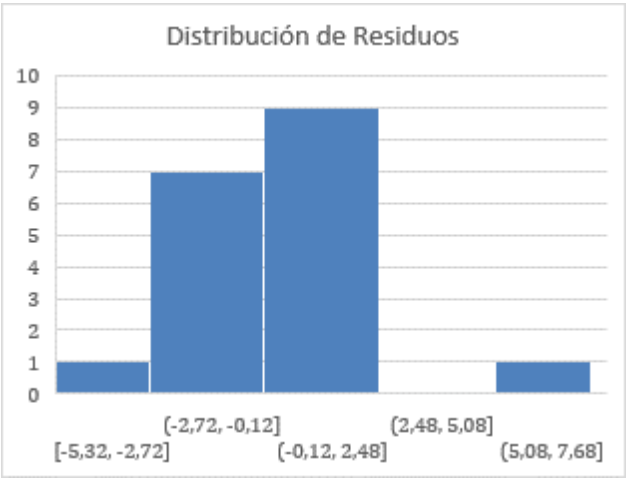


Tabla 2: Prueba de normalidad Kolmogorov-Smirnov.

Como el p-valor resulta >= 0.05 en las poblaciones (dos por cada año 2018, 2019, 2020) entonces se acepta la H_0 y se rechaza la H_1, es decir los datos de las poblaciones de los dos dictados: 2018, 2019 y 2020, **tienen una distribución normal**.

2.- Homocedasticidad

Para el análisis de homogeneidad de varianzas se realizó la prueba de Levene.

Planteo de hipótesis

H_0: Las varianzas de las 3 poblaciones son homogéneas.

H_1: La varianza de al menos una población no es homogénea.

Nivel de significancia

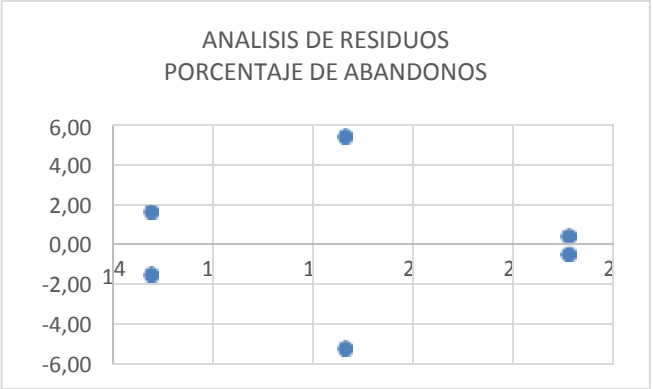
Alfa = 5% (0.05)

Criterio de decisión

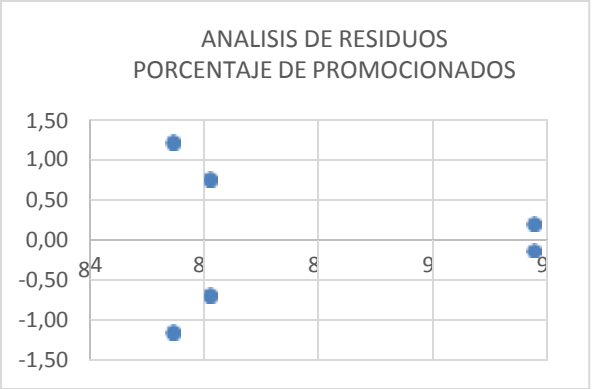
Si p-valor >= 0.05 se acepta la H_0 y se rechaza la Ha.

Si p-valor < 0.05 se rechaza la H_0 y se acepta la Ha.

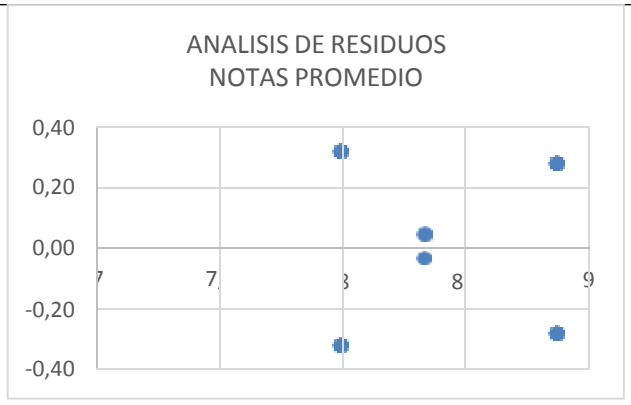
Decisión y conclusión



F= 143,5708938 Fc= 161,4476388



F= 49,10283351 Fc= 161,4476388



F= 64
Fc= 161,447639

Tabla 3: Prueba de homogeneidad de varianzas.

Como el p-valor resulta >= 0.05 en las poblaciones (comprendidos los 2 dictados para 2018, 2019 y 2020) entonces se acepta la H_0 y se rechaza la H_1, es decir **las varianzas de las poblaciones (dos por año) 2018, 2019 y 2020 son homogéneas**.

A) ANÁLISIS DE CANTIDAD DE RETENCIONES

Planteo de hipótesis

H_0: Las medias de las 3 poblaciones son iguales. H_0) μCL=μFLP=μFLV

H_1: Al menos una de las medias es distinta. H_1) Al menos un promedio es distinto.

Nivel de significancia

Alfa = 5% (0.05)

Criterio de decisión

Si p-valor >= 0.05 se acepta la H0 y se rechaza la Ha.

Si p-valor < 0.05 se rechaza la H0 y se acepta la Ha.

Desarrollo

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
CL	2	37,248383	18,6241915	56,59353218
FLP	2	46,1709212	23,0854606	0,394185274
FLV	2	29,48389639	14,74194819	4,235417263

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	69,72595624	2	34,86297812	1,708323738	0,31968305	9,552094496
Dentro de los grupos	61,22313472	3	20,40771157			
Total	**130,949091**	**5**				

Decisión y conclusión

ANALISIS: CANTIDAD DE RETENCIONES		
Porcentaje de abandonos		
CL	FLP	FLV
13,30	22,64	16,20
23,94	23,53	13,29

Tabla 4: resumen % desgranamiento promedio para las 3 metodologías, c/u en sus 2 dictados.

Resultado: F< F_Crítico no se rechaza H_0) con un nivel de significancia de 0.05

Los promedios de desgrane / abandono son iguales.

B) ANÁLISIS DE CANTIDAD DE PROMOCIONADOS

Planteo de hipótesis

H_0: Las medias de las 3 poblaciones son iguales.

H_1: Al menos una de las medias es distinta.

Nivel de significancia

Alfa = 5% (0.05)

Criterio de decisión

Si p-valor >= 0.05 se acepta la H_0 y se rechaza la H_a.

Si p-valor < 0.05 se rechaza la H_0 y se acepta la H_a.

Desarrollo y consideraciones complementarias

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
CL	2	170,8929226	85,44646131	2,818897432
FLP	2	172,1790405	86,08952024	1,04738663
FLV	2	183,5321225	91,76606126	0,05740804

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	48,38265756	2	24,19132878	18,496351	0,02054522	9,552094496
Dentro de los grupos	3,923692102	3	1,307897367			
Total	**52,30634967**	**5**				

Prueba Tukey

Réplicas	r= 2
Compara.	k= 3
Varianza	CME= 1,30789737
	α_E= 0,05
De tablas	q= 5,88
Estadístico	DHS= 4,755

	$\|\mu_i - \mu_j\|$	*LI*	*LS*	¿Diferente de cero?	
CL vs FLP	0,6431	-4,112	5,398	No	Clase FL presenciales tiene igual porcentaje de promocionados que las clases clásicas (normales) presenciales.
CL vs **FLV**	6,3196	1,565	11,075	Sí	Clase **FL virtual tiene mayor porcentaje de promocionados** que las clases normales presenciales.
FLP vs **FLV**	5,6765	0,922	10,432	Sí	Clase **FL virtual tiene mayor porcentaje de promocionados** que las clases FL presenciales

Decisión y conclusión

ANALISIS: CANT.PROMOCIONADOS		
Porcentaje de aprobados		
CL	FLP	**FLV**
86,63	85,37	**91,60**
84,26	86,81	**91,94**

Tabla 5: resumen % de aprobados para 2 dictados, de cada metodología: 2018, 2019, **2020**.

Resultado: F> F_Crítico se rechaza H0) con un nivel de significancia de 0.05
Al menos un promedio es diferente (los dos obtenidos en 2020).

C) ANÁLISIS DE RENDIMIENTO SEGÚN LA NOTA PROMEDIO

Planteo de hipótesis

H_0: Las medias de las 3 poblaciones son iguales.

H_1: Al menos una de las medias es distinta.

Nivel de significancia

Alfa = 5% (0.05)

Criterio de decisión

Si p-valor >= 0.05 se acepta la H_0 y se rechaza la H_a.

Si p-valor < 0.05 se rechaza la H_0 y se acepta la H_a.

Desarrollo

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
CL	2	16,66	8,33	0,0032
FLP	2	17,74	8,87	0,1568
FLV	2	15,98	7,99	0,2048

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	0,787733333	2	0,393866667	3,239035088	0,17807461	9,552094496
Dentro de los grupos	0,3648	3	0,1216			
Total	**1,152533333**	**5**				

Decisión y conclusión

ANALISIS DEL RENDIMIENTO		
Según Nota promedio		
CL	FLP	FLV
8,37	9,15	7,67
8,29	8,59	8,31

Tabla 6: análisis según la nota promedio para evaluar rendimiento de los dos dictados de cada modalidad.

Resultado: F< F_Crítico no se rechaza H0) con un nivel de significancia de 0.05

Los promedios de notas finales, son equiparables (iguales).

CONCLUSIONES

Indicadores y Resultados considerados, primariamente:

Históricamente el porcentaje de promocionados rondaba un 75% (en años previos al 2017-2016-2015). Durante los cuatro dictados: 2018-2019, se mejoró el índice de promocionados a un 85% promedio. La versión FL-totalmente virtual (durante los dos dictados 2020), mejora el porcentaje de promocionados, llevándolo de un 85% a un 91%. El objetivo del estudio y análisis estadístico formal es darle verosimilitud a este apriorismo enunciativo.

El análisis estadístico inferencial-formal[5], permite concluir:

1. La modalidad **FL-virtual (2020) genera mayor porcentaje de aprobados** que la FL-presencial (dictados del 2019) y que las "clásicas normales" históricas (2018 y dictados previos).

2. En lo que respecta a las modalidades con presencialidad, las clases "clásicas normales" tienen el **mismo porcentaje de aprobados** que las clases FL-con momentos presenciales.

3. Las **notas promedio son iguales,** tanto en "clases presenciales normales" como en las dos modalidades de FL-Presencial y Virtual.

4. Siendo que las clases FL, tanto totalmente virtuales como presenciales, no disminuyen el promedio respecto a las "clásicas normales" (años 2018 y previos), **puede adoptarse la modalidad FL**, ya que son más eficientes en términos demostrados fácticamente e inferenciables con certidumbre comprobada[6].

Si bien el incremento en la cantidad de promocionados, durante los dictados FL-virtuales de 2020, se correspondería con una disminución en el promedio del "desgrane" tardío (existe cierta correlación entre ambas variables), lo que demostraría incluso mayor efectividad del FL-virtual sobre su versión con momentos de encuentro presenciales (FL-presencial); o -al menos- que las versiones implementadas desde la virtualidad tienden a un mayor seguimiento individualizado y organizan el tiempo del estudiante más metódicamente. Lo que a la postre redunda en mayor rendimiento.

[5] *Se invita a presenciar/leer detalles del plano didáctico-pedagógico que serán expuestos y publicados también con ISBN durante las I°JoVInEdU-Jornadas Virtuales de Innovación Educativa de la UNSa (que se desarrollan en paralelo a las presentes).*

[6] *Por lo menos para la realidad y la población de los alumnos de Informática de las cuatro Ingenierías "clásicas" que se imparten en la Facultad de Ingeniería de la UNSa.*

Como perspectiva de futuras indagaciones: si bien pareciera haber una disminución en los abandonos iniciales (alumnos que nunca concurrieron o asistieron exiguamente durante el primer mes), parecería ser FL-virtual para INFORMÁTICA, tiende a mantener la cantidad de inscriptos con los que terminan cursando la asignatura.

Valores agregados colaterales: es también significativo resaltar que se puede transformar el paradigma de aprendizaje - enseñanza "clásico" (centrado en el post-estudio), en uno de pre-estudio, donde el estudiante valora su rol activo y toma conciencia de ser artífice de su conocimiento.

Además, los docentes pueden producir estos cambios de una manera no traumática netamente práctica y escalable, incluso en momentos críticos.

BIBLIOGRAFÍA

1. Bergman, J., Santiago, R.: Aprender al revés. Flipped Learning 3.0 y metodologías activas en el aula. 1ra Ed. Paidós, Buenos Aires (2018).
2. Bloom, B.: Taxonomía de los objetivos de la educación. 1ra Ed. Marfil, NY (1979)
3. Diez, S., Andía, L.: Flipped classroom. 33 experiencias que ponen patas arriba el aprendizaje. 1ra Edn. UOC, Barcelona-España (2018).
4. Diez, S. A.: Y mientras tanto en Ingeniería. https://www.theflippedclassroom.es/15578-2/. Accedido el 10/02/2019.
5. Grados Mitteen, A.: Flippeando en la Educación Universitaria. https://www.theflippedclassroom.es/flipeando-en-la-educacion-universitaria. Accedido el 10/02/2019.
6. Prado, A., Lara, L.: Herramientas TIC para la enseñanza de programación empleando aula invertida. En: XIII Congreso sobre Tecnología en Educación & Educación en Tecnología, 217-226. Redunci, Misiones-Argentina (2018).
7. De Zubiría, J. (2020). La educación en tiempos de cuarentena. Revista Semana. https://bit.ly/3fNcOWj. Accedido el 15/03/2019.
8. López M. (2017). Aula invertida o Flipped classroom. Educación, Habilidades & competencias. https://bit.ly/2CMp9vb. Accedido el 17/03/2019.
9. Anderson, L. W., & Krathwohl, D. R. (2001). A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition. New York.

10 y 11 de Diciembre de 2020

Educación virtual en tiempos de pandemia COVID-19: Percepción de alumnos de ingeniería de la Universidad Nacional de Salta

Héctor Rubén Tarcaya, Jorge Oscar Roig Aranda

Facultad de Ingeniería, IIDISA, Facultad de Ciencias Exactas, Universidad Nacional de Salta. Salta. Argentina

rutaunsa@gmail.com

RESUMEN

En tiempos de la pandemia COVID-19, las medidas de aislamiento han ocasionado la suspensión de actividades tradicionales en varios países, entre ellas las clases presenciales en los establecimientos educativos. En la Facultad de Ingeniería de la Universidad Nacional de Salta, seguramente al igual que en otras tantas universidades nacionales, repentinamente iniciaron actividades remotas con los recursos disponibles para afrontar la situación. Este trabajo presenta un relevamiento de percepción de los alumnos sobre la evolución de la educación virtual y sobre los aspectos a mejorar. Para ello se realiza una encuesta online, con opciones de respuestas múltiples y se presentan los resultados en gráficas y tablas estadísticas. Esta información conforma un marco referencial que puede ser de utilidad para oportunidades de mejora en la educación virtual.

Palabras Claves: percepción, educación virtual, universidad, pandemia, COVID-19.

Keywords: perception, virtual education, university, pandemic, COVID-19.

INTRODUCCION

La pandemia por el COVID-19 ha marcado nuevas dinámicas en la salud, la economía, la vivencia social, como así también en la educación en todos sus niveles. Esta enfermedad epidémica mundial ha provocado una situación sin precedentes en todos los ámbitos de actividad, con un estado de confinamiento que ha afectado a todos los niveles educativos (García Peñalvo et al, 2020). Una de las primeras medidas de amplio espectro ha sido el cierre de los centros educativos de todos los niveles (Zubillaga y Gortazar, 2020), lo que ha afectado al 91,3% del total de estudiantes matriculados en el mundo: más de 1.500 millones de personas se han quedado sin poder asistir a sus clases presenciales (UNESCO, 2020). En este contexto, la educación virtual, es decir enseñanza mediante actividades no presenciales, aparece como la alternativa para continuar con las actividades académicas.

Previo a la pandemia COVID-19 ya había nuevos paradigmas en la educación universitaria, entre ellos las redes, la globalización y la sociedad del conocimiento, que configuran un escenario que requiere de nuevas modalidades de comunicación y de intercambio (Tarcaya, 2019b). Si bien la enseñanza virtual no era el común en las universidades públicas (Torrecillas, 2020), algunas publicaciones sobre experiencias planificadas de la enseñanza virtual hacen hincapié en los cambios en las metodologías docentes, en la infraestructura tecnológica y en la evolución de la participación de docentes y alumnos desde el año 2006 (León de Mora et al, 2008).

Durante la pandemia, la suspensión de las actividades docentes presenciales en las universidades ocasionó el repentino surgimiento del dictado de las asignaturas en un formato online, que no puede pretenderse que sea similar en experiencia, planificación y desarrollo a las propuestas que están específicamente diseñadas desde su concepción para impartirse online (Hodges et al, 2020).

El dictado no presencial en épocas de la pandemia COVID-19, surgió repentinamente y con escasa planificación, por lo que ha puesto de manifiesto y magnificado la existencia de tres brechas (Fernández Enguita, 2020):

- Una brecha de acceso y conectividad.
- Una brecha de uso, relacionada al uso compartido de dispositivos e internet en una familia.
- Una brecha de competencias, relacionada con la destreza de docentes y alumnos en la virtualidad.

Ante la situación de aislamiento por pandemia, surgieron respuestas de las universidades públicas y privadas, con actividades no presenciales, y como todo sistema, al tener cambios en los valores de entrada, surgen nuevos riesgos (Tarcaya et al, 2019a) y experiencias según su contexto. Diversas publicaciones muestran que la pandemia COVID-19 ha evidenciado la urgente transformación que demandan los sistemas educativos tradicionales y la importancia de poseer una estrategia educativa virtual, así como un alumnado y un profesorado con habilidades y competencias para la enseñanza y el aprendizaje en el ciberespacio (Tejedor et al, 2020). De hecho, en varios países latinoamericanos, la COVID-19 influyó negativamente en el proceso de enseñanza-aprendizaje, debido a que no se invirtió durante muchos años en la adecuación de los campus virtuales, los sitios web institucionales, las revistas científicas digitales y en la capacitación de los docentes y alumnos en el manejo de las TICs, teniendo que improvisar soluciones tecnológicas (Ríos Campos, 2020).

Algunos relevamientos de percepciones de los estudiantes sobre la educación virtual en España. Italia y Ecuador, muestran valoraciones negativas en el paso a la virtualidad, pues lo asocian, de forma recurrente, con un incremento de la carga lectiva (Tejedor et al, 2020). En lo que respecta a la experiencia de la Facultad de Ingeniería de la Universidad de Salta, un relevamiento de percepción de los alumnos a principios del aislamiento, abril 2020, mostraba también brechas similares a las anteriormente descriptas y una valoración negativa sobre las clases virtuales (Tarcaya et al, 2020c). Con los antecedentes mencionados, este trabajo tiene el objetivo de relevar la percepción de los alumnos sobre la evolución de la educación virtual y sobre los aspectos a mejorar.

El concepto de opinión pública es el de una tendencia, preferencia o postura que una parte de la sociedad o comunidad posee sobre un determinado evento o sobre una situación, o sobre hechos sociales de interés. Las investigaciones sobre opinión pública evidenciaron dos ideas fundamentales: una teórica y otra experimental, alrededor de las cuales se han formado dos escuelas: la clásica y la empírica (Rivadeneira Prada, 1992). La última de ellas, se ocupa de los datos que se extraen del estudio de una determinada población y es la que se utilizará en el presente trabajo.

METODOLOGIA

Para el marco teórico, se realizó un relevamiento bibliográfico en Internet, a través de palabras clave identificadas en español e inglés, así como sus sinónimos y sus combinaciones a través de los operadores booleanos correspondientes (AND, OR, NOT). Se establecieron como criterios de inclusión: utilizar documentos que contengan datos útiles para la investigación, trabajos preferentemente con referencias bibliográficas que contribuyan a sustentar su autenticidad.

El estudio se estructura a los efectos de conocer la opinión pública mediante una encuesta auto administrada. Experiencias que comparaban las encuestas presenciales con las encuestas auto administradas en internet en poblaciones acotadas, arrojaron resultados similares en la tasa de respuesta y con escasas diferencias en la calidad de respuesta (Díaz de Rada, 2012), por lo que la realizada vía internet es una metodología apropiada para este estudio. Este estudio de opinión cuantitativo se diseñó con “preguntas cerradas” (Hentschel, 2002). La técnica se basa en el relevamiento de la percepción mediante encuestas de opinión (Martínez et al, 2003), considerando las recomendaciones de los Códigos ESOMAR. El diseño de la encuesta contempla 5 preguntas relacionadas desde la mejora de la enseñanza virtual respecto al primer cuatrimestre del corriente año 2020, hasta indagar sobre los aspectos a mejorar. Las encuestas fueron analizadas calculando los porcentajes en cada pregunta. Los datos se presentan en gráficas estadísticas que facilitan la visualización de los resultados obtenidos.

La Ficha técnica de las encuestas de opinión se resume en los siguientes datos:

- Población: Alumnos de quinto año de la Facultad de Ingeniería de la Universidad Nacional de Salta, Argentina.
- Modalidad de consulta: vía internet, utilizando Google Forms.
- Periodo de toma de muestra: 30/11/2020 al 03/12/2020.

Se eligió la población de alumnos de quinto año de ingeniería debido a que son los que más participaron de las clases virtuales, según lo relevado en la encuesta de abril 2020.

DESARROLLO

Durante el periodo de toma de muestra, se recibieron 69 respuestas de los 105 alumnos que cursan el quinto año en las diferentes carreras de la Facultad de Ingeniería, lo que representa un 66%.

Pregunta 1: ¿A su criterio, la enseñanza virtual ha mejorado en este segundo cuatrimestre 2020 con respecto al primero?

Tabla 1 - Respuestas a la pregunta 1

Alternativa de respuestas	Porcentual
Sí, bastante	38 %
Si, un poco	43 %
No	16 %
No sabe / No contesta	3 %
	100 %

¿A su criterio, la enseñanza virtual ha mejorado en este segundo cuatrimestre 2020 con respecto al primero?

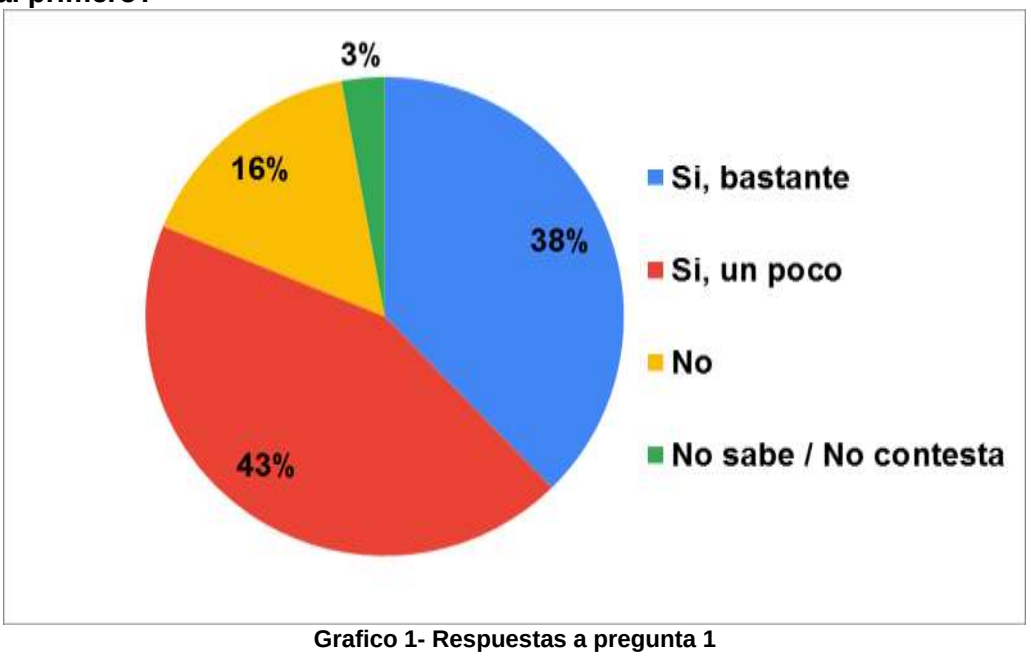


Grafico 1- Respuestas a pregunta 1

Puede verse en la tabla y el gráfico el predominio (81 %) de opinión que mejoraron las clases virtuales, donde un 38 % percibe que mejoraron bastante.

Pregunta 2: ¿Cómo cree que se están dictando las materias de manera virtual?

Tabla 2 –¿Cómo cree que se están dictando las materias de manera virtual?

Alternativa de respuestas	Porcentual (diciembre 2020)	Porcentual (abril 2020)
Bien	52%	44%
Regular	42%	35%
Mal	6%	17%
No sabe / No contesta	0%	4%
	100 %	**100%**

En comparación con el relevamiento realizado en abril, y en concordancia con la respuesta 1, puede observarse en la reciente encuesta que las clases virtuales mejoraron.

Pregunta 3: Comparación de clases virtuales versus presenciales

Tabla 3 - Comparación de clases virtuales versus presenciales

Alternativa de respuestas	Porcentual
Son mejores las clases presenciales	**57 %**
Son iguales	**26 %**
Son mejores las clases virtuales	**16 %**
No sabe / No contesta	**1 %**
	100 %

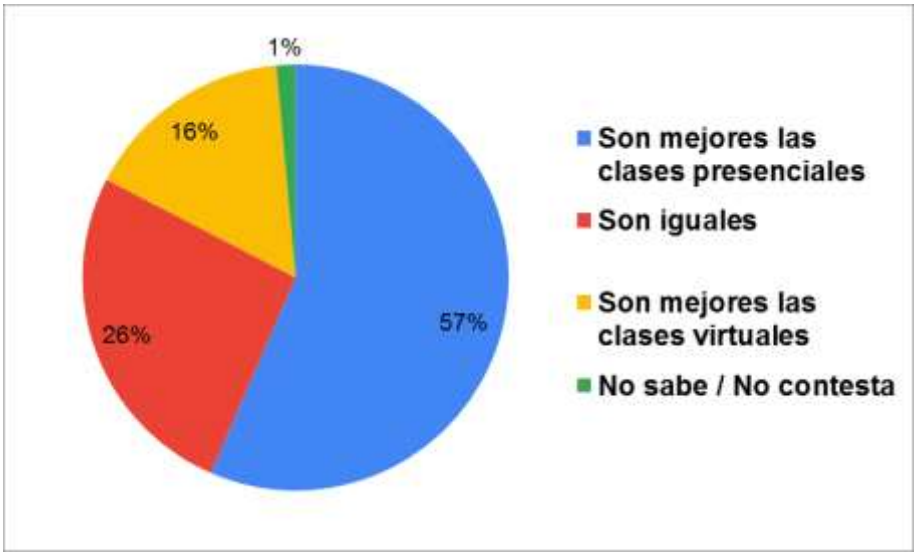


Grafico 2 - Comparación de clases virtuales versus presenciales

Se observa un predominio (57%) de opinión que son mejores las clases presenciales en comparación con las virtuales.

Pregunta 4: Comparación de evaluaciones virtuales versus presenciales

Tabla 4 - Comparación de evaluaciones virtuales versus presenciales

Alternativa de respuestas	Porcentual
Son mejores las evaluaciones presenciales	39 %
Son iguales	29 %
Son mejores las evaluaciones virtuales	28 %
No sabe / No contesta	4 %
	100 %

En lo que respecta a evaluaciones, están más repartidas las opiniones con una leve mayoría que coincide que son mejores las evaluaciones presenciales.

Pregunta 5: ¿A su criterio, qué debería mejorar en la enseñanza virtual?

Tabla 5 – La planificación de las clases

Alternativa de respuestas	Porcentual
Si, bastante	30 %
Si, algo	35 %
No, nada	35 %
No sabe / No contesta	0 %
	100 %

Tabla 6 – La conectividad por parte de los docentes

Alternativa de respuestas	Porcentual
Si, bastante	7 %
Si, algo	49 %
No, nada	44 %
No sabe / No contesta	0 %
	100 %

Tabla 7 – La destreza de los docentes en la virtualidad

Alternativa de respuestas	Porcentual
Si, bastante	35 %
Si, algo	55 %
No, nada	10 %
No sabe / No contesta	0 %
	100 %

Tabla 8 – La participación de los alumnos

Alternativa de respuestas	Porcentual
Si, bastante	40 %
Si, algo	51 %
No, nada	6 %
No sabe / No contesta	3 %
	100 %

Tabla 9 – La modalidad de la evaluación

Alternativa de respuestas	Porcentual
Si, bastante	22 %
Si, algo	46 %
No, nada	32 %
No sabe / No contesta	0 %
	100 %

El relevamiento muestra una significativa coincidencia en deben mejorar la destreza de los docentes en la virtualidad (90%) y la participación de los alumnos (91%).

En menor medida, existe una coincidencia de opinión que deben mejorar la planificación de las clases (65%) y la modalidad de la evaluación (68%). La conectividad de los docentes, al parecer es bastante aceptable, ya que el 44% de las opiniones coincide en que no necesita mejorar.

CONCLUSIONES

La encuesta como herramienta de investigación permitió conocer la percepción de los alumnos respecto a la evaluación de la educación virtual y de los aspectos a mejorar, alcanzado el objetivo planteado.

La educación virtual como respuesta para afrontar el contexto de aislamiento por pandemia de a poco va evolucionando en el camino de la mejora continua. Se observa una coincidencia de los alumnos en opinar que la enseñanza virtual ha mejorado respecto al primer cuatrimestre y una leve tendencia creciente de que las clases virtuales se están desarrollando bien. El relevamiento también muestra los aspectos a mejorar, donde sobresalen la destreza de los docentes en la virtualidad y la participación de los alumnos. Esta percepción coincide en parte con las opiniones de estudiantes en otras universidades públicas argentinas y de otros países.

Este es un relevamiento puede ser complementado con muestras mayores y con exploraciones en otros ítems relacionados a la educación virtual en busca de identificar oportunidades que posibiliten planes de acciones para mejorar.

BIBLIOGRAFIA

Díaz de Rada (2012). "Ventajas e inconvenientes de la encuesta por Internet". *Papers*, 97(1), pp. 193-223.

ESOMAR. (2016). Código internacional para la práctica de la investigación de mercados, opinión social y del análisis de datos. ICC. Francia

ESOMAR-WAPOR. (1998). Guía para la realización de sondeos de opinión. ICC. Francia.

Fernández Enguita, M. (2020). Una pandemia imprevisible ha traído la brecha previsible. Recuperado de https://bit.ly/2VT3kzU

García-Peñalvo, F., Corell, A., Abella-García, V., Grande, M. (2020). La evaluación online en la educación superior en tiempos de la COVID-19. Education in the Knowledge Society 21 (2020) article 12. Ediciones Universidad de Salamanca | https://doi.org/10.14201/eks.23013

Hentschel, H. (2002). Encuestas y opinión pública. Aspectos metodológicos. Edivern. Buenos Aires. Argentina.

Hodges, C.,Moore, S.,Lockee, B.,Trust, T. y Bond, A. (2020). The difference between emergency remote teaching and online learning. Educause Review. Recuperado de https://bit.ly/3b0Nzx7

León de Mora, C., Camarillo Casado, J., Ramos Gómez, M., Sanchez Aguilar, M. (2008). La enseñanza virtual en la Universidad de Sevilla. Pixel-Bit. Revista de Medios y Educación. Nº 32 Marzo 2008 pp. 7-20.

Martínez, M. – Valentín, C. (2003). Diseño de encuestas de opinión. Editorial RA-MA. Barcelona. España.

Ríos Campos, C. (2020). COVID-19 y Educación Superior Universitaria Pública del Perú. Revista Clake Education, 1(02), 1-1. Recuperado de: http://revistaclakeeducation.com/ojs/index.php/Multidisciplinaria/article/view/16

Rivadeneira Prada, R. (1992). La opinión pública. Análisis, estructura y métodos para su estudio. Editorial Trillas. México.

Tarcaya, H.R., Arenas A.N., Plaza, G. (2019a). Evolution and trends in management systems based on international standards. *Impactos das tecnologias nas ciências sociais aplicadas 3*. [online]. Vol.3 pp. 108-114. [ISBN 978-85-7247-213-5]. DOI 10.22533/at.ed.135192703. Atena Editora. Brasil.

Tarcaya, H.R., Rodríguez, H.I. (2019b). Jornadas de estadística aplicada como estrategia para contribuir al desarrollo de competencias en alumnos de ingeniería. El enfoque por competencias en la ciencias básicas: casos y ejemplos en educación en Ingeniería / Marisa Battisti... [et al.]; compilado por Uriel Cukierman ; Guillermo Kalocai. - 1a ed.-Ciudad Autónoma de Buenos Aires : edUTecNe ; Buenos Aires : CONFEDI - CIIE, 2019. ISBN 978-987-4998-16-3

Tarcaya H.R., Rodríguez H.I., Arenas A.N., Pistán H.D. y Kolodziej S.F. (2020c). La enseñanza virtual durante la pandemia COVID-19. Experiencias en la Universidad Nacional de Salta, Argentina. Salão de Conhecimento 2020. XXVIII Seminário de Iniciação Científica. Ijui. Brasil.

Tejedor, S., Cervi, L., Tusa, F. y Parola, A. (2020). Educación en tiempos de pandemia: reflexiones de alumnos y profesores sobre la enseñanza virtual universitaria en España, Italia y Ecuador. Revista Latina de Comunicación Social, 78, 1-21. https://www.doi.org/10.4185/RLCS-2020-1466

Torrecillas, C. (2020). El reto de la docencia online para las universidades públicas españolas ante la pandemia del Covid-19. Instituto Complutense de Estudios Internacionales (ICEI), N°16. Universidad Complutense de Madrid. Recuperado de https://www.ucm.es/icei/file/iceipapercovid16

UNESCO. (2020). COVID-19 Impact on Education. Recuperado de https://bit.ly/2yJW4yy

Zubillaga, A. y Gortazar, L. (2020). COVID-19 y educación: Problemas, respuestas y escenarios Madrid, España: Fundación Cotec para la Innovación. Recuperado de https://bit.ly/3auXnP8

10 y 11 de Diciembre de 2020

"Estadística aplicada para la conformación de la Sala de Situación de Salud en el Hospital Nuestra Señora Del Carmen - Jujuy"

Autores: Mg. Ana María Chalabe, Esp. Susana Angélica Chalabe, Dr. Javier Altamirano, Lic. Blanca Zumbay
Instituciones: FHyCS - FI - Universidad Nacional de Jujuy - Hosp.N.Sra, del Carmen - Ministerio de Salud
Datos de contacto:chalabeana@hotmail.com – Tel.: 388 4718554

RESUMEN

Las Salas de Situación de Salud son herramientas muy útiles para procesar información acerca de eventos de salud-enfermedad ya que permiten orientar las intervenciones sociales y sanitarias puntuales que tiendan a disminuir o eliminar factores de riesgo específicos. El desarrollo de las mismas conduce a identificar áreas, población o grupos donde priorizar el uso de los recursos existentes a efectos de preservar la salud de las comunidades. Esta publicación tiene como objetivo presentar el diseño e implementación de una Sala de Situación de Salud en el municipio de El Carmen, procurando relacionar datos de las familias que manifiestan situaciones de salud o enfermedad con un enfoque espacial que facilite los procesos de análisis y de toma de decisiones. Tiene el propósito de revisar los procedimientos contando con la colaboración de las cátedras de Estadística y Epidemiología de la FHyCS para establecer una serie de protocolos con el fin de transformar datos dispersos y sectoriales en información integrada que permita prever problemas de salud. Con las nuevas prácticas y herramientas espaciales incorporadas, el lema de Epidemiología: "información para la acción", se concreta, los datos están disponibles y facilitarán la toma de decisiones en beneficio de la salud comunitaria.

Palabras Claves: Estadística – Epidemiología - Sala de Situación – Vigilancia de la Salud.

INTRODUCCIÓN

La extensión universitaria tiene el espíritu de promover el vínculo entre la Universidad y la comunidad generando espacios de intercambio, transferencia de conocimientos y cooperación entre los distintos actores sociales que componen la sociedad. Esta vinculación es imprescindible para favorecer el abordaje y la búsqueda de soluciones que permitan superar problemáticas sociales, especialmente aquellas relacionadas con la salud de la población, por ello, en este

trabajo se exponen resultados alcanzados a partir de la investigación aplicada en proyectos de extensión.
En éste contexto, se firma un Acta Acuerdo entre la Facultad de Humanidades y Ciencias Sociales de la Universidad Nacional de Jujuy y el Hospital Nuestra Señora de El Carmen, perteneciente al Área Programática V del Ministerio de Salud de la provincia para desarrollar el proyecto de extensión "Diseño e implementación de la Sala de Situación de Salud – Hospital Nuestra Señora del Carmen" que tiene como objetivos: a) Relevar y validar la información, b) Revisar y dilucidar las necesidades de protocolos para el tratamiento de la información, c) Establecer un programa de capacitación para el manejo estadístico de información y de herramientas lógicas y geoespaciales, d) Desarrollar, seleccionar y describir indicadores a utilizar, e) Poner en marcha y actualizar la Sala de Situación y f) Facilitar a los tomadores de decisiones el producto alcanzado g) Probar y ajustar procedimientos en el Hospital a los fines que este continúe ejecutándose al finalizar el proyecto. Éste proyecto se genera desde las cátedras de Estadística de la carrera de Licenciatura en Antropología y la cátedra Epidemiología y Estadística de la Licenciatura en Educación para la Salud, de la Facultad de Humanidades y Ciencias Sociales de la Universidad Nacional de Jujuy
El sistema de Salud en la provincia de Jujuy se organiza en cuatro regiones: Quebrada, Puna, Yungas, Valles; en esta última región que comprende los departamentos de El Carmen y San Antonio se asienta el Hospital Nuestra Señora de El Carmen, que como se mencionó, es el sitio donde se ejecuta el proyecto.
La ciudad de El Carmen es la capital del departamento homónimo, tiene una población de 24.800 habitantes según censo del año 2019 realizado por el departamento de Atención Primaria de la Salud (APS) El 80 % de esta población es estable. Asimismo, de un total de 34 barrios, en seis de ellos se asienta una población joven de alta natalidad, en cuatro se observa una población estacionaria o envejecida.
En general, son familias estables y tradicionales cuyos apellidos vienen de varias generaciones resultando una sociedad muy arraigada a sus costumbres, cerrada a los cambios que imponen las nuevas generaciones; se observa con frecuencia que los jóvenes se marchan en busca de mejores oportunidades. La ciudad de El Carmen se caracteriza por ser una de las poblaciones de la Provincia de Jujuy con mayor número de adultos.

Justificación y Relevancia
Existían en el Hospital Ntra. Sra. Del Carmen múltiples sistemas inconexos (Bioestadística, Atención Primaria de la Salud, Recursos Humanos, Laboratorio, Maternidad, Epidemiología, etc…), donde cada uno definía sus propias variables, instrumentos de registro y vías de circulación; dependían y reportaban a oficinas diferentes sin voluntad de compartir la información; asimismo, la integración de datos era un tema complejo ya que el tratamiento de los mismos estaba diseñado a medida de las necesidades singulares de cada área operativa, sin considerar la totalidad y en general, produciendo informes parciales; por otra parte no siempre estaban accesibles o disponibles al momento de la toma de decisiones.
Todo éste sistema particionado no permitía detectar situaciones de riesgo, ya sea epidemiológico, social o simplemente conocer cabalmente el estado de salud de la población, no era posible responder a preguntas como ¿en ésta semana que patologías deberíamos esperar?, ó ¿cuantas embarazadas tendrían que internarse en éste mes para el parto?, ¿entre las embarazadas existe alguna en riesgo de cualquier tipo?, otras cuestiones de mayor envergadura como ¿de qué se enferma la población?, ¿cuáles son las principales causas de derivación?. Y se podría estar enumerando dudas sin cubrir la totalidad de la realidad, pero las premisas que alentaron éste trabajo de extensión fue organizar y validar las diferentes fuentes de información para poner en la superficie datos conectados que respondan a las preguntas enunciadas. Se hizo muy evidente la

necesidad de contar con información de calidad, que por ejemplo permita conocer la ubicación e identificación rápida de las personas de mayor riesgo de enfermar. Era necesario trabajar con indicadores diferenciales; como: indicadores demográficos, de cobertura, de riesgo; la dimensión, estructura, evolución y características de la población. Se abordó desde la definición de los datos, la información como insumo del análisis estadístico y epidemiológico y el monitoreo de la situación de salud, para fortalecer y crear mecanismos institucionales que garantizarán a futuro el flujo del dato -información sanitaria del Hospital de El Carmen.

Marco teórico

Entre los antecedentes más destacados en América, se distingue Brasil que incorporó el Sistema Único de Salud (SUS) como el resultado de un proceso social que definió a la salud como un derecho colectivo y como una responsabilidad de estado de proveerla, bajo los principios de la universalidad, integralidad y equidad de la atención, y toma como elemento central el proceso de gestión, a través del cual se organiza y sistematiza la información estratégica para fundamentar las decisiones en las políticas de salud.

Un desafío impostergable del SUS brasilero, es el de promover la sistematización y la democratización de la información sanitaria, buscando ofrecer elementos para el proceso de toma de decisión en los gestores públicos de salud, para el efectivo mejoramiento de las condiciones de vida y salud de la población.[1]

En nuestro país, se implementa el Proyecto VIGI+A (2000), fue una iniciativa del Ministerio de Salud de la Nación que tuvo como objetivos: fortalecer la vigilancia de la salud, el control de enfermedades transmisibles y no transmisibles y la promoción de la salud[2]. En este contexto, planteó la difusión e implementación de las Salas de Situación a nivel nacional y jurisdiccional, para lo cual se sugirió la construcción de estructuras en los ministerios provinciales.

La provincia de Jujuy inicia el proceso mediante la designación de referentes epidemiológicos a nivel Área Programática. El Área Programática V, Hospital Nuestra Señora del Carmen incorpora a su organización el Servicio de Epidemiología que tiene como finalidad desarrollar e implementar una cultura de análisis y uso de la información sanitaria para identificar los principales problemas, sus determinantes y posibles intervenciones, de forma a reorientar las acciones de salud hacia la atención de las prioridades identificadas.

METODOLOGÍA

El trabajo se inició desarrollando el encuentro de los actores intervinientes con las autoridades de la institución para elaborar el listado de preguntas guía. Se procedió a la preparación y puesta a punto de sistemas de información lógicos y geográficos que incorporaron la información de la visita domiciliaria por parte de agentes sanitarios, con los diagnósticos de atención, de derivaciones entre otros, se consolidaron y profundizaron las acciones de salud, integrando y proyectando desde el individuo, hacia la familia y a la comunidad.

Se relevó la información que se recaba en la institución, se validó la misma, con responsables y producciones existentes.

Se realizó el diseño de la Sala de Situación de Salud acorde a jefes de servicio y dirección y se establecieron protocolos de tratamiento de la información, diferenciando los ejes conductores de una Sala de Situación local. Fue necesaria la capacitación básica para el manejo de información y herramientas lógicas y geoespaciales, con el planteo de los fundamentos teóricos de construcción

[1] Salas de Situación de Salud de Brasil – OPS – 2010 - http://www.msal.gob.ar/saladesituacion/Biblio /Sala_Situacion _de_Salud.pdf

[2] SALA DE SITUACIÓN DE ARGENTINA – 2013- http://www.msal.gob.ar/saladesituacion/Biblio/experiencias-exitosas-sala-situacion-argentina.pdf

de indicadores y su análisis. Se recuperaron y analizaron las bases de datos existentes, se seleccionaron y consolidaron las mismas.
Se definieron responsables directos e indirectos y roles para la actualización continua de la Sala de Situación, se programaron las reuniones de difusión hacia adentro de la organización hospitalaria y hacia afuera – organizaciones sociales y gubernamentales – de la información de la Sala de Situación, poniendo a disposición de los tomadores de decisiones del producto concluido.

DESARROLLO

Se seleccionaron 4 ejes principales:

1. Derivaciones
2. Traumas y lesiones
3. Enfermedades denunciables
4. Corredores endémicos y brotes

1 - Derivaciones

Se categorizaron las derivaciones de la guardia del nosocomio, catalogaron los formularios utilizados, y de los datos estadísticos, resumidos en la Tabla 1, emergen las lesiones (tratadas en forma particular en otro eje), las derivaciones por causas ginecológicas que luego se desglosan en la Tabla 2, y por su envergadura quedan pendientes los tumores y derivaciones neurológicas. Ambas categoría exceden a ésta primera etapa de implementación de la Sala de Situación, quedando pendientes para la siguiente.

TABLA 1 -TIPO DE DERIVACIONES – AÑO 2019

LESIONES	512
QUIRURGICO	209
OBSTETRICO - GINECOLOGICO	194
ALTA MEDICA - QTA -COBETURA - DOMIC	173
CARDIACO	158
NEUROLOGICO	149
DIGESTIVO - CLINICO - URINARIO	146
RESPIRATORIO	133
METABOLICO	45
INFECCIOSO	41
INTERCONSULTA - OBITO	37
SOCIALES	28
ABORTO	27
TUMORES	23
Total general – Contrarreferencia??? Fallecidos?	**1875**

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

Con respecto a las derivaciones gineco-obstétricas se recalcan los abortos, que demandan un registro más pormenorizado a los efectos de incorporar a esas mujeres a los programas de salud reproductiva funcionando en la institución. Se observa con preocupación el número de parturientas en “alto riesgo” y “gestosis” por la gravedad de la patología.

TABLA 2: DIAGNÓSTICO DE DERIVACIONES GINECÓLOGICAS – OBSTÉTRICAS

CATEGORÍA	ALISOS	CHAMICAL	EL CARMEN	EL CARMEN RURAL	MONT. RURAL	MONTERRICO	PERICO	PERICO RURAL	SAN ANTONIO	TOTAL
ABORTO	1	2	12	3	1	6		2		27
ALTO RIESGO	3	6	36	5	9	15	5	7	4	90
GESTOSIS		1	5		2	2			4	14
NEONATAL			9	2	1	5			1	18
OBSTETRICA	1	1	22	10	2	5		4	5	50
QUIRURGICO			7	3	4	7	1	3		25
SOCIALES			2	1				1		4
TOTAL	5	10	93	24	19	40	6	17	14	228

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

Asimismo se observa que éste hospital en su servicio de guardia recepciona pacientes de otras localidades de la provincia en un 30%, involucrando gasto en infraestructura y recursos, tanto humanos como económicos.

TABLA 3 - DERIVACIONES SEGÚN LOCALIDAD Y DESENLACE – AÑO 2019

LOCALIDAD	CANTIDAD	DESENLACE	CANTIDAD
EL CARMEN	1141	H.N.S.C.	836
SAN ANTONIO	147	SORIA	425
EL CARMEN RURAL	125	H.M.I.	336
CHAMICAL	63	DOMICILIO	116
MONTERRICO	99	H.S.R.	32
MONT. RURAL	46	H.W.G.	9
PERICO RURAL	45	ZABALA	3
PERICO	44	SAN JOSE	27
S.S. DE JUJUY	24	LOS LAPACHOS	28
ALISOS	23	OTRO	12
ALTO COMEDERO	18	FATIMA	27
OTRA PCIA	11	ROSARIO	9
PALPALA - SAN PEDRO	7	LAVALLE	8
OTROS - QTA	82	NORTE	7
TOTAL	1875	TOTAL (118)	1875

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

2 - Traumas y lesiones

La tasa de incidencia por lesiones por transporte para El Carmen, Jujuy, supera la media provincial y el hospital local da respuestas a la emergencia; se incorpora al Sistema de Vigilancia de Lesiones, (SIVILE) y conforma un grupo de estudio con la universidad local con el propósito de utilizar herramientas SIG (Sistemas de Información Geográficos) para el análisis de los accidentes, considerando que el contexto espacial y geográfico es esencial para la compresión del evento que ocasiona la lesión.
Por ello, este proyecto tuvo como objetivo consolidar un sistema de vigilancia de lesiones definiendo los modelos conceptuales, lógicos y físicos necesarios de implementar para llevar a cabo el proceso de georreferenciación de las variables estandarizadas por SIVILE, de forma tal que los datos sean interoperables para que puedan ser examinados y observados en un marco de integración de distintos factores como por ejemplo los socioeconómicas, las configuraciones del espacio urbano-rural y las geometrías de la redes viarias.
Como resultado, la Unidad Centinela de Lesiones (UCL) que pertenece al SIVILE, dispone de información espacial adecuada, confiable y oportuna para la vigilancia epidemiológica de lesiones, con los formatos adecuados en un sistema (SIG) que permite una evaluación colectiva del entorno, posibilitando la interacción con otras instituciones como el municipio local a los fines de conjugar acciones para mejorar la capacidad de respuesta, revisar o proponer nuevas legislaciones y en definitiva, trabajar en forma integrada para reducir las lesiones consolidando entre todos los programas de vigilancia.
Se analizaron los registros de guardia y de la Unidad Centinela para dimensionar el número absoluto de eventos y el porcentaje de cobertura de la Unidad, requiriéndose en un futuro la revisión de los componentes para aumentar su envergadura.
Los datos enunciados se observan en la Tabla 4, Tabla 5 y Tabla 6.

TABLA 4 - TRAUMAS Y LESIONES – AÑO 2019

	TOTAL	AMBUL	SIVILE	ODA	NINI	% NO
HOGAR	582	59	136	531	387	66%
IP	116			116	116	100%
S.E.	59	4	9	58	45	76%
VIAL	1975	273	432	1855	1266	64%
TOTAL	2732	336	577	2559	1820	67%

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

Referencias: HOGAR: Accidente en el hogar - I.P: Informe Policial – S.E: Sin especificar – VIAL: Evento vial

TABLA 5 - UNIDAD CENTINELA DE LESIONES -2019

Nombre	Total	%
18 - Mordedura de perro	256	28%
01 - Lesiones por transporte	199	22%
04 - Caída en el mismo nivel	162	18%
03 - Golpe (por objetos, por personas)	92	10%
05 - Caída de un nivel a otro	61	7%
06 - Trauma con objeto punzo-cortante	61	7%
17 - Contacto traumático con animal/planta	44	5%
02 - Envenenamiento o intoxicaciones	10	1%
11 - Contacto con sustancias calientes	8	1%
22 - Otra lesión, especifique	25	3%
Total	**918**	**100%**
TOTAL DE LESIONES: 2732 – TOTAL FICHAS: 918		

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

TABLA 6 - DISTRIBUCIÓN DE LESIONES DE CAUSA EXTERNA POR SEVERIDAD Y SEXO

SISTEMA DE VIGILANCIA DE LESIONES DE CAUSA EXTERNA (SIVILE)							
HOSP NUESTRA SEÑORA DEL CARMEN – JUJUY – AÑO 2019							
	Severidad				*Fallecidos*		
Sexo	Leve	Moderado	Severo	Total	En la unidad	En el lugar	Total
Masculino	331	199	14	**544**	**4**	-	**4**
Femenino	251	112	3	**366**	-	-	-
Desconocido	-	-	-	**4**	-	-	-
Total	**582**	**311**	**17**	**914**	**4**	-	**4**

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

Otro aspecto a considerar es la intencionalidad de las lesiones, siguiendo las categorías del sistema nacional de lesiones resultan los datos volcados en la Tabla 7, resaltando los intentos de suicidio entre los 15 a 30 años de edad, población muy joven, económicamente activa.

TABLA 7 - DISTRIBUCIÓN DE LESIONES DE CAUSA EXTERNA POR INTENCIONALIDAD Y SEXO

HOSP. NUESTRA SEÑORA DEL CARMEN – JUJUY – AÑO 2019				
	F	M	Total	Deriv
01 - Intencional - Interpersonal	8	21	29	
02 - Intencional - Autoinflingida	6	10	16	26
03 - No Intencional	351	513	864	-
04 - Intervención por agente legal	2	1	3	
05 - Intención no determinada	3	3	6	
Total	**370**	**548**	**918**	
Intentos de suicidio 15 a 30 años	**9**	**12**	**21**	**21**

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

3 - Enfermedades denunciables

Las enfermedades de notificación obligatoria se denuncian según semana epidemiológica en el SISA – Sistema Integrado de Información Sanitaria Argentina – siendo los servicios de Bioestadística el encargado de recabar el dato, dicho dato se corrobora desde el servicio de Epidemiología local y procede a la denuncia online, se puede observar el peso de las enfermedades respiratorias y en segundo lugar las lesiones por causa externa. El resumen elaborado se presenta en la Tabla 8.

TABLA 8 – DISTRIBUCIÓN SEGÚN CATEGORÍAS SISA – AÑO 2019

SISA Sistema Integrado de Información Sanitaria Argentino — Cobertura Universal de Salud — Ministerio de Salud Presidencia de la Nación

Semanas	1	2	3	4	,,,	35	36	Total
Enfermedades de transmisión sexual						1	1	24
Eventos provinciales JUJUY - Hoja de tabaco verde - mordedura de can	16	15	6	13	,,,	11	13	317
Gastroentéricas	70	85	85	90	,,,,	53	48	2010
Inmunoprevenibles				1	,,,,		1	20
Lesiones por causas externas	87	93	96	103	,,,,	92	70	2399
Respiratorias	46	36	30	71	,,,,,	106	99	3391

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

4 - Corredores endémicos y brotes

El operativo "Alto Al Dengue" se configuró como la planificación estratégica a fin de focalizar acciones de vigilancia, prevención y control, búsqueda, control clínico, monitoreo focal entomológico y saneamiento ambiental con participación comunitaria, a partir de 2006.
La mecánica se organizó desde una mesa gerencial gestionada desde el servicio de Epidemiología del nosocomio, con representantes multisectoriales, Equipo de Salud, Municipio, Policía, Consejo deliberante, Prensa y Difusión, con el acompañamiento de la Universidad, que se reúnen en forma diaria durante el tratamiento de brote y mensualmente en períodos intermedios.
Para la construcción de la Sala de Situación de Salud se analiza los brotes controlados durante el año 2019.
La metodología empleada fue revisada y aplicada en el brote del año 2020 con 63 sospechosos y 6 casos confirmados de Dengue. Se puede visualizar la mecánica de trabajo con buffer de acciones de vigilancia activa, saneamiento ambiental y educación para la salud en el Gráfico 1 y la Tabla 9.

GRÁFICO 1 - BROTES CONTROLADOS – AÑO 2019

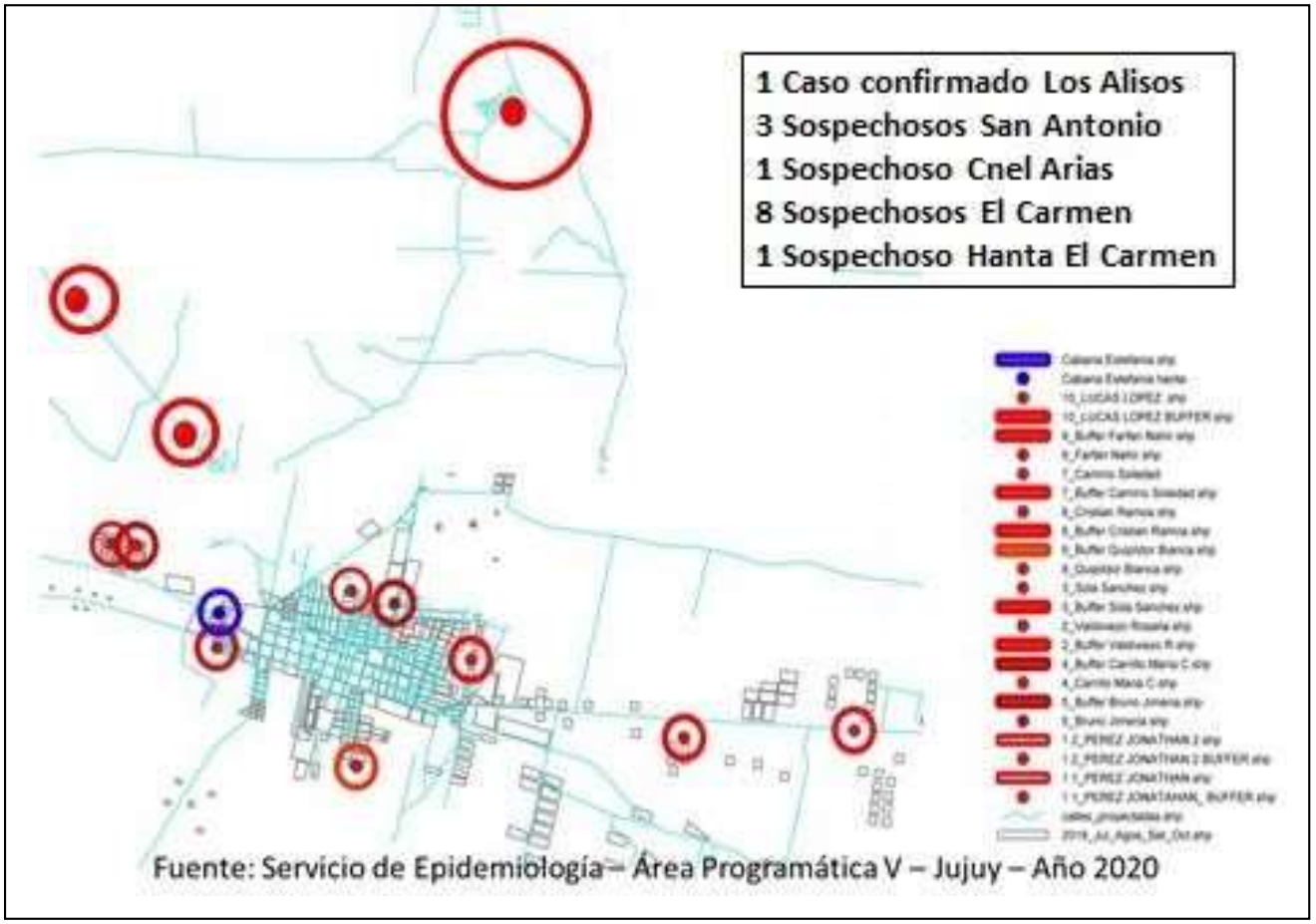


TABLA 9 – MONITOREO DE DENGUE – AÑO 2019

GRUPO	MANZANAS VIGILADAS	CASAS VISIT.	FOCOS	HAB	CASAS TOT.	% CASAS VISIT.	RECIPIENTES	BALDIOS	CONSTRUC
CEMENTERIO	1	1	50	0	1	100%	1150	0	0
VILLA MARIA	20	382	18	1343	415	92%	2278	37	10
16 DE JULIO	8	148	14	703	155	95%	962	0	0
SAN CARLOS	15	222	12	862	237	94%	1364	13	10
APAZA	7	119	7	485	126	94%	1346	14	6
CCC	1	29	6	125	30	97%	282	0	0
SANTA ANA	7	128	5	506	132	97%	1015	1	0
LA LOMA	1	24	5	112	24	100%	294	0	0
TUPAC - SANTA ANA	4	81	4	318	85	95%	825	2	0
MALVINAS ARG.	3	41	4	180	41	100%	560	0	0
ROSALIA	25	234	4	829	250	94%	1234	58	21
TOTAL	528	5406	162	19119	5836	93%	38325	1105	543

IND. VIV. = 3% (0,5%)- IND. BREATAU= 3% (0,5%)- IND. RECIP.= 0,4% (0%) -VIV: 22

BARRIOS : 50 - FINCAS : 427 - RURAL: 5 ZONAS (206 VIV) – ESP. PUBL. E INST. 104

Fuente: Servicio de Epidemiología – Área Programática V – Jujuy – Año 2020

Al finalizar el año de desarrollo del proyecto de extensión queda pendiente trabajar en protocolos, información e indicadores de Tuberculosis, Sífilis y H.I.V., Dengue y Vigilancia Ambiental, Control de Rabia y Causas de morbilidad y mortalidad.

Estrategias de Sostenibilidad a Futuro

Para lograr la sostenibilidad a futuro es necesaria una profunda transformación tanto en el proceso de gestión de información de la salud — donde la información sea realmente utilizada–, como en lo que se refiere a compartir y difundir la misma.

Frente a este desafío, se ve la necesidad de estimular la mantención de mecanismos instituciones que permitan el flujo de la información, identificando responsables en las distintas áreas de manera que el equipo de la Sala de Situación pueda acceder en tiempo y forma a la información necesaria para realizar los análisis pertinentes.

El flujo de información hacia y desde la Sala de Situación, además de hacer que el trabajo sea más dinámico, permite el uso de diversas fuentes provenientes de distintos programas y áreas de salud, por lo que evita que los análisis se realicen repetidos o desactualizados.

CONCLUSIONES

Se organizó una Sala de Situación de Salud como una herramienta estratégica para la gestión. El Análisis del Situación de Salud de la Ciudad de El Carmen nos permite conocer la realidad de sus habitantes con sus características socio-culturales-económicas de la región, ambiente, historia presente y pasada, esto nos permitió efectuar un análisis de salud con otra mirada, rompiendo el paradigma hegemónico de salud donde la mirada se efectúa desde la enfermedad sin tener en cuenta al individuo y su historia personal y el contexto en el cual se desenvuelve. Se resolvieron una serie de obstáculos, especialmente referido a las dificultades en transformar el paradigma sobre la apropiación de los datos, hecho que repercute en superposiciones de trabajos, tiempos

perdidos, mayores costos e inconsistencias difíciles de resolver. Y se instauró una cultura informacional compartimentada del uso de los datos en las instituciones de salud. No se logró el desarrollo, selección y descripción de los indicadores a utilizar en la Sala de Situación debido a que desde el mes de Enero de 2020 el Área Programática se ve inmersa en un brote de Dengue, seguido por el brote de Coronavirus.

Esperamos haber contribuido a la mejora de la calidad del proceso de generación y transferencia del dato-información sanitaria y profundizar el análisis producido por la Sala de Situación, poniendo énfasis en nuevas técnicas analíticas para la priorización de problemas coyunturales y no coyunturales; teniendo como objetivo, desde el inicio, el establecer y mejorar los mecanismos de difusión de los análisis producidos por la Sala de Situación hacia los diferentes efectores en salud, los distintos niveles de gestión y otros actores interesados, que garanticen la dinámica del proceso y establecer la dinámica de redefinición permanente de la información necesaria con los distintos niveles de gestión.

BIBLIOGRAFIA

"Estadística para las Ciencias del Comportamiento". Robert R. Pagano, 2008, Ed. Internacional Thomson Editores.

"Estadística Para Las Ciencias Sociales", Ferris J. Ritchey, 2008, Ed. Mc. Graw Hill. México.

“Estadística Aplicada a través de Excel”, Perez C., 2007, Ed. Pearson Prentige Hall.

“Probabilidad y Estadística”. Boaglio L. et al., 2012, Editorial Científica Universitaria. Córdoba.

“Introducción a la Epidemiología”. Almeida Fhilo N. et al. 2008. Bs. As. Ed. Lugar.

"Epidemiología sin números", Naomar de Almeida Filho. Serie PALTEX. Nº 28. 2012. OPS

“Sistemas de Información Geográfica aplicados a la gestión del territorio”, Aliaga, Gastón. (2006). Juan Peña Llopis.. Revista de geografía Norte Grande, (36), 97-101. Recuperado en 28 de noviembre de 2015, de http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-34022006000.

“Vulnerabilidad y Riesgos: Acciones de salud que favorecen la toma de decisiones en la gestion de desastres”, Chalabe s.; Chalabe A.;. Jujuy. Argentina. Foro Internacional de Peligros Geológicos. INGEMMET. Arequipa. Perú. 14 al 16 de octubre de 2013.

Sistema Nacional de Vigilancia de la Salud – SNVS. 2010. Ministerio de Salud. Acceso Internet Octubre 2019, en https://www.snvs.msal.gov.ar/

“Experiencias Exitosas de Sala de Situación de Salud en Argentina” – Ministerio de Salud. 2013. Acceso Internet Octubre 2019 - http://www.msal.gob.ar/saladesituacion/Biblio/experiencias-exitosas-sala-situacion-argentina.pdf-

“Sala de Situación de Salud – Guía Metodológica” - Ministerio de Salud. 2013. Acceso Internet Octubre 2019 – http://www.msal.gob.ar/municipios/images/stories/4-recursos/pdf/2013-09_guia-metodologica-Sala-Situacion-2013.pdf

10 y 11 de Diciembre de 2020

CLASES VIRTUALES EN EL CONTEXTO DE LA PANDEMIA COVID-19 (EXPERIENCIAS DE ALUMNOS DE LA FACULTAD DE INGENIERIA DE LA UNIVERSIDAD NACIONAL DE SALTA)

Mg. Ing. Héctor Iván Rodríguez – Ing. Gisella Mautino
Facultad de Ingeniería
Universidad Nacional de Salta – Salta Argentina
ivan@ing.unsa.edu.ar - Cel +5493874129731

RESUMEN.
Ante la cuarentena obligatoria impuesta por el estado que derivó en la necesidad de impartir clases de manera virtual, el decano de la Facultad de Ingeniería, en la necesidad de conocer la opinión de los alumnos de la facultad, promueve realizar una encuesta a efectos de tomar decisiones orientadas a la ayuda de los alumnos. El objetivo de este trabajo es exponer la experiencia de la Facultad de Ingeniería que buscaba contar con información que permita conocer, no solo la opinión, sino también el volumen de respuesta, a efectos de tomar decisiones como así también evaluar el funcionamiento de la estructura comunicacional de la Institución con sus estudiantes. Se realizó la encuesta a todo alumnado de la Facultad de Ingeniería de la Universidad Nacional de Salta, sobre las principales necesidades y dificultades respecto a las clases virtuales. También se consultó sobre beneficios y grado de acuerdo con el dictado virtual.

Palabras Claves: Clases Virtuales; Pandemia, Covid19

INTRODUCCIÓN

El nivel de respuesta de los alumnos fue satisfactorio y permitió a la Facultad determinar sobre acciones a seguir, justificar programas de becas para ayudar con los problemas de conectividad de los alumnos. Orientar a los profesores sobre la problemática del estudiante, seleccionar y diseñar contenidos, en cursos y talleres de capacitación para el cuerpo docente, en lo referente a la mejora del dictado y toma de exámenes virtuales.

METODOLOGIA
Fecha de consulta: 14 al 17 de abril del 2020
Universo: Alumnos Inscriptos de la Facultad de Ingeniería de la Universidad Nacional de Salta en el año 2020.
Tipo de relevamiento: Censo, consulta realizada a través del Email, plataforma Moodle y el Facebook de la Facultad de Ingeniería de la UNSa, a toda la población estudiantil definida en el Universo.
Instrumento de recolección: Cuestionario estructurado para auto llenado.
Total de alumnos inscriptos en el 2020: 3812 alumnos.
Total de alumnos que se estima para el 2020: 3420 alumnos.
Cantidad de respuestas obtenidas: 1725.

COMPOSICION DE LA POBLACION ESTUDIANTIL RESPECTO A LA ENCUESTA

Consulta realizada a la población estudiantil de la Facultad de Ingeniería, inscripta en el primer cuatrimestre del año 2020. Se envió el cuestionario vía mail, plataforma Moodle y Facebook de la Facultad a través del departamento de cómputos, durante los días 14, 15, 16 y 17 de abril, obteniéndose 1725 respuestas de un total estimado de 3420 alumnos registrados. Este resultado constituye un nivel de respuesta del 50,4%, significa que los datos que refleja la encuesta representan a la mitad de la población estudiantil de la Facultad de Ingeniería que toma clases en el primer cuatrimestre.

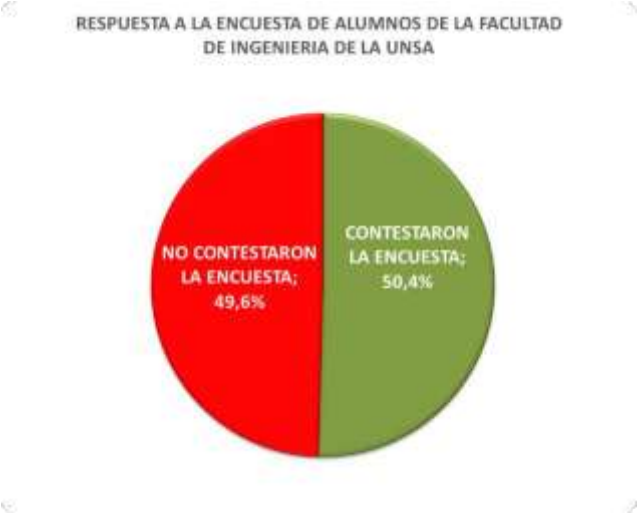


Gráfico 1

Es necesario aclarar que el 49,6% de la población que no respondió la encuesta, podría estar constituido por:

- Alumnos con voluntad de responder pero que no leyeron el mail o lo hicieron de manera tardía.
- Alumnos que leyeron el mail, pero decidieron no contestar.
- Alumnos a los que no les llegó el mail.

En los puntos que refieren a la falta de respuesta por razones que hacen a la estructura comunicativa, es donde se hace necesario trabajar para asegurar que la comunicación llegue en tiempo y forma.

COMPOSICION DE LA POBLACION QUE RESPONDIO LA ENCUESTA
A continuación, los datos reflejan la opinión del 50,4% de la población estudiantil de la Facultad de Ingeniería de la UNSa.

Dificultades para el cursado

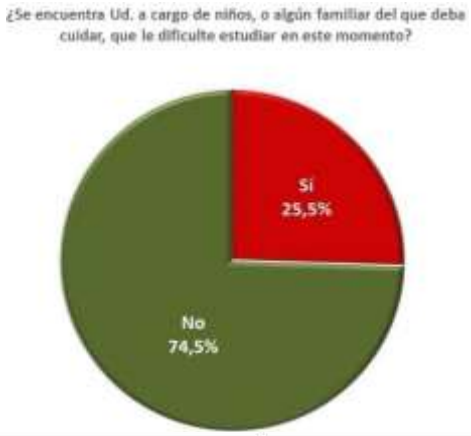
El 25,5% de los alumnos manifiesta dificultad para estudiar, argumentando tener alguna persona a cargo para cuidar al momento de la encuesta, esta cuantía se observa tanto para ingresantes como para no ingresantes.

Cuando se consulta sobre la disponibilidad económica para el acceso a las clases virtuales, el porcentaje de alumnos con dificultades aumenta al 31,3%, habiendo, ahora sí, diferencia entre los ingresantes, de los cuales el 40,6% manifiesta alguna imposibilidad económica para el acceso a clases virtuales. **En este punto es oportuno aclarar, que el porcentaje de alumnos con dificultades podría ser mayor, ya que la encuesta fue contestada por aquellos estudiantes que recibieron el cuestionario vía internet, quedando excluidos los que no tienen conectividad**.

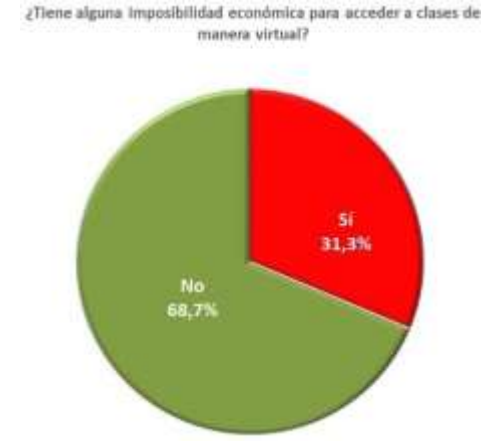
De ambos datos se puede cuantificar que el 54,4% de los alumnos no tiene ningún impedimento para tomar clases virtuales, 11,2% se encuentra con dificultades económicas y además debe asistir a un tercero; 20,1% solo con dificultades económicas para el acceso a clases virtuales y 14,3% sin inconvenientes económicos, pero con la restricción de tener que cuidar a un tercero.

Las dificultades económicas afectan más a ingresantes, y de fuera de la ciudad de Salta, que a los alumnos que ya están en carrera y viven en la Capital. La restricción de tener personas a cuidado, impuesta por la pandemia afecta a ingresantes y no ingresantes por igual.

(Ver cuadros y gráficos de la página siguiente)



		¿Es ingresante 2020?	
		Sí	No
¿Se encuentra Ud. a cargo de niños, o algún familiar del que deba cuidar, que le dificulte estudiar en este momento?	Sí	25,1%	25,6%
	No	74,9%	74,4%
	Total	100,0%	100,0%



		¿Es ingresante 2020?	
		Sí	No
¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	Sí	40,6%	26,8%
	No	59,4%	73,2%
	Total	100,0%	100,0%

Gráfico 2

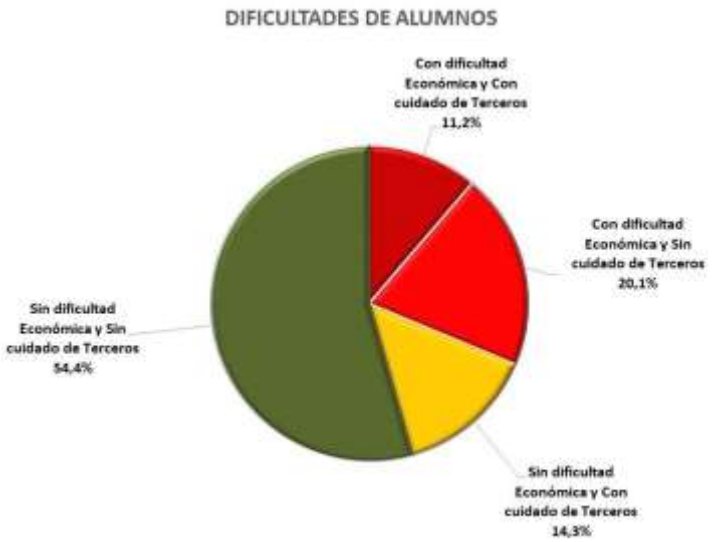


Gráfico 3

		¿Es ingresante 2020?	
		Sí	No
DIFICULTADES	Con dificultad Económica y Con cuidado de Terceros	**13,4%**	10,2%
	Con dificultad Económica y Sin cuidado de Terceros	**27,3%**	16,6%
	Sin dificultad Económica y Con cuidado de Terceros	11,8%	15,4%
	Sin dificultad Económica y Sin cuidado de Terceros	47,6%	57,8%
	Total	100,0%	100,0%

		¿Cuál es su lugar de residencia fuera de épocas de clases?				
		Ciudad de Salta	Municipios periféricos (cerrillos, La caldera, Quijano, San Luis)	Interior Provincia de Salta	Otra Provincia	En el extranjero
¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	Sí	24,1%	**41,7%**	**42,3%**	**40,3%**	**57,1%**
	No	75,9%	58,3%	57,7%	59,7%	42,9%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 1

SITUACION DEL ALUMNO ANTE LA CUARENTENA

El estado actual de situación, en la que están suspendidas las clases presenciales debido a la cuarentena obligatoria dispuesta por el Gobierno Nacional, ha motorizado a docentes y autoridades de la Facultad de Ingeniería que se abocaron al dictado virtual de clases y actividades relacionadas. Ante esta situación se observan tres tipos de conducta, a saber:

- **82,1% de los alumnos declaran que se encuentran cursando y pendientes** de las indicaciones por mail o por la plataforma de la Facultad para saber que deben hacer y estudiar. Este porcentaje aumenta al 86,5% entre no ingresantes.

- **16,3% decidieron dejar de cursar hasta que se normalice y pase la cuarentena**. Este porcentaje es mayor entre ingresantes (23%). Puede decirse que la gran mayoría de los que decidieron dejar de cursar declararon inconvenientes económicos y/o por cuidado de terceros (71,9%). Solo el 28,1% de este grupo decidieron dejar de cursar por otras razones.

- **1,6% no indican estar cursando, pero si pendientes de instrucciones por parte de la Facultad**. Este último grupo está constituido mayoritariamente por ingresantes y una pequeña parte que podrían ser alumnos que terminaron de cursar la carrera y quedaron pendientes de exámenes o trámites.

		¿Es ingresante 2020?	
		SI	No
SITUACION DEL ALUMNO	Cursando y pendiente de indicaciones	73,1%	86,5%
	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	23,0%	13,0%
	No Cursan y estan pendientes de indicaciones.	3,9%	0,5%
	Total	100,0%	100,0%

Tabla 2

Gráfico 4

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
DIFICULTADES	Con dificultad Económica y Con cuidado de Terceros	8,5%	**25,6%**	3,6%
	Con dificultad Económica y Sin cuidado de Terceros	17,7%	**31,0%**	32,1%
	Sin dificultad Económica y Con cuidado de Terceros	14,1%	**15,3%**	10,7%
	Sin dificultad Económica y Sin cuidado de Terceros	59,7%	**28,1%**	53,6%
	Total	100,0%	100,0%	100,0%

Tabla 3

CANTIDAD DE MATERIAS QUE LOS ALUMNOS AFIRMAN CURSAR
Observaciones: se registran valores atípicos debido a que hay alumnos que contestaron sobre el total de materias realizadas a la fecha en lugar de materias cursando a la fecha.

El promedio de la cantidad de materias que el alumno cursa está en **tres,** con una desviación estándar que indica que el 85,7% de los alumnos se encuentra cursando una cantidad entre 2 y 4 materias, dichas materias pueden ser de un mismo año o no. La distribución de la cantidad de materias que se cursan presenta una importante simetría, siendo en promedio, y coincidente con el máximo valor observado.
Si analizamos la distribución dentro de cada año se observa que los promedios varían respecto a la media general, siendo menores entre las materias de primero a tercer año, y mayores para cuarto y quinto año.

Gráfico 5

Media	3,0563
Mediana	3,
Moda	
Desviación estándar	
Asimetría	
Error estánda	
Curtosis	
Err	

Tabla 4

		Cusando Materias de				
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	Quinto Año
Media		**2,71**	**2,79**	**2,72**	**3,27**	**3,61**
Mediana		3,00	2,00	2,00	3,00	3,00
Moda		3	1	1	3	3
Desviación estándar		1,310	1,909	1,968	2,168	2,202
Asimetría		-0,664	1,240	1,096	0,369	0,835
Error estándar de asimetría		,076	,098	,105	,133	,153
Curtosis		2,088	1,587	1,755	,347	-,239
Error estándar de curtosis		,153	,196	,210	,266	,305
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	Quinto Año
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
CANTIDAD DE MATERIAS CURSANDO	1	18,3%	28,7%	30,5%	21,6%	17,4%
	2	20,9%	24,1%	29,2%	22,8%	15,0%
	3	50,0%	25,0%	19,1%	24,6%	30,0%
	4	2,0%	9,1%	6,5%	9,9%	13,0%
	5	1,1%	1,8%	3,2%	2,7%	1,2%
	6	6,9%	4,5%	4,5%	8,7%	9,9%
	7	,3%	,2%	,7%	,6%	3,2%
	8	,5%	6,3%	5,0%	6,9%	8,3%
	9		0,3%	1,3%	2,4%	2,0%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 5

Analizando los coeficientes de asimetría podemos ver los siguientes comportamientos:

- **Materias de Primer año**: La conducta es elegir hasta tres materias para cursar, con un valor modal de 3 materias en el 50% de los casos y un promedio de 2,71. (simetría negativa).

- **Materias de segundo y tercer año**: Mayores promedios en una materia y de ahí decrece levemente hasta 3, para caer significativamente de 4 en adelante. (simetría positiva).

- **Materias de cuarto y quinto año**: Si bien se mantiene que la mayoría está entre 1 y 3, pero aparece la conducta de anotarse en más de tres materias. (simetría positiva leve).

Grafico 6

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
CURSANDO	Primer Año	59,9%	63,0%	0,0%
	Segundo Año	37,4%	31,0%	0,0%
	Tercer Año	33,0%	25,3%	0,0%
	Cuarto Año	20,8%	13,9%	0,0%
	Quinto Año	15,1%	13,9%	0,0%
	No Indica	0,0%	11,7%	100,0%
	Total	100,0%	100,0%	100,0%

Tabla 6

		SIN CURSAR MATERIAS DE AÑOS ANTERIORES				
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	TOTAL
CURSANDO	Segundo Año	14,6%				35,8%
	Tercer Año	20,1%	12,9%			31,2%
	Cuarto Año	11,8%	10,7%	5,9%		19,4%
	Quinto Año	7,5%	7,3%	6,2%	5,2%	14,7%

		SIN CURSAR MATERIAS DE AÑOS ANTERIORES				
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	
CURSANDO	Segundo Año	40,8%				100,0%
	Tercer Año	64,3%	41,3%			100,0%
	Cuarto Año	60,8%	55,1%	30,2%		100,0%
	Quinto Año	51,0%	49,8%	42,3%	35,6%	100,0%

Tabla 7

Las tablas deben leerse por filas. El 14,6% de los alumnos que cursan segundo año declaran no estar cursando primer año, podría deberse a que no deben materias de segundo, este porcentaje representa el 40,8% sobre el total de alumnos cursando materias de segundo (ver segunda tabla).
En la segunda tabla se puede ver que de los que cursan materias de tercer año, 64,3% no cursan materias de primero y 41,3% no cursa materias de segundo año. Es llamativo los datos de cuarto y quinto año. Obsérvese que del 100% de alumnos cursando materias de quinto, la mitad afirma no estar cursando materias de primer año, cuando se esperaría que ese valor sea del 100% en la segunda tabla y de 14,7% en la primera tabla.
OBSERVACIONES: en la segunda tabla los valores del 100% en el margen derecho hacen referencia que los datos son sobre el total de la categoría por fila, la suma en la fila supera el 100% porque los alumnos hacen más de una materia.

Gráfico 7

Puede observarse que 59,4% son alumnos cursando materias de primer año, donde el 37,5% cursan solo materias de primero, y 21,9% cusan materias de Primero junto con materias de otros años.

		CURSAN MATERIAS DE					
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	Quinto Año	
CURSANDO	Primer Año	37,5%	21,2%	11,1%	7,6%	7,2%	59,4%
	Segundo Año	21,2%	6,6%	18,3%	8,7%	7,4%	35,8%
	Tercer Año	11,1%	18,3%	6,7%	13,5%	8,5%	31,2%
	Cuarto Año	7,6%	8,7%	13,5%	4,2%	9,4%	19,4%
	Quinto Año	7,2%	7,4%	8,5%	9,4%	4,6%	14,7%

		CURSAN MATERIAS DE					
		Primer Año	Segundo Año	Tercer Año	Cuarto Año	Quinto Año	
CURSANDO	Primer Año	63,1%	35,6%	18,7%	12,8%	12,1%	100,0%
	Segundo Año	59,2%	18,3%	51,2%	24,3%	20,6%	100,0%
	Tercer Año	35,7%	58,7%	21,4%	43,3%	27,1%	100,0%
	Cuarto Año	39,2%	44,9%	69,8%	21,6%	48,8%	100,0%
	Quinto Año	49,0%	50,2%	57,7%	64,4%	31,2%	100,0%

Tabla 8

OPINION RESPECTO A LAS CLASES VIRTUALES

Cantidad de materias que dictan clases virtuales: Se consulta al segmento de estudiantes que afirmaron estar cursando y pendientes de recibir instrucciones y novedades por parte de la Facultad y sus profesores. Este segmento constituye el 82,1% de la población estudiantil de la Facultad de ingeniería de la UNSa.

La asimetría negativa indica, como se puede ver en el gráfico, que la mayoría afirma estar recibiendo a lo sumo 3 clases virtuales, cuyo promedio está en 2,45 clases, perfil este similar en las cuatro Ingenierías, salvo en la TUTA donde se observa un 33% de alumnos que afirman no estar recibiendo ninguna clase virtual.

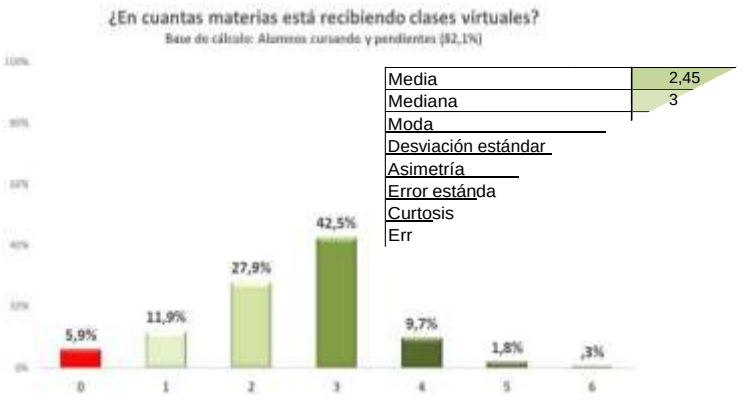


Gráfico 8

		Alumnos de la Carrera				
		Ing. Civil	Ing. Electromecánica	Ing. Industrial	Ing. Química	TUTA (Total)
¿En cuantas materias está recibiendo clases virtuales?	0	4,8%	8,4%	5,0%	4,8%	**33,3%**
	1	12,6%	11,2%	9,7%	14,0%	16,7%
	2	31,9%	23,2%	25,8%	29,6%	5,6%
	3	38,1%	42,1%	42,8%	48,4%	44,4%
	4	11,2%	12,3%	12,9%	2,9%	0,0%
	5	1,1%	2,8%	2,9%	,3%	0,0%
	6	,2%	0,0%	,9%	0,0%	0,0%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%
PROMEDIO		**2,43**	**2,47**	**2,61**	**2,32**	**1,61**

Base de cálculo: Alumnos cursando y pendientes (82,1%)

Tabla 9

		TIPOLOGIA DE ALUMNOS					
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años
¿En cuantas materias está recibiendo clases virtuales?	0	9,0%	3,1%	1,0%	1,4%	1,5%	5,6%
	1	9,2%	22,4%	30,4%	13,0%	19,1%	8,1%
	2	26,9%	39,8%	34,3%	27,5%	23,5%	26,2%
	3	54,3%	25,5%	22,5%	34,8%	39,7%	39,4%
	4	,4%	8,2%	11,8%	23,2%	16,2%	15,8%
	5	0,0%	1,0%	0,0%	0,0%	0,0%	4,5%
	6	,2%	0,0%	0,0%	0,0%	0,0%	,5%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
PROMEDIO		**2,29**	**2,16**	**2,14**	**2,65**	**2,50**	**2,67**

Base de cálculo: Alumnos cursando y pendientes (82,1%)

Tabla 10

MEDIOS TECNOLOGICOS PARA LA TOMA DE CLASES VIRTUALES

Medios usados por los docentes

Los alumnos que afirman estar cursando y pendientes de recibir clases e instrucciones por parte de sus profesores y la Facultad, afirman estar recibiendo clases virtuales principalmente por la plataforma Moodle, le siguen las video conferencias tipo Zoom o similares, y en tercer lugar el WhatsAapp. Estas son las tres herramientas más usadas por todas las carreras de la Facultad, donde se puede observar que Ingeniería Civil usa más que el resto la modalidad de clases virtuales, y TUTA menos que el resto dicha modalidad. El uso de Mail y otros medios también forman parte de las herramientas de comunicación con los alumnos, pero en menor medida. Si bien los alumnos que cursan materias de primer año son los que menos afirman recibir clases por video conferencias, es de destacar que el 17,1% respondieron si recibirlas, lo que constituye una cifra importante dado la particularidad del primer año, donde hay ingresantes y mayor cantidad de alumnos.

Gráfico 9

		TIPOLOGIA DE ALUMNOS					
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Por cuáles de los siguientes medios recibe Ud. las clases virtuales? (puede señalar más de una opción)	Clases por Zoom o similar de video conferencias.	17,1%	61,1%	82,2%	86,8%	88,1%	66,8%
	Tareas en Moodle.	96,4%	95,8%	94,1%	91,2%	82,1%	95,8%
	Tareas por mail.	11,8%	14,7%	18,8%	30,9%	46,3%	18,4%
	Whatsapp.	22,2%	53,7%	22,8%	52,9%	50,7%	41,2%
	Otros medios.	0,6%	6,3%	2,0%	2,9%	9,0%	4,4%

Base de cálculo: Opinan alumnos que están cursando y pendientes (82,1%)

Tabla 11

Medios disponibles por los alumnos

El celular se constituye en la herramienta masiva que disponen los alumnos para recibir clases virtuales, 89,1% afirman disponer de los mismos. Este porcentaje se mantiene tanto para alumnos con dificultades económicas para recibir clases, como para quienes decidieron dejar de cursar hasta que se normalice la situación de pandemia.

El principal inconveniente es la disposición de señal de internet, 48,9% afirma tener acceso limitado, a los que se les suma un 6% sin acceso.

El 99% dispone por lo menos de algún dispositivo para recibir clases virtuales, este valor tan elevado puede deberse al hecho que la encuesta fue realizada a través de mail y la plataforma Moodle de la Facultad. Por lo tanto, esta pregunta no debe extrapolarse al total de la población, sino solo al segmento de alumnos que reciben comunicación por internet desde la Facultad. De estos alumnos, 35,8% poseen solo teléfono celular para recibir clases virtuales, 53,3% poseen además de celular algún otro dispositivo (Notebook, Netbook, Pc de escritorio, Tablet), y 9,9% no posee celular, pero si algún otro dispositivo para tomar clases virtuales.

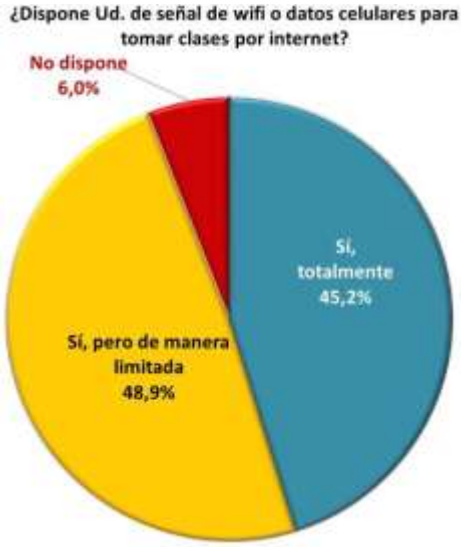


Gráfico 10

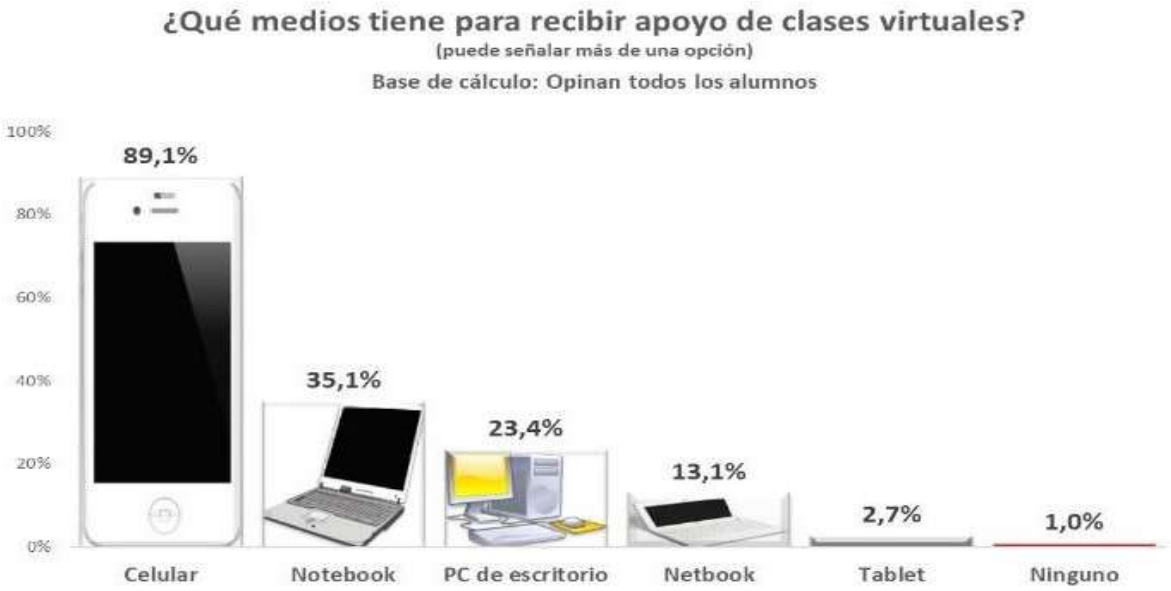


Gráfico 11

Las cifras más elevadas de alumnos con restricciones para el acceso a clases virtuales, (*con dificultades de acceso a señal wifi, que cuentan únicamente con un celular, o no disponen de ningún dispositivo*), se dan entre los que decidieron dejar de cursar; los que afirman tener dificultades económicas; y entre los ingresantes. Estos tres grupos se detallan a continuación:

1. Alumnos que decidieron dejar de cursar (16,3% del total)
    - 59,8 tienen solo celular para tomar clases virtuales.
    - 4,6% no disponen de ningún dispositivo para tomar clases virtuales.
    - 75,8% con wifi restringido o nulo.
2. Alumnos con dificultades económicas (31,3% del total)
    - 66,9 tienen solo celular para tomar clases virtuales.
    - 3,3% no disponen de ningún dispositivo para tomar clases virtuales.
    - 91,3% con wifi restringido o nulo.
3. Ingresantes (32,6% del total)
    - 47,4 tienen solo celular para tomar clases virtuales.
    - 1,6% no disponen de ningún dispositivo para tomar clases virtuales.
    - 62,7% con wifi restringido o nulo.

Puede observarse también la existencia de correlación entre la cantidad disponible de dispositivos con que cuenta el alumno y el grado de avance en la carrera. A medida que el alumno es más avanzado, disminuye el porcentaje de los que tienen solo celular, aumentando así la disponibilidad de múltiples recursos. (Ver segundo gráfico y primera tabla).

Gráfico 12

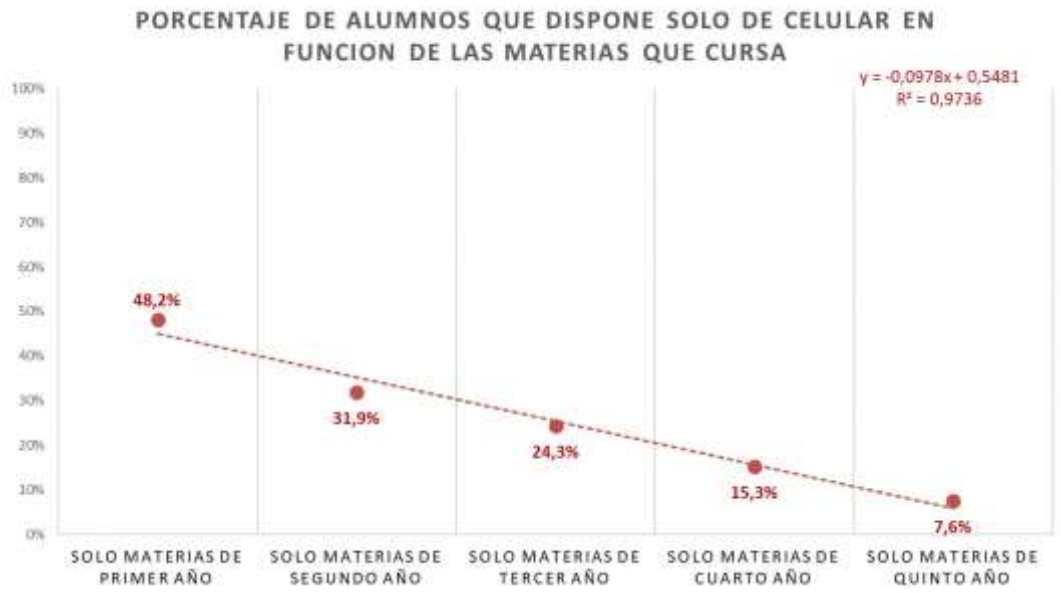


Gráfico 13

Tabla 12

		¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	
		Sí	No
		% del N de columna	% del N de columna
¿Qué medios tiene para recibir apoyo de clases virtuales?	Solo Celular	**66,9%**	21,6%
	Celular y Computadora, tablet	23,1%	67,1%
	Computadora y/o tablet	6,7%	11,3%
	Ninguno	**3,3%**	0,0%
	Total	100,0%	100,0%

Tabla 13

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
		% del N de columna	% del N de columna	% del N de columna
¿Qué medios tiene para recibir apoyo de clases virtuales?	Solo Celular	31,2%	**59,8%**	25,0%
	Celular y Computadora, tablet	58,1%	28,5%	64,3%
	Computadora y/o tablet	10,5%	7,1%	7,1%
	Ninguno	0,3%	**4,6%**	3,6%
	Total	100,0%	100,0%	100,0%

Tabla 14

		¿Es ingresante 2020?	
		Sí	No
		% del N de columna	% del N de columna
¿Qué medios tiene para recibir apoyo de clases virtuales?	Solo Celular	47,4%	30,0%
	Celular y Computadora, tablet	41,7%	59,0%
	Computadora y/o tablet	9,3%	10,2%
	Ninguno	1,6%	0,8%
	Total	100,0%	100,0%

Tabla 15

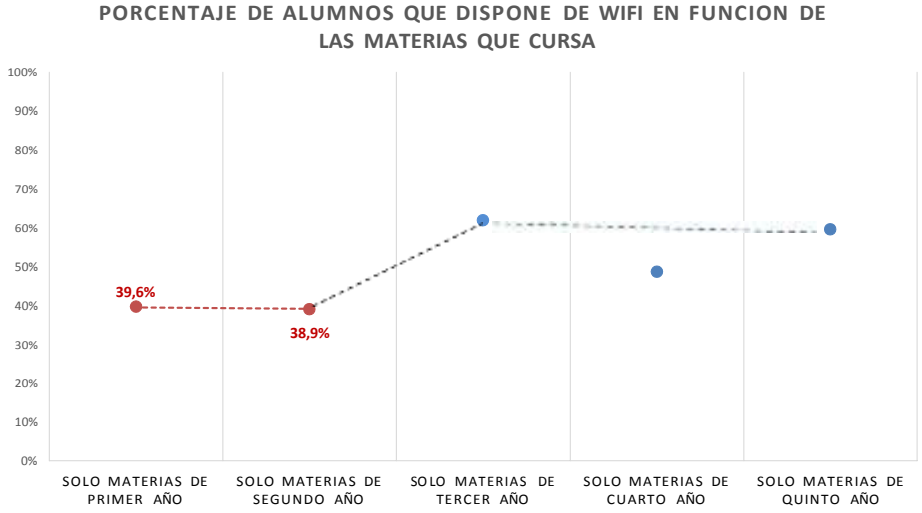


Gráfico 14

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Dispone Ud. de señal de wifi o datos celulares para tomar clases por internet?	Sí, totalmente	39,6%	38,9%	61,7%	48,6%	59,5%	47,3%	39,3%
	Sí, pero de manera limitada	52,6%	54,0%	37,4%	51,4%	39,2%	48,1%	39,3%
	No dispone	7,9%	7,1%	0,9%	0,0%	1,3%	4,5%	21,3%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 16

		¿Es ingresante 2020?	
		Sí	No
		% del N de columna	% del N de columna
¿Dispone Ud. de señal de wifi o datos celulares para tomar clases por internet?	Sí, totalmente	37,3%	49,0%
	Sí, pero de manera limitada	52,2%	47,3%
	No dispone	10,5%	3,7%
	Total	100,0%	100,0%

Tabla 17

		¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	
		Sí	No
		% del N de columna	% del N de columna
¿Dispone Ud. de señal de wifi o datos celulares para tomar clases por internet?	Sí, totalmente	8,7%	61,8%
	Sí, pero de manera limitada	74,4%	37,2%
	No dispone	16,9%	1,0%
	Total	100,0%	100,0%

Tabla 18

MEDIOS DE COMUNICACIÓN VIRTUAL USADOS POR LOS ALUMNOS

El WhatsApp es el medio por excelencia que usan los alumnos para comunicarse, solamente el 19,7% usa mail, y el resto de las aplicaciones en el orden del 10%. Nótese que el uso de mail cobra importancia a medida que el alumno es más avanzado.

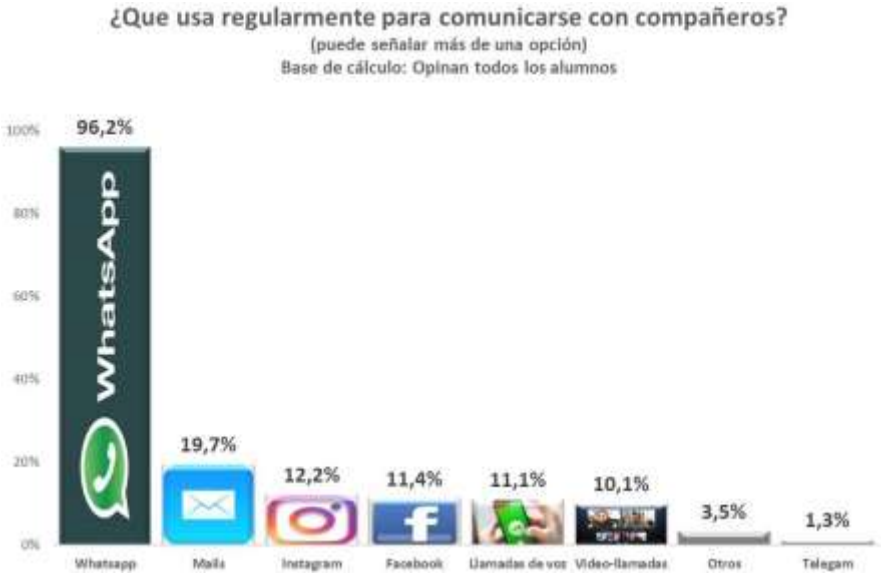


Gráfico 15

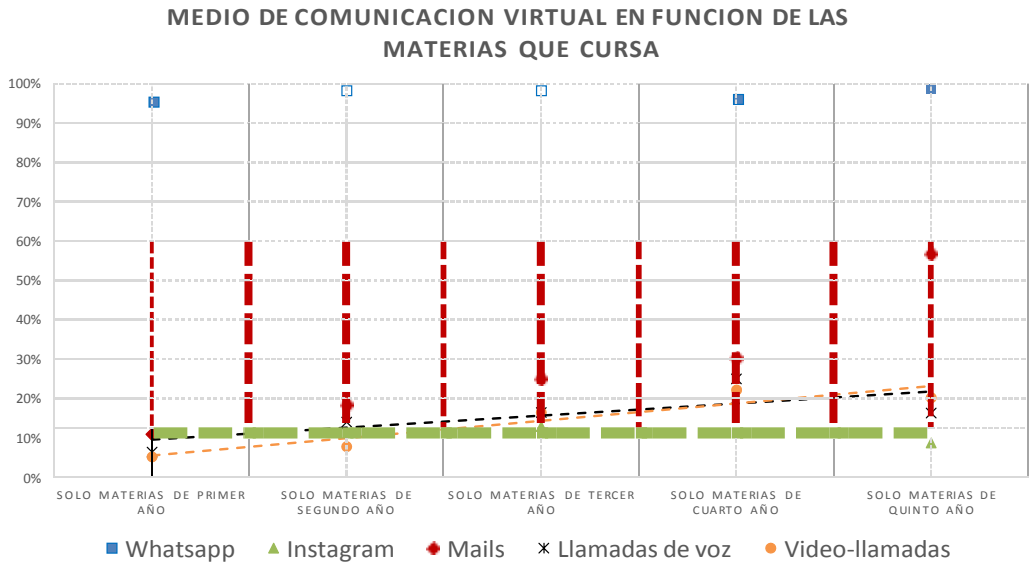


Gráfico 16

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Que usa regularmente para comunicarse con compañeros? (puede señalar más de una opción)	Whatsapp	94,7%	98,2%	98,3%	95,8%	98,7%	98,0%	82,0%
	Facebook	11,3%	8,8%	8,7%	12,5%	17,7%	11,0%	18,0%
	Instagram	10,7%	12,4%	13,0%	11,1%	8,9%	14,1%	13,1%
	Telegam	1,2%	1,8%	3,5%	2,8%	1,3%	0,9%	0,0%
	Mails	11,0%	18,6%	25,2%	30,6%	57,0%	22,1%	16,4%
	Llamadas de voz	6,5%	14,2%	16,5%	25,0%	16,5%	11,9%	11,5%
	Video-llamadas	5,3%	8,0%	16,5%	22,2%	20,3%	11,9%	8,2%
	Otros	3,1%	1,8%	6,1%	2,8%	2,5%	3,6%	6,6%

Base de cálculo: Opinan todos los alumnos

Tabla 19

NECESIDADES DE CAPACITACION PARA LA TOMA DE CLASES VIRTUALES

En promedio 51% de los alumnos afirman necesitar capacitación para el acceso a toma de clases virtuales, este valor es mayor entre alumnos que cursan los primeros años. El 63,5% de alumnos cursando materias de primer año afirman necesitar algún tipo de capacitación. La necesidad de capacitación decrece a medida que los alumnos cursan materias de años superiores, los alumnos cursando materias de quinto año que necesitan alguna capacitación es del 22,1%.

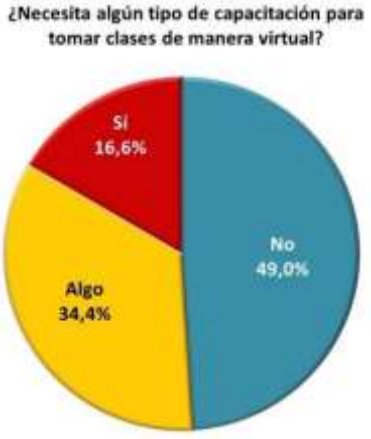


Gráfico 17

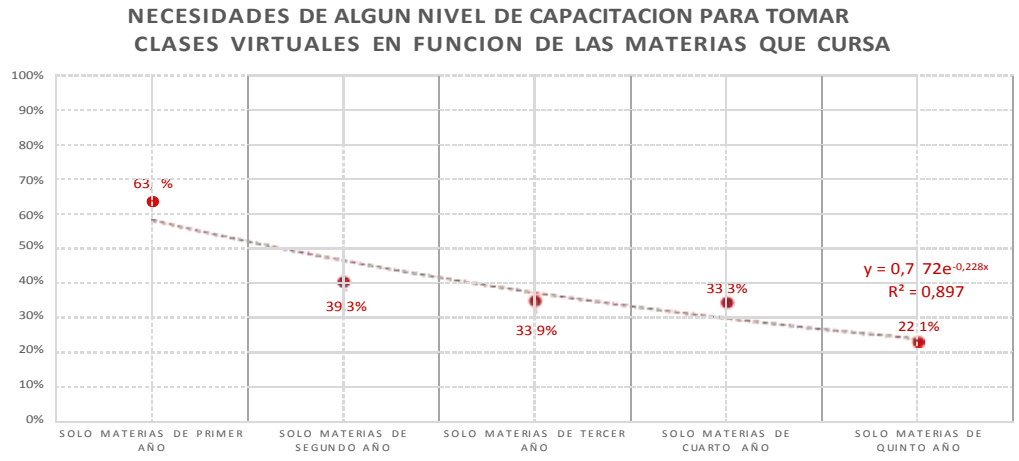


Gráfico 18

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Necesita algún tipo de capacitación para tomar clases de manera virtual?	Sí	24,0%	5,4%	4,3%	8,3%	5,2%	14,0%	34,4%
	Algo	39,6%	33,9%	29,6%	25,0%	16,9%	33,6%	31,1%
	No	36,5%	60,7%	66,1%	66,7%	77,9%	52,4%	34,4%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 20

No hay diferencias significativas por carrera, si las hay por ingresantes; entre los que tiene restricciones económicas y los que decidieron no cursar hasta que pase la cuarentena.

		Alumnos de la Carrera				
		Ing. Civil	Ing. Electromecánica	Ing. Industrial	Ing. Química	TUTA (Total)
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Necesita algún tipo de capacitación para tomar clases de manera virtual?	Sí	16,3%	16,2%	15,2%	19,5%	19,0%
	Algo	35,3%	36,3%	33,8%	32,2%	28,6%
	No	48,4%	47,5%	51,0%	48,4%	52,4%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 21

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
		% del N de columna	% del N de columna	% del N de columna
¿Necesita algún tipo de capacitación para tomar clases de manera virtual?	Sí	13,0%	33,2%	32,1%
	Algo	34,4%	34,6%	32,1%
	No	52,6%	32,1%	35,7%
	Total	100,0%	100,0%	100,0%

Tabla 22

		¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	
		Sí	No
		% del N de columna	% del N de columna
¿Necesita algún tipo de capacitación para tomar clases de manera virtual?	Sí	30,6%	10,2%
	Algo	38,2%	32,7%
	No	31,2%	57,1%
	Total	100,0%	100,0%

Tabla 23

		¿Es ingresante 2020?	
		Sí	No
		% del N de columna	% del N de columna
¿Necesita algún tipo de capacitación para tomar clases de manera virtual?	Sí	27,3%	11,4%
	Algo	39,0%	32,1%
	No	33,7%	56,5%
	Total	100,0%	100,0%

Tabla 24

TIPO DE CAPACITACION REQUERIDA POR ALUMNOS

El segmento del 51% de alumnos que requiere capacitación para la toma de clases virtuales especifica necesitar aprender como tomar clases usando programas de video conferencias, y manejo de la plataforma Moodle para, subir archivos, leer lo que los profesores mandan, ver videos y hacer consultas. No se observan diferencias significativas por carrera ni por situación económica y particular del alumno.

Gráfico 19

Gráfico 20

Gráfico 21

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Qué tipo de capacitación necesita para tomar clases de manera virtual? (puede señalar más de una opción)	Moodle para subir archivos, leer lo que los profesores mandan, ver videos y hacer consultas a los profesores.	64,5%	34,9%	47,4%	25,0%	38,9%	48,5%	72,5%
	Como usar WhatsApp, u otras redes sociales para comunicarse con los profesores.	10,6%	7,0%	0,0%	0,0%	16,7%	8,3%	15,0%
	Como tomar clases usando programas como por ejemplo el programa Zoom o Jitsi meet.	54,7%	69,8%	57,9%	83,3%	38,9%	64,5%	60,0%
	Otra	18,2%	25,6%	31,6%	12,5%	44,4%	20,6%	15,0%

Base de cálculo: sobre el 51% de alumnos que requieren capacitación

Tabla 25

VALORACION DEL DICTADO DE CLASES VIRTUALES

Comparando las valoraciones dadas por alumnos sobre el dictado virtual, y la opinión de los docentes, en la encuesta que se les realizó durante el 7 y 8 de abril del 2020, sobre el grado de efectividad que creen se puede lograr con las clases virtuales, puede observarse que el cuerpo docente tiene mejores expectativas que la actual apreciación del alumnado. **Hay que señalar que la valoración dada por los alumnos cursando, es mejor que la de los que decidieron no cursar hasta que finalice la pandemia**. **También buena parte de la valoración negativa se debe a la falta de disponibilidad de internet.**

Mejora Significativamente la valoración positiva dada por los alumnos de cuarto y quinto año y disminuye en los alumnos de los primeros años. De todas maneras, prima la calificación “Regular” dada por los alumnos, cuestión que es normal debido que el dictado virtual es un proceso iniciado súbitamente y se encuentra en evolución, además altos valores de una calificación “regular” suele ser indicador de desconcierto o falta de información.

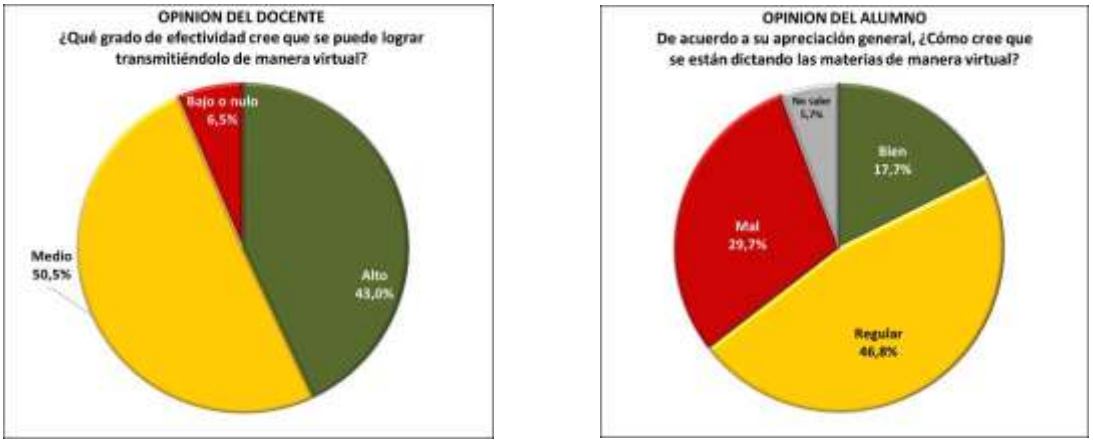


Gráfico 22

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	17,0%	17,1%	13,9%	37,5%	44,3%	14,3%	11,5%
	Regular	47,0%	47,7%	52,2%	43,1%	35,4%	49,1%	29,5%
	Mal	30,1%	33,3%	28,7%	16,7%	16,5%	32,4%	24,6%
	No sabe	5,9%	1,8%	5,2%	2,8%	3,8%	4,2%	34,4%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 26

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
		% del N de columna	% del N de columna	% del N de columna
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	20,1%	5,0%	25,0%
	Regular	50,0%	31,8%	35,7%
	Mal	27,5%	41,1%	28,6%
	No sabe	2,4%	22,1%	10,7%
	Total	100,0%	100,0%	100,0%

Tabla 27

		¿Dispone Ud. de señal de wifi o datos celulares para tomar clases por internet?		
		Sí, totalmente	Sí, pero de manera limitada	No dispone
		% del N de columna	% del N de columna	% del N de columna
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	25,3%	11,9%	7,8%
	Regular	51,7%	44,9%	25,5%
	Mal	18,7%	37,2%	52,0%
	No sabe	4,2%	6,1%	14,7%
	Total	100,0%	100,0%	100,0%

Tabla 28

		¿Cuánto diría Ud. que está participando en las clases virtuales?		
		Mucho	Poco	Nada
		% del N de columna	% del N de columna	% del N de columna
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	25,5%	13,3%	7,5%
	Regular	45,1%	52,9%	26,7%
	Mal	29,0%	31,0%	26,7%
	No sabe	0,4%	2,8%	39,0%
	Total	100,0%	100,0%	100,0%

Tabla 29

		¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	
		Sí	No
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	11,1%	20,7%
	Regular	40,8%	49,6%
	Mal	38,4%	25,8%
	No sabe	9,6%	4,0%
	Total	100,0%	100,0%

Tabla 30

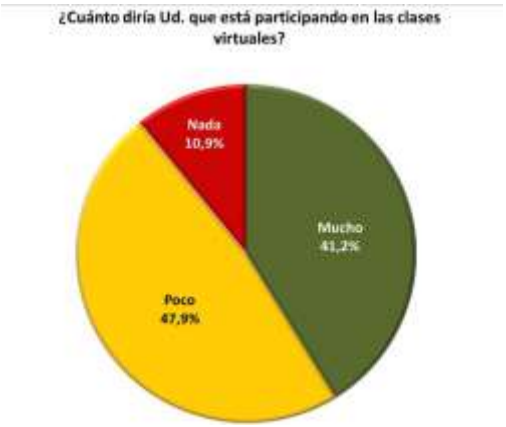
		¿Es ingresante 2020?	
		Sí	No
		% del N de columna	% del N de columna
De acuerdo a su apreciación general, ¿Cómo cree que se están dictando las materias de manera virtual?	Bien	14,4%	19,2%
	Regular	45,6%	47,5%
	Mal	33,5%	27,8%
	No sabe	6,4%	5,4%
	Total	100,0%	100,0%

Tabla 31

PARTICIPACION DE ALUMNOS EN CLASES VIRTUALES

Se observa una importante coincidencia entre la respuesta de los docentes y la de los alumnos respecto a la participación en clases virtuales. El 89,1% de los alumnos responden estar participando, de los cuales prácticamente la mitad afirman estarlo haciendo de manera significativa.

Opinión de Alumnos



Opinión de Docentes

Gráfico 23

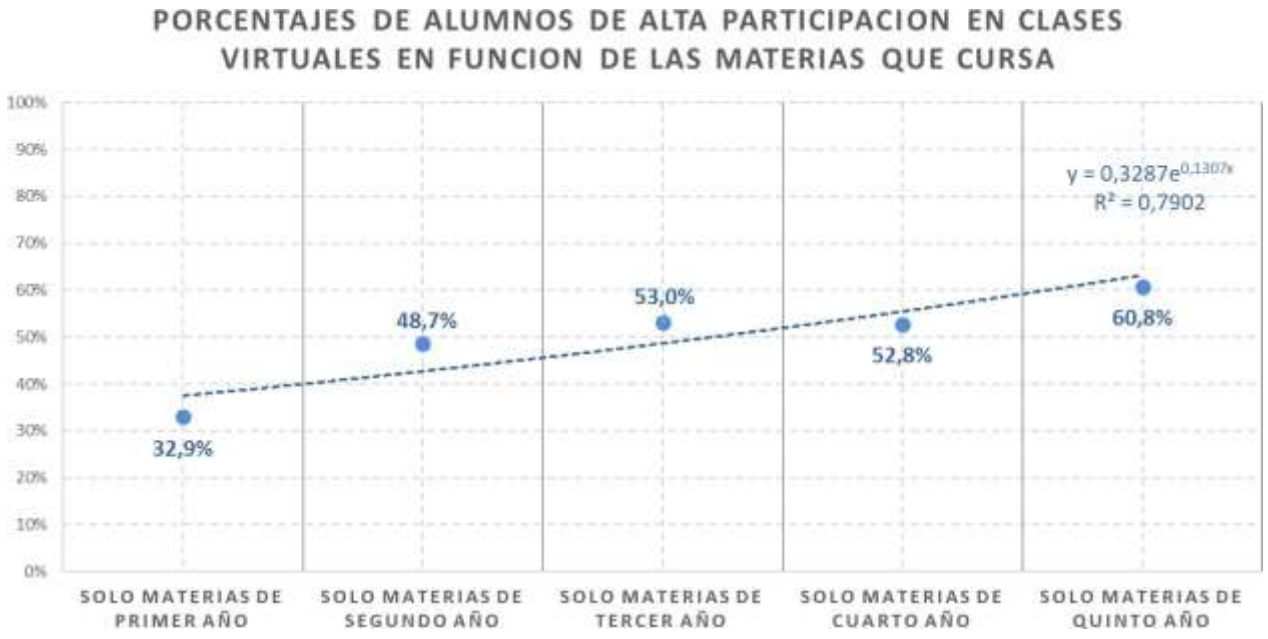


Gráfico 24

		TIPOLOGIA DE ALUMNOS						
		Solo materias de Primer año	Solo materias de Segundo año	Solo materias de Tercer año	Solo materias de Cuarto año	Solo materias de Quinto año	Mezcal de materias de distintos años	No Contesta
		% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna	% del N de columna
¿Cuánto diría Ud. que está participando en las clases virtuales?	Mucho	32,9%	48,7%	53,0%	52,8%	60,8%	45,1%	11,5%
	Poco	56,1%	43,4%	40,0%	40,3%	26,6%	45,8%	44,3%
	Nada	11,0%	8,0%	7,0%	6,9%	12,7%	9,1%	44,3%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla 32

		¿Tiene alguna imposibilidad económica para acceder a clases de manera virtual?	
		Sí	No
		% del N de columna	% del N de columna
¿Cuánto diría Ud. que está participando en las clases virtuales?	Mucho	28,1%	47,1%
	Poco	55,7%	44,4%
	Nada	16,1%	8,5%
	Total	100,0%	100,0%

Tabla 33

		¿Dispone Ud. de señal de wifi o datos celulares para tomar clases por internet?		
		Sí, totalmente	Sí, pero de manera limitada	No dispone
		% del N de columna	% del N de columna	% del N de columna
¿Cuánto diría Ud. que está participando en las clases virtuales?	Mucho	49,6%	36,2%	18,4%
	Poco	42,2%	53,1%	48,5%
	Nada	8,2%	10,7%	33,0%
	Total	100,0%	100,0%	100,0%

Tabla 34

		SITUACION DEL ALUMNO		
		Cursando y pendiente de indicaciones.	Decidieron dejar de cursar hasta que se normalice y finalice la cuarentena	No Cursan y estan pendientes de indicaciones.
		% del N de columna	% del N de columna	% del N de columna
¿Cuánto diría Ud. que está participando en las clases virtuales?	Mucho	47,9%	9,3%	21,4%
	Poco	45,9%	57,3%	57,1%
	Nada	6,2%	33,5%	21,4%
	Total	100,0%	100,0%	100,0%

Tabla 35

COMPARATIVO DE CLASES PRESENCIALES VS VIRTUALES

Tanto docentes como alumnos, opinan en mayoría, que las clases presenciales son mejores a las virtuales. La opinión de los alumnos es compartida en todas las distintas segmentaciones que se analizaron en el presente estudio.

Gráfico 25

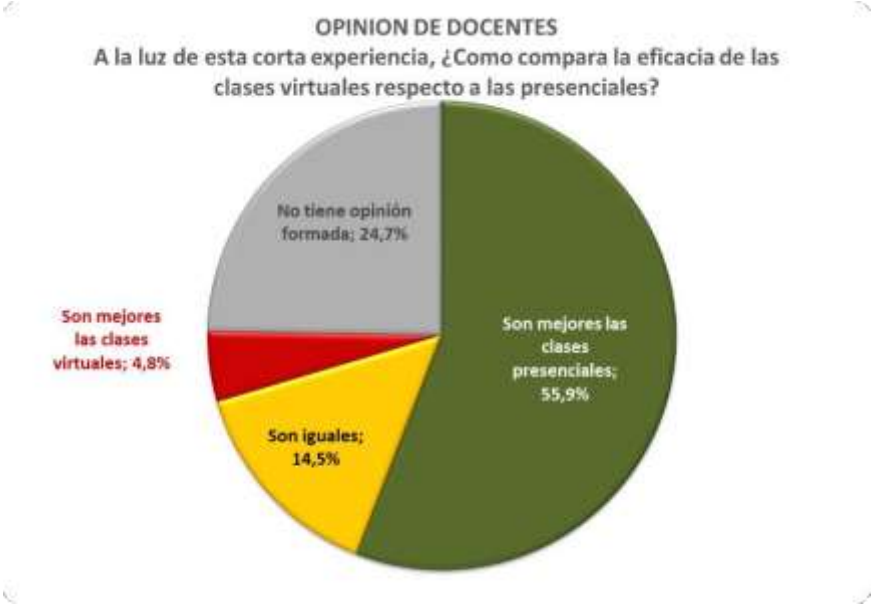


Gráfico 26

PRINCIPALES CONCLUSIONES

Los alumnos de la Facultad de ingeniería se constituyen por 32,6% de ingresantes, 40,9% del interior o de otras provincias o extranjeros, siendo la carrera de Ingeniería Química la que más alumnos del Interior de la Provincia de Salta tiene en su distribución interna, mientras que Ingeniería Electromecánica tiene más de alumnos de otras provincias.

El 25,5% de los alumnos manifiesta dificultad para estudiar por tener alguna persona a cargo para cuidar.

31,3% los alumnos manifiestan dificultad económica para recibir clases virtuales, siendo mayor este porcentaje entre alumnos ingresantes (40,6%).

En resumen 54,4% de los alumnos no tienen ningún impedimento para tomar clases virtuales, 20,1% solo con dificultades económicas para el acceso a clases virtuales; 11,2% se encuentran con dificultades económicas y además debe asistir a un tercero; y 14,3% sin inconvenientes económicos, pero con la restricción de tener que cuidar a un tercero.

Las dificultades económicas afectan más a ingresantes, y de fuera de la ciudad de Salta, que a los alumnos que ya están en carrera y viven en la Capital. La restricción de tener personas a cuidado, impuesta por la pandemia afecta a ingresantes y no ingresantes por igual.

82,1% de los alumnos declaran que se encuentran participando de clases virtuales y están pendientes de las indicaciones por mail o por la plataforma de la Facultad para saber que deben hacer y estudiar. Este porcentaje aumenta al 86,5% entre no ingresantes.

16,3% decidieron dejar de cursar hasta que se normalice y pase la cuarentena, básicamente por razones económicas y/o por cuidado de terceros. Este porcentaje es mayor entre ingresantes (23%).

85,7% de los alumnos se encuentran cursando una cantidad entre 2 y 4 materias, siendo en gran mayoría una mezcla de materias de diferentes años.

59,4% son alumnos cursando materias de primer año, donde el 37,5% cursan solo materias de primero, y 21,9% cusan materias de Primero junto con materias de otros años. Entre este 59,4% se encuentran alumnos cursando hasta materias de quinto año, lo cual indica un régimen de correlatividades que permite avanzar sin haber completado el primer año, o dicho de otra manera se avanza son completar años anteriores. En el caso de alumnos cursando materias de quinto año, la mitad de ellos deben por lo menos una de primer año.

Respecto a las clases virtuales, la mayoría afirma estar recibiendo a lo sumo 3 clases virtuales, cuyo promedio está en 2,45 clases, perfil este similar en las cuatro Ingenierías, salvo en la TUTA donde se observa un 33% de alumnos que afirman no estar recibiendo ninguna clase virtual.

Los alumnos afirman estar recibiendo clases virtuales principalmente por la plataforma Moodle, le siguen las video conferencias tipo Zoom o similares, y en tercer lugar el WhatsAapp. Estas son las tres herramientas más usadas por todas las carreras de la Facultad. El uso de Mail y otros medios

también forman parte de las herramientas de comunicación con los alumnos, pero en menor medida. El celular se constituye en la herramienta masiva que disponen los alumnos para recibir clases virtuales, 89,1% afirman disponer de los mismos. Este porcentaje se mantiene tanto para alumnos con dificultades económicas para recibir clases, como para quienes decidieron dejar de cursar hasta que se normalice la situación de pandemia.

El WhatsApp es el medio por excelencia que usan los alumnos para comunicarse, solamente el 19,7% usa mail, y el resto de las aplicaciones en el orden del 10%.

El principal inconveniente es la disposición de señal de internet, 48,9% afirma tener acceso limitado, a los que se les suma un 6% sin acceso

Existe correlación entre la cantidad disponible de dispositivos con que cuenta el alumno y el grado de avance en la carrera. A medida que el alumno es más avanzado, disminuye el porcentaje de los que tienen solo celular, aumentando así la disponibilidad de múltiples recursos.

51% de los alumnos afirman necesitar capacitación para el acceso a toma de clases virtuales, especifican necesitar aprender como tomar clases usando programas de video conferencias, y manejo de la plataforma Moodle para subir archivos, leer lo que los profesores mandan, ver videos y hacer consultas.

La necesidad de capacitación es mayor entre alumnos que cursan los primeros años. El 63,5% de alumnos cursando materias de primer año afirman necesitar algún tipo de capacitación. La necesidad de capacitación decrece a medida que los alumnos cursan materias de años superiores, los alumnos cursando materias de quinto año que necesitan alguna capacitación es del 22,1%.

BIBLIOGRAFÍA

[1] Greimas, A. J., *Semántica estructural. Investigaciones metodológicas* [1966], Madrid, 1971; "Elementos para una teoría de la interpretación del relato mítico", en *Communications*, nº 8 (trad.: *Análisis estructural del relato*, pp. 45-86, y luego recogido en *Du Sens. Essais sémiotiques*, París, 1970, pp. 185-230 (también en *En torno al sentido*, Madrid, 1973, pp. 219-269, con el título de "Contribución a la teoría de la interpretación del relato mítico"); *La semiótica del texto: ejercicios prácticos. Análisis de un cuento de Maupassant*, Barcelona, Buenos Aires, 1983.
[2] Rodríguez Gómez G., Gil Flores J. y García Jiménez E. (1996) *Metodología de la Investigación Cualitativa. España.* Aljibe.
[3] Garrat, D. (2003) My qualitative research journey. Cresskill, New Jersey: Hampton Press Inc.
Briones G. “Métodos y Técnicas de Investigación”. Trillas 1995.
[4] Hernández, Fernández Baptista. “Metodología de la Investigación”. McGraw Hill 1994. Colombia.
[5] Padua J. “Técnicas de Investigación” FCE-Colegio de México 1982, México.
Sabino, Carlos A. El Proceso de Investigación. Buenos Aires: Edit. Lumen.1996
Salkind, Neil J. Métodos de Investigación. México: Prentice Hall. 1999.
[6] Sierra Bravo R. Tecnicas de investigación Social Teoría y ejercicios, Décima edición, Editorial Paraninfo 1995 Madrid.
[7] Taylor, S.J. y R. Bogdan. Introducción a los métodos cualitativos de investigación. Barcelona: Paidós. 1987.

10 y 11 de Diciembre de 2020

La estadística aplicada a la Seguridad Pública (policial y penitenciaria)

Autores: Marcela Carolina Calvó

Institución: Instituto Universitario Provincial de Seguridad

Datos de contacto: caro_calvo@hotmail.com – teléfono 388-4135975.

RESUMEN

En este escrito se realiza un recorrido por la experiencia presentada en el desarrollo del espacio curricular "Estadística aplicada a la seguridad" correspondiente al curso de ascenso obligatorio al grado inmediato superior policial y penitenciario en la jerarquía de oficiales jefes de la Policía y del Servicio Penitenciario de la Provincia de Jujuy- Argentina., con la intención de destacar el carácter procesual, la construcción y apropiación de contenidos de otra disciplina adaptada a la seguridad. Se busca dar una relectura al concepto de estadística y al de seguridad, un sentido especial al enunciado de lo situacional que acompaña cada proceso de aprendizaje por parte de estos cursantes. Ello también con la idea de que a partir de la aproximación crítica de situaciones que se problematizan, que se hacen visibles y conscientes tornando posible la construcción de proyectos o líneas de acción que busquen superar dichas situaciones problemáticas.

Palabras Claves: Estadística aplicada-Seguridad pública/ciudadana- Análisis de datos- Prevención

INTRODUCCION

El presente escrito pretende dar a conocer la modalidad de trabajo y los resultados del espacio curricular Estadística aplicada a la Seguridad; en el desarrollo del curso (2019) de ascenso al grado inmediato superior para el personal que se desempeña profesionalmente en la Policía y en el Servicio Penitenciario de la Provincia de Jujuy-Argentina, curso obligatorio y con un plan de estudios prefijado con antelación y con materias obligatorias.

Se realiza un recorrido por las características que presente el actual Instituto Universitario Provincial de Seguridad (IUPS), por la oferta educativa del mencionado curso, se dan a conocer las características de los cursantes y la modalidad de trabajo en espacio curricular.

Posteriormente se sintetizan dos propuestas una policial y una penitenciaria a los fines de socializar las producciones contratadas en el 2019, se recortaron los gráficos y extracto de los escritos de los grupos, algunas tablas que manifiestan la diversidad e inventiva para mostrar los datos. Si bien eran 41 cursantes y eran trabajos grupales, se considera que los plasmados en el presente escrito dan a conocer la esencia de lo conceptualizado en las aulas y finalmente en la conclusión, pensada como una instancia que abre caminos, poder rescatar las reflexiones de los comisarios y alcaides mayores en cuanto a la importancia de la estadística aplicada al ámbito de la seguridad.

METODOLOGIA

Las clases del espacio curricular Estadística aplicada a la seguridad se fueron desarrollando de manera teórico-práctica, bajo la modalidad semipresencial y con uso de plataforma MOODLE; en el dictado de las clases presenciales se utilizaba el proyector, con datos vinculados a la Seguridad, representados en tablas, gráficos de torta o gráficos de barra, en frecuencias, etc.

Se explicaban los contenidos de manera sencilla y con un lenguaje claro, considerando el docente que los cursantes no tenían nociones de estadística; se trataba de clases muy participativas y dialogadas ya que el docente, al ser civil, también iba incorporando los conceptos de las fuerzas de seguridad e interrelacionando con los propios de la estadística.

Se sistematizan más adelante; las producciones finales de los cursantes, tanto jefes policías como penitenciarios, a los fines de dar a conocer la implicancia y la importancia del análisis de datos en las fuerzas de seguridad, con diferentes finalidades.

Los trabajos socializados son el resultado del proceso de recolección de datos, concretos y reales de las comisarias o unidades penales, actualizados o de los últimos cuatro años, que con el desarrollo de los contenidos los datos presentados de manera suelta o sin conexión; se fue construyendo la información y analizando, volviéndola contextualizada y situada[1], con ello también; la confeccionaron gráficos pertinentes para la exposición en el examen final y la socialización del análisis al que fueron arribando.

[1] La noción de conocimiento situado ha venido figurando crecientemente en los discursos académicos y políticos, a donde ha ingresado fundamentalmente por la vía de las diferentes teorías, es que todo conocimiento se produce en situaciones históricas y sociales particulares, por mucho que se quiera hacer aparecer el verdadero conocimiento científico como universal, neutral y por lo tanto desprovisto de relaciones directas con determinados factores políticos, culturales y sociales. (PIAZZINI SUÁREZ 2014:12)

DESARROLLO

Contexto institucional

El Instituto Superior de Seguridad Pública (ISSP), fue creado mediante el Decreto Nº 3126-G-01 en el año 2001 y avalado por Ley N° 5348/02; tiene como finalidad la formación inicial, la capacitación permanente (Resolución N° 626-G-03), actualización y reconversión de los recursos humanos que se desempeñan en el ámbito de la Seguridad Pública como ser personal de la Policía y del Servicio Penitenciario en la Provincia, sumándose los vigiladores pertenecientes a empresas de Seguridad Privada (Ley N° 5436-G-04). Depende orgánica y funcionalmente del Ministerio de Seguridad, mientras que el Ministerio de Educación realiza el seguimiento, asesoramiento y evaluación de los aspectos académicos de las carreras que se dictan, titula a los egresados de la tecnicatura en Seguridad. Es el único en la Provincia y se encuentra ubicado en un barrio de la ciudad capital; funciona en su edificio propio desde el año 2010.

Actualmente se encuentra en proceso de transición y transformación a Instituto Universitario Provincial, establecido con la Ley Nº 6120 del año 2019, cuenta con el principio rector de la autonomía académica e institucional, autarquía económica y financiera. Se rige igualmente por la Constitución Nacional, por la Constitución de la Provincia de Jujuy, por su Ley de creación y por la Ley de Educación Superior Nº 24.521.

A continuación, se presenta información del espacio curricular en cuestión y otros elementos que se consideran fundamentales para representar los usos de la estadística en el ámbito de la seguridad.

Alumnos del curso de ascenso en el grado de oficial jefe para ser oficial superior que es la jerarquía inmediata superior- comisario (profesional policial) y alcaide mayor (profesional penitenciario), tienen entre 42 a 49 años y con una antigüedad en la fuerza de 23 a 25 años, recorrido por las diferentes dependencias policiales o áreas y localidades, y el personal penitenciario por diferentes unidades y áreas. Al ser oficiales jefes delegan todo tipo de tarea administrativa como ser la redacción de notas u escritos específicos, la carga de información a cargo de personal de menor jerarquía, lo que genera dificultades al momento de escribir, hablar y de usar las nuevas tecnologías, puesto que presentan ciertas falencias, desconocimiento total o parcial de ciertos programas o aplicaciones, utilidades, variedades. Esta materia permitió, más allá de los datos transcriptos por subalternos que ellos se vieran obligados a aprender el manejo de una computadora, de los programas básicos de Word y Excel y otras herramientas. El confeccionar soporte mediante diapositivas para los exámenes finales fue de gran ayuda para la mejora de la oralidad, dicción, lenguaje científico y coherencia en las secuencias lógicas.

Espacio curricular

El Curso Obligatorio de Ascenso al grado inmediato superior, para la jerarquía de oficiales jefes de la Policía y del Servicio Penitenciario de la Provincia de Jujuy- Argentina, en el que se comenzó a implementar en el 2016 el espacio curricular de “Estadística aplicada a la seguridad” y luego se retomó su dictado en el 2018 ya dentro del plan de estudios en la denominada “Diplomatura en Gestión de las Instituciones de Seguridad” la cual es en el marco de un Convenio con la Universidad Nacional de Jujuy (UNJu)- Argentina. Dicho espacio curricular se aborda con una modalidad particular, los cursantes realizan la recolección de datos que luego son analizados, tarea que no termina ahí ya que parte de las producciones implican la elaboración de líneas de acción.

Los Contenidos mínimos son: Nociones generales de estadística. Estadística descriptiva e inferencial. Población y muestra. Tablas de frecuencia e histogramas. Software específico. Contó la materia con siete encuentros presenciales de dos horas reloj y un encuentro presencial de

examen final, más una carga horaria virtual en plataforma educativa. El perfil docente prevé Ingeniero, Contador Público Nacional, Licenciado en áreas de las Ciencias Sociales, Licenciado en Seguridad, con experiencia en áreas de operaciones policiales.

Conceptos vertebradores

Es condición sine qua non comenzar a abordar los conceptos centrales o ejes vertebradores que le dan contexto a todo lo que se trabajó en torno a los cursos de ascenso en general y en Estadística aplicada en particular, lejos de la idea de definir de manera aislada los que se presenta a continuación, es vincularlos a la posición que se adoptó para abordar la materia y los sentidos epistemológicos que buscaban; no expertos en estadística sino constructores de información a partir de los datos que le presentaba la realidad, que dejaran de ser solo una cantidad para que pasaran a ser insumos de una política de acción e intervención, que pudiera cobrar vida lo plasmado en un dato estadístico y frio número.

En primera instancia vamos a desarrollar sucintamente ***Estadística aplicada,*** reconociendo que la enseñanza de la estadística ha ido tomando relevancia en el contexto actual debido a la complejidad de la realidad social, al gran crecimiento cuantitativo de datos[2] o información inconexa, avances vertiginosos en la tecnología, el componente fundamental tiene que ver también en el contribuir a la formación integral de las personas que se capacitan con diferentes planes de estudio y en distintos niveles educativos.

"*La Estadística es la ciencia que se encarga de recoger, organizar e interpretar los datos, … es la ciencia de los datos* (Gorás García 2011). Al ser una ciencia en si misma se pueden aplicar las metodologías a cualquier ámbito del conocimiento, con las debidas adaptaciones del caso, por ello y *"dado el desarrollo de la ciencia y tecnología, se hace también hincapié en cómo el análisis de los datos y su posterior interpretación estadística apoyan la toma de decisiones en cualquier área del conocimiento",* siendo fundamental los resultados que se pueden obtener y partiendo de la idea de que el "*diagnóstico de una organización de manera sistemática y metodológica es una práctica estratégica para los procesos de elaboración de un proyecto de toma de decisiones*" Se puede definir la estadística aplicada como *"un conjunto de procedimientos para reunir, medir, clasificar, codificar, computar, analizar y resumir información numérica adquirida sistemáticamente"* (Villegas Zamora 2019), donde la importancia está en evitar el riesgo de caer en sólo la recolección de datos, profundizando el análisis e interpretación para acciones futuras a partir de dicha información.

La **Estadística aplicada a la Seguridad,** en primer término implica concientizar sobre la importancia en la implementación de políticas públicas, en evaluarlas, en la ventaja de crear estadísticas referidas a la seguridad ciudadana, se van a considerar como parte del proceso el diagnóstico, la institucionalización de la producción estadística, uso de metodologías particulares y/o estandandares, generación de registros administrativos, automatización de la producción de datos estadísticos, identificación de fuentes de información, y oficialización de la información obtenida, con ello poder identificar fortalezas y generar políticas publicas tanto para el ámbito policial como penitenciario.

Toma relevancia así, el ***Análisis de datos;*** haciendo necesaria la mención de las etapas de un proceso de análisis de datos; las cuales son: El Problema o necesidad. La Recolección y recopilación de datos, definiendo métodos y técnicas. El Procesamiento de datos. La selección y el Filtrado de datos útiles. La representación de los datos en gráficos. El Análisis, mirando población y muestra, y su vinculación de la hipótesis, manipulación de diferentes técnicas como:

[2] Se lo puede "*definir como aquella información extraída de la realidad que tiene que ser registrada en algún soporte físico o simbólico, que implica una elaboración conceptual y que se pueda expresar a través de alguna forma de lenguaje"* (Abritta)

la predicción, la clasificación o los métodos de causa-efecto, entre otros. La Conclusión, plasmada en los resultados del análisis[3].

Es imperioso contextualizar los conceptos ***Seguridad pública y seguridad ciudadana,*** puesto que sería el campo de conocimiento donde toma protagonismo la estadística aplicada. En las últimas décadas, los delitos aumentaron considerablemente, presentando un impacto negativo en la imagen que se tiene de la policía y la penitenciaria, lo que genera inestabilidad en la seguridad pública en general, con actos y situaciones de indefensión e inseguridad.

Ante la necesidad de dar respuestas a los desafíos de estos momentos y a la superación progresiva de las deficiencias institucionales, se contrapone una postura crítica en la que se afirma *"(…) que la alta conflictividad social y en particular el aumento del delito responden (…) a la extensión de la violencia social que deriva del crecimiento de la pobreza, la marginalidad y la desintegración social*" (Saín 2010), que la policía o el sistema penitenciario como instituciones tienen muy poca o nula injerencia en revertir estas problemáticas cada vez más agudas y que corresponden al orden estructural.

Al definir seguridad pública se incorporan diferentes elementos que se vinculan con las políticas de intervención, los fines y las funciones del Estado, se la considera como un "*conjunto de acciones públicas (normativas, intervenciones, desarrollos institucionales) orientadas a producir y garantizar determinadas condiciones de convivencia , a la persecución de delitos, a la reducción de formas de violencia y a la protección de los bienes y de la integridad física de las personas"* en el cual influyen de manera determinante los ámbitos de competencia del Estado y los niveles de decisión, sin perder de vista que hay más actores encargados del abordaje de la seguridad pública -la comunidad, las organizaciones, los investigadores, los sistemas penitenciarios, entre otros- para dar lugar a la constitución de denominada seguridad ciudadana.

Los cambios en la concepción de seguridad antes mencionados tienen un fundamento diferente si el concepto es considerado desde los diferentes modelos de Estado del país. Esto significa que en los tiempos actuales se volvió la mirada a lo social, a lo comunitario, a la intervención y a la prevención. Implica tomar a la seguridad pública considerando el contexto social y principalmente como *"(…) categoría social (…) a los cuales se incorporan indicadores sociales y económicos (…)" (*Peñaloza 2006).

A partir de que el Estado generara instancias de participación, entendida como algo que se construye paulatinamente, donde hay mediación social, reflexión y discusión, donde los ciudadanos pasan a ser actores, es entonces cuando se comienza a hablar de un nuevo modelo asentado en una perspectiva democrática de la seguridad, en la que ésta se vuelve pública y no privada. "*En una democracia participativa los representantes no están para sustituir a los representados sino para convocarlos y generar marcos que profundicen su participación"*. La participación comunitaria preventiva es "*un núcleo constituyente del modelo democrático de seguridad ciudadana*" (Ministerio de Seguridad 2011).

Desde lo expuesto en el párrafo anterior, aparece el último factor clave que es la ***Prevención,*** como se dijera, en esta etapa se requiere de la participación de la ciudadanía, es alarmante cómo hay perdida permanente del espacio público a raíz de la comisión de delitos de diferente índole y envergadura, pero es necesario el abordaje de manera conjunta y no se resuelve con un modelo reactivo y de persecución tanto en el ámbito penitenciario como policial, sino en delimitar las causas de dichos delitos y trabajar en evitarlas. "*La problemática de seguridad que padecen muchas ciudades … denota, entre otros aspectos el abandono de espacios públicos, la desintegración comunitaria, la desconfianza ciudadana hacia las autoridades, el incremento del consumo de drogas ilícitas y la ausencia de los valores cívicos, (…) afrontar el fenómeno delictivo y los problemas de seguridad pública desde la perspectiva de la prevención del delito y la*

participación ciudadana. Lo anterior tiene el propósito de construir entornos seguros y devolver a la ciudadanía la confianza en sus instituciones públicas, para hacer un frente común y atacar los factores que originan la criminalidad"[4]

"*La prevención del delito se ha convertido en un componente cada vez más importante de muchas estrategias nacionales de seguridad pública. El concepto de prevención se basa en la idea de que el delito y la victimización se ven favorecidos por numerosos factores causales o de fondo, los cuales son resultado de una amplia gama de elementos y circunstancias que influyen en la vida de las personas y las familias a medida que pasa el tiempo, y de los entornos locales, así como situaciones y oportunidades que facilitan la victimización y la delincuencia. Determinar qué factores están asociados a los diferentes tipos de delitos puede dar lugar a la elaboración de una serie de estrategias y programas para cambiar estos factores y prevenir o reducir la incidencia de tales delitos" (Naciones Unidad 2011).*

Producciones finales

- ✓ **Penitenciaria**

Actividad: Diagnóstico para la prevención del delito e infracciones disciplinarias de internos (2017-2018)

-Delitos o infracciones disciplinarias de internos:

- Solo agresiones verbales, no se registran agresiones físicas
- Agresiones físicas entre internos (golpes de puño)
- Secuestros de elementos de no permitidos, tenencia de teléfono celular.
- Secuestros de sustancias no permitidas, tenencia de hojas de coca y practica de coqueo.

-Días con mayor frecuencia: la mayor cantidad de hechos ocurrieron durante los días miércoles, sábados y domingos en las actividades en común posteriores a la visita de familiares, y en los horarios de prácticas deportivas.

-Los motivos de las agresiones son variados: rivalidades entre grupos (liderazgo), rechazo a someterse a los reglamentos penitenciarios

-INFRACCIONES

Ley 23.737 de estupefacientes. En lo que va del año 2017-2018 se registraron actuaciones administrativas por secuestro de estupefacientes, en el Establecimiento Penitenciario N° X de Internos Mayores varones del Servicio Penitenciario de Jujuy

-OTROS PROBLEMAS – RIESGOS Y AMENAZAS

Alojamientos diferenciados por resguardo de integridad física, se producen generalmente en los sectores de alojamiento, en razón de que los internos han agotado todos los espacios disponibles para su alojamiento, tornándose la convivencia peligrosa para su integridad física por lo cual debe alojarse en sectores diferenciados con un régimen especial a fin de no vulnerar derechos humanos fundamentales del mismo. También influye el tipo de causas con la que ingresa al Establecimiento carcelario, siendo las de abuso generalmente un condicionante en la normal convivencia con el resto de los alojados.

Rivalidades entre bandos; en ocasiones logran conformarse bandos internos, a fin de disputarse el liderazgo de un sector, o suelen expresarse en las actividades en común como las deportivas o educativas donde concurren a un mismo lugar de encuentro.

Atentados contra la autoridad, existen antecedentes de conflictos que afectaron directamente al personal de servicio de guardia y oficinista, resultando solo agresiones verbales en el periodo 2017 -2018, ante llamados de atención por ejemplo de eventuales infracciones disciplinarias, requisas de

rutina ante actitudes sospechosas, también cuando en señal de protesta por alguna medida se niegan a recibir el racionamiento de almuerzo o presentan escritos de habeas corpus.
-FACTORES CRIMINOGENOS, entendemos como tales a los elementos que contribuyen a favorecen o facilitar:

- la comisión u omisión de una conducta de resultado antisocial.
- la aparición del crimen (aunque por sí solos son incapaces de producirlo).

Elementos comunes

- malas compañías: en los grupos de alojados, existen internos influyentes sobre los más manipulables, quienes son inducidos a cometer actos de indisciplina o a cometer actos en contra del orden y la disciplina en el sector, a fin de que sea escuchados en sus reclamos o peticiones.
- aglomeración de internos, que favorecen la actividad de los indisciplinados, en las actividades en común (educativas, recreativas, sectores).
- desarraigo familiar: en menor proporción algunos internos son abandonados por sus familiares, y en escaza rara vez reciben visitas de algún allegado.

-Cantidad de efectivos penitenciarios: Posee (14) efectivos, de los cuales (13) se encuentran en servicio activo y (01) en disponibilidad previo retiro:

- ✓ Director: 01
- ✓ Jefe de Tratamiento: 01.
- ✓ Guardia A: 03 (01 Oficial de turno y 02 Celadores)
- ✓ Guardia B: 03 (01 Oficial de turno y 02 Celadores)
- ✓ Guardia C: 03 (01 Oficial de turno y 02 Celadores)
- ✓ Servicio Intendencia: 01
- ✓ Judicial. 01
- ✓ Consejo Correccional: 01

-**La problemática en gráficos**- fuente: libro de registro de expedientes administrativos

Gráfico N°1 Tipo de faltas cometidas por los internos 2017

Gráfico N°2 Tipo de faltas cometidas por los internos 2018

Gráfico N°3 Tipo de requisas 2017

Gráfico N°4 Tipo de faltas cometidas por los internos 2018

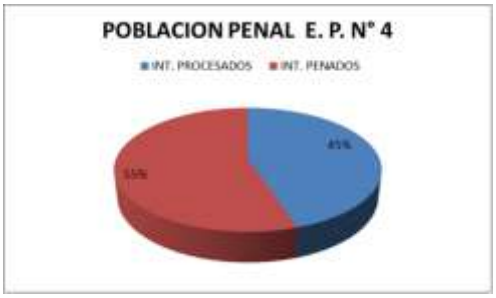


Grafico N° 5 Tipo de población penal

Otros tipos de infracciones

SECUESTRO AÑO 2017-2018

Nº	INTERNO	PAB. Nº	VISITA	PARENTESCO	FECHA	ORDEN	DIAS	SECUESTRO		REQUISADOR
1	Gustavo	5	Carmen	Hermana	31-12-17	04-D.E.P.Nº1/18	7	30grs de hoja de coca-fondo de la bolsa mercaderia	INGRESO	Subayte. Fatima
2	Daniel	enfermeria	Rebeca	Madre	07.01.18	32 - D E.P.Nº1/18	7	(01) celular - bolsillo del pantalon	INGRESO	Subayte. Laura
3	Facundo	1	Cleotilde	Madre	10.01.18	45 - D E.P.Nº1 /18	5	$100-bolsillo del pantalon	INGRESO	Subayte. Rina
4	Sergio	enfermeria	Juana	Madre	10.01.18	46 - D E.P.Nº1/18	10	(06) encendedores, (01) cable, $1.400	INGRESO	Ayte 2da Brenda
5	Mariano	4	Gabriel	Hermano	10.01.18	47 - D E.P.Nº1/18	5	(01) bateria de celular - bolsillo del pantalon	EGRESO	Ayte de 2da Adrian
6	Mauricio	5	Raul	Padre	14.01.18	66 - D E.P.Nº1/18	7	(01) cargador, (02) encendedores	INGRESO	
7	Enzo	5	Edic	Madre	17.01.18	68 - D E.P.Nº1/18	10	(01) cargador de fabricacion casera- Bolsa de yerba	INGRESO	Subayte. Barbara
8	Cristian	3	Vicente	Padre	17.01.18	69 - D E.P.Nº1 /18	10	(01) envoltorio de "marihuana"- bolsa de fideo	INGRESO	Subayte Rebeca
9	Cristian	enfermeria	Luciana	Concubina	21.01.18	89 - D E.P.Nº1 /18	15	$1.500 - bolsa de mercaderia	INGRESO	Ayte 2da Brenda

Tabla N° 1 Modelo de registro de elementos secuestrados-tipo de infracción (extracto)

Grafico N° 6 Tipos de elementos secuestrados 2018 (a partir de tabla N°1

Propuestas de mejora

Para la Función de seguridad deben tenerse en cuenta ciertos elementos de información:

- Esta función debe ser "proactiva". No debemos esperar a que ocurran los delitos o las infracciones para recién actuar, sino que previamente debe existir mayor presencia en los sectores de alojamiento de internos, acentuándose la actividad de observación permanente por parte de los celadores sobre las actividades diarias de los internos individuales y colectivos.
- De igual forma, como en el resto de los Establecimientos Penitenciarios existe una escasa cantidad de efectivos penitenciarios en relación a la cantidad de internos alojados (proporción).
- Es notable la falta de respeto a la autoridad por parte de los alojados sin embargo en esta unidad en particular se ha disminuido en este último tiempo.

Medidas de seguridad alternativas: el número de aparatos portátiles de comunicación, teléfonos corporativos ha permitido una mayor fluidez en la comunicación y en la transmisión de novedades o eventos de gravedad que pudieran ocurrir, haciéndolo a la brevedad posible se ha mejorado en este aspecto.

Sobre el sistema de custodia

a. Para un mejor control, incrementar el personal penitenciario destinado a la función de celadores; personal de guardia en contacto directo con los internos.
b. Mantener un régimen sanitario: aseo, limpieza de las instalaciones, permanentemente. Al respecto, los sectores de alojamiento cuentan con ducha e inodoros.
c. Se dispone de horarios de visitas para familiares, amigos, abogados. Implica la vigilancia y posterior requisa de los detenidos.
d. Disponer de alimentos para la totalidad de los alojados (desayuno-almuerzo-merienda-cena).
e. Debe disponerse de personal para control o tratamiento médico intramuros, entrevistas o audiencias con servicios de asistencia social, gabinete y directivos.
f. Recorridos, deberían ejecutarse por: A) De a pares ante eventuales procedimientos, B) Celdas o espacios de aglomeración de internos.
g. Potenciar las actividades educativas, recreativas, deportivas. Por el cual se dispone de personal para el control de las actividades y la vigilancia de los mismos.

- ✓ **Policial**

Problemática: Denuncias contra personal policial de la provincia de Jujuy por mala resolución en el manejo verbal ante conflictos entre terceros durante el año 2018

Población en estudio: personal policial de la policía de la provincia de Jujuy denunciado por mala resolución en el manejo verbal ante conflictos entre terceros, durante el año 2018.

Variables:

1.-Seccional
2.-Unidad Regional
3.-Jerarquía
4.-Sexo
5.-Edad
6.-Mes de Ocurrido
7.-Lugar de Origen
8.-Denuncia por cada efectivo

Sistematización de datos

JERARQUIA	FREC.ABSOL.	FREC.RELAT.	FREC.ABSOL.ACUM.	FREC.REL.ACUM.
Comisario	2	4,55%	2	4,55%
Sub Crio	2	4,55%	4	9,10%
Of.Insp.	2	4,55%	6	13,65%
Of.Sub-Insp.	2	4,55%	8	18,20%
Of.Ayte	1	2,27%	9	20,47%
S.O.P.	2	4,55%	11	25,02%
Sgto.Ayte.	5	11,36%	16	33,38%
Sgto.1º	5	11,36%	21	47,74%
Sgto.	4	9,10%	25	56,84%
Cabo 1º	8	18,18%	33	75,02%
Cabo	6	13,64%	39	88,66%
Agente	5	11;36%	44	100,00%
TOTALES	44	100%		

Tabla N°2 Porcentajes de personal denunciado por jerarquía

Tabla N°3 Cantidad de denuncias por jerarquía

	MES DE OCURRIDO										
SECCIONAL	Enero	Febrero	Marzo	Abril	Junio	Julio	Agosto	Septiembre	Noviembre	Diciembre	TOTALES
A	2	2		1				2			7
B		1	1			1	1			1	5
C		1			1			1			3
D	4	1						2			7
E		1					3		1	1	6
F		1									1
G		1						1			2
H				1			2				3
I								1		1	2
J	1	1									2
K	1	1	1					2		1	6
TOTALES	8	10	2	2	1	1	6	9	1	4	44

Tabla N°4 Mes en el cual se realiza la denuncia al personal policial

Propuestas de mejora

- ✓ Capacitación al personal policial en materia de relaciones con la comunidad.
- ✓ Mejoras en vínculos con la comunidad.
- ✓ Distribución del personal policial, según aptitudes favorables para la resolución de conflictos.

UN FINAL QUE ABRE CAMINOS

El abordaje de la Estadística aplicada en el ámbito de la Seguridad, de cara a las diferentes problemáticas debería constituirse en una herramienta tendiente a la mejora en la gestión de las instituciones de este tipo.

Este aporte en la formación de jefes brindaría conceptos básicos y herramientas a poner en práctica para el relevamiento de datos, recolección de datos de diversas fuentes, y a partir de ello el análisis y sistematización a la postre de que a partir de la interpretación son factibles la generación de estrategias de gestión, el poder de decisión que ellos tienen facilitaría la ideación de políticas tendientes a la mejora permanente y continua.

Actualmente las áreas policiales y penitenciarias generan estadísticas en manera individual y cada uno por su lado, en otras ocasiones no le prestan atención a la importancia de la construcción de información a partir de estos datos que brinda la realidad. Para abrir caminos se concluye este escrito con la frase de *"No confíes en lo que la estadística te dice hasta haber considerado con cuidado que es lo que no dice."* William W. Watt[5]

BIBLIOGRAFIA

ABRITTA G. P. (2014). NOCIÓN Y ESTRUCTURA DEL DATO disponible en http://metodoscomunicacion.sociales.uba.ar/wpcontent/uploads/sites/219/2014/09/Abritta.pdf Consultado en septiembre 2016

BLOGG https://conceptosclaros.com/como-analizar-dato. Consultado en noviembre del 2020.

Conferencia Estadística de las Américas Quito, (2012). Estadísticas de Seguridad Ciudadana en los Sistemas Estadísticos Nacionales. Disponible en https://www.cepal.org/sites/default/files/presentations/inec_estadisticas-de-seguridad.pdf . Consultado en noviembre del 2020.

GORAS GARCIA J. (2011) Estadística básica para estudiantes de ciencias. Disponible en https://webs.ucm.es/info/Astrof/users/jaz/ESTADISTICA/libro_GCZ2009.pdf. Consultado en noviembre del 2020

PEÑALOZA, J. (2006). Seguridad Pública: La crisis de un Paradigma. Disponible en http://www.criminologiaysociedad.com.mx/wp-content/uploads/2017/12/SeguridadPublica-LacrisisdeunParadigma.pdf. Consultado en agosto del 2016.

PIAZZINI SUÁREZ C. E. (2014). Conocimientos situados y pensamientos fronterizos: una relectura desde la universidad. Instituto de Estudios Regionales Universidad de Antioquia Disponible en https://revistas.ucm.es/index.php/GEOP/article/view/47553. Consultado en septiembre del 2017.

SAIN, M. (2010). "La reforma policial en América Latina. Una mirada crítica desde el progresismo". Buenos Aires. Prometeo.

Secretaria de Seguridad Pública- México, Informe de labores. Disponible en https://pdba.georgetown.edu/Security/citizensecurity/mexico/evaluaciones/InformeLabores-prevencion.pdf. Consultado en el mes de octubre del 2015

[5] en Gorás García op. cit.

VILLEGAS ZAMORA, D.A. (2019) La importancia de la estadística aplicada para la toma de decisiones en Marketing. Disponible en http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2521273720190002000004&lng=es&nrm=iso. Consultado en noviembre del 2020.

10 y 11 de Diciembre de 2020

Deserción escolar en los niveles obligatorios durante el periodo de pandemia dentro de la Provincia de Jujuy

Autores: Castro, Cristian Eduardo; Mamaní, Gabriela; Maraz, María del Rosario; Tejerina, Guillermo Fernando; Farfán, José Humberto; Rodríguez, Mariela Ester

Institución: Facultad de Ingeniería, Universidad Nacional de Jujuy. San Salvador de Jujuy.

Datos de contacto: Datos de contacto: cecpersonal@gmail.com; gbrlmamani@gmail.com,; marazmaria39@gmail.com; guille.fert@gmail.com; jhfarfan@fi.unju.edu.ar, mariela.rodriguez@fi.unju.edu.ar

RESUMEN

En la actualidad el mundo se encuentra atravesando un cambio radical en todos sus ámbitos (social, educativo, económico, etc.) debido a la propagación global de la declarada pandemia causada por el COVID-19. ¿Es posible medir estos efectos en el ámbito educativo y dentro de los niveles obligatorios de escolaridad en la Provincia de Jujuy?

El estado de confinamiento en la provincia de Jujuy impuso que sean los padres y/o tutores, los educadores para los estudiantes de los niveles inicial, primario y secundario en muchas ocasiones reemplazando el rol de los maestros y profesores. Esta situación inesperada, ha encontrado a padres que no cumplen adecuadamente el rol de educadores, situación que perjudica el aprendizaje de los estudiantes y provoca deserción escolar. El presente trabajo intenta descubrir patrones, tendencias y reglas que expliquen el comportamiento en el caso de estudio de “Estudiantes Salidos sin Pase”, con datos proporcionados por las Unidades Educativas de los Niveles Obligatorios de la provincia de Jujuy, aplicando técnicas de minería de datos para cumplir este objetivo.

Este estudio realiza un Análisis Inteligente para obtener conclusiones y validar de esta manera resultados, además proporciona información fundamental para la toma de decisiones a nivel ministerial, convirtiéndose en un recurso estratégico para la gestión escolar.

Palabras Claves: Minería de Datos, Deserción Escolar, Unidades Educativas, Análisis Inteligente, Jujuy.

INTRODUCCIÓN

El actual trabajo se centra sobre la situación informada de "Estudiantes Salidos sin Pase". Entiéndase a los mismos a aquellos alumnos regulares que abandonan la escuela en los Niveles Obligatorios antes de finalizar el ciclo lectivo, sin solicitar el certificado necesario (como requisito oficial) para inscribirse en otro establecimiento y continuar allí sus estudios con regularidad. En la realización del trabajo se utilizó la metodología KDD (metodología usada dentro de la minería de datos) y el programa informático Rapidminer, el cual permite procesar grandes cantidades de datos a través de la implementación de distintos modelos con algoritmos de aprendizaje automáticos (o *machine learning*). Una vez que son ejecutados dichos algoritmos se clasifica el conjunto de datos iniciales denominados "Estudiantes Salidos sin Pase" para identificar los patrones más relevantes que de aquí se deriven, así como aquellos modelos que proporcionen la mayor eficiencia sobre el registro propuesto. Se explicitan además las limitaciones encontradas, así como las clases obtenidas al aplicar los distintos modelos, las performances alcanzadas con grandes cantidades de datos en tiempos reducidos aprovechando la gran capacidad de cálculo y la clasificación que se le puede aplicar a este caso de estudio para finalmente validar los modelos que muestran los mejores resultados, implementando meta operadores de Rapidminer que refrendan los modelos propuestos.

METODOLOGÍA

Como metodología adoptada para este estudio se emplea la metodología KDD (Descubrimiento de Conocimiento en Base de Datos) "KDD está basada en un bien definido proceso KDD de múltiples pasos, para el descubrimiento de conocimiento en grandes colecciones de datos. El proceso KDD es iterativo por naturaleza, y depende de la interacción para la toma de decisiones, de manera dinámica" [1] a fin de encontrar y descubrir patrones o reglas que expliquen el comportamiento de los datos de acuerdo a la información registrada en "Estudiantes Salidos sin Pase". KKD se basa en 5 pasos [2] y [3] que se detallan en el siguiente apartado (Figura 1).

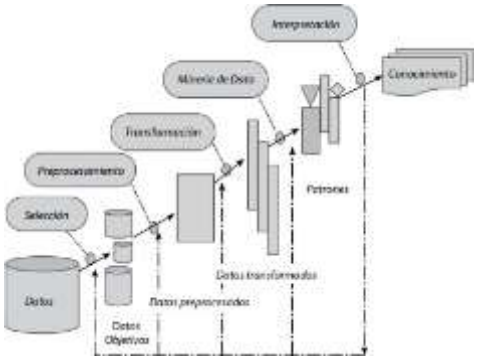


Figura 1. Etapas del proceso KDD

DESARROLLO
APLICACIÓN DE LA METODOLOGÍA KDD

1. Selección de datos:

El punto de partida lo constituyen los **8202** registros que conforman el dataset centralizado e inicial de datos recolectados, con un total de **25 atributos**.
Identificando como etiqueta al atributo **motivo_baja** definido como "label" (atributo de estudio principal) a alcanzar, En este caso, se procede a seleccionar los atributos con los que se seguirá el estudio (no considerando id_unidad_servicio, ciclo_act_2020 y tipo_baja) y tampoco los apellidos, nombres y nro_documento de los estudiantes por considerarlos datos sensibles en el contexto educativo y para proteger la identidad de los mismos.
Los objetivos establecidos son:

Objetivo Principal: Analizar el atributo motivo_baja para determinar cuáles fueron los motivos principales por los cuales los estudiantes salieron sin el certificado correspondiente de las unidades educativas durante el periodo lectivo 2020 en la Provincia de Jujuy.

Objetivos Específicos:

- Determinar Región Educativa donde hubo mayor número "Estudiantes Salidos sin Pase".
- Establecer en qué Nivel Educativo se dieron más casos de "Estudiantes Salidos sin Pase".
- Determinar rango etario de los "Estudiantes Salidos sin Pase" que presentan más incidencia en el registro total
- Escuelas en las que se dio mayor cantidad de "Estudiantes Salidos sin Pase".
- Cuáles son las localidades de la provincia en las que se concentran más casos de "Estudiantes Salidos sin Pase".
- Establecer clases que permitan definir patrones entre los datos registrados.

2. Limpieza y pre-procesamiento:

Como herramienta de Data Mining se emplea el software informático Rapidminer, el cual "es una plataforma de análisis que permite acelerar la creación, entrega y mantenimiento de analíticas predictivas de alto valor"[4]
En esta etapa de limpieza se procede a trabajar sobre el DataSet para garantizar que los mismos queden limpios, por lo que el tratamiento incluye tratar valores outlier (remover los valores atípicos), eliminar atributos, eliminar o filtrar registros no útiles para enriquecer y garantizar la utilidad de los mismos. Para la limpieza se trabajó con un operador de subproceso que contiene los operadores que permiten limpiar los datos del DataSet (Figura 2).
Lo primeros tres operadores permiten generar un atributo nuevo llamado "edad" que se obtiene a partir del campo fecha_nacimiento del estudiante (mediante el operador Generate Atributes de la Figura 3), el operador Real to Integer denominado Edad de Real a Entero y el operador Discretize by user Specification llamado Disc Edad (Figura 4).

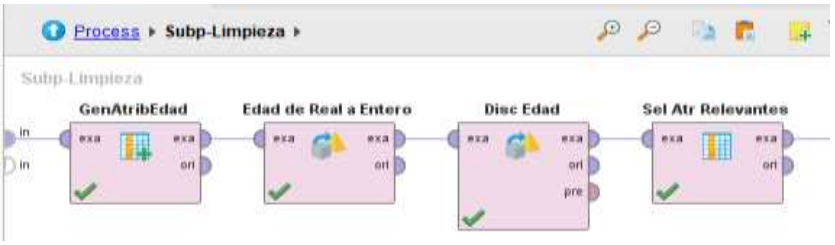


Figura 2. Operadores del Subprocess Subp-Limpieza para discretizar edad

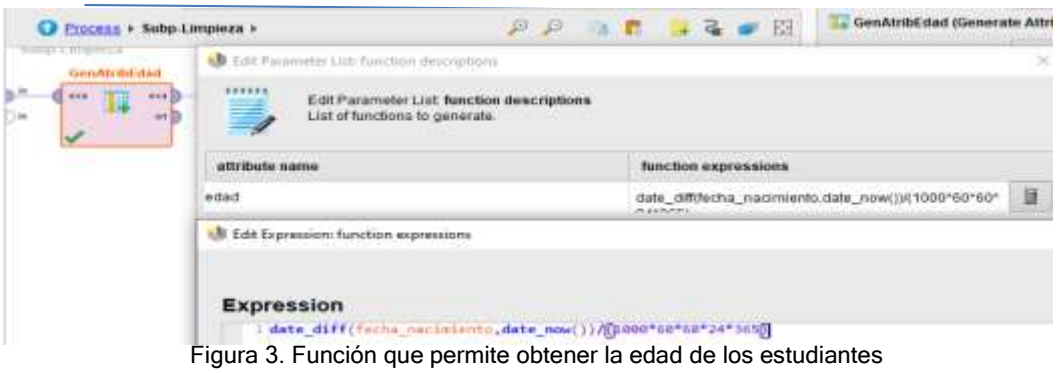


Figura 3. Función que permite obtener la edad de los estudiantes

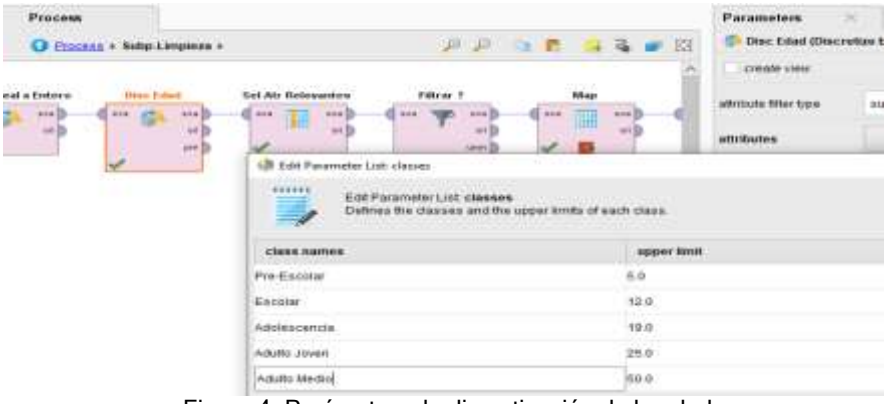


Figura 4. Parámetros de discretización de la edad

Para discretizar la edad se considera el rango etario según las etapas del desarrollo psicosocial de Erikson que "... afirma que los seres humanos con un desarrollo sano deben pasar a través de ocho etapas entre la infancia y la edad adulta" [5] (Figura 5).

Etapa	Edad	Conflicto	Figura Representativa	Virtud	Malignidad
Infante	0 - 2 años	confianza vs. desconfianza	madre	esperanza y fé	distorsión sensorial
Bebé	2 - 3 años	autonomía vs. vergüenza/duda	padres	voluntad y determinación	impulsividad y compulsión
Pre-escolar	3 - 6 años	iniciativa vs. culpa	familia	propósito y coraje	crueldad e inhibición
Escolar	7 - 12 años	laboriosidad vs. inferioridad	escuela + vecinos	competencia	virtuosidad unilateral
Adolescencia	12 - 19 años	identidad yoica vs. confusión de roles	grupos	fidelidad y lealtad	fanatismo y repudio
Adulto Joven	20 - 25 años	intimidad vs. aislamiento	colegas + amigos	amor	promiscuidad y exclusividad
Adulto Medio	25 - 50 años	generatividad vs. autoabsorción	hogar + trabajo	cuidado	sobrextensión y rechazo
Adulto Viejo	50 - ... años	integridad vs. desesperación	humanos + "mios"	sabiduría	presunción y desesperanza

Figura 5. Tabla utilizada para discretizar edad de estudiantes

Después se seleccionan los atributos con los que se continúa el estudio en función de los objetivos específicos propuestos se filtran los registros cuyo motivo_baja y localidad no son valores perdidos. Debido a que el atributo motivo_baja es considerado el objetivo del estudio en curso utilizando el operador Filter con el renombrado como "Filtrar ?". (Figura 6).

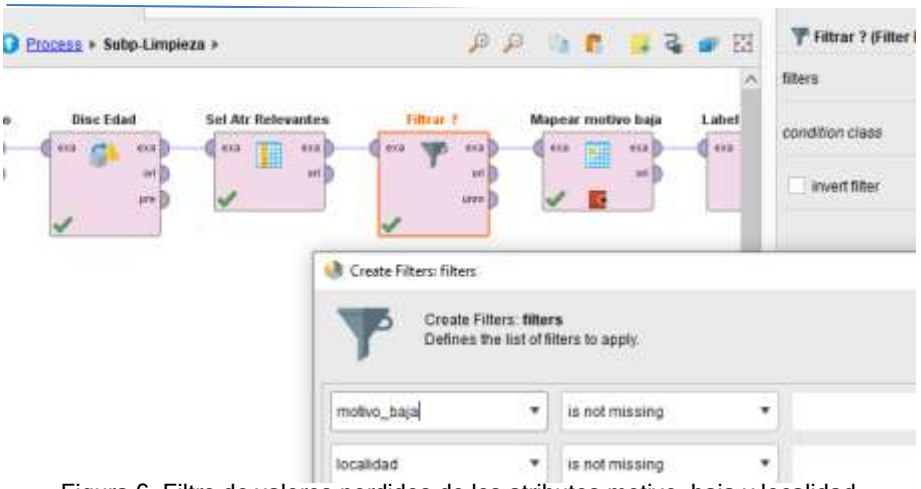


Figura 6. Filtro de valores perdidos de los atributos motivo_baja y localidad

Se mapean los valores del atributo motivo_baja a fin de considerar distintos aspectos de este motivo. Se emplea el operador Map renombrado como “Mapear motivo baja” (Figura 7).

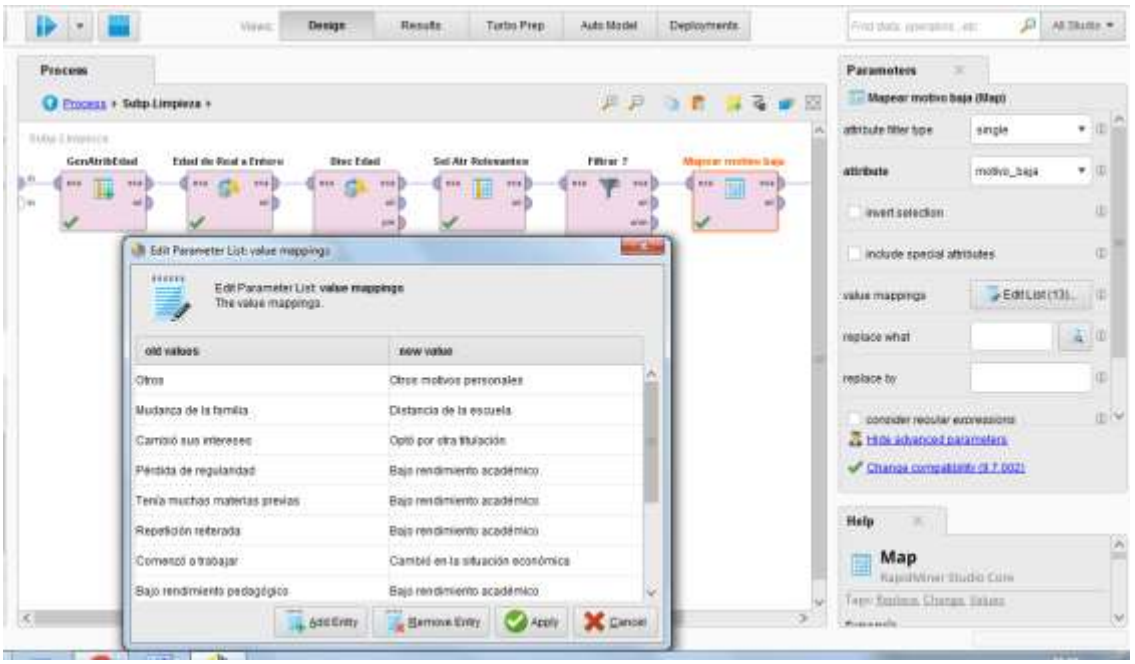


Figura 7. Mapeo de atributo motivo_baja

Como último paso se establece como label al motivo-baja, para ello se utiliza el operador Set Role denominado Label motivo_baja.

3. Transformación de los datos:

La mejora de la calidad de los datos que esta etapa busca se realiza con la transformación de la edad de los estudiantes mediante la conversión de valores reales a numéricos (Figura 3) y con la Selección de Atributos, a través de la cual se redujo la dimensionalidad del archivo a trabajar (Figura 7).

Mientras que, con el fin de conocer la concentración de los estudiantes salidos sin pase según la localidad de residencia declarada se aplican técnicas de transformación para posteriormente aplicar como modelo el cluster con los siguientes operadores (Figura 8):

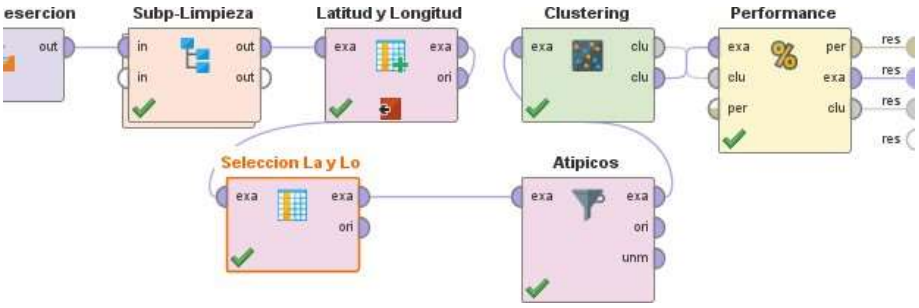


Figura 8. Operadores aplicados para obtener la latitud y longitud según localidad de residencia de los estudiantes

Como primer paso, se generan los atributos latitud y longitud a partir de la localidad declarada de los estudiantes (Figura 9):

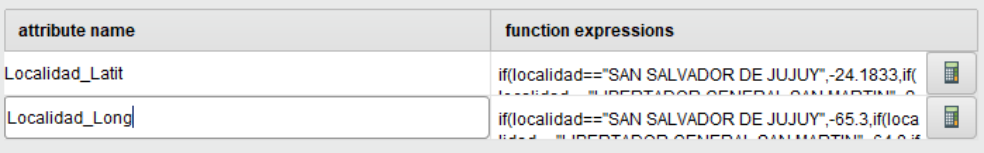


Figura 9. Funciones utilizadas para generar latitud y longitud

Para la correcta funcionalidad del operador cluster (K-Means) solo se seleccionan los atributos latitud y longitud mediante el operador Select Attribute llamado “Seleccion La y Lo”. Por último se filtran aquellas localidades donde la cantidad de estudiantes es inferior a 10 (Figura 10):

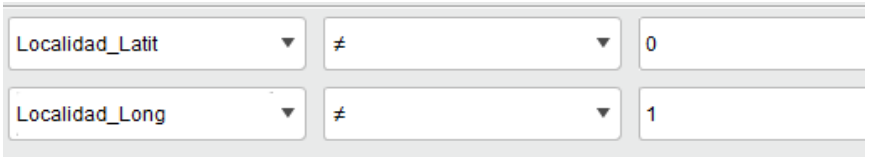


Figura 10. Funciones utilizadas para generar latitud y longitud

4. Minería de datos: Elección de tareas de Minería de Datos

Con el fin de continuar con el estudio correspondiente se determina que los modelos a utilizar en primera instancia son los árboles de decisión debido a que los mismos se utilizan en problemas de aprendizaje supervisado para determinar patrones desde el DataSet luego de ser limpiado y transformado para garantizar la calidad de los datos. También se trabaja con meta operadores los cuales validan los resultados obtenidos por los modelos propuestos, y finalmente se emplean clusters para mostrar la distribución de los estudiantes según las localidades de la provincia.

Para tal fin se incluye en este apartado el análisis estadístico realizado:

1. Cantidad de estudiantes salidos sin pase por Región (Figura 11)

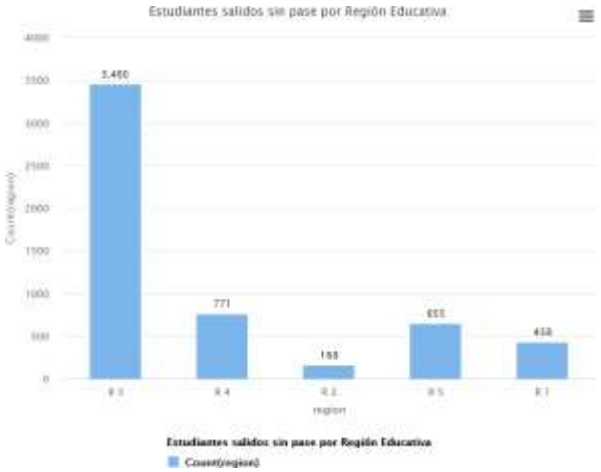


Figura 11. Cantidad de estudiantes salidos sin pase por Región

Donde las Regiones se conforman de la siguiente manera:

- Región 1: Departamentos de Yavi, Santa Catalina, Rinconadas y Cochinoca.
- Región 2: Departamento de Humahuaca, Tilcara y Tumbaya.
- Región 3: Departamento Manuel Belgrano, Palpalá, San Antonio y El Carmen.
- Región 4: Departamento de San Pedro y Santa Barbará.
- Región 5: Departamento de Ledesma y Valle Grande.

A continuación, se muestran los gráficos de:

- Motivos de baja recurrentes (Figura 12).
- Estudiantes salidos sin pase por sector educativo (Figura 13).
- Estudiantes salidos sin pase según rango etario (Figura 14).
- Estudiantes salidos sin pase según sector educativo (Figura 15).

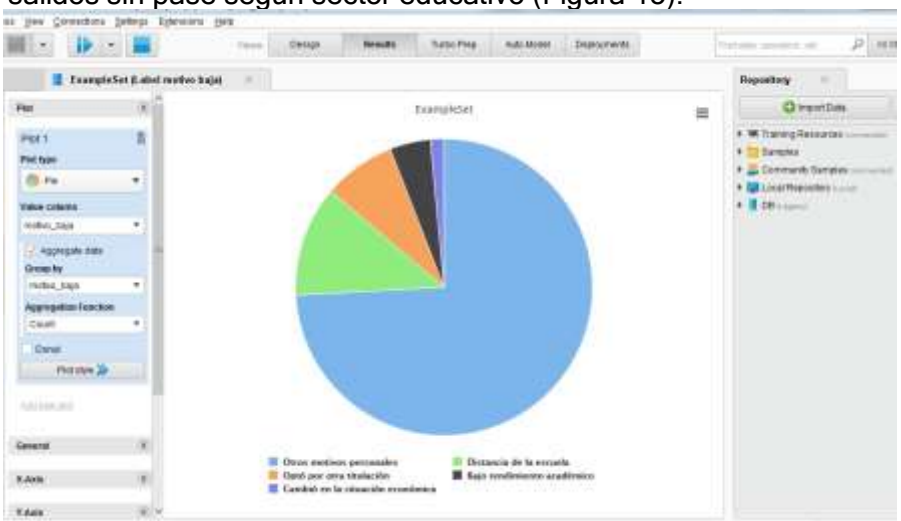


Figura 12. Motivos de baja recurrentes

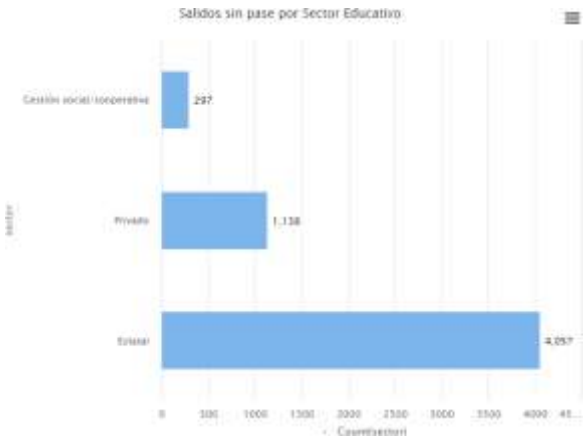


Figura 13: Estudiantes salidos sin pase por sector educativo.

Figura 14: Estudiantes salidos sin pase según rango etario

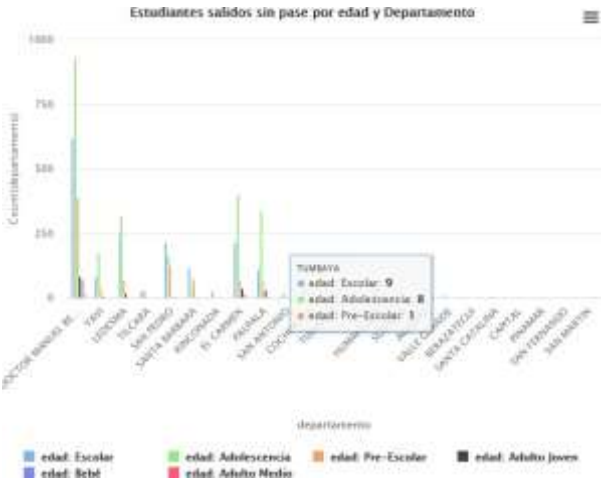


Figura 15: Estudiantes salidos sin pase por rango etario y departamento

Continuando, se muestra la aplicación de los diferentes modelos utilizados, en primer lugar árboles de clasificación, luego reglas de asociación y finalmente la aplicación de clusters. Considerando para ello las condiciones necesarias para su utilización.
En este caso, se incluye el Subproceso denominado “Subp Modelo” que contiene el modelo y los operadores requeridos para obtener la performance de cada modelo (Figura 16):

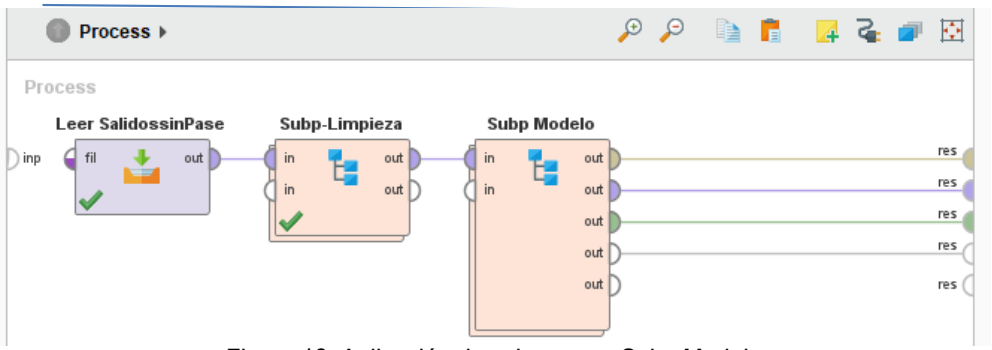


Figura 16: Aplicación de subproceso Subp Modelo

El Operador Subp Modelo contiene los siguientes operadores:

- Split Data: separa los valores del dataset para prueba y entrenamiento.
- W-J48: es el árbol de decisión más eficiente y eficaz para nuestra clasificación.
- Apply Model : aplicamos el modelo para su prueba.
- Performance: obtenemos las performance de clasificación para nuestros datos

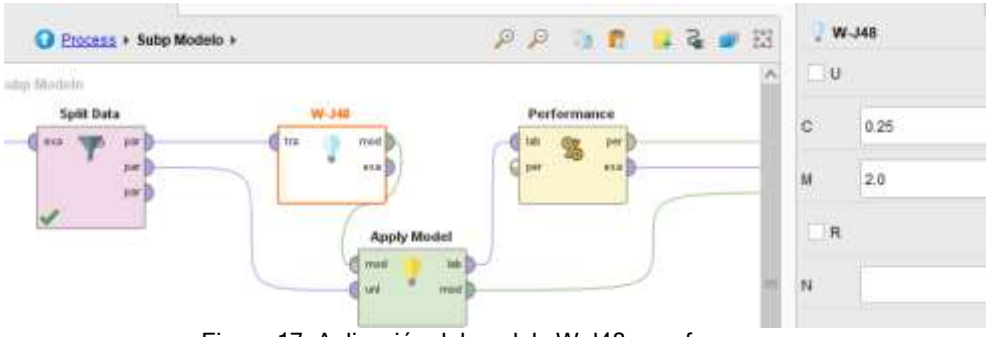


Figura 17: Aplicación del modelo W-J48 y performance

Para la técnica J48 con una confianza del 0,25 se obtiene una performance de 91,36% (figura 17).

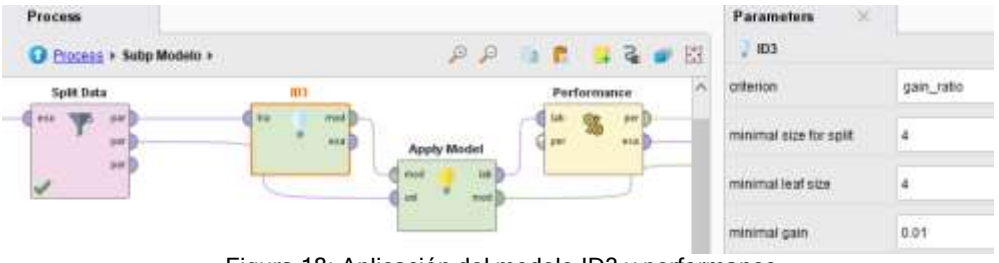


Figura 18: Aplicación del modelo ID3 y performance

Con el operador ID3 (árbol de decisión sin podar a partir de datos nominales para su clasificación) se obtiene una performance de 90.00% (figura 18).

Reglas de Clasificación:

Se aplica W-OneR, operador de Weka que usa particiones derivadas de un sólo atributo para asignar valores a los individuos que tienen ese atributo (figura 19) tal como se muestra a continuación:

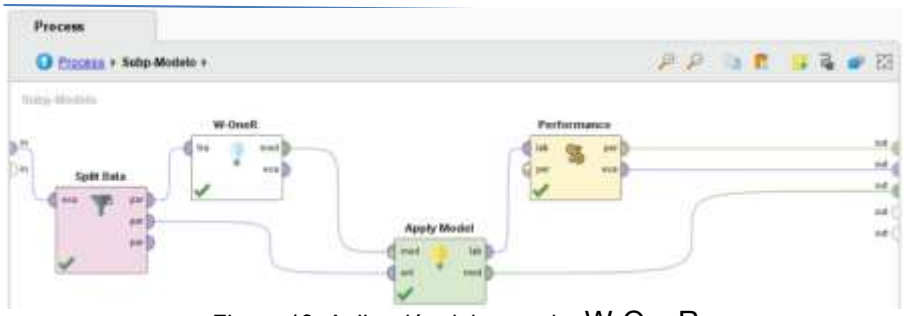


Figura 19: Aplicación del operador W-OneR

Se obtienen los siguientes resultados (figura 20):

Figura 20: Obtención de la eficiencia del operador W-OneR

También se aplican Reglas de asociación con Zero-R, el cual es un operador de Weka y es el método de clasificación más simple que existe y que solo depende del target u objetivo ignorando todos los predictores [FR].

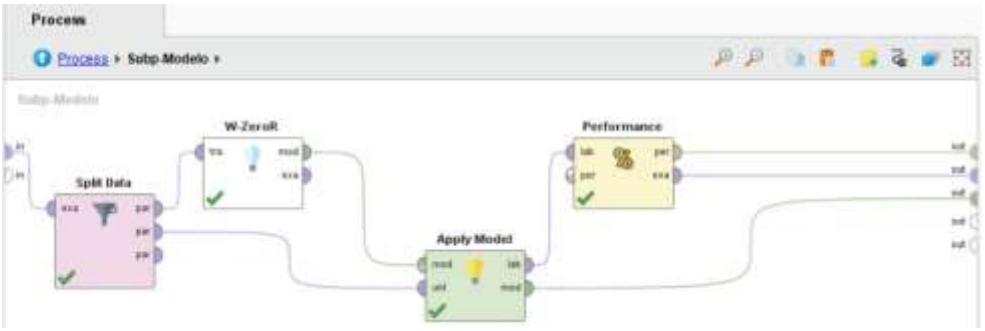


Figura 21: Operadores para el análisis de eficiencia del modelo Zero-R

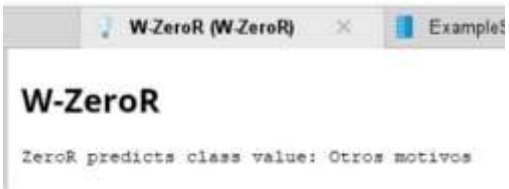


Figura 22: Resultado obtenido de Zero-R

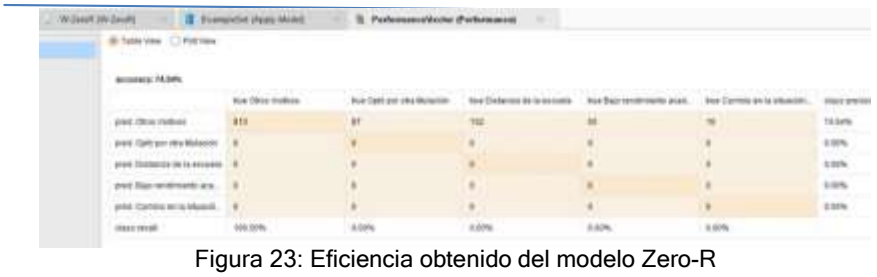


Figura 23: Eficiencia obtenido del modelo Zero-R

A continuación se aplica Prism, tal como se muestra en la figura 24, el cual es un algoritmo de tipo cobertura que, en cada paso identifica una regla que cubre algunas de las instancias.

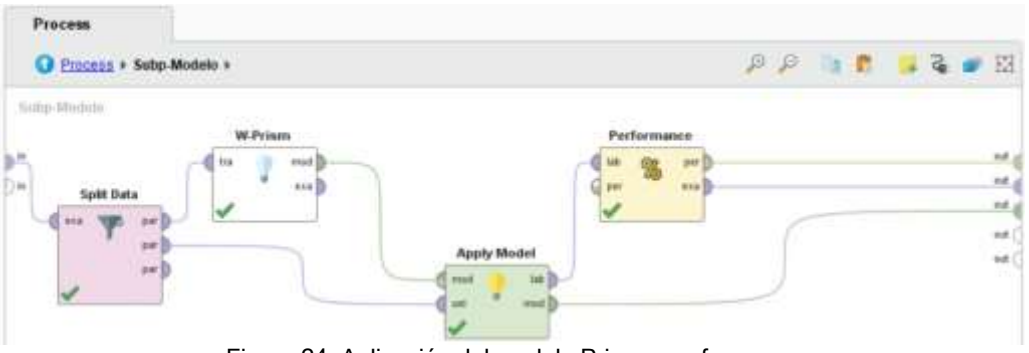


Figura 24: Aplicación del modelo Prism y performance

Como puede observarse en la figura 25, se obtiene una performance del 88.62%.

Figura 25: Eficiencia del modelo Prism

Finalmente, se procede a aplicar técnicas de clustering, en este caso el operador K-Means (figura 26), el cual es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es el más cercano.

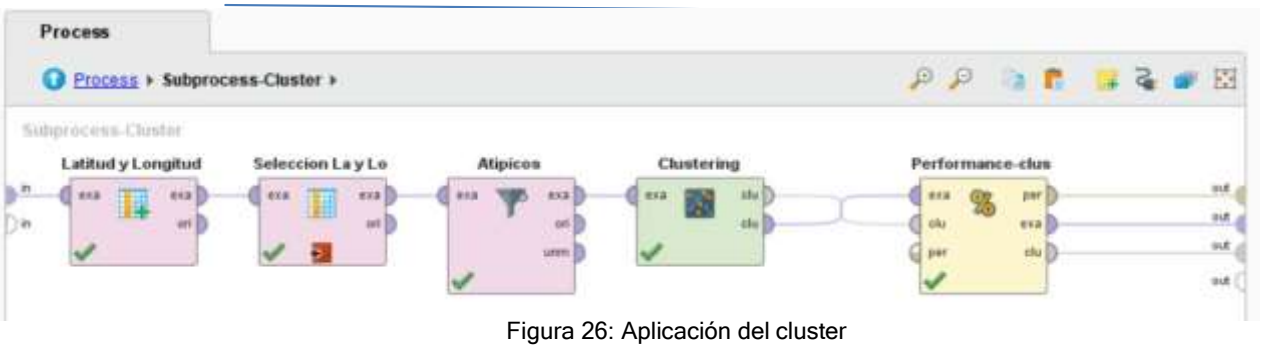


Figura 26: Aplicación del cluster

Resultado del Cluster:

Como se observa en la figura 27, se utiliza un K=2 ya que los valores más alejados corresponden a la localidad de La Quiaca (color verde):

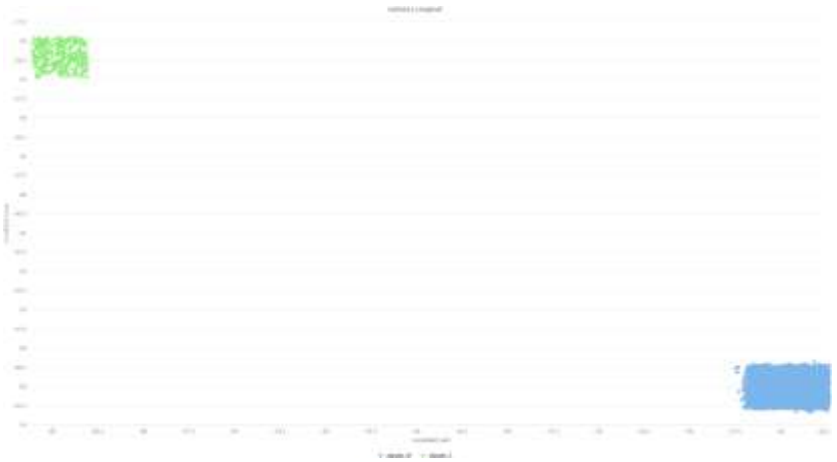


Figura 27: Resultados del cluster con K=2

y la distancia obtenida se muestra en la figura 28:

Figura 28: Resultados del cluster

Se observa que la mayor cantidad de estudiantes se concentra en el cluster 0, entonces se procede a filtrar esa localidad (figura 29):

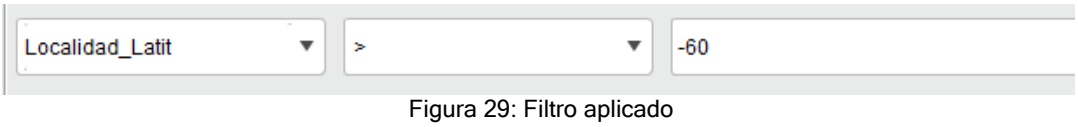


Figura 29: Filtro aplicado

Continuando, se aplica nuevamente las técnicas de cluster con K=8 ya que es la cantidad óptima de separaciones para las localidades y se obtiene una distancia muy pequeña y aceptable (figura 30 y 31):

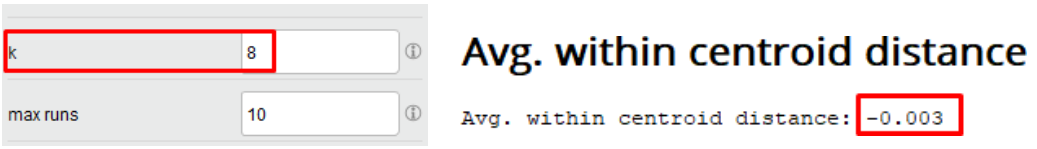


Figura 30: Distancia obtenida con K=8

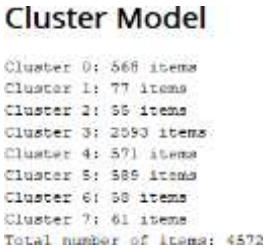


Figura 31: Items por cluster

El resultado es el que se puede observar en la figura 32:

Figura 32: Cluster conformados

De donde se deduce que la mayor cantidad de estudiantes sin pase se concentra en las coordenadas latitud -24 y longitud -65 aproximadamente que corresponden a los departamentos de el Carmen, Palpalá y Dr. Manuel Belgrano.

5. Implementación- Entendimiento del conocimiento:

Los resultados obtenidos muestran coherencia en cuanto a la exploración de los datos y la comparación con los resultados obtenidos en los modelos, así como los obtenidos por el análisis estadístico y de visualización gráfica nos permite detallar las conclusiones abajo obtenidas.

CONCLUSIONES

En primera instancia se destaca la necesidad e importancia que conlleva la limpieza de los datos del

DataSet origen a fin de que la misma garantice la calidad de los resultados.
Otro aspecto es la recomendación de desagregación o clasificación necesaria para detallar el motivo de salida ya que cuando este tome el valor otro_motivo permita brindar más información referida al objetivo bajo estudio. Por lo que en los próximos relevamientos que se realicen es conveniente contemplar esta sugerencia.
Realizada una exploración visual de datos se obtiene que los tres departamentos con mayor cantidad de estudiantes salidos sin pase son el departamento Dr. Manuel Belgrano, El Carmen y Ledesma pertenecientes los dos primeros a la Región Educativa 3 y el último a Región Educativa 5.
En lo que se refiere al tratamiento realizado sobre los datos durante el transcurso de la investigación, conforme se contextualizan los datos ya completos, se realiza la selección de las características más destacadas y pertinentes para cada caso de esta manera obtener una buena performance en post de alcanzar el objetivo principal propuesto.
Luego de aplicado los modelos propuestos y obtenidos los resultados se concluye que con árboles de decisión aplicando reglas de Asociación, con el operador One-R se puede obtener el nombre de cada escuela con el motivo de deserción de alumnos más predominante en dicha institución educativa, aplicando este operador se obtiene una precisión del 89,88%. Al aplicar el operador Zero-R se analizan todas las reglas y se muestra la más sobresaliente, que en este caso permite inferir que la deserción de alumnos en el año 2020 fue mayormente por otros motivos. Obteniendo una precisión del 74,11%, también se aplica el operador Prism donde se muestra cada una de las reglas generadas por el modelo, detallando establecimiento, región educativa, departamento y motivo de deserción predominante, con este operador se obtiene una precisión del 87,88%.
Respecto a los clusters, se puede concluir que, para este trabajo, el algoritmo K-Means es un buen clasificador si se obtiene la cantidad óptima de K, además permite verificar que con este algoritmo la cantidad de estudiantes salidos sin pase se encuentran concentrados en la localidad de San Salvador de Jujuy y alrededores. Se utilizan varios operadores además de K-Means como K-Medoids y DBScan, pero estos dos últimos requieren un tiempo excesivo en la ejecución y los resultados obtenidos no muestran mejoras a los obtenidos con K-Means.
En los clústeres se puede observar la conglomeración de datos en las coordenadas correspondientes a las localidades de San Salvador de Jujuy, Palpalá, El Carmen (Región Educativa 3) las cuales son ciudades de mayor densidad poblacional y donde se cuenta con mayor disponibilidad en cuanto a infraestructura tecnológica y servicios de conectividad sin embargo se observa que en el periodo comprendido de registro fueron las localidades donde más salidas de alumnos hubo, aún considerando que ese mismo periodo las clases se desarrollan de manera remota por la pandemia.
Mientras que en la localidad de La Quiaca y Abra Pampa también se dieron casos de estudiantes que salieron, pero en esas localidades a pesar de tener densidad demográfica baja hubo un elevado número de deserción lo cual puede considerarse el análisis de la brecha digital existente al no contar con la conectividad necesaria.
Como consideraciones para futuros trabajos se propone registrar el momento fehaciente en el que sucedieron las salidas sin pase ya que el registro no incluye esta fecha. Aplicar índices poblacionales por localidad a fin de normalizar conglomeración de datos según registro de localidad de residencia de los estudiantes.

BIBLIOGRAFÍA

[1] SK Gupta, V. B. (1997). “A proposal for Data Mining Management System. Sk Wasan.”

[2] Hernandez Orallo, (2005). “Introducción a la Minería de Datos”. España. Editorial Pearson Educación S.A.

[3] José Farfán, Mariela Rodriguez. (2020). “Técnicas de Minería de Datos”. Obtenido de https://jhfarfan.wixsite.com/datamining/tecnicas-dm-1, consultado en Noviembre 2.020.

[4] MarTech Forum. (2019). “Rapidminer. MarTech Forum”. Obtenido de https://www.martechforum.com/herramienta/rapidminer//, consultado en Noviembre 2.020.

[5] Web del Maestro cmf. “La teoría de Erik Erikson: Las etapas del desarrollo psicosocial. Web del Maestro cmf”. Obtenido de https://webdelmaestrocmf.com/portal/la-teoria-de-erik-erikson-las-etapas-del-desarrollo-psicosocial, consultado en Enero 2.020.

10 y 11 de Diciembre de 2020

Aplicación de mapas auto-organizados para la identificación de patrones de comportamiento de los conductores en España

Autores: Almudena Sanjurjo de No, Blanca Arenas Ramírez, José Manuel Mira y Francisco Aparicio Izquierdo

Institución: Instituto Universitario de Investigación del Automóvil Francisco Aparicio Izquierdo (INSIA-UPM), Escuela Técnica Superior de Ingenieros Industriales (ETSII-UPM), Universidad Politécnica de Madrid (UPM), Madrid (España)

Datos de contacto: almudena.sanjurjo.no@gmail.com y +34 910677278

RESUMEN

Los accidentes de tráfico son un problema socioeconómico clave en las sociedades y una de las principales causas de muerte no natural. En los últimos años, España ha alcanzado una posición destacada entre los países del mundo con mejores niveles de seguridad vial. Sin embargo, lograr una nueva mejora desde la posición actual podría requerir un mayor conocimiento sobre los comportamientos de los diferentes colectivos de conductores para discriminar, en lo posible, las acciones más eficaces para cada uno de ellos. Por ello, en esta investigación, se propone el uso de la metodología de mapas auto-organizados de conglomerados con el objetivo de identificar patrones de comportamiento entre los conductores por género y edad y en relación a las infracciones cometidas. Los resultados desvelan diferentes patrones multivariantes de comportamiento de los conductores, lo que ayuda en la toma de decisiones de política vial que podrían enfocarse en el desarrollo de medidas de prevención específicas para los diferentes colectivos de conductores y en la focalización de los recursos por parte de las oficinas reguladoras de seguridad vial.

Palabras Claves: comportamiento del conductor; identificación de patrones, mapas auto-organizados.

1. INTRODUCCIÓN

Las estadísticas de la mayoría de los países reflejan una mayor tasa de accidentes de hombres frente a mujeres y de jóvenes y ancianos frente a conductores de mediana edad, tanto en la participación de accidentes, como en la comisión de infracciones y en los comportamientos de riesgo. Estos hechos requieren un análisis exhaustivo de las causas y factores de influencia, y una comprensión más profunda de estos colectivos para mejorar las políticas de aplicación, educación y prevención enfocadas en la gestión de riesgos de los conductores por género y grupos de edad, para contribuir a obtener mejores cifras de seguridad vial.

Las diferencias de comportamiento entre distintos colectivos de conductores ha sido objeto de estudio entre los investigadores de seguridad vial desde hace muchos años. En la literatura, se han encontrado numerosos análisis que ponen su foco sobre el género como una variable importante y de interés, y los resultados indican que es un factor relevante en la conducción y el riesgo de resultar implicado en un accidente de tráfico, como han señalado los autores de las referencias (1-4).

La accidentalidad con resultado de víctimas mortales es mayor entre los conductores del género masculino (2, 4-8). Estas tasas en ambos sexos están influenciadas por la edad, siendo los conductores jóvenes (2, 3, 6, 8-10) y los más mayores (6, 8, 9) los que reportan una mayor cantidad de accidentes.

Por otro lado, los autores del trabajo (11) concluyen que los hombres muestran menor percepción del riesgo durante la conducción. Estos también conducen con velocidades mayores que las mujeres (8, 10) y, en general, las investigaciones realizadas, concluyen que los hombres y los conductores jóvenes cometen más infracciones de distinto tipo, desde una menor tasa de utilización del cinturón de seguridad hasta un mayor consumo de alcohol y/o drogas (5, 7, 8, 10-12) y son más propensos a infringir las normas de tráfico (5, 10). Las mujeres, por su parte, son más propensas a distraerse y a cometer errores perceptivos (8).

Dado el interés por el análisis de los temas relacionados con la seguridad vial, considerando el género y la edad de los conductores, así como las diferencias significativas encontradas en diversas investigaciones, en este trabajo se propone el uso de la metodología estadística de aprendizaje no supervisado de conglomerados conocida con el nombre de mapas auto-organizados (SOM, por sus siglas en inglés: Self-Organizing Maps) con el objetivo de identificar patrones de carácter multivariante en relación al género y la edad de los conductores con respecto a las infracciones cometidas y el estado de dichos conductores.

Las técnicas de conglomerados de SOM se han aplicado a diferentes campos. En accidentes de tráfico, se han encontrado algunos trabajos relevantes: se propone un modelo probabilístico en (13) para predecir accidentes de tráfico mediante la aplicación de monitoreo 3D de vehículos. En el trabajo de (14) se estudiaron los accidentes de peatones, aplicando técnicas de agrupación con el fin de identificar patrones para diseñar medidas preventivas y asignar recursos a los problemas identificados.

Para llevar a cabo esta investigación se ha utilizado la base de datos del registro de conductores implicados en accidentes entre dos vehículos en España entre el año 2004 y 2013. Esta base de datos, que fue proporcionada por la Dirección General de Tráfico (DGT), fue tratada y depurada previamente para estudiar colisiones entre turismos, en zona interurbana y para los tipos de accidentes: frontal, frontolateral, lateral y de alcance. Por lo que la base de datos finalmente utilizada contiene un total de 145.904 conductores. La base de datos utilizada contempla la información de todos los conductores, así como de las infracciones cometidas, el estado de los mismos o el estado de sus vehículos. Entre todas las variables, se seleccionaron las infracciones

y el estado de los conductores para identificar los patrones por género y edad en relación a estas variables. Estas son: infracción del conductor, infracción de velocidad, infracción administrativa, defecto físico, estado del vehículo, consumo de alcohol y/o droga, enfermedad súbita, sueño, cansancio y/o preocupación.

2. METODOLOGÍA

Para llevar a cabo esta investigación se han empleado la técnica de conglomerados de aprendizaje no supervisado conocida con el nombre de mapas auto-organizados o Self-Organizing Maps (SOM).

El propósito de SOM es representar datos multidimensionales en un espacio de menor dimensión (2D o 3D), típicamente 2D, de modo que la agrupación se pueda visualizar en un llamado mapa, manteniendo la estructura topológica, es decir, de tal manera que los datos que están cerca en el espacio original seguirán estando cerca en el espacio de dimensión reducida. Esta característica de reducción de dimensionalidad es una gran ventaja de SOM porque el mapa 2D proporcionará un agrupamiento visible y, por lo tanto, analizable muy rápidamente (15- 17).

El resultado final de SOM es, por lo tanto, un mapa de un número finito de nodos, indexados por pares de números enteros, donde cada dato de muestra del espacio original (de alta dimensión) se asigna a un solo nodo del mapa de dimensión proyectada (de baja dimensión). Los nodos en el mapa tienen una representación en el espacio de alta dimensión original que se denominan sus pesos.
El algoritmo SOM procede secuencialmente, es decir, cada dato de la muestra se asigna a su vez al nodo más cercano en el mapa, y posteriormente se actualizan tanto las coordenadas del nodo ganador como las vecinas. Esta actualización de los nodos vecinos se denomina etapa de aprendizaje cooperativo, es la que proporciona la preservación de la topología mencionada anteriormente y es la diferencia esencial, junto con la reducción de dimensión, con respecto a la metodología estadística de aprendizaje no supervisado de agrupamiento estándar conocida con el nombre de k-medias (k-means).

3. DESARROLLO

En esta sección, se ha aplicado la metodología estadística de conglomerados de mapas auto-organizados (aprendizaje no supervisado) a las distintas infracciones o posibles estados desfavorables del conductor con el objetivo de identificar patrones multivariantes de comportamiento de los conductores en función del género y de las distintas franjas de edad de los mismos.

Por lo tanto, se realizará un análisis de conglomerados de SOM para las 8 variables de infracciones y estado de los conductores: el agrupamiento de los datos de los conductores en el espacio inicial de 8 dimensiones se proyecta en un mapa bidimensional, preservando la topología, es decir, las distancias entre observaciones. El mapa se muestra en la Figura 1, donde la notación que se utiliza es C_{ij} para cada nodo (conglomerado) donde “i” y “j” representan filas y columnas del nodo en el mapa, respectivamente. Los grupos son lo más homogéneos posible en el espacio de 8 dimensiones original (homogeneidad multivariante) y esta trata de mantenerse en el espacio proyectado (de dimensión reducida). A cada variable de infracción o estado del conductor se le asigna un color diferente, como se especifica en la leyenda del mapa (18).

Para la construcción del mapa auto-organizados (SOM) ha sido necesaria la codificación de los valores de las variables para que estas pasasen de ser categóricas a tener un valor numérico, hecho necesario para su tratamiento en el software estadístico R. Así, se estableció el criterio que se muestra en la Tabla 1.

Valor categórico de las variables	Valor numérico de las variables
Sí está presente la infracción o defecto	2
No está presente la infracción o defecto	0
Se ignora si existe dicha infracción o defecto	1

Tabla 1: Transformación numérica de las variables categóricas

El mapa SOM de infracciones es el que se muestra en la Figura 1.

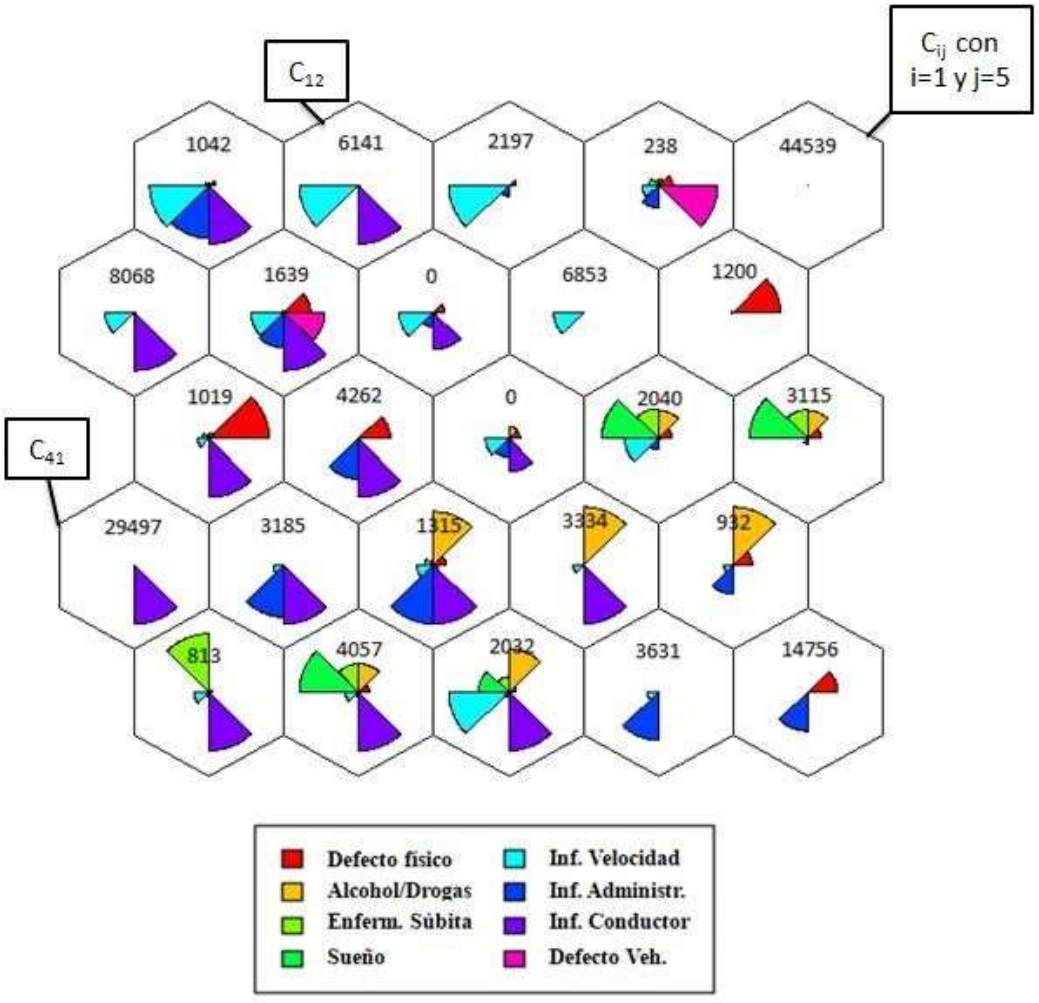


Figura 1: Mapa auto-organizados (SOM) de infracciones

En primer lugar, es importante señalar que se toma el valor intermedio de 1 para los casos en los que se desconoce el valor de la variable. Esto es una hipótesis que se realiza, al considerar inicialmente que si se desconoce el valor de una variable, será igualmente probable que dicha variable pueda estar o no presente. Se requieren análisis estadísticos adicionales de validación de esta hipótesis, dado que es cuestionable el hecho de que el valor que deba ser asignado, para los casos en los que se desconoce el valor de la variable, esté entre medias de los dos valores extremos. Un análisis más profundo que ya tiene en cuenta esta cuestión es elaborado y presentado en (19).

El tamaño de cada sector de color mide el promedio de los datos que caen en ese conglomerado para la variable correspondiente a ese color. Así, cuanto mayor es el tamaño de cada sector, mayor es el número de conductores (de los que caen en ese nodo) que han cometido la infracción o presentan el defecto (valor 2) que indica el color de la variable que está representando. Por otro lado, si los sectores de colores tienen un tamaño medio, querrá decir que su valor estará en torno a 1, que indica que el valor medio de los conductores que caen en ese nodo presentarán la infracción que represente el color de ese sector, en estado desconocido, dado que dicho valor está en torno a 1. Finalmente, si un sector de color no puede ser visualizado en un nodo, eso es porque la media de esa variable para los conductores que caen en dicho nodo está en torno a 0 y, por lo tanto, dicha variable no está presente en los conductores de ese conglomerado o nodo. Así, por ejemplo, el nodo C_{12}, que incluye un total de 6.141 conductores, contiene los sectores que representan a las variables infracción de velocidad e infracción del conductor en tamaño máximo (valor 2 para las dos variables). Por lo tanto, la totalidad de conductores que caen en ese conglomerado habrán cometido sólo dichas infracciones.

Adicionalmente, en el SOM, se observa un grupo particularmente llamativo de conductores, que es el que se encuentra en el nodo C_{15}, dado que incluye una proporción bastante importante de los conductores (un total de 44.539 conductores) y es el conglomerado que se caracteriza por estar formado por aquellos conductores que no han cometido ningún tipo de infracción ni presentan ningún estado desfavorable para la conducción. A su vez, también destaca el nodo C_{41}, que incluye a los conductores que han cometido algún tipo de infracción de las denominadas infracciones del conductor. Los nodos C_{55} y C_{41} representan un total de 50,74% del total de los conductores de la base de datos. Por ello, son conglomerados a los que hay que prestar atención, dado que más de la mitad de los conductores están repartidos entre estos dos nodos.

3.1. Análisis del mapa en función del género del conductor

En esta subsección se va a llevar a cabo el análisis del reparto de conductores, en función del género de los mismos, a lo largo del mapa auto-organizados (SOM) de la Figura 1. El objetivo de este apartado es identificar patrones en relación al género de los conductores y en función de las variables de infracción analizadas.

La Figura 2 muestra el reparto, a lo largo del mapa, de los conductores en función del género del mismo. En el mapa se indica los porcentajes de hombres y mujeres que caen en cada uno de los conglomerados. Estos porcentajes han sido obtenidos con respecto al total de conductores hombres y mujeres, respectivamente.

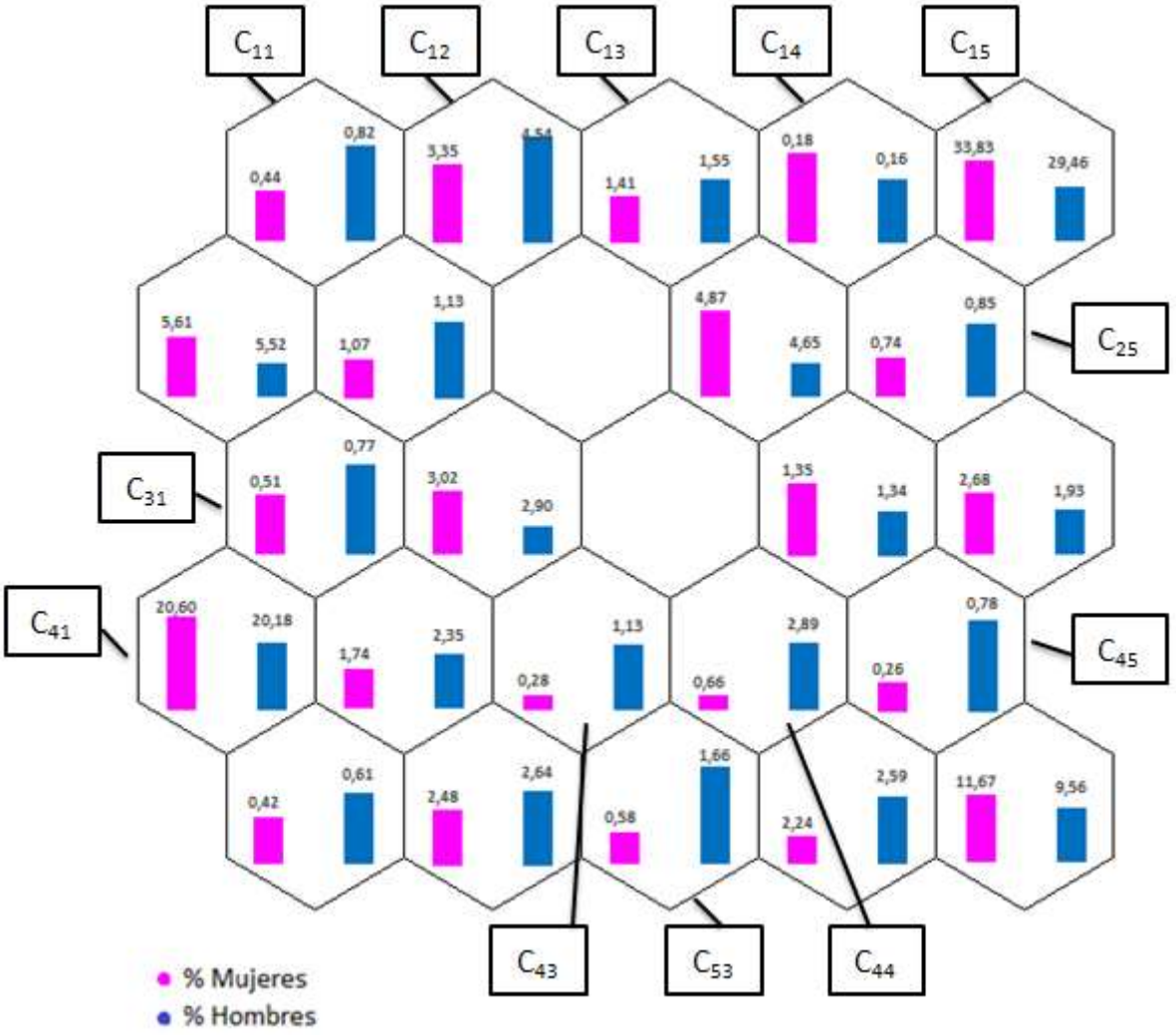


Figura 2: Distribución de los conductores en el mapa por género

Comparando el reparto de conductores por género (Figura 2) con el mapa auto-organizados (SOM) de infracciones (Figura 1), puede realizarse un análisis desagregado por tipo de infracción. Así, se observa que los hombres cometen un mayor número de infracciones de distinto tipo, a pesar de que globalmente las infracciones del conductor resultan relativamente igualadas entre hombre y mujeres (C_{41}) e incluso las mujeres parecen ligeramente superar a los hombres en este tipo de infracciones. Una posible causa de esto es que se cree que las mujeres, en general, cometen un mayor número de infracciones por distracción, siendo estas, la causa más frecuente de infracciones del conductor, tal y como señalan algunas investigaciones como la realizada por (8). No obstante, será necesario llevar a cabo un análisis complementario de las infracciones del conductor para dictaminar una causa más precisa del por qué se produce este hecho en este tipo de infracciones.

Por otro lado, puede observarse como el porcentaje de hombres aumenta con respecto al de las mujeres cuando se tienen en cuenta las infracciones de velocidad (C_{13}). Además este aumento es más acentuado cuando las infracciones de velocidad van combinadas con algún otro tipo de

infracción, especialmente las infracciones del conductor y el consumo de alcohol y/o drogas (C_{11}, C_{12} y C_{53}). Con respecto a dicho consumo de alcohol y/o drogas, otra observación importante del análisis es que el porcentaje de hombres es notablemente mayor que el de las mujeres cuando este consumo es positivo, tanto cuando esta infracción aparece de manera prácticamente aislada (C_{45}), como cuando aparece combinada con otro tipo de infracciones (C_{43}, C_{44} y C_{53}). Otro tipo de infracciones, como las administrativas, también parecen ser más frecuentes entre hombres que mujeres, aunque claros patrones multivariantes no se han identificado con respecto a esta infracción. Por otro lado, parece que los defectos físicos tienden también a estar más presentes entre los conductores masculinos (C_{25} y C_{31}), mientras que se han identificado más defectos en los vehículos entre los conductores del género femenino (C_{14}), aunque particularmente esta última condición no parece ser muy representativa, dado que afecta a una proporción muy pequeña de los conductores analizados.

Finalmente, el nodo C_{15} incluye a todos los conductores que no han cometido ningún tipo de infracción. Este nodo se caracteriza por un porcentaje de mujeres superior al de los hombres, lo que refuerza la idea de que los hombres, en general, tienden a cometer un mayor número de infracciones al volante en comparación con los conductores del género femenino.

3.2. Análisis del mapa en función de la edad del conductor

En esta subsección, se realiza el análisis conjunto de los conductores en relación a los dos mapas creados anteriormente (Figura 1y 2) y con respecto a las diferentes franjas de edad a las que pertenecen los conductores.

En la Figura 3 se muestra la distribución en el mapa de los conductores en función de la franja de edad a la que estos pertenecen.

En primer lugar, se puede observar que las diferencias en proporciones por grupos de edad en relación a las infracciones o el mal estado de los conductores no son tan significativas como las observadas por género.

El análisis conjunto de las Figuras 1 y 3 muestra que los conductores mayores de 75 años cometen un mayor número de infracciones de conductor seguido de los conductores más jóvenes, mientras que los conductores cuya edad oscila entre 30 y 54 años, son los menos representados en este tipo de infracción (C_{41}). No obstante, análisis estadísticos adicionales son requeridos para evaluar si estas diferencias son estadísticamente significativas. Además, en el nodo C_{15}, que es el que incluye a todos los conductores que no han cometido ningún tipo de infracción, puede apreciarse que el porcentaje de conductores mayores de 75 años es notablemente más bajo que el de otras franjas de edad. En este nodo puede verse también que los conductores entre 30 y 54 años son los que en mayor proporción se encuentran aquí representados, lo cual está en consonancia con lo indicado anteriormente.

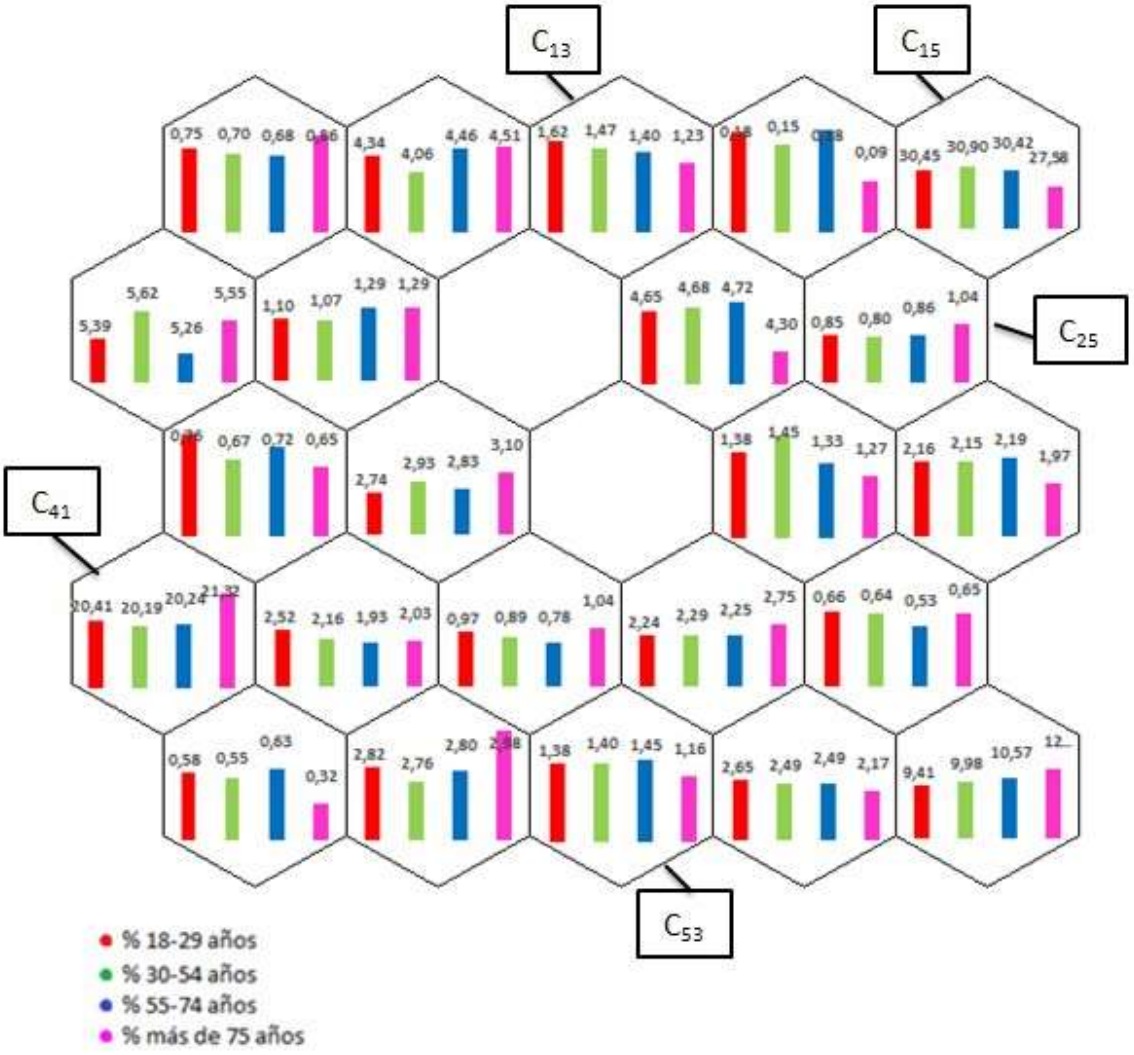


Figura 3: Distribución de los conductores en el mapa por franjas de edad

Por otro lado, en general los conductores más mayores son los que presentan un menor consumo de alcohol y/o drogas, principalmente cuando esta infracción está combinada con la comisión de infracciones de velocidad (C_{53}), donde, como se ha indicado anteriormente, la presencia de conductores masculinos es notablemente superior a la de conductores del género femenino. Se requieren análisis estadísticos adicionales para evaluar si las diferencias son estadísticamente significativas.

Adicionalmente, los conductores más jóvenes cometen un mayor número de infracciones de velocidad con respecto a los conductores del resto de franjas de edad (C_{13}). Esto está en consonancia con los resultados obtenidos de trabajos previos, donde los conductores más jóvenes, particularmente hombres, buscan la emoción y tienen menos experiencia (2, 3, 8-11). Además, en el nodo C_{13} se puede observar que, a medida que aumenta la edad del conductor, las infracciones de velocidad tienden a disminuir.

Finalmente, cuando únicamente el defecto físico está presente, hay una mayor proporción de

conductores mayores, dado que a medida que aumenta la edad de los conductores, estos son más propensos a sufrir algún tipo de deterioro físico (C_{25}).

4. CONCLUSIONES

La metodología de mapas auto-organizados (SOM) de conglomerados, aplicada a los datos de conductores implicados en accidentes de tráfico, permite realizar un análisis descriptivo multivariante de los conductores con el objetivo de identificar patrones de comportamiento de los mismos que nos ayuden a entender mejor un fenómeno tan complejo como es el caso de los accidentes de tráfico. En particular, en esta investigación se ha llevado a cabo el análisis de la distribución de los conductores en el mapa auto-organizados creado a partir de las infracciones cometidas por dichos conductores, así como por el estado de los mismos. El mapa obtenido fue sometido a un análisis desagregado por género y franjas de edad, lo que permite la detección de los patrones mencionados. En primer lugar debe indicarse que, como trabajos de investigación futuros, los resultados obtenidos deben ser sometidos a valoraciones estadísticas adicionales para evaluar si son estadísticamente significativas las diferencias. Por otro lado, las hipótesis de partida deben ser contrastadas para mejorar la calidad de los resultados obtenidos.

Así, se ha concluido que las diferencias identificadas por grupos de edad son menores que las identificadas por el género del conductor, especialmente entre los conductores de edades intermedias. En general, se ha observado que los hombres cometen un mayor número de infracciones en comparación con las mujeres. Sin embargo, se ha visto que las infracciones del conductor aparecen en porcentajes similares para ambos sexos. Por lo que se requerirán análisis adicionales de este tipo de infracciones para encontrar la explicación más plausible de este fenómeno. Por otro lado, los hombres tienden a cometer un mayor número de infracciones de velocidad, especialmente si estas van combinadas con la comisión de otro tipo de infracciones. A su vez, el consumo de alcohol y/o droga es mayor entre la población masculina. Adicionalmente, se ha percibido que los defectos físicos están más presentes entre el colectivo masculino, mientras que el colectivo femenino se caracteriza en mayor proporción por tener un estado del vehículo desfavorable.

El análisis realizado con las franjas de edad incorporadas muestra que los conductores más mayores y los más jóvenes cometen, en general, un mayor número de infracciones. No obstante, son muy diferentes el tipo de infracciones que ambos grupos cometen. Los más mayores cometen más infracciones del conductor y presentan mayores defectos físicos, presentando, en general, un menor consumo de alcohol y/o drogas. Sin embargo, los conductores más jóvenes cometen un mayor número de infracciones de velocidad y a medida que aumenta la edad del conductor, las infracciones de velocidad tienden a disminuir.

Estos resultados son muy interesantes porque proporcionan un enfoque multivariante a un problema extremadamente complejo como es el caso de los accidentes de tráfico.

BIBLIOGRAFÍA

(1) Al-Balbissi A., “Role of Gender in Road Accidents”, Traffic Injury Prevention; 4, 64-73, (2003).

(2) Turner C. and McClure R., “Age and gender differences in risk-taking behavior as an explanation for high incidence of motor vehicle crashes as a driver in young males”. Injury Control and Safety Promotion; 10 (3), 123-130, (2003).

(3) Jiménez J.J., Lardelli P., Luna J.D., García M., Bueno A., Gálvez R., "Efecto de la edad, el sexo y la experiencia de los conductores de 18 a 24 años sobre el riesgo de provocar colisiones entre turismos", Gaceta Sanitaria; 18 (3), 166-176, (2004).

(4) Durán M., Cantón D., Castro C., "Changing Patterns in Women's Driving", International Journal of Psychological Research; 2 (1), 54-66, (2009).

(5) Laapotti S., Keskinen E., Rajalin S., "Comparison of young male and female drivers' attitude and self-reported traffic behavior in Finland in 1978 and 2001", Journal of Safety Research; 34, 579-587, (2003).

(6) Williams A., "Teenage drivers: Patterns of risk", Journal of Safety Research; 34, 5-15, (2003).

(7) Bose D., Segui-Gomez M., Crandall J., "Vulnerability of Female Drivers Involved in Motor Vehicle Crashes: An Analysis of US Population at Risk", American Journal of Public Health; 101 (12), 2368-2373, (2011).

(8) Massie D., Campbell K., Williams A., "Traffic Accident Involvement Rates by Driver Age and Gender", Accident Analysis and Prevention; 27 (1), 73-87, (1995).

(9) Lardelli P., Luna J.D., Jiménez J.J., Bueno A., García M., Gálvez R., "Age and Sex Differences in the Risk of Causing Vehicle Collisions in Spain, 1990 to 1999", Accident Analysis and Prevention; 35, 261-272, (2003).

(10) Özkan T., Lajunen T., "What causes the differences in driving between young men and women? The effects of gender roles and sex on young drivers' driving behavior and self-assessment of skills", Transportation Research Part F; 9, 269-277, (2006).

(11) Harris C., Jenkins M., "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?", Judgment and Decision Making; 1 (1), 48-63, (2006).

(12) DeJoy D., "An Examination of Gender Differences in Traffic Accident Risk Perception", Accident Analysis and Prevention; 24 (3), 237-246, (1992).

(13) Hu W., Xiao X., Xie D., Tan T., Maybank S., "Traffic Accident Prediction Using 3-D Model-Based Vehicle Tracking", IEEE Transactions on Vehicular Technology; 53 (3), 677-694, (2004).

(14) Giacomo C., Gitelman V., Bekhor S., "Mapping patterns of pedestrian fatal accidents in Israel", Accident Analysis and Prevention; 44, 56-62, (2012).

(15) Chen Y., Zhang Y., Hu J., Yao D., "Pattern Discovering of Regional Traffic Status with Self-Organizing Maps", Presented at Proceeding of the 9th International Conference on Intelligent Transportation Systems, Toronto, 647-652, (2006).

(16) Kohonen T., "The self-organizing map", Neurocomputing; 21, 1-6, (1998).

(17) Yin H., "The Self-Organizing Maps: Background, Theories, Extensions and Applications", Computational Intelligence: A Compendium; 115, 715-762, (2008).

(18) Arenas-Ramírez B., Sanjurjo-de-No A., Mira-McWilliams J.M. & Aparicio-Izquierdo F. (2020, April 27-30). Driver's involvement and traffic offences by gender and age group in two-car collisions in Spain. Transport Research Arena 2020, Helsinki, Finland. (Conference canceled).

(19) Sanjurjo-de-No A., Arenas-Ramírez B., Mira J., Aparicio-Izquierdo F., "Driver Pattern Identification in Road Crashes in Spain", IEEE Access; 8, 182014-182025, (2020).

10 y 11 de Diciembre de 2020

Índice de riesgo para niñas, niños y adolescentes en Paraguay

Autores: José Marcelo Orrego Otazú, Anselmo Maciel, Mercedes Benítez

Facultad de Ciencias Exactas y Naturales

Universidad Nacional de Asunción

San Lorenzo-Paraguay.

marceloorrego@facen.una.py +595981739757

RESUMEN

En el mundo, más de la mitad de la población infantil están expuestos a conflictos, pobreza discriminación. Este trabajo clasifica según mayor o menor gravedad, las amenazas que enfrenta la infancia en el Paraguay y trata de describir los factores que privan a niños y niñas de la posibilidad de disfrutar de su infancia y que le impiden alcanzar su pleno potencial. Para así recomendar dónde y con qué urgencia ir poniendoatención para la promoción y protección integral de derechos.

El índice compara los distintos departamentos del Paraguay en función de un conjunto de indicadores que representan acontecimientos que tienen un profundo impacto en la vida y que representan la interrupción del disfrute de la niñez: (1) La niña o el niño muere,
(2) La niña o el niño sufre malnutrición grave, (3) La niña o el niño no va a la escuela, (4)La niña o el niño contrae matrimonio, (5) La niña o adolescente tiene una hija o un hijo,
(6) El adolescente o la adolescente muere, (7) La niña, niño o adolescente sufre violencia,
(8) La niña, niño o adolescente vive en situación de pobreza, estos indicadores coincideny crean entornos dañinos para la infancia.

Palabras Claves: Índice, riesgo, niñez, peligro.

INTRODUCCION

Este trabajo con una propuesta metodológica de una versión aproximada del índice de peligros para la niñez en el Paraguay. Se toma como referencia la publicación realizada “Save the Children, END OF CHILDHOOD REPORT 2018” donde se compara el peligro de la niñez entre 175 países, ubicándole a Paraguay en el puesto número 99, con una puntuación de 806 de un máximo de 1000 puntos, calificándolo entre los países con “algunos niños en peligros”

En ese reporte mundial, se definieron 8 indicadores asociados con: (1) tasa de mortalidad de NNA, (2) Retraso de Crecimiento (porcentaje), (3) NNA en edad no escolarizada(porcentaje), (4) NNA que trabajan, (5) NNA que se casan, (6) Partos en adolescentes, (7) Población desplazada, y (8) tasa de homicidios en NNA. En particular, para este trabajo no se disponen de datos para las variables mencionada arriba, por lo que se decidió calcular indicadores para las variables disponibles que están asociados con el peligro para la niñez. Para este reporte, se definieron y valoraron en una mesa interinstitucional ocho indicadores asociados con:

(1) La niña o el niño muere, **(2)** La niña o el niño sufre malnutrición grave, **(3)** La niña o el niño no va a la escuela, **(4)** La niña o el niño contrae matrimonio, **(5)** La niña o adolescente tiene una hija o un hijo, **(6)** El adolescente o la adolescente muere, **(7)** La niña, niño o adolescente sufre violencia, **(8)** La niña, niño o adolescente vive en situación de pobreza

METODOLOGIA

Se seleccionaron ocho indicadores con la participación de FACEN, CDIA, MINNA, DGEEC, MEC, MSPBS, UTGS, UNICEF y el Ministerio de Hacienda, porque son los que mejor representan estos peligros, están disponibles y se actualizan de forma regular. Los datos se obtuvieron de fuentes confiables, exclusivamente de organismos nacionales. Como los indicadores se miden en diferentes escalas, primeramente, se normaliza usando una técnica de escalado lineal:

$$Zn = (X - Peor) / (Mejor - Peor)$$

Dónde:

Zn es el valor normalizado

X es el valor real del índice

Observación: Lo peor es el valor más alto observado para el indicador. El mejor es el valor más bajo posible para el indicador.

Los valores normalizados oscilan entre 0 y 1, y que todos los indicadores están codificados de manera positiva (es decir, mayor las puntuaciones indican un mejor rendimiento).

Todos los indicadores se ponderan por igual (1/8). Lo que, en la práctica, reduce a la puntuación final en un promedio simple de los índices. Los puntajes fueron multiplicados por 100 y redondeado a tres dígitos para establecer el ranking. Los departamentos se clasificaron de mayor a menor por este puntaje general del índice (siendo lo cercano 100 el mejor). Para este trabajo se valida y se adecua la clasificación realizada en el trabajo original en 5 categorías o grupos.

Clasificación	Clasificación Original (Esc.1000)	Adecuación Local (Esc.100)
Relativamente pocos niños	≥ 940	≥ 95
Algunos niños	760 to 939	76 to 94
Muchos niños	600 to 759	60 to 75
La mayoría de los niños	380 to 599	38 to 59
Casi todos los niños	≤ 379	≤ 37

Tabla 1: Clasificación del ranking: ¿Cuántos niños están en peligro?

DESARROLLO

	1	2	3	4	5	6	7	8
	2018	2018	2018	2017	2018	2018	2018	2017
Departamentos	**Tasa de mortalidad, menores de 5 años por 1.000 nacidos vivos**	**Pocentaje de niños y niñas de 0 a 5 años con desnutricón crónica**	**Tasa bruta de escolarización**	**Tasa del NNA (10-17 años) que han contraido matrimonio o viven en pareja por cada 1000**	**Tasa del NNA (10-19 años) que tienen hijo/as por cada 1000**	**Tasa de defunciones adolescentes (14-17 años), por cada 10000**	**Tasa de violencia del NNA (0-18 años) sexual hacia NNA por cada 10000**	**Porcentaje Pobreza 0 a 17 años[1]**
Concepción	17,0	18,8	56,3	13,3	14,7	4,8	7,6	54,7
San Pedro	12,6	17	58,3	22,7	15,2	1,7	9,7	51,4
Cordillera	13,8	11,5	67,2	7,5	11,4	5,4	11,2	36,1
Guairá	17,4	10,3	59,6	17,5	10,9	3,1	6,0	41,3
Caaguazú	15,1	13	55,9	15,7	12,4	2,9	9,8	52,7
Caazapa	17,3	11,8	53,7	9,0	11,5	7,2	10,9	56,7
Itapua	14,9	11,5	43,7	27,4	11,0	3,8	11,5	42,9
Misiones	12,3	7,4	68,5	3,7	12,8	9,6	14,8	34,4
Paraguari	17,3	13	62,9	0,0	9,8	1,0	12,1	44,5
Alto Parana	16,0	12,7	56,6	13,5	15,8	2,9	12,0	28,2
Central	14,9	12,6	60,5	5,3	11,3	2,2	21,4	21,7
Ñeembucú	12,5	7,5	63,4	4,5	6,2	0,0	8,5	24,5
Amambay	18,0	23,6	39,7	21,7	16,0	8,2	6,4	21,4
Canindeyu	16,8	14,7	48,5	39,5	16,8	3,8	6,3	47,3
Pte. Hayes	19,6	7,7	46,6	28,9	20,6	10,4	8,9	38,7
Boquerón	27,6	8,2	40,8	46,5	27,6	4,3	23,2	26,7
Alto Paraguay	21,9	12	42,2	7,1	21,9	6,8	7,0	50,2
Asunción	14,3	9,9	101,4	39,7	10,0	1,6	21,5	20,1
Mejor	12,3	7,4	101,4	0,0	6,2	0,0	6,0	20,1
Peor	27,6	23,6	39,7	46,5	27,6	10,4	23,2	56,7

Tabla 2: Indicadores Proxi de Peligros para la Niñez según Departamento

Z(normalizado)	Zn = (X - Peor) / (Mejor - Peor)								Puntuacion
Peso	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	De 100 puntos
Concepción	0,689	0,296	0,269	0,714	0,603	0,537	0,908	0,055	51
San Pedro	0,977	0,407	0,301	0,512	0,581	0,835	0,790	0,145	57
Cordillera	0,901	0,747	0,446	0,839	0,757	0,483	0,701	0,563	68
Guairá	0,668	0,821	0,323	0,623	0,779	0,697	1,000	0,421	67
Caaguazú	0,817	0,654	0,263	0,662	0,712	0,723	0,780	0,109	59
Caazapa	0,672	0,728	0,227	0,806	0,754	0,310	0,716	0,000	53
Itapua	0,830	0,747	0,065	0,411	0,776	0,632	0,680	0,377	56
Misiones	1,000	1,000	0,467	0,921	0,692	0,073	0,493	0,609	66
Paraguari	0,673	0,654	0,376	1,000	0,833	0,899	0,649	0,333	68
Alto Parana	0,758	0,673	0,274	0,709	0,552	0,724	0,655	0,779	64
Central	0,828	0,679	0,337	0,885	0,762	0,791	0,105	0,956	67
Ñeembucú	0,989	0,994	0,384	0,904	1,000	1,000	0,854	0,880	88
Amambay	0,625	0,000	0,000	0,533	0,544	0,210	0,980	0,964	48
Canindeyu	0,703	0,549	0,143	0,152	0,504	0,629	0,982	0,257	49
Pte. Hayes	0,525	0,981	0,112	0,378	0,327	0,000	0,831	0,492	46
Boquerón	0,000	0,951	0,018	0,000	0,000	0,586	0,000	0,820	30
Alto Paraguay	0,373	0,716	0,041	0,848	0,267	0,341	0,942	0,178	46
Asunción	0,868	0,846	1,000	0,146	0,822	0,845	0,103	1,000	70

Tabla 3: Calculo normalizado y ponderado de los indicadores de peligro para la niñez

CONCLUSIONES

En general, casi todos los departamentos presentan alto índice de peligros para la niñez (por debajo de 95 puntos). Según el ranking, el departamento con peor (más bajo) el índice de peligros para la niñez es el departamento de Boquerón, donde casi todos los niños están en situación de peligro. Los demás departamentos del Chaco: Alto Paraguay y Presidente Hayes también presentan alto índice de peligros. Concepción, Canindeyú, Amambay y Central, también figuran entre los departamentos con alto índice de peligro. Sobre sale también otros departamentos de la región oriental con alto índice de peligros en la que las mayorías de los niños están en situación de peligros. Los mejores departamentos en cuanto al índice peligros, es el departamento de Ñeembucú (88) y Asunción (Capital) (70).

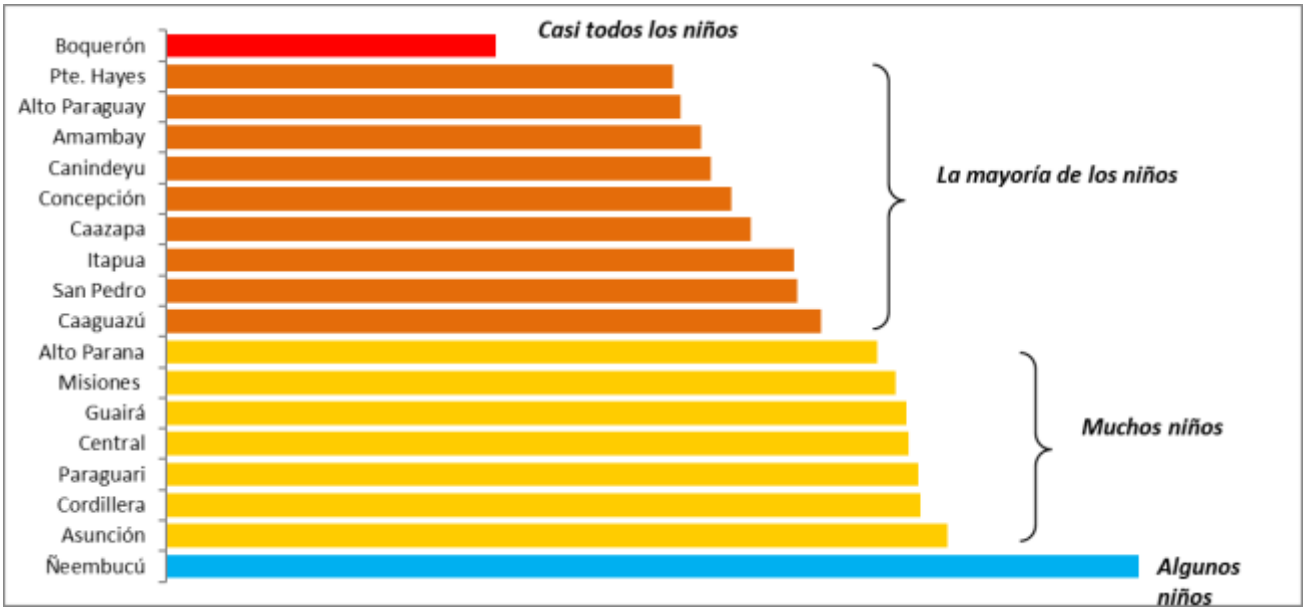


Gráfico 1: Índice de Peligros de para la Niñez, Según Departamento

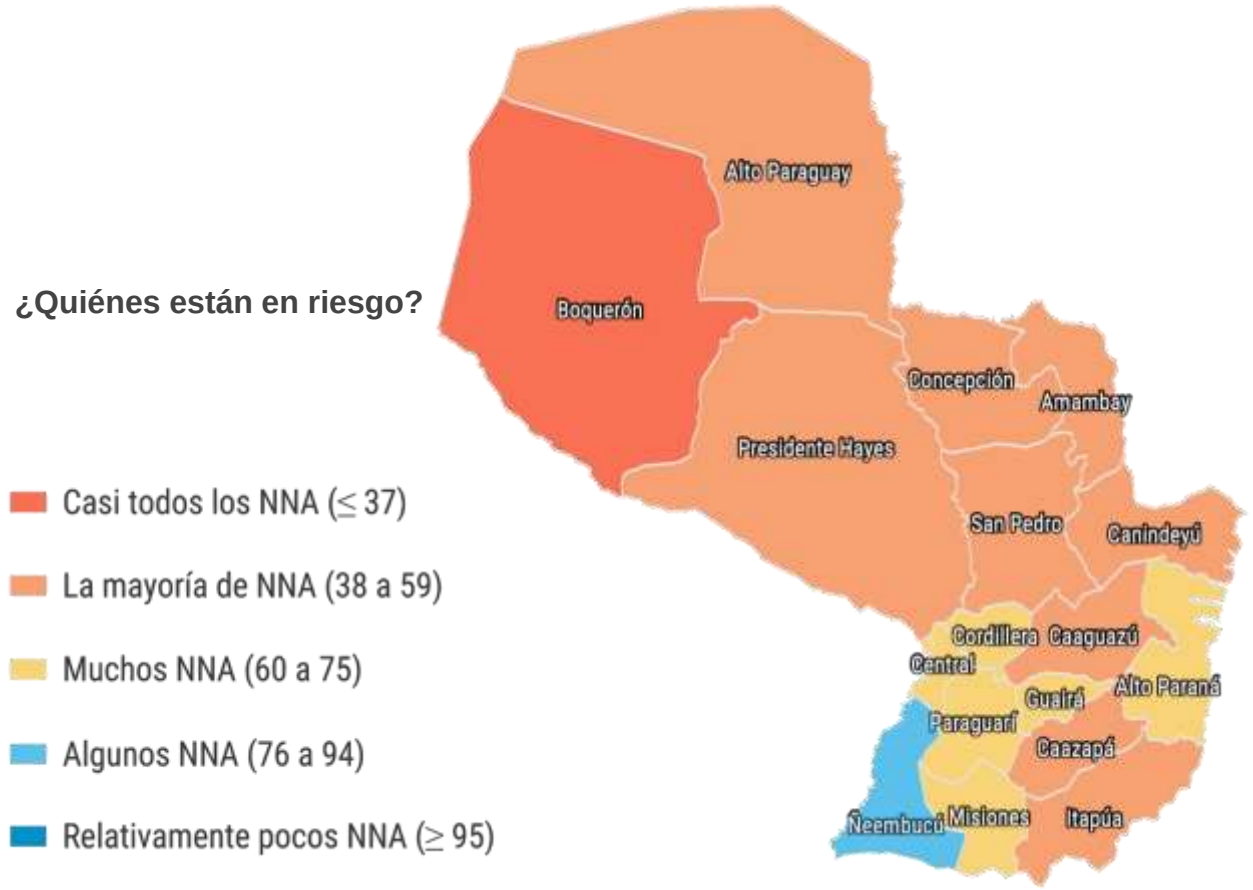


Gráfico 2: Distribución del Índice de Peligro, según departamento.

Consideraciones Finales

- ➔ Los indicadores de riesgo colaboran a identificar dónde y con qué urgencia ir poniendo atención para la promoción y protección integral de derechos.
- ➔ Nos permiten tener una aproximación de la situación de los derechos de la niñez y adolescencia a nivel local y por lo tanto, que las gobernaciones y municipios, y sus sistemas de protección integral, tomen las decisiones correspondientes.
- ➔ Pueden apoyar al monitoreo y seguimiento de la situación de la niñez y adolescencia, a nivel territorial.
- ➔ Reconocer avances e identificar obstáculos.

BIBLIOGRAFIA

“Save the Children” 501 Kings Highway East, Suite 400 Fairfield, Connecticut 06825 Estados Unidos (800) 728-3843 ISBN: 1-888393-34-3

Acrónimo	Significado
UNA	Universidad Nacional de Asunción.
FACEN	Facultad de Ciencias Exactas y Naturales.
CDIA	Coordinadora por los Derechos de la Infancia y la Adolescencia
MINNA	Ministerio de la Niñez y la Adolescencia
DGEEC	Dirección General de Estadística, Encuestas y Censos
MSPBS	Ministerio de Salud Pública y Bienestar Social
UTGS	Unidad Técnica de Gabinete Social

Web Scrapping y *Data Mining* empleados en *Google Scholar* para recuperar publicaciones científico-académicas

José Federico Medrano

Facultad de Ingeniería - Visualización y Recuperación Avanzada de Información (VRAIn)
Universidad Nacional de Jujuy
San Salvador de Jujuy, 4600

jfmedrano@fi.unju.edu.ar – 388 4845386

RESUMEN

El interés puesto en los motores de búsqueda académicos de libre acceso es cada vez mayor al ser considerados una excelente alternativa como bases de datos bibliográficas, tal es el caso de *Google Scholar* (GS), el buscador preferido entre la comunidad científica por su amplia cobertura y facilidad de uso. Si bien este motor es gratuito, no posee una API para poder recolectar las publicaciones de un investigador, primer paso en cualquier estudio bibliométrico, otro problemas que presenta GS es que no permite realizar consultas automáticas, por ello la necesidad de apelar a otros mecanismos como el web scraping y la minería de datos para sacar partido de los resultados entregados por el mismo. En este trabajo se emplean las técnicas mencionadas para recuperar las publicaciones de un investigador a partir de consultas realizadas a *Google Scholar*, para luego comprobar la correcta autoría de las mismas resolviendo así los problemas de consistencia que posee este motor. De este modo, los resultados entregados, un conjunto de publicaciones de un autor, facilitarán los análisis bibliométricos sobre este tipo de motores.

Palabras clave: web scraping, data mining, google scholar, recuperación de información, bibliometría

INTRODUCCIÓN

Las bases de datos bibliográficas actúan como repositorios de material científico-académico en

forma de publicaciones tales como artículos, actas de congresos, tesis y libros entre los tipos de publicaciones más abundantes. Las bases de datos tradicionales en este aspecto y que han dominado el mercado durante décadas han sido sin lugar a dudas *Scopus*[1] y *Web of Science*[2]. Estas bases siguen siendo el origen de datos para cualquier tipo de estudio bibliométrico a diferentes escalas, tanto para un investigador en particular, como una universidad y hasta un país entero. Ambas poseen muchísimas similitudes entre las que se destacan la cobertura sesgada a un conjunto determinado de revistas de alto impacto, a ciertos campos del conocimiento y a un cierto tipo de publicaciones, y otra de las similitudes más importantes es que no son gratuitas, es necesario abonar una suscripción a las mismas, a veces solo alcanzable por grandes corporaciones o universidades.

En este aspecto, bases de datos bibliográficas de libre acceso como *Google Scholar*[3] (GS), *Microsoft Academic*[4] (MA) y *SemanticScholar*[5] han empezado a cobrar relevancia como alternativas a las bases tradicionales, principalmente porque son gratuitas, de fácil uso y sobre todo por la amplia cobertura de contenido bibliográfico que ofrecen. Al respecto, GS ofrece contenido de todo lo que considera académico, es decir, publicaciones y materiales que provienen de universidades o instituciones académicas. GS recopila la producción científica de un investigador y la ofrece agregada en una página web, añadiendo información sobre el número de citas de cada referencia.

Este tipo de motores académicos ofrecen como resultados a las búsquedas un conjunto de publicaciones que cumplen los filtros y criterios de búsquedas aplicados, estos dependerán de las opciones que cada herramienta ofrezca, por ejemplo: *Google Scholar* permite realizar búsquedas genéricas y avanzadas, esta última permite buscar por palabras, frase exacta, indicando la existencia o ausencia de ciertas palabras, buscando las palabras dentro del artículo o solamente en el título, indicando el autor de la misma, el lugar de publicación y un rango de años. Por su parte, *Microsoft Academic* solo permite una búsqueda única pero los algoritmos de recomendación e indexación sugieren al ir escribiendo que lo que uno ingresa puede ser un campo de estudio, autor, revista, universidad, departamento, o combinaciones de estos. Por último *SemantiScholar* también posee una única búsqueda pero se puede elegir el campo del conocimiento en donde buscar, y al igual que *Microsoft Academic* los algoritmos de recomendación pueden detectar de qué trata la búsqueda: un autor, una institución, un campo o tema, o una publicación.

Si bien *Microsoft Academic* y *SemanticScholar* ofrecen una interfaz de programación de aplicaciones (API) (Hug & Brändle, 2017; Betts, Power, & Ammar, 2019) para poder recolectar las publicaciones según criterios de búsquedas específicos, no es el caso de *Google Scholar*. El gigante de las búsquedas a día de hoy no solo no posee una API para consultar sus registros, sino que tampoco permite consultas o procesos automáticos que intenten recolectar sus registros o resultados. Las razones detrás de estas limitaciones no son claras para la comunidad científica, sin embargo al ser GS tan extremadamente amplio en la cobertura de material científico-académico, los estudios sobre este motor no escasean, algunos de estos estudios comparan la cobertura de GS frente a las bases de datos tradicionales y frente a su gran competidor gratuito, Microsoft Academic (Harzing, 2019; Moed, Bar-Ilan, & Halevi, 2016; Walters, 2011; Martín-Martín, Orduna-Malea, Thelwall, & López-Cózar, 2018).

Frente a estas limitaciones, la utilización de Google Scholar como origen de datos para un estudio bibliométrico requiere el empleo de técnicas que permitan por un lado extraer los resultados para

[1] https://www2.scopus.com/
[2] http://apps.webofknowledge.com
[3] https://scholar.google.com/
[4] https://academic.microsoft.com/
[5] https://www.semanticscholar.org/

luego procesarlos y comprobarlos (Medrano, 2017). La minería de texto es una forma de extraer información de un conjunto de datos, ésta, integra la minería de contenido web, que contiene 4 formas de extracción: minería de datos no estructurada, minería de datos estructurada, minería de datos semi- estructurada, extracción de datos multimedia. En la "Minería de datos no estructurada" está la minería de páginas web, que utiliza la técnica de *web scraping* (Murillo & Saavedra, 2017). El *Web Scraping* es una técnica que consiste en la extracción de una o varias páginas web de un sitio web que estén relacionadas mediante enlaces, para su manipulación, procesar parte de su contenido y análisis posterior de los datos. Para hacer *web scraping* es necesario analizar aspectos cómo: Accesibilidad de los datos de origen, análisis de patrones de los datos, frecuencia de extracción de los datos con el objetivo de buscar la vía más óptima para obtener los datos (Borrego, 2018). Se utilizó este tipo de mecanismo de recuperación debido a que a día de hoy GS no posee una API para poder recuperar datos de forma automática o al menos semiautomática. El *crawler* construido posee un funcionamiento sencillo, realiza las consultas mediante peticiones HTTP *request* a la URL que entrega GS al momento de hacer las consultas por autor, por ejemplo:

https://scholar.google.com/scholar?as_sdt=1,5&q=autor:daniel+ autor:torres+autor:salinas&hl=es&oe=ASCII&num=20&as_vis=1

En estos casos, los de las consultas por autor, GS retorna un listado de publicaciones agrupados bajo un mismo nombre, es decir, si el autor posee un perfil creado en *Google Scholar Citations* (GSC), estas publicaciones aparecerán bajo la forma del nombre del autor (como se observa en la Figura 1), en caso que este perfil no esté creado se ofrece un listado igualmente, pero aunque la consulta sea por autor, GS retorna resultados (publicaciones) que no necesariamente pertenecen a dicho autor, en algunos casos pertenecen a otros autores que comparten parte del nombre del autor buscado, algún nombre o apellido.

Figura 1: Perfil creado en GSC

Es necesario resolver estos casos de inconsistencia para poder entregar un listado de publicaciones correcto donde se pueda asegurar que pertenecen al autor buscado para luego calcular indicadores por ejemplo, representarlos visualmente o realizar algún tipo de informe. Por todo esto, este trabajo plantea dar solución a la problemática planteada por medio del *web scraping* y minería de datos determinando la correcta autoría de una publicación en los casos que no sea posible identificar el autor en primera instancia.

METODOLOGÍA

Google Scholar no es un generador de contenido, sino que funciona como un indexador de contenido proveniente de la web, recorre el ciberespacio creando índices de contenido a recursos, para ello dispone de robots o *spiders* que rastrean sitios web en busca de material, de este modo el buscador arma su propia base de datos con contenido relacionado, pero no genera contenido por sí mismo. Este motor académico cumple el mismo propósito y sigue la misma filosofía que el motor

de propósito general *Google*, solo que circunscrito al ámbito académico.
Google Scholar al ser un motor de libre acceso y gratuito no posee muchos controles de consistencia, en su intento de indexar todo lo que puede pierde precisión, por ello es que al momento de realizar una búsqueda de un autor puede devolver registros que no pertenecen al mismo, esto representa un gran problema puesto que si no es tenido en cuenta aumentaría los indicadores de productividad de un investigador con trabajos de otro investigador, o por el contrario, si no se es capaz de recuperar los trabajos que si pertenecen a un investigador se estarían reduciendo dichos indicadores. Como lo indica (Lee, On, Kang, & Park, 2005), el problema mencionado posee dos orígenes bien diferenciados. El primero de ellos denominado *Mixed Citation* (MC) se presenta cuando dos o más autores poseen exactamente el mismo nombre, son homónimos, con lo cual los datos de las citas se fusionarán erróneamente en un único autor derivando en un análisis incorrecto de las mismas. El segundo, *Split Citation* (SC) o también más conocido como *Name matching* (Giles, Zha, & Han, 205), ocurre cuando en una misma biblioteca o base de datos, un autor posee diferentes variantes o etiquetas para su nombre cuando de hecho se refieren al mismo autor, por ejemplo, para un autor llamado "Ji-Woo K. Li", algunos posibles errores pueden ser las abreviaturas ("J. K. Li"), alternación de nombres ("Li, Ji-Woo K."), de tipo ("Ji-Woo K. Lee" o "Jee-Woo K. Lee"), contracciones ("Jiwoo K. Li"), omisiones ("Ji-Woo Li") o combinaciones de estas.
La segunda variante, muchas veces tiene que ver con en el modo en que los motores académicos indexan dicho trabajo y al respecto, *Google Scholar* es especialista en indexar nombres en múltiples formas, ejemplo de ello son las variantes de nombres encontradas para el autor español "Emilio Delgado López Cózar", las cuales se listan en la Tabla 1.

Tabla 1: Variantes de nombre de autor en GS

Variantes del nombre en Google Scholar
Delgado López-Cózar E
Delgado-López-Cózar
E Delgado López Cózar
E Delgado Lopez-Cozar
E Delgado López-Cozar
E Delgado López-Cózar
E Delgado-Lopez-Cozar
E Delgado-López-Cózar
E López-Cózar
ED Lopez-Cozar
ED Lopez-Cózar
ED López-Cózar
EL Cózar
LCE Delgado
DEDL Cózar
EDL Cózar

En cualquier base de datos de autores es necesario tener un estricto control de como los nombres son almacenados o tener la suficiente capacidad de identificar las múltiples variantes, al primera

opción es muy costosa ya que requiere un esfuerzo considerable a medida que aumenta de tamaño la base de datos, en cambio con la segunda opción se pueden apelar a mecanismos inteligentes o basados en reglas para poder identificar de forma biunívoca a un autor.
Otro problema que presenta *Google Scholar* es que en el espacio del registro de resultado destinado para los nombres de autor de una publicación (ver Figura 2) a veces no hay espacio para incluir a todos los autores, por ello son omitidos algunos de estos nombres, esto representa otro problema puesto que si decidiera aplicar un esquema estricto en donde se indicara que si el nombre de autor no aparece en la sección correspondiente el registro será descartado, esto provocaría que una publicación que efectivamente pertenece al autor no fuese tenida en cuenta.

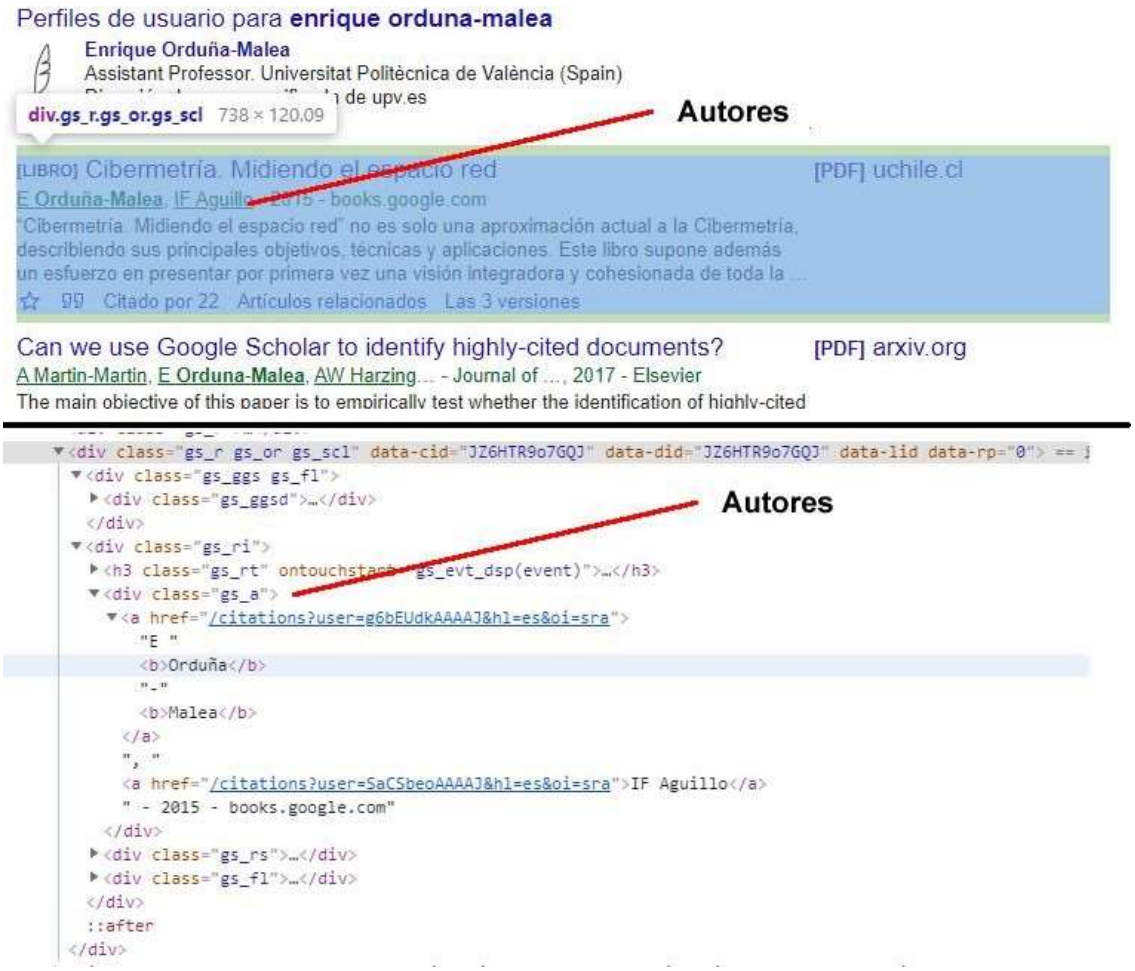


Figura 2: Registro de resultado en GS y código HTML del mismo

El *web scraping* permite identificar secciones dentro del código HTML de una página web, por ello en este trabajo se busca identificar cuál es la sección que corresponde a un registro de resultado y de este modo recuperar los autores del mismo. En la Figura 2 se observa en la parte superior la sección que corresponde a los autores de un registro, del mismo modo, en la parte inferior de la figura se observan las etiquetas HTML que delimitan el espacio para los nombres de autor, entonces teniendo en cuenta lo anterior es posible realizar una extracción de datos de *Google Scholar* a partir de consultas mediante un *crawler* creado para tal propósito. Las publicaciones se

irán almacenando al igual que los autores de la misma, para luego poder identificar si pertenecen al autor objeto de la búsqueda.
Una vez que las publicaciones han sido recolectadas, el esquema que plantea este trabajo consta de identificar todas las posibles variantes de nombres de autor a partir del nombre completo ingresado, de este modo se generan las diferentes posibilidades y se comparan con las recolectadas. Para los casos en que el nombre del autor buscado no aparezca en el lugar correspondiente, lo que se hace es una recuperación del archivo al que apunta el recurso, es decir, se navega a través de la URL al recurso. Hay casos en los que GS ofrece el texto completo del recurso (eso se logra ver cuando aparece a la derecha el tipo de archivo que trata: PDF, DOC, HTML), cuando esto no es posible, el enlace apunta a la publicación dentro de la revista donde se encuentra publicado, para acceder al texto completo es necesario poseer las credenciales necesarias y las suscripciones que correspondan. Sin embargo en este último caso, la página si ofrece información resumida del recurso: título de la publicación, nombre de los autores y resumen de la misma (ver Figura 3).

Figura 3: Publicación con restricciones de acceso

Para estos casos se recurre nuevamente al *web scraping* para recuperar la página completa, se procesa la misma eliminando las etiquetas HTML dejando solamente el texto limpio. El paso siguiente es buscar dentro del texto, en la parte que corresponde a los nombres de autor, algunas de las variantes del nombre del autor, el mismo esquema se aplica cuando es posible recuperar el documento completo de la publicación. Es necesario identificar las partes del texto ya que los nombres de autor siempre van al inicio, antes del resumen, si se buscara el nombre de autor en cualquier parte del texto se podría cometer el error de encontrar una coincidencia en una cita o en un referencia bibliográfica y esto no sería correcto. Si el nombre del autor o una de sus variantes es encontrada, el registro es marcado como perteneciente al autor, caso contrario se descarta.

DESARROLLO
Se realizaron pruebas con distintos autores, con nombres cortos, con nombres compuesto (dos nombres y dos apellidos), autores muy productivos y poco productivos. En todos los casos los resultados fueron favorables, es decir, si bien *Google Scholar* indica la cantidad de registros recuperados para un autor, este número no es del todo real por los problemas mencionados, la propuesta que aquí se presenta reduce este número (cantidad de registros) pero es un número

mucho más cercano a la realidad, los registros entregados por el esquema que aquí se presenta pueden ser empleados para realizar estudios bibliométricos, puesto que al momento de realizar la recuperación se identificaron y almacenaron las otras secciones del registro además de los autores, las cuales son: título de la publicación, cantidad de citas, lugar de publicación, año de publicación y resumen.
En las pruebas para el autor "Daniel Torres Salinas", el registro siguiente (ver Figura 4) fue descartado porque efectivamente no pertenece al autor, como se observa, los nombres del autor aparecen repartidos entre los nombres de los diferentes autores de la publicación, por un lado "Torres" y por otro "Salinas", sin embargo, como ninguna de las variantes del nombre aparecen en la sección de nombres de autor, se recuperó el registro completo (el archivo PDF) y se comprobó que la publicación en cuestión no era de su autoría.

Figura 4: Registro de GS para DT Salinas descartado

Los problemas aumentan con nombres y apellidos largos y comunes, puesto que aumenta la probabilidad de encontrar homónimos (personas con exactamente el mismo nombre), para este tipo de problemas, que no es tratado en el presente trabajo, es necesario aplicar esquemas adicionales para poder detectar y separar a dos o más personas con el mismo nombre.
En la Tabla 2, a modo de comparación, se ofrecen los resultados (cantidad de registros) entregados por *Google Scholar* al realizar una búsqueda por autor, los resultados según el esquema propuesto en este trabajo y el porcentaje de reducción en la cantidad de registros devueltos entre ambos esquemas. Dichas pruebas se realizaron para diferentes autores con distintos niveles de producción científico-académica y distintos niveles de popularidad. Se puede observar, como se había mencionado anteriormente, que la cantidad de registros de la autoría del investigador buscado se reduce entre un 4% y 10%. Esto si bien pareciera contradictorio o un despropósito no lo es, ya que el esquema propuesto realiza un preprocesamiento de los resultados en crudo entregados por *Google Scholar*, entregando un conjunto de resultados limpio y que bajo los parámetros descritos en la Metodología de este trabajo, se puede asegurar que pertenecen al autor buscado y no se incluyen resultados/publicaciones con los errores antes mencionados.

Tabla 2: Comparación entre el número de registros entregados por GS y el esquema propuesto

	GS	*Esquema propuesto*	*% de reducción*
Daniel Torres-Salinas	366	347	5,19
Emilio Delgado López Cózar	403	386	4,22
Enrique Orduña-Malea	180	173	3,89
Isidro F. Aguillo	206	193	6,31
José Luis Alonso Berrocal	159	142	10,69

Esta limpieza de resultados resulta notablemente satisfactoria al momento de analizar los niveles de producción de distintos científicos, ya que los indicadores serán calculados sobre un conjunto de datos, si bien menor, mejor aproximado al conjunto de publicaciones reales del autor buscado.

CONCLUSIONES

El esquema aquí presentado resuelve una parte del problema relacionado con la incorrecta asignación de crédito a un investigador a partir del recuento de publicaciones y citas, originado principalmente por la imposibilidad o incapacidad de identificar correctamente al autor de una publicación. La idea que se expone se vale del web scraping y la minería de datos para asignar correctamente una publicación a un autor.

El trabajo presentado es especialmente útil por dos motivos si se decide emplear *Google Scholar* como origen de datos. El primer motivo, puesto que GS no posee una API de recolección, contar con una herramienta o esquema que permita realizar estas recolecciones de manera automática o semiautomática ya es una ventaja. El segundo motivo, ahorra muchísimo trabajo al momento de comprobar la autoría de los trabajos recolectados puesto que se realiza automáticamente. En un sentido más amplio, el conjunto de datos resultante, es un conjunto de datos limpio y estructurado listo para ser utilizado para estudios más detallados.

Una tarea que quedaría a futuro sería la de emplear esquemas para resolver la identificación y detección de homónimos, una de las técnicas que se plantea emplear es el uso de redes de coautoría, de este modo se identificarían los coautores más frecuentes con los que suele trabajar el autor en cuestión, formando tantas agrupaciones como conjunto de coautores no relacionados posea el autor.

BIBLIOGRAFÍA

Betts, C., Power, J., & Ammar, W. (2019). *GrapAL: Querying Semantic Scholar's Literature Graph.* arXiv preprint arXiv:1902.05170.

Borrego, F. (2018). *Alternativas para realizar web scraping*. Recuperado el 20 de Noviembre de 2019, de http://felicianoborrego.com/alternativas-para-realizar-web-scraping/

Giles, C. L., Zha, H., & Han, H. (205). Name disambiguation in author citations using a k-way spectral clustering method. *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)* (págs. 334-343). IEEE.

Harzing, A.-W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, 1-9.

Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics, 113*(3), 1551-1571.

Lee, D., On, B. W., Kang, J., & Park, S. (2005). Effective and scalable solutions for mixed and split citation problems in digital libraries. *Proceedings of the 2nd international workshop on Information quality in information systems* (págs. 69-76). ACM.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160-1177.

Medrano, J. F. (2017). *Evaluación de la producción científica mediante motores de búsqueda académicos y de acceso libre.* Universidad de Salamanca: Tesis doctoral.

Moed, H., Bar-Ilan, J., & Halevi, G. (2016). A new methodology for comparing google scholar and

scopus. *Journal of Informetrics, 10*(2), 533-551.

Murillo, D., & Saavedra, D. (2017). Web Scraping de los Perfiles y Publicaciones de una Afiliación en Google Scholar utilizando Aplicaciones Web e implementando un Algoritmo en R. *Memorias de Congresos UTP*, (págs. 8-15).

Walters, W. (2011). Comparative recall and precision of simple and expert searches in google scholar and eight other databases. *Portal: Libraries and the Academy, 11*(4), 971-1006

.

ANÁLISIS ESTADÍSTICO SOBRE EL RENDIMIENTO DE ESTUDIANTESQUE AUTO GESTIONARON SU EVALUACIÓN VIRTUAL

Marcia Mac Gaul – Claudio Vargas – Martín Díaz

Consejo de Investigación de la Universidad Nacional de Salta, Salta, Argentina

mmacgaul@gmail.com +549 387 4419224

RESUMEN

El presente trabajo se enmarca en el Proyecto de Investigación CIUNSa N° 2497, denominado "Tecnologías de Inteligencia Artificial aplicadas a la construcción de un Motor de Aprendizaje en el campo de la Programación". Expone la metodología aplicada por una asignatura inicial de Programación, para la evaluación de acreditación basada en el momento que voluntariamente elige el estudiante para ser evaluado, durante la cursada completamente virtual en la pandemia por COVID-19.

Los resultados obtenidos corresponden al análisis de la varianza (ANOVA) aplicado a tres grupos de interés, tomados de una muestra de 54 estudiantes, reconocidos como *no voluntarios*, *voluntarios* y *rezagados* que se evalúan luego de participar en un plan de contingencia específico para ellos. Se presentan los supuestos del modelo a través de gráficos y tablas elaborados para la comprobación de normalidad, independencia yhomocedasticidad.

Analizadas las notas obtenidas en el parcial, se concluye una diferencia estadísticamente significativa entre voluntarios y no voluntarios, a favor de los primeros. El análisis en profundidad de este experimento alienta a la cátedra a extender la experiencia de evaluación auto gestionada por el estudiante, aun en contexto de presencialidad.

Palabras clave: Evaluación – ANOVA – Estadística Inferencial – COVID-19

1. INTRODUCCIÓN

El presente trabajo se enmarca en el Proyecto de Investigación N° 2497 del Consejo de Investigación de la Universidad Nacional de Salta, denominado "Tecnologías de Inteligencia Artificial aplicadas a la construcción de un Motor de Aprendizaje en el campo de la Programación", cuyo período de ejecución es 2019-2022.

La investigación en curso prevé la construcción de un motor de aprendizaje inteligenteque soporta la estrategia de enseñanza y aprendizaje de Programación en estudiantes iniciales de carreras de Informática. Se espera que este motor sostenga un modelo de Tutoría Inteligente que colabore con el docente en el seguimiento sostenido del proceso de aprendizaje y brinde la posibilidad de una enseñanza personalizada y acorde a lasnecesidades y estilos de cada estudiante.

En el contexto de la pandemia del COVID 19, la asignatura Elementos de Programación, correspondiente al primer cuatrimestre del primer año de las carreras Licenciatura en Análisis de Sistemas y Tecnicatura Universitaria en Programación, de la Universidad Nacional de Salta, se desarrolla de manera completamente virtual entre marzo y septiembre del año 2020.

La naturaleza del estudiante inicial, marcada por su dificultad para adherir a la agenda universitaria, orientó a los docentes de la cátedra, asesorados por los especialistas delproyecto de investigación, a diseñar un sistema de evaluaciones de acreditación centrado en la autogestión.

El presente trabajo expone la metodología aplicada por la cátedra, para la toma de decisiones vinculadas con la evaluación de acreditación de la asignatura, enmarcadas en el estudio de las analíticas de aprendizaje recabadas desde el proyecto de investigación. Esta metodología se fundamenta en los conceptos actuales y abarcativos sobre el tema evaluación virtual; en la planificación flexible de contenidos objeto de la evaluación adaptada a la virtualidad; en la evaluación de proceso apoyada sobre la inspección de algunas analíticas de aprendizaje y finalmente –eje central de la metodología- en el aspecto afectivo de los estudiantes, manifestados a través de su compromiso con la propia evaluación.

1. METODOLOGÍA

1.1. Objetivo

El propósito de este trabajo es exponer el análisis estadístico de los resultados del sistema de evaluación aplicado, centrado en la voluntad del estudiante por ser evaluado. En este estudio por observación o diseño no experimental, se espera asociar las relaciones entre las respuestas y las condiciones del tratamiento. Característico de este tipo de estudios es que los grupos identificables sencillamente existen en sus circunstancias particulares. Claro está que en la pandemia, dichas circunstancias no parecen ser las mismas que en la presencialidad, especialmente en lo relativo a la evaluación de acreditación. Por tanto, la motivación para esta investigación es reconocer nuevos tratamientos no convencionales y analizar estadísticamente los resultados.

1.2. Modalidad de cursado virtual

Se expone una breve descripción de la asignatura Elementos de Programación, indicando las adecuaciones efectuadas en el contexto de la pandemia. Tradicionalmente, la asignatura se ofrece bajo la modalidad *blended learning*. Integra semanalmente 10 horas de clases presenciales. Las actividades virtuales se desarrollan en el Aula Virtual (AV) montada sobre la plataforma Moodle. El AV incluye un conjunto de actividades, algunas de las cuales son obligatorias para acceder a las evaluaciones de acreditación. Las actividades desarrolladas quedan registradas en la plataforma. El estudio de los indicadores derivados de ese registro es siempre un soporte al proceso de evaluación de proceso aplicado en la cátedra.

Al inicio de la pandemia, siguiendo la organización tradicional, los estudiantes se encontraban agrupados en el AV tal como estaban separados en las comisiones de trabajos prácticos presenciales. Al poco tiempo de iniciar el cursado virtual, estos grupos se fueron redefiniendo bajo el criterio de actividad, distinguiendo al alumnado en tres segmentos: muy activos, activos y no activos. Transcurridos los primeros meses y ante la reformulación del calendario académico de la Facultad, extendiendo la fase virtual, se redefinió una nueva segmentación entre activos y rezagados. La difusión de la extensión de la fase virtual dio lugar a la reinserción de alumnos no activos. La categoría de rezagados se corresponde entonces con estudiantes que se integraron a un plan de recuperación denominado Plan de Contingencia (PC), debido a que, por diversas razones, estos estudiantes no reunían las condiciones para ser evaluados.

1.3. Muestras

La población está formada por los alumnos de Elementos de Programación de la Universidad Nacional de Salta, cursantes durante la pandemia por COVID-19 (marzo- octubre, 2020). La muestra es aleatoria, de tamaño 54. La unidad experimental es cada estudiante. Se tomó a los alumnos en orden alfabético, sin considerar turnos de clase ni consulta a los que estuvieran asignados, para evitar el efecto de la interrelación con los docentes. La situación de pandemia -de hecho- facilita mitigar ese efecto ya que los alumnos tuvieron la libertad de conectarse a diversas clases y consultas virtuales, no propia de la modalidad tradicional, en la que cada estudiante tiene acceso a una única comisión de práctica. Las observaciones son la nota obtenida en la evaluación virtual de

acreditación de la materia. La muestra posee una sola nota por cada alumno, pudiendoser esta la del parcial o su recuperación.
Los tres grupos, objeto de estudio para determinar si existen diferencias estadísticamente significativas, son:

1. GANV. Grupo Activo y No Voluntario
2. GAV. Grupo Activo y Voluntario
3. GR. Grupo Rezagado (estudiante del Plan de Contingencia)

Interesa comparar estos tres grupos. Se asume que en un cursado tradicional, todos los estudiantes son sometidos a evaluación en una fecha determinada, una vez concluido el desarrollo del contenido a evaluar. Por tanto, todos los estudiantes son del grupo GANV. Son Activos porque poseen cumplimentados los requisitos para acceder al parcial y son No Voluntarios porque no eligen su fecha de evaluación.
En el cursado en pandemia, el segundo grupo, GAV, surge como una estrategia metodológica más acorde a la virtualidad, centrada en dos elementos, el principal fue dar oportunidad a estudiantes en situaciones más desfavorables respecto al acceso ala tecnología y la conectividad y el segundo, la limitante tecnológica y de recursos humanos de la cátedra, que permitía aplicar evaluaciones de a lotes, a lo largo de un período de dos meses.
El tercer grupo. GR, surge por la implementación del Plan de Contingencia para estudiantes con baja conectividad, rezagados en las actividades y toda otra razón de índole personal, notoriamente presente en las circunstancias de la pandemia. Este Plan de Contingencia estuvo fuertemente apoyado desde la tutoría virtual, proponiendo un trabajo intensivo de consulta, práctica y desarrollo de actividades extras. En el GR no hay alumnos voluntarios porque al momento de reunir los requisitos exigidos para la evaluación, ya no se contaba con más que un breve período de evaluación previo a lafinalización del cursado.

3. DESARROLLO

3.1. La evaluación auto gestionada como estrategia

Corresponde fundamentar nuestro posicionamiento como docentes frente a un proyecto de innovación educativa, considerando la evaluación como parte central, en el marco de propuestas para revisar la enseñanza en el ámbito universitario. En general se considera la evaluación en los estudios superiores como un apéndice del proceso de enseñanza, situada desde la transmisión de conocimientos descontextualizados. El alumno solo debe demostrar en el marco de una secuencia prestablecida, los conocimientosadquiridos de manera acumulativa. En esta propuesta se repensaron otras formas de evaluar, situando la misma en un proceso integrado al complejo proceso de enseñanza. Al finalizar el desarrollo de los contenidos, se da inicio a una fase en la que los estudiantes activos podían postularse como voluntarios para ser evaluados. Los rezagados, en cambio, no podían anotarse como voluntarios. Continuaban tomando clases de apoyo, conectándose para consultas y cumpliendo con las actividades obligatorias necesarias para ingresar a las nóminas de alumnos autorizados para la evaluación.
Se habilita semanalmente, una Consulta de Moodle con los días y horarios de evaluación. Los primeros anotados en cada fecha se publican posteriormente para confirmar su inclusión en el turno de parcial elegido. Como otra adecuación al reglamento de cátedra convencional se aumenta una instancia extra de recuperación para regularizar la asignatura. Es decir, el alumno tiene que aprobar el parcial o algunade un máximo de dos recuperaciones.

3.1. Resultados

La motivación para el análisis surge de los primeros datos observados, como por ejemplo, el porcentaje de aprobados del GAV (figura 1) y la cantidad de instancias de evaluación o recuperación necesarias para regularizar la asignatura (figura 2).

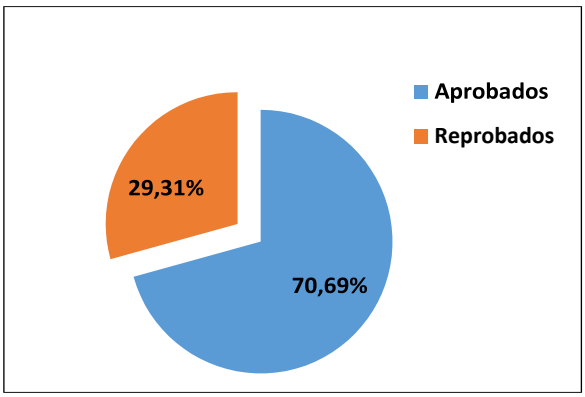


Fig.1. GAV - Porcentaje de aprobación

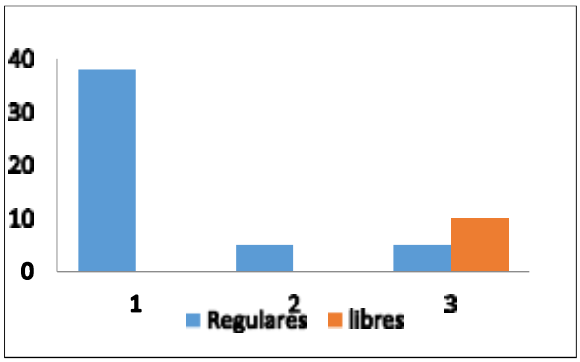


Fig.2. GAV - Cantidad de instancias necesarias para aprobar

El análisis de la varianza (ANOVA) es una técnica estadística para comparar más de dos grupos, es decir un método útil cuando la variable de estudio o variable respuestaes numérica y se desean comparar más de dos tratamientos.
ANOVA necesita satisfacer los supuestos del modelo: Normalidad, Independencia y Homocedasticidad. Para ello utilizamos procedimientos gráficos y analíticos.

Hipótesis de Normalidad: se analizan la normalidad de las notas y la normalidad de los residuos. Se aplica el contraste de Shapiro-Wilk que es adecuado cuando las muestrasson pequeñas. No todos los p-valores (Sig.) son mayores que el nivel de significación
0.05. Concluyendo que las muestras de las notas podrían no distribuir de forma normal

	Grupo	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Nota	Activos No Voluntarios	,203	18	,047	,890	18	,039
	Activos Voluntarios	,153	18	,200*	,934	18	,224
	Rezagados	,222	18	,020	,810	18	,002

*. Este es un límite inferior de la significación verdadera.

a. Corrección de la significación de Lilliefors

Tabla 1. Pruebas de normalidad

El estudio de la Normalidad de los residuos, se realiza mediante procedimientos gráficos: Histograma y Gráfico probabilístico Normal y procedimientos analíticos: Contraste de Kolmogorov-Smirnov. Para ello, se calcularon previamente los residuos.

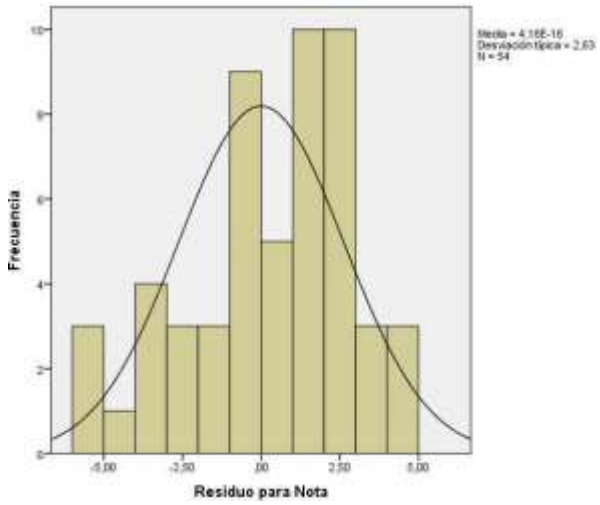


Fig.3. Histograma de Residuo para Nota

Aunque podemos observar en el histograma resultante algunas desviaciones de la normalidad, éstas no implican necesariamente la ausencia de normalidad de los residuos.

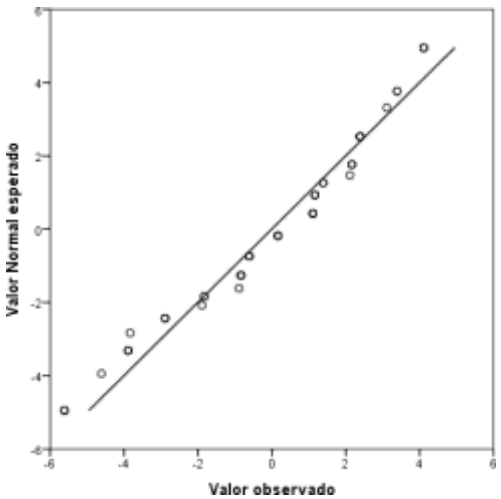


Fig. 4. Gráfico Q-Q Normal de Residuo para Nota

Podemos apreciar en el gráfico de la figura 4 que los puntos aparecen próximos a la línea diagonal. Esta gráfica no muestra una desviación marcada de la normalidad.

		Residuo para Nota
N		54
Parámetros normales[a,b]	Media	,0000
	Desviación típica	2,62986
Diferencias más extremas	Absoluta	,145
	Positiva	,076
	Negativa	-,145
Z de Kolmogorov-Smirnov		1,067
Sig. asintót. (bilateral)		,205

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Tabla 2. Prueba de Kolmogorov-Smirnov

El valor del p-valor en la tabla 2, es mayor que el nivel de significación 0.05, no rechazándose la hipótesis de normalidad.

Hipótesis de Independencia: para comprobar que se satisface el supuesto de independencia realizamos un gráfico de los residuos frente a los valores pronosticados o predichos por el modelo. Para ello, se calcularon previamente los valores predichos. El empleo de este gráfico es útil puesto que la presencia de alguna tendencia en el mismo puede ser indicio de una violación de dicha hipótesis.

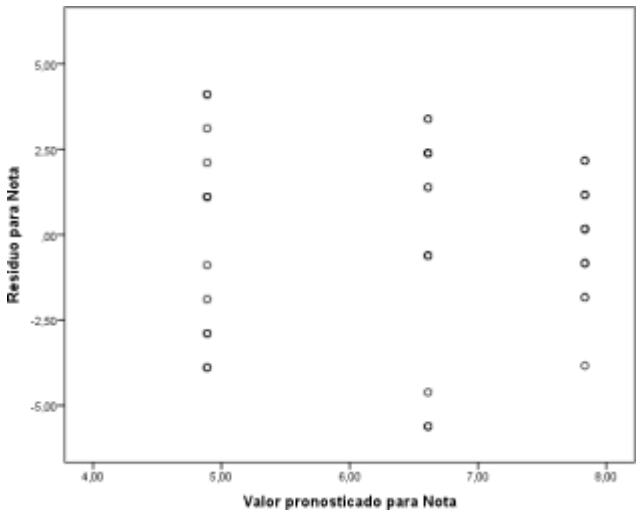


Fig. 5. Dispersión de los residuos y las predicciones

No observamos ninguna tendencia sistemática que haga sospechar del incumplimientode la suposición de independencia.

Hipótesis de Homocedasticidad: iniciamos la comprobación gráficamente.

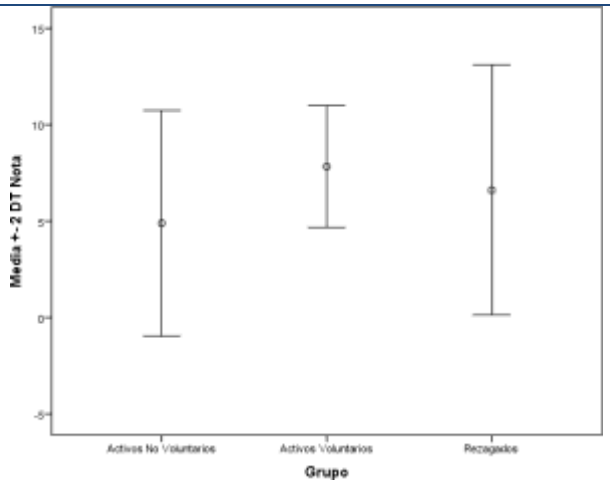


Fig. 6. Barra de error

Cada grupo tiene su promedio, identificado con el círculo en cada una de las barras y dos desviaciones típicas a la izquierda y dos desviaciones típicas a la derecha del promedio (multiplicador=2). Observamos que en GANV y en GR hay más dispersión que en GAV, aunque entre ellos son muy similares. Del gráfico no se deduce directamente si hay homogeneidad en estas varianzas, por lo que se decide analizar la heterocedasticidad analíticamente mediante el test de Levene.

Estadístico de Levene	gl1	gl2	Sig.
6,401	2	51	,003

Tabla 3. Prueba de homogeneidad de varianzas

El p-valor es 0.003 por lo tanto se rechaza la hipótesis de homogeneidad de las varianzas y se concluye que los grupos no tienen varianzas homogéneas. Debido a que la homocedasticidad no se cumple, se realizan pruebas alternativas para realizar los contrastes de Comparaciones Múltiples como contrastes Post-hoc.

ANOVA y Comparaciones Múltiples: debido a no haber verificado la hipótesis de homocedasticidad, se utilizan las pruebas que no asumen varianzas iguales.

Nota

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	78,778	2	39,389	5,480	,007
Intra-grupos	366,556	51	7,187		
Total	445,333	53			

Tabla 4. ANOVA de un factor

En la tabla 4, ANOVA, el valor del estadístico de contraste de igualdad de medias, F=5.480 deja a su derecha un p-valor de 0.007, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias. Esdecir, existen diferencias significativas en las notas obtenidas por los estudiantes entrelos tres grupos GANV, GAV y GR.

El modelo propuesto fue desarrollado y validado, comprobando la verificación de las hipótesis básicas del modelo, es decir, si las perturbaciones son variables aleatorias independientes con distribución normal de media 0 y varianza constante (homocedasticidad). Al no haberse verificado este último supuesto, se efectuaron las comparaciones múltiples adecuadas para estos casos, a través de intervalos de confianza para las diferencias entre las medias.

	(I) Grupo	(J) Grupo	Diferencia de medias (I-J)	Error típico	Sig.
Tamhane	Activos No Voluntarios	Activos Voluntarios	-2,944*	,784	,003
		Rezagados	-1,722	1,029	,279
	Activos Voluntarios	Activos No Voluntarios	2,944*	,784	,003
		Rezagados	1,222	,849	,413
	Rezagados	Activos No Voluntarios	1,722	1,029	,279
		Activos Voluntarios	-1,222	,849	,413
T3 de Dunnett	Activos No Voluntarios	Activos Voluntarios	-2,944*	,784	,003
		Rezagados	-1,722	1,029	,275
	Activos Voluntarios	Activos No Voluntarios	2,944*	,784	,003
		Rezagados	1,222	,849	,404
	Rezagados	Activos No Voluntarios	1,722	1,029	,275
		Activos Voluntarios	-1,222	,849	,404
Games-Howell	Activos No Voluntarios	Activos Voluntarios	-2,944*	,784	,002
		Rezagados	-1,722	1,029	,230
	Activos Voluntarios	Activos No Voluntarios	2,944*	,784	,002
		Rezagados	1,222	,849	,337

	Rezagados	Activos No Voluntarios	1,722	1,029	,230
		Activos Voluntarios	-1,222	,849	,337
C de Dunnett	Activos No Voluntarios	Activos Voluntarios	-2,944*	,784	
		Rezagados	-1,722	1,029	
	Activos Voluntarios	Activos No Voluntarios	2,944*	,784	
		Rezagados	1,222	,849	
	Rezagados	Activos No Voluntarios	1,722	1,029	
		Activos Voluntarios	-1,222	,849	

Tabla 5. Comparaciones múltiples

Se construyen los intervalos de confianza, con confianza del 95% que corroboran los resultados de las pruebas aplicadas. Las diferencias de medias entre tratamientos (marcadas con * en la tabla 5), corresponden a intervalos que en ningún caso contiene el cero. Por ejemplo, observando C de Dunnet, el intervalo de confianza para la diferencia de medias entre GANV y GAV es (-4,96, -,93).

Los resultados exponen las diferencias estadísticamente significativas entre los grupos GANV y GAV, señalando que los voluntarios muestran un mejor rendimiento. Esto se corrobora a través las comparaciones múltiples y de subconjuntos homogéneos. Explica también las conjeturas formuladas respecto a los alumnos voluntarios, al observar el alto porcentaje de aprobados (fig. 1), la alta tasa de aprobación sin necesidad de llegar a instancias de recuperación (fig. 2), la mayor media de notas con la menor varianza (fig. 6).

3. CONCLUSIONES

Circunstancias extraordinarias como la pandemia por COVID-19 interpelan a la sociedad. La educación en todos sus niveles debió hacer adecuaciones que movilizaron profundamente su estructura y su dinámica. En el nivel superior, presumiblemente más cercano a modalidades mixtas de educación presencial y virtual, la pandemia generó también muchos interrogantes, algunos de los cuales nos motivan a aproximar una respuesta desde este trabajo.
A la luz de documentos emanados desde organismos nacionales, como el Consejo Federal de Educación, las universidades nacionales y específicamente la Universidad Nacional de Salta, la evaluación virtual se instala en el centro del debate. Numerosos aspectos instrumentales, de logística y de tecnología apropiada dieron lugar a sistemas de evaluación que se diseñaron y aplicaron, pero que ahora interesa validar en profundidad. En esa línea de análisis de la experiencia, apoyado estadísticamente en métodos cuantitativos, el proyecto de

investigación CIUNSa 2497 concluye que, el proceso de evaluación flexible, centrada en la voluntad del alumno por ser evaluado, luego de cumplir con actividades que lo preparan para esa instancia, resulta una estrategia adecuada, cuyos resultados de rendimiento son estadísticamente superiores en comparación con los resultados obtenidos por otros alumnos que no se presentaron voluntariamente a rendir el parcial.

Desde un análisis más de tipo cualitativo, otro aspecto muy interesante de la estrategia es que el actor central de la misma es el propio estudiante, que debió realizar un proceso de construcción y análisis metacognitivo más allá de los contenidos aprendidos. Poner a este alumno inicial, frente al desafío de construir su propia agenda de trabajo, entendemos que es un paso definitivamente positivo en su filiación a la vida universitaria.

Es claro que en un estudio por observación como el aquí presentado, nos limitamos a asociar las relaciones entre las respuestas y las condiciones del tratamiento. Las circunstancias de la pandemia nos llevaron de manera específica a seleccionar los trestratamientos y a efectuar la prueba la hipótesis sobre las medias de dichos tratamientos, pero se evita extender las conclusiones a otros tratamientos no considerados. No obstante, el análisis en profundidad de este experimento alienta a la cátedra a extender la experiencia de evaluación auto gestionada por el estudiante, aun en contexto de presencialidad y más aún, en aquello que se espera sea la inminente "nueva normalidad", momento en que la educación superior fortalezca la modalidad *blended learning* mejorando específicamente el sistema de evaluación a distancia.

3. BIBLIOGRAFÍA

Hines, W y Montgomery, D. (1996). *Probabilidad y Estadística para ingeniería yAdministración*. Tercera Edición. Compañía Editorial Continental.

Kuehl, R. (2001). *Diseño de Experimentos*. 2da. Edición. Ed. Thomson.

Montgomery, D. C. (2004). *Diseño y Análisis de Experimentos*. 2da. Edición. Ed. Limusa Wiley.

Artículos en la Web:

Amat Rodrigo, J. (2016). *ANOVA análisis de varianza para comparar múltiples medias.* https://www.cienciadedatos.net/documentos/19_anova.html

Fallas, J. (2012). *Análisis de Varianza*. https://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/analisis_de_varianza_2012.pdf

Secretaría Académica, Universidad Nacional de Salta (2020). *La evaluación en entornos virtuales*. https://drive.google.com/file/d/1mqz3qqLWpEvPOSDFnzZb4Q3TWG8SWDe_/view CIN-RUEDA (2020). *Sugerencias para los exámenes finales y parciales a distancia enlas universidades nacionales en el contexto del COVID-19*. https://drive.google.com/file/d/1y2BbNZ8TDTa4gfEtB6jM3t7NCnDubKIw/view

10 y 11 de Diciembre de 2020

IMPLEMENTACION DE ANALITICAS DE APRENDIZAJE EN LEXING, UNA PLATAFORMA VIRTUAL PARA EL APRENDIZAJE DEL LÉXICO EN INGLÉS

Sosa Chasampi Cintia, Jais Carlos, Murua Javiera

Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Catamarca. Catamarca.

Argentina

cintia@exactas.unca.edu.ar

RESUMEN

Este trabajo presenta la implementación de Analíticas de Aprendizaje en el diseño del software LEXING, diseñado especialmente para el apoyo en el aprendizaje del léxico en inglés en los cursos de lectura comprensiva en el nivel superior. Este software es usado como herramienta de apoyo en las clases de inglés con fines específicos de las carreras de la Facultad de Ciencias Exactas de la Universidad Nacional de Catamarca. Su implementación ha logrado un impacto positivo en los logros de aprendizaje y en el rendimiento académico de los alumnos. LEXING permite adaptar estrategias de enseñanza y evaluación usadas en el aula de manera virtual, incentivando a los alumnos con distintas actividades y promoviendo la gestión de sus propios aprendizajes. La Incorporación de las analíticas de aprendizaje a este software, permitió registrar la actividad de los estudiantes en la utilización de la plataforma, con el propósito de analizar estos datos y poder llevar un registro de rendimiento y de esta manera poder realizar correcciones en caso de ser necesario en el proceso educativo propuesto. Se espera que esta información, presentada en forma de tablas y gráficos estadísticos, permita favorecer la toma de decisiones que mejoren la práctica docente desarrollada por los integrantes de la cátedra.

Palabras Claves: Analíticas del Aprendizaje, herramientas informáticas, léxico en inglés, educación.

INTRODUCCIÓN

En el proceso de comprensión de textos científicos en inglés intervienen no sólo las competencias lingüísticas del lector sino también las competencias cognitivas que derivan de su formación disciplinar. El desarrollo léxico en el aprendizaje de una lengua cumple un rol fundamental para poder expresarse y comprender lo que se nos presenta. Autores como Laufer (1994), Marconi (2000), Tréville (2001) y Gómez Molina (2004), definen las unidades léxicas como signos lingüísticos constituidos por elementos gramaticales, referenciales, discursivos entre otros. Diversos estudios han demostrado que los estudiantes que asisten a los cursos de inglés con Fines Específicos, carecen de un repertorio de vocabulario, lo que les dificulta la lectura de los textos académicos.

En la era digital el docente e investigador asume muy diversos roles, que pasan por serorientador, generador y evaluador de contenidos multimedia y multiformato, o ser analistade datos, encaminado a la gestión y el análisis de la información obtenida con objeto deconocer mejor su propia acción docente, la tipología de sus estudiantes y sus resultados, olas actitudes y el compromiso que con el programa formativo adquieren. (Villasol , 2019). Una de las estrategias planteadas para lograr que los alumnos de inglés con finesespecíficos, adquieran léxico para optimizar el procesamiento de información de los textoscientíficos, es la creación de un entorno virtual de aprendizaje como apoyo en los procesosde enseñanza y aprendizaje, para que el alumno se motive y tenga a su alcance unaherramienta con múltiples ventajas, que además permita recibir una retroalimentación deforma síncrona o asíncrona sobre la evolución del aprendizaje gracias al uso de lasanalíticas del aprendizaje.

Las analíticas de aprendizaje son herramientas que se incorporan a las plataformas educativas o redes sociales con el fin de registrar la actividad de los estudiantes, creando grandes bases de datos. Las analíticas de aprendizaje son definidas como un campo de investigación, con su consecuente herramental técnico para la recolección de datos.Suthers y Verbert (2013). En el que se pretende, como campo de indagación, recolectar y analizar datos sobre las acciones que realizan los estudiantes en un entorno virtual de aprendizaje, con el fin de mejorar y adaptar las propuestas educativas virtuales.

LEXING es un software diseñado especialmente para el apoyo en el aprendizaje del léxico en los cursos de lectura comprensiva en el nivel superior. Esta plataforma es usada como herramienta de asistencia en las clases de inglés con fines específicos de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca, y su implementación ha logrado un impacto positivo en los logros de aprendizaje y en el rendimiento académico de los alumnos. Asimismo, las actividades de los alumnos pueden ser monitoreadas, permitiendo de esta manera la medición, recopilación, análisis y reportede datos sobre las mismas.

En tal sentido, las analíticas de aprendizaje tienen el desafío de recoger datos que muestran la articulación entre los espacios cerrados (aula virtual) y abiertos (medios sociales), entendiendo que el aprendizaje se produce a partir del complemento de ambos entornos. Gabriela Sabulsky (2019).

METODOLOGÍA

Para el diseño del entorno virtual de aprendizaje se utilizó el modelo ADDIE[1]. Este modeloincluye cinco etapas: Análisis, Diseño, Desarrollo, Implementación y Evaluación.

[1] ADDIE: acrónimo en inglés de 5 fases de las que se compone el Sistema de Diseño Instruccional de la formación e-learningcon el objetivo de potenciar las habilidades y conocimientos de los estudiantes.

Análisis: Se analizaron las necesidades al comienzo de cualquier tarea con el fin de determinar:

- Si se requiere adaptar los conocimientos y habilidades del equipo de profesionales que integran la cátedra para aplicarlos en material diseñado para el entorno virtual de aprendizaje. Para esta etapa se realizó una capacitación al grupo docente e investigadores del proyecto, con el fin de unificar conocimientos sobre el desarrollo de laplataforma.
- Características de los alumnos: El diseño del material presentado dependerá de lascaracterísticas más importantes de los alumnos (por ejemplo, sus conocimientos y habilidades previas, el origen geográfico, el contexto de aprendizaje y el acceso a tecnología). Para ello se realizaron encuestas a estudiantes actuales y personal docentede la misma área.
- Análisis de las tareas: se identificaron las labores que los alumnos deben aprender o mejorar en su trabajo, así como los conocimientos y habilidades que requieren mayor desarrollo y refuerzo.
- El análisis de temas: se llevó a cabo para identificar y clasificar los contenidos del entorno.
- Análisis de datos en crudo: se analizaron qué datos se deben generar y cómo almacenarlos para ser utilizados en las analíticas de aprendizaje.

Diseño: En la etapa de diseño se formuló un conjunto de objetivos necesarios para lograr el objetivo general que es el diseño de una plataforma de apoyo al aprendizaje del léxico en inglés. Además de definir el orden en el cual se deben lograr los objetivos (secuencia), y seleccionar estrategias pedagógicas, de recursos, de evaluación y entrega.

- El resultado de la etapa de diseño se utilizó como referencia para llevar a cabo el entorno. Este plan de acción refleja la estructura del plan de estudios (por ejemplo, su organización en unidades, lecciones, actividades); los objetivos de aprendizaje asociados con cada unidad, los métodos y formatos pedagógicos (por ejemplo, materiales interactivos para seguir a un ritmo individual, actividades conjuntas sincrónicas y/o asincrónicas) para impartir cada unidad.
- En esta etapa además se determinaron los análisis y modelos se deben implementarsobre los datos crudos obtenidos, de manera que otorguen aplicaciones educacionales.

Desarrollo: (Producción del contenido del entorno e-learning). En esta etapa se determinó el contenido de la plataforma, formada por recursos simples (es decir, aquellos con muy poca o ninguna interactividad o multimedia, como documentos en formato PDF organizados), en combinación con otros recursos (por ejemplo, archivos de audio o video),tareas y pruebas.

El desarrollo de contenido interactivo multimedia está compuesto por tres pasos principales:

- Desarrollo de contenidos: escribir o recopilar todo el conocimiento y la información requerida.

Desarrollo del guion gráfico: integrar los métodos pedagógicos (todos los elementospedagógicos necesarios para apoyar el proceso de aprendizaje) y los elementos multimedia.

- Desarrollo de programas pedagógicos: desarrollo de componentes multimedios e interactivos; producción del curso en distintos formatos e integración de los elementos delcontenido en una plataforma de aprendizaje a la que puedan acceder los alumnos.

Implementación: En esta etapa se presenta el entorno a los alumnos. Los recursos pedagógicos se instalaron en un servidor y se encuentran a disposición de los mismos, esta etapa también incluye administrar y facilitarles las actividades.

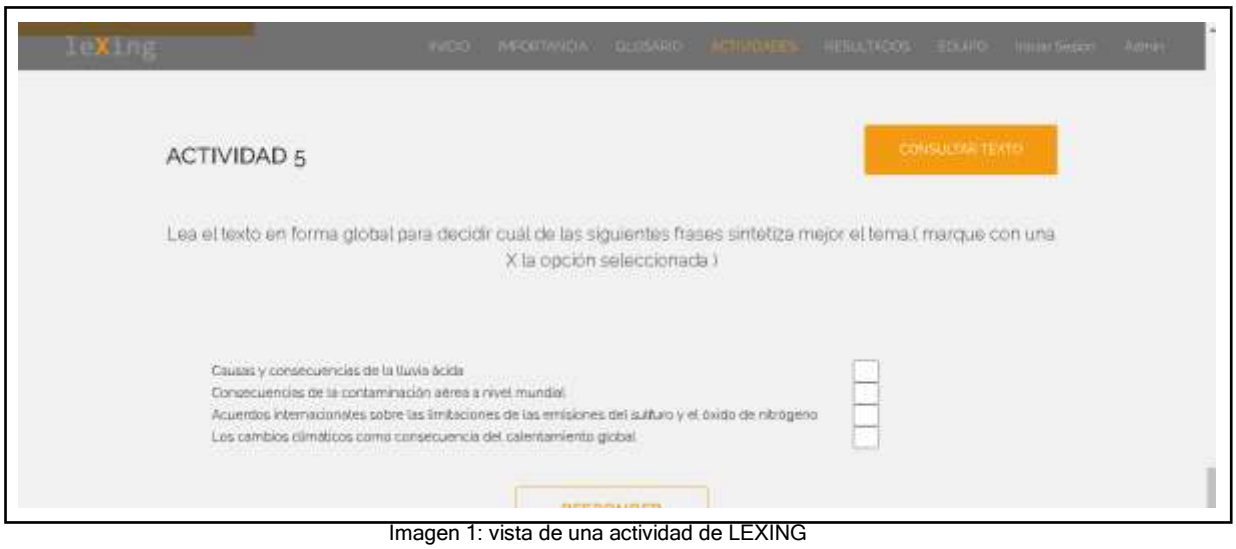


Imagen 1: vista de una actividad de LEXING

DESARROLLO

La participación en proyectos de e-learning además del conocimiento específico en inglés, requiere capacidades en ciertas áreas como habilidades tecnológicas y relacionadas con los multimedios. El proyecto cuenta con integrantes docentes pertenecientes a las cátedras específicas de inglés y docentes especialistas en el área informática, ambos correspondientesa la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Catamarca. Además, se ha procedido a la incorporación de alumnos de los últimos años con la finalidad de incorporarlos a actividades de investigación y desarrollo en las áreas específicas del proyecto
.

IMPLEMENTACION DE ANALITICAS DE APRENDIZAJE UTILIZADAS EN LEXING

Ruipérez-Valiente(2020) propone un proceso de implementación de analíticas de aprendizaje basado en cinco etapas las cuales fueron incluidas al proceso de Análisis, Diseño, Desarrollo, Implementación y Evaluación de LEXING:

1. **Entornos de aprendizaje**: El primer paso del proceso sucede en el entorno y contexto deaprendizaje y con los usuarios que intervienen en el mismo Ruipérez-Valiente (2020). Con la implementación de LEXING fue posible recolectar datos del rendimiento de los estudiantes.

2. **Recolección de datos en crudo**.
Si bien en la recolección de datos es posible recopilar todo tipo de información (sensores de audio, video, señales biométricas, etc), LEXING almacena información básica sobre el rendimiento de los alumnos de manera que permita retroalimentar el accionar de la cátedra al respecto. A medida que

Los estudiantes trabajan en LEXING, se guardan estos datos para que posteriormente podamos generar la analítica:

- Cantidad de veces que consultó el texto base.
- Tiempo en realizar las actividades.
- Respuestas Correctas / Incorrectas.
- Cantidad de Intentos.
- Participación en foros de consulta.
- Otros

3. **Manipulación de datos e ingeniería de características**:
En esta etapa se procesa la información almacenada en crudo para poder transformarla en datos que servirán para la construcción, tanto de gráficas como de reportes y sobre la cual se podrá tomar decisiones. Por ejemplo, en la resolución de trabajos prácticos se guardará el tiempo que demoró y el resultado del estudiante en cada actividad, pero esta información no será muy útil hasta que algorítmicamente se calculemos el tiempo total que demoró en terminar el trabajo completo y el rendimiento de todos los demás alumnos y de esta manera reportar cual fue la actividad que más les costó a los estudiantes.

4. Análisis y modelos

El análisis y modelado que se realice sobre los datos recopilados, son clave para su comprensión. Existen algoritmos de aprendizaje supervisado que son capaces de modelar el futuro en función de conjuntos de datos históricos que se suelen usar en el área de analítica de aprendizaje. Sin embargo, en esta instancia se puede aplicar cualquier algoritmo que permita obtener los resultados deseados. Cabe destacar que “el proceso de manipulación de datos e ingeniería de características, en conjunto con esta fase de análisis y modelado, es un proceso iterativo que se puede repetir hasta que se alcancenlos resultados deseados”.Ruipérez-Valiente (2020).

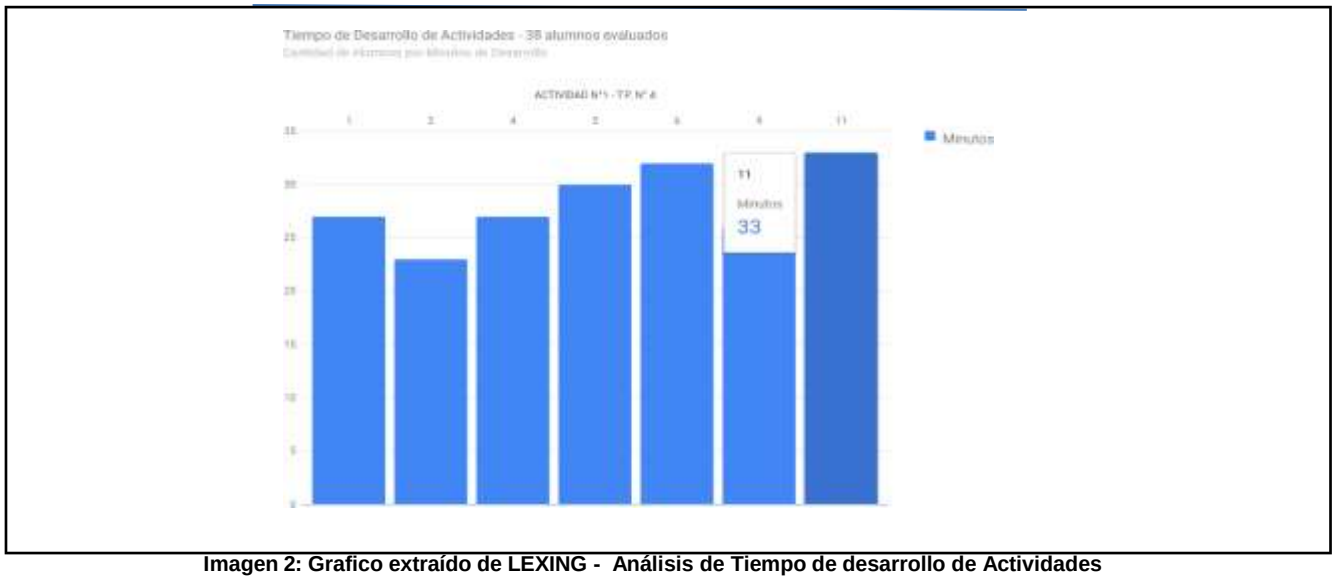


Imagen 2: Grafico extraído de LEXING - Análisis de Tiempo de desarrollo de Actividades

5. **Aplicación educacional**.

Estas aplicaciones deberían generar una retroalimentación, en los contextos educativos donde se generaron los datos, que permita mejorar el proceso de aprendizaje, y con ello se cerraría el ciclo de la implementación de analíticas de aprendizaje. Además, el efecto de estos cambios debería ser evaluado, a través de una metodología de analítica de aprendizaje. Esta evaluación es imprescindible para ser capaces de medir el impacto de los cambios introducidos en el contexto educativo Ruipérez-Valiente (2020).

CONCLUSIONES

La necesidad de analizar las interacciones de los estudiantes en entornos virtuales de aprendizaje hizo posible en los últimos tiempos el crecimiento de la analítica de aprendizaje en el área de investigación educacional. Esto motivó a la implementación de las mismas en el software LEXING desarrollado comoherramienta de apoyo al aprendizaje del léxico.

Dados los múltiples beneficios que esta herramienta aportó en el rendimiento de los alumnos en el áreade inglés, se suman los beneficios que aportan las analíticas de aprendizaje, facilitando la toma de decisiones educativas y también de carácter administrativo.

Otro objetivo pendiente será la re-educación de los usuarios, para que, en el caso de los docentes, aprendan a incorporarlas en sus prácticas pedagógicas y en el caso de los alumnos, las adopten como parte de las herramientas para soportar la auto-regulación de su aprendizaje.

BIBLIOGRAFÍA

- Beatriz Fainholc (2008). “Cómo Las Tics Podrían Colaborar En La Innovación Socio-Tecnológico-Educativa En La Formación Superior Y Universitaria” . Revista Iberoamericana deEducación a Distancia.
- Covadonga de la Iglesia Villasol (2019).”Learning Analytics and Education: clasificación, descripción y predicción del aprendizaje de los estudiantes”. Revista iberoamericana deEducación. ISSN: 1022-6508
- Gabriela Sabulsky (2019). “Analíticas de Aprendizaje para mejorar el aprendizaje y la comunicación a través de entornos virtuales”. Revista iberoamericana de Educación. ISSN:1022-6508
- GÓMEZ MOLINA, J. R. (2000), “La competencia léxica en la enseñanza-aprendizaje deespañol como L2 y Le”, Mosaico, 5: 23-29.
- LAUFER, B. (1994), “Apropiation du vocabulaire: mots faciles, mots dificiles, mots impossibles”,L´acquisition du lexique d´une langue étrangère, AILE (Acquisition et Interaction en Langue Etrangère), 3: 97-113. París, Association Encrages.
- Leydier Argüelles (2006).” Concepción y diseño de sistemas e-learning. Visión desde una plataforma para la enseñanza de idiomas”. Revista de Universidad y Sociedad del Conocimiento.Tercera Edición.
- Long, P. & Siemens, G. (2011). “Penetrating the Fog: Analytics in Learning and Education”.Educause Review., pp. 31-40.

.

- Ruipérez-Valiente, J. A. (2020). “El Proceso de Implementación de Analíticas de Aprendizaje”. *RIED. Revista Iberoamericana de Educación a Distancia, 23*(2), pp. 88-101.doi: http://dx.doi.org/10.5944/ried.23.1.26283
- Suthers, D., & Verbert, K. (2013). “Learning analytics as a middle space. In Proceedings of the Third”.International Conference on Learning Analytics and Knowledge (pp. 1-4). ACM. Recuperado de https://bit.ly/2UCEYHY

- TREVILLE, M.C. (2001), “Le développement du vocabulaire en L2: point de vue pédagogique”.En C. Cornaire y P. M. Raymond, Regards sur la didactique des langues secondes. Québec: Les Éditions Logiques, pp. 271-294.

10 y 11 de Diciembre de 2020

Violencia contra las Mujeres en tiempos de aislamiento social

Rodriguez, Mariela [1]; Laureano, Nazarena [2]; Soria, Micaela [3]; Castro Norma[4]; Vargas, Gerardo Leon[5]; Farfan, José Humberto[6].

Institución: Facultad de Ingenieria - UNJu [1, 5, 6] M, Facultad de Humanidades y Ciencias Sociales- UNJu [2], Universidad Catolica de Santiago del Estero[3] , Universidad Catolica de Salta [4]

Datos de contacto: mariela.rodriguez@fi.unju.edu.ar

RESUMEN

La Violencia contra las Mujeres es una problemática social compleja que impacta en todos los sectores de la sociedad. Se trata de una violencia que afecta a las mujeres por el mero hecho de serlo. Constituye un atentado contra la integridad, la dignidad y la libertad de las mujeres, independientemente del ámbito en el que se produzca. La víctima transita por una serie de daños a su integridad que van desde la violencia psicológica, físicas hasta el punto de ocasionar la muerte.

Ante esta situación, es imprescindible que los organismos encargados de trabajar en esta problemática, puedan contar con información confiable que permita diseñar estrategias de prevención y abordaje intersectorial, que contribuyan a la erradicación de todo tipo de violencia.

A partir de este trabajo se procura analizar y describir los principales atributos que caracterizaron la Violencia contra la Mujer y comparar los datos obtenidos a partir de la los llamados a la línea de emergencia 911 y la denuncias efectuadas en comisarías, en tiempo de confinamiento en la Provincia de Jujuy en los años 2019 y 2020.

Para realizar el estudio se aplicó las técnicas de minería de texto, reglas de clasificación, que permitieron generar información relevante para los especialistas en Violencia contra la Mujer.

Palabras Claves: violencia contra las mujeres, minería de datos, minería de textos, confinamiento social.

1. INTRODUCCIÓN

Según la Ley Nº 26.485 de Protección Integral a las Mujeres, se entiende por violencia contra las mujeres toda conducta, acción u omisión, que de manera directa o indirecta, tanto en el ámbito público como en el privado, basada en una relación desigual de poder, afecte su vida, libertad, dignidad, integridad física, psicológica, sexual, económica o patrimonial, como así también su seguridad personal [1].

La ley además puntualiza que, se considera violencia indirecta, a los efectos de la presente ley, toda conducta, acción, omisión, disposición, criterio o práctica discriminatoria que ponga a la mujer en desventaja con respecto al varón. La violencia de género puede adoptar diversas formas, lo que clasifica al delito, de acuerdo con el tipo: violencia física, psicológica, sexual, económica - patrimonial y simbólica; y según la modalidad que involucra: violencia doméstica, institucional, laboral, contra la libertad reproductiva, obstétrica y/o mediática. [2]

Según la ley, lo que diferencia a este tipo de violencia de otras formas de agresión y coerción, es que el factor de riesgo o de vulnerabilidad, acontece por el solo hecho de ser mujer.

Para la ONU “el aumento de la violencia interpersonal en tiempos de crisis es un hecho bien documentado”, por lo que la violencia contra la mujer en tiempos de aislamiento social en la provincia de Jujuy, no es una situación desvinculada de ello. Las medidas de confinamiento en casa llevan a la "tormenta perfecta" de puertas adentro, afirma la Sra. Mlambo-Ngcuka, ya que exacerba las tensiones acerca de la seguridad, la salud y el dinero.

En el mes de abril de este año, el Secretario General de la ONU, António Guterres hizo un llamamiento a la paz en los hogares de todo el mundo, e instó a todos los Gobiernos a incluir la prevención y la reparación de los casos de violencia contra las mujeres en sus planes nacionales de respuesta contra el COVID-19. Más de 140 gobiernos han apoyado su llamamiento. [3]

Por lo que trabajar en la eliminación de la violencia en todos los órdenes de la vida y desarrollar políticas públicas de carácter interinstitucional, es una tarea que compromete a todos los sectores, por lo tanto, es necesario que los organismos competentes, especialmente los vinculados a garantizar la seguridad personal, contribuyan a: estudiar la complejidad de esta problemática, suministrar información confiable y relevante a la sociedad y a los organismos estatales competentes y contribuir a la definición políticas públicas intersectoriales de largo plazo enfocadas hacia la sensibilización, prevención, sanción y erradicación de esta problemática.

El análisis inteligente de datos estadísticos permitirá construir un diagnóstico preliminar sobre cómo esta problemática se presenta en la Provincia de Jujuy.

La minería de datos es la ciencia y tecnología de explorar datos en orden para descubrir patrones desconocidos, es una parte general del proceso de descubrimiento del conocimiento. La minería de datos cuenta con una rama muy extendida para el análisis de textos que es la minería de textos permite trabajar con base de datos no estructuradas. Hoy en día se tiene masivas cantidades de información respecto a delitos que suceden en el ámbito provincial que pueden dar patrones que ayuden a diseñar trabajos operativos estratégicos para el área de seguridad. [4].

2. METODOLOGÍA

2.1 Datos

Como se menciona precedentemente, el DataSet del presente trabajo se obtiene de las llamadas que se registraron en la línea de emergencia 911 de la provincia de Jujuy – Argentina en el periodo de enero a octubre de los años 2019 y 2020, se cuenta con 10.595 registros en estos periodos, cabe aclarar que se preservaron los datos de identidad de personas.

La provincia cuenta con una proyección poblacional de 762.440 habitantes para el año 2019 según información del INDEC [5]. La población afectada, en el periodo de análisis, resulta el 1,38% del total, sin embargo, existen casos que se pueden haber registrado más de dos veces las llamadas en el mismo o distinto año.

Las variables con las que se cuenta son: Día y Hora en que ocurre el hecho, Modalidad de Violencia, el Tipo de Violencia, Factores de riesgo, Estados de la víctima o agresor, Protagonistas del hecho,, Reacciones que generan los hechos, Barrios con mayor incidencia, Lugar del Hecho y Resultados de la intervención del 911.

2.2 Metodología utilizada

La minería de datos es "un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos" [6], se pretende aplicar técnicas de la misma en la problemática de Violencia contra las Mujeres. De los relatos que se obtienen de las llamadas telefónicas al Centro de Emergencia obtiene una gran cantidad de información oculta, de importancia estratégica. El descubrimiento de dicha información resulta posible gracias a la minería de datos (Data Mining), pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) es el que se encarga de la preparación de los datos y de la interpretación de los resultados obtenidos, los cuales darán un significado a los patrones encontrados [7], cabe destacar que KDD es producto del rápido desarrollo de la minería de datos y la gran aplicación de tecnologías de información y bases de datos, formulan un proceso basado en la extracción de conocimiento (KDD), definen [8] una secuencia iterativa de cuatro pasos: la definición del problema, el pre - procesamiento de datos (que incluye la preparación de datos), data mining, y el post data mining.

La fuente de datos contiene variables parametrizadas y una gran parte y una variable en particular a estudiar que es el "relato" para el análisis de esta variable se utilizarán técnicas de minería de texto.

Finalmente se pretende como resultado final de este proyecto de investigación generar un modelo de software, que se desarrolle utilizando metodologías ágiles (tales como Scrum o similares) y trabaje bajo ambientes de desarrollo integrado (IDE), cuyo fin es mostrar información tabulada y gráfica que permita un eficiente análisis de una red de este tipo por parte de los usuarios que hacen uso de la misma, principalmente en el área de seguridad gubernamental.

2.3 Limpieza de Datos

El proceso de limpieza se realiza por etapas, en principio se procede a la limpieza de los datos que utilizarán técnicas de minería de datos. Se inicia este proceso con la limpieza de la variable "barrios" donde se clasificó los barrios en general para unificar criterios de las zonas calientes. Se siguió por las fechas y horas del hecho en el que se discretizo los valores para obtener la cantidad de hechos por día y por hora.

La limpieza del cuerpo de textos (relatos) aptos para ser parte del entrenamiento de minería de texto. El proceso inicia con una etapa de Tokenización donde los diferentes textos son separados en palabras individuales. A partir de aquí, se emplea Lematización, que consiste en llevar cada palabra a su lema o forma de diccionario. Este paso es particularmente útil debido a que colabora en reducir palabras que parecen distintas para una máquina pero en esencia representan lo mismo como es el caso de los verbos conjugados. Otro proceso es el de remover palabras con errores de

ortografía, este procedimiento puede tener efectos negativos porque son datos que son eliminados, su aplicación debe ser considerada dado el contexto del proyecto, calidad de los datos de origen y viabilidad aplicación de un proceso manual para corregir y no eliminar [9].

Para concluir, se eliminan signos de puntuación y Palabras Vacías (Stop Words), estas últimas representan palabras que no aportan valor para su análisis, por ejemplo, preposiciones, artículos, etc.

3. DESARROLLO

3.1 Visualización de datos

El inicio de esta investigación requiere de la visualización de la información que se tiene en la fuente de datos. El Centro de Monitoreo y Sistema de llamados de Emergencia 911 es un área que funciona dentro de la órbita del Ministerio de Seguridad, en la cual se atienden los llamados de emergencia por diversas situaciones vinculadas a la seguridad, teniendo una cobertura en los departamentos Dr. Manuel Belgrano y Palpalá. A continuación se visualizará la información que se releva al momento que el ciudadano realiza una llamada.

3.1.1 Comparativo general 2019 - 2020

Durante el periodo comprendido entre enero a octubre de 2019 y 2020, el 911 recibió un total de **10.595** llamadas por situaciones de Violencia contra la Mujer. Siendo 6.061 en el año 2020 y 4.534 en el año 2019, por lo que evidencia un incremento del 14,41% respecto al año anterior (Figura 1).

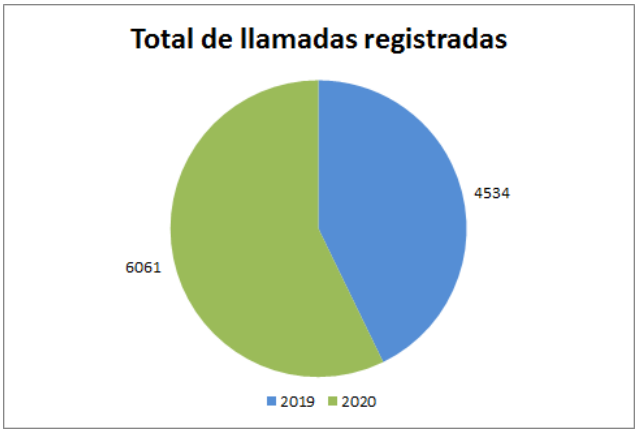


Figura 1: Llamadas registradas en el 911

En el análisis de datos detallado por mes se manifiesta que julio, agosto y septiembre son los meses en donde el número de llamadas se redujo significativamente en relación al año anterior (Figura 2), periodo que coincide con el inicio de la flexibilización de la cuarentena en localidades consideradas como zonas verdes y en los que las víctimas de violencia podían acercarse a radicar la denuncia en las comisarías.

Un dato curioso se produce en el mes de mayo, donde se evidencia un incremento significativo de llamados respecto al mismo mes del año anterior. Este dato coincide con el aumento de las flexibilizaciones en tiempo de pandemia, sin embargo, el hecho de que las comisarías recibían denuncia solo por casos de emergencia (como violencia de género) puede haber significado una falta de difusión de este tipo de información y/o puede estar vinculado con el hecho de que la mujer tuvo intenciones de efectuar la denuncia, pero las restricciones vigentes, terminaban desalentando

esta posibilidad. El tiempo de convivencia con el agresor significaba a su vez la reproducción constante del ciclo de la violencia, donde al vivir la fase de "explosión" de la violencia, en la que la víctima, por lo general (en períodos de circulación normal) recurre a radicar una denuncia, queda desalentada de hacerla por las restricciones de circulación aún vigente, de esta manera se ve involucrada en la continuidad de la fase de "arrepentimiento" donde la mujer suele auto-convencerse del "arrepentimiento" del agresor y del cambio de conducta prometido, negando las fases anteriores (Ley Micaela) [10].

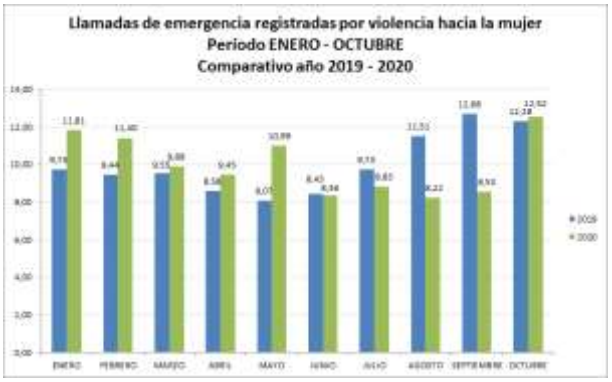


Figura 2

3.1.2 Días y horarios de ocurrencia de los hechos

Una de las variables a considerar, al momento de tomar decisiones, es la afluencia de hechos en días de la semana y los horarios de ocurrencia del hecho, el conocimiento de estos datos permiten determinar los días y horarios en que es necesario estar más alerta y programar operativos preventivos de seguridad. Se puede observar que los días de la semana en los que se produjo una mayor cantidad de llamadas por violencia, fueron los fines de semana, días en los que las personas habitualmente dedican su tiempo al esparcimiento y en el que dejan de realizar sus tareas laborales habituales, si relacionamos con el contexto de confinamiento, podemos presumir que, las situaciones de violencia no dejaron de producirse a pesar de la restricción en las actividades recreativas. La violencia contra las mujeres, no es un fenómeno que ha dejado de ser reproducido en contexto de aislamiento social en tiempos de pandemia.

Respecto a los horarios, se observa un mayor registro de incidencias a partir de horas 19 con un ascenso notable a horas 22 y un descenso progresivo durante la madrugada (Figura 4).

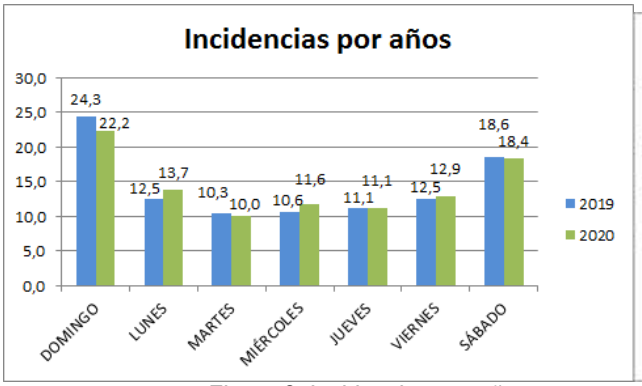


Figura 3: Incidencias por año

Figura 4: Horario de ocurrencia de los Hechos

3.1.3 Barrios con mayor incidencia de llamadas al sistema de emergencia

Se considera fundamental poder descubrir los lugares de donde provienen con mayor frecuencia los llamados de emergencia por Violencia contra la Mujer, ya que esto nos permite poder contextualizar la situación y orientar la toma de decisiones.

Del total de llamados producidos durante el año 2020, se incluye en el gráfico todos aquellos barrios que superan el 1%, de lo cual se puede visualizar que se evidencia llamados de emergencia en casi todos los barrios del Dpto. Dr. Manuel Belgrano y Palpalá (área de cobertura del sistema de emergencia), sin embargo, el pico más elevado se manifiesta en los barrios Mariano Moreno, Malvinas y Punta Diamante; seguidos por San José de Chijra, Alberdi y el Sector B6 de Alto Comedero, en menor número, le siguen, Villa Lidia, Cuyaya, San Fco. De Álava y Luján, coincidentemente estos son barrios caracterizados por un alto índice delictivo (C.I.A.C., 2020) [11] y altos niveles de vulnerabilidad social, con lo cual se puede concluir que la mayor ocurrencia de Violencia contra la Mujer se manifiesta más en comunidades de mayor vulnerabilidad social (exclusión y marginalidad).

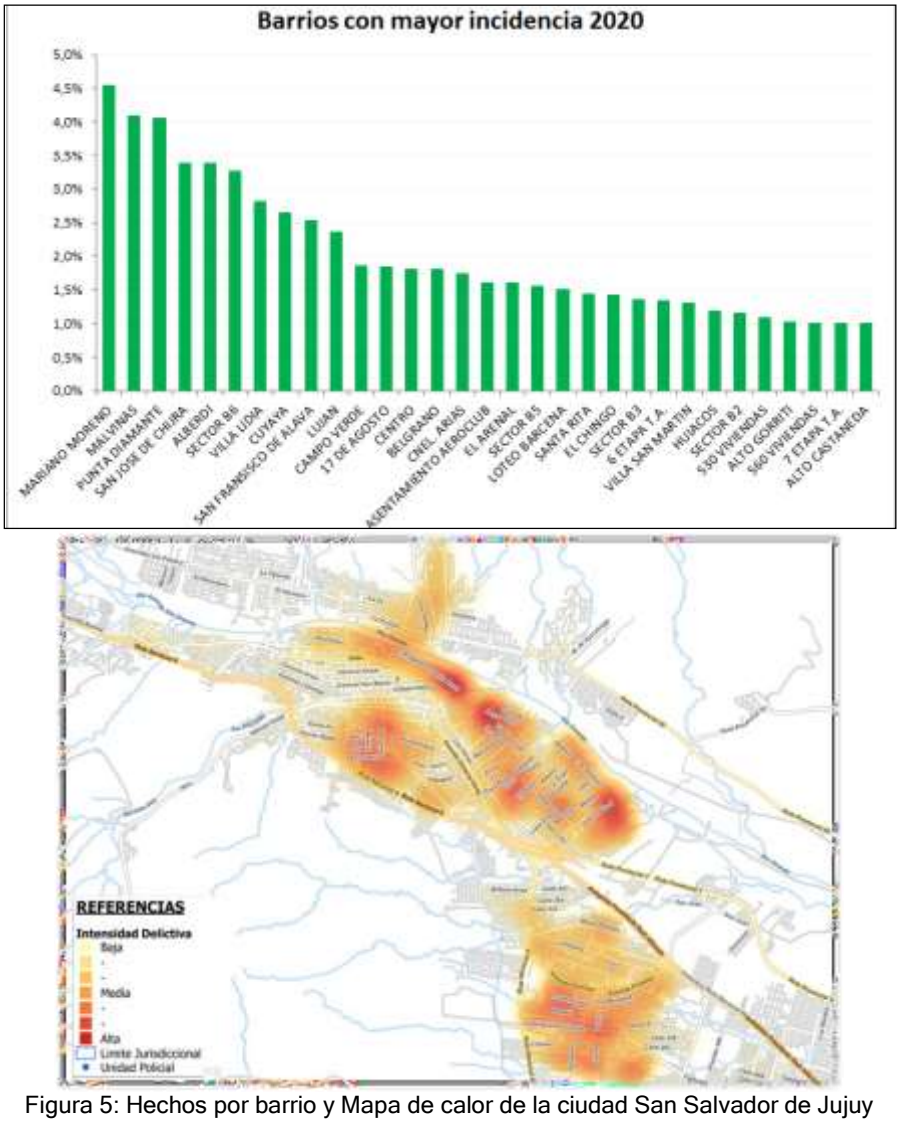


Figura 5: Hechos por barrio y Mapa de calor de la ciudad San Salvador de Jujuy

3.1.4 Resultado de las intervenciones del 911

Al analizar el resultado de las intervenciones por parte del Sistema de Emergencia 911, se evidencia que casi el 90% de las actuaciones refieren a que se realizaron Acciones Preventivas, es decir se acercaron al lugar a verificar el hecho que produjo la llamada de emergencia. Cabe aclarar que existe, en la base de datos, la dificultad para definir si la verificación del hecho tuvo resultado positivo o negativo, siendo un aspecto a mejorar en el sistema de registro en la base de datos.

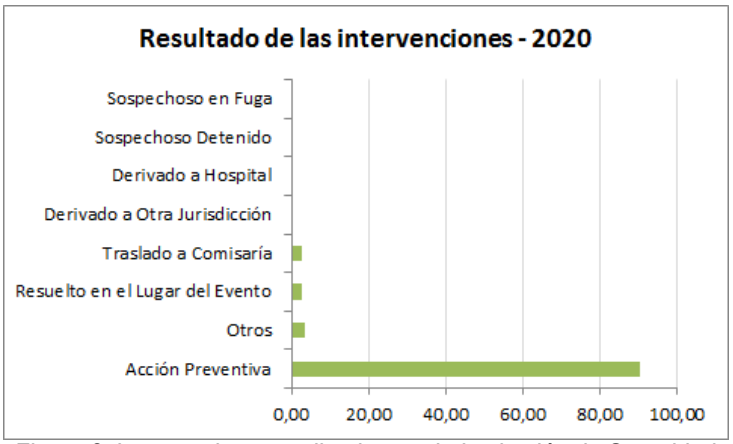


Figura 6: Intervenciones realizadas por la Institución de Seguridad

3.2 Clasificación de Textos

El modelo de clasificación se realizó de acuerdo a variables necesarias para estudiar el comportamiento de las emergencias cuando ocurre un hecho de violencia. En primera instancia se procedió a la Tokenización permitiendo dividir el texto en una serie de tokens. La configuración de token que se dio fue la división a partir de caracteres que no sean letras, que dio como resultado token que son solo una palabra y luego se construye el vector de palabras. Para unificar palabras se procedió a llevar al vector a un solo tipo de letras en este caso se utilizó el formato “lower case”, es decir utilizar letras en minúscula. Para finalizar el proceso de transformación del texto se filtró las palabras que sean “stopwords” configurado en español, que permitió excluir aquellas palabras que no son relevantes para la investigación.

3.2.1 Tipos y modalidades de violencia

En la siguiente figura las palabras relevantes que aparecen se vinculan específicamente al tipo de violencia, la modalidad de la misma y el vínculo de la víctima con el agresor. Las palabras “masculino” y “femenino” evidencian fehacientemente que la violencia es ejercida por parte del hombre hacia una mujer, es decir, basada en una relación desigual de poder.

La palabras “agredió”, “agresivo” y “físico” nos remiten al tipo de violencia física, entendida como “la que se emplea contra el cuerpo de la mujer produciendo dolor, daño o riesgo de producirlo y cualquier otra forma de maltrato agresión que afecte su integridad física” (Ley 26.485) [1]. Este tipo de violencia apareció en 5.177 datos, seguida por la violencia psicológica, con un número significativamente menor, 334 según tabla 1.

Aparece “domicilio” y “casa” como los ámbitos predominantes en donde se ejerce la violencia (la ley los presenta como “modalidades”). De esta manera la violencia doméstica es aquella por la cual se produce mayor cantidad de llamados, siendo este ejercida contra las mujeres por un integrante del grupo familiar, independientemente del espacio físico donde ésta ocurra. Según la Ley 26.485, se entiende por grupo familiar el originado en el parentesco sea por consanguinidad o por afinidad, el matrimonio, las uniones de hecho y las parejas o noviazgos. Incluyendo a las relaciones vigentes o finalizadas, no siendo requisito la convivencia. Relacionado a esta modalidad, se evidencia el vínculo de la víctima con el agresor, en la palabra “pareja” se define que es éste el principal agresor del grupo familiar.

Figura 7: wordcloud de relatos de incidencias del 911

3.2.3 Clasificación de los Tipos de Violencia

La violencia que se ejerce a la víctima está clasificado la Ley Nacional Nº 26.485 según el modo que se produce, los cuales se clasifican en Violencia Física: la que se produce sobre el cuerpo de la víctima produciendo dolor, daño y riesgo; Violencia Psicológica: la que causa daño emocional y disminución del autoestima; Violencia sexual: Cualquier acción que implique la vulneración en todas sus formas, con o sin acceso carnal; Violencia Económica y Patrimonial, Simbólica y Vicaria. De acuerdo a la clasificación de las palabras se encontró palabras que daban indicio a los tipos de violencia que ejercen los victimarios. Con mayor frecuencia en los relatos se escucha decir que hay “agresión”, la palabra debe ser analizada para la violencia física o psicológica, depende del entorno que se forma la oración, para este análisis se consideró en una sola categoría. Respecto a las palabras que propiamente alertan que ocurre violencia física son física, golpea, pelea, fuerza, pateando, matar, entre otras, todas ellas alarman a los servicios de seguridad, ya que se debe actuar con rapidez y evitar que los hechos sean fatales.

Cuando se trata de violencia psicológica se escuchan palabras como amenaza, insulta, verbalmente que llegan a un total de 334 que tiene una representación menor al 10 % de la violencia física, sin embargo la violencia física muchas veces se acompaña de la violencia verbal, formando una categoria de Violencia Fisica y Psicologica, como se pudo analizar en “Predicción de Factores de Tipos de Violencia contra las Mujeres” [12]. En el análisis de las palabras no se pudo analizar este comportamiento.

Clasificador	Palabras Genericas	Total
Violencia Fisica	agredio	3302
Violencia Fisica	fisica	1115
Violencia Fisica	golpea	352
Violencia Fisica	pelea	186
Violencia Fisica	fuerza	136
Violencia Fisica	pateando	55
Violencia Fisica	matar	14
Violencia Fisica	homicidio	7
Violencia Fisica	maltrato	5
Violencia Fisica	abuso	3
Violencia Fisica	ahogarla	1
Violencia Fisica	ahorcarla	1
Violencia Psicologica	amenaza	193
Violencia Psicologica	insulta	72
Violencia Psicologica	verbalmente	69

Tabla 1: Clasificador de Tipo de Violencia

Tipo de Violencia	TOTAL
Violencia Fisica	5177
Violencia Psicológica	334

Tabla 2: Total de palabras por Tipo de Violencia

3.2.3 Protagonistas de los Hechos de Violencia

Si consideramos a los protagonistas desde una dimensión psicosocial, ésta permite analizar aspectos concretos de los roles asignados a hombres y mujeres y sus efectos diferenciales (en la sociedad), las actividades que desarrollan, los espacios que habitan, los rasgos que los definen y el poder que detentan. [13]

En los hechos de violencia, aparecen las palabras “pareja”, “masculino”, “femenino” e “hijo/a” con mayor frecuencia. Estas permiten identificar los protagonistas del acto violento, con claridad se evidencia que la violencia contra la mujer es un acto realizado por el hombre quien desde su posición de autoridad reconocida hace abuso de poder en la relación de pareja, en figura 8 se refleja la cantidad de datos que refieren a esta situación.

La “femenino” e “hijo/a” son otros protagonistas importantes que aparece en la escena, además de revelar que el acto violento ocurre con mayor frecuencia en el ámbito doméstico, nos permite deducir que se encuentran presentes otros miembros del grupo familiar (hijos/as) que se constituyen en herederos de la violencia, quienes al vivenciar la situación, se configuran como potenciales protagonistas (salvo resignificación) reproductores de patrones violentos.

Esta reproducción se realiza a través de la transmisión transgeneracional que implica que los hijos hacen suyos los deseos, tristezas, miedos y situaciones traumáticas de sus padres, identificándose con sus necesidades y recuerdos (conscientes o no), cumpliendo sus funciones y tareas inconclusas[13].

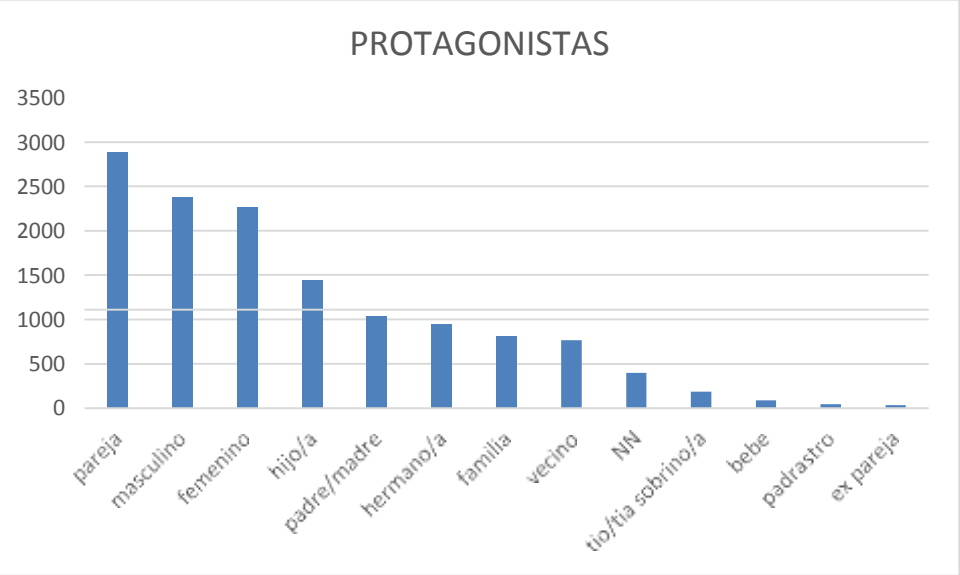


Figura 8: Protagonistas de las Incidencias

3.2.4 Lugar donde ocurren los hechos

Si analizamos con mayor detenimiento el ámbito donde se produce la violencia, observamos nuevamente que se presentan 4797 datos que revelan que es el ámbito doméstico el que prevalece. Le sigue en una cantidad significativamente menor los espacios en donde se concreta el acto violento como ser: “calle”, “ruta” “escuela”, “negocio”, espacios que puntualmente no se encuentra especificado en la ley, pero que puede encuadrarse dentro de la modalidad doméstica, por referirse a la violencia ejercida por cualquier miembro del grupo familiar. En este estudio, predomina como agresor, la pareja, quien buscaría a su víctima en cualquiera de estos lugares.

CLASIFICADOR	TOTAL
domicilio / hogar / casa	4797
calle	463
asentamiento	104
ruta	62
escuela	53
negocio	52
inquilinato	60
hospital	12

Tabla 3: Clasificador de Lugar del Hecho

3.2.5 Factores de riesgo que influyen en los hechos de violencia

Se define factor de riesgo a cualquier característica de una persona o grupo de personas, que asociado a otros, tiene una probabilidad de padecer o estar expuesto a un problema (Pita, Vila y Carpente, 1997). De allí, que es necesario investigar cuáles son los factores de riesgo para que ocurra la violencia contra la mujer, ya que el término riesgo implica la presencia de una característica o de varios que producen la violencia hacia la mujer. Al respecto la Organización Mundial de la Salud

(OMS, 2013; citado por Mejía, Ochoa, Ríos, Yaulema & Veloz, 2019) considera una serie de factores individuales, familiares y sociales que protegen o exponen a la mujer a una situación de riesgo de producir violencia. Diversos estudios asocian el consumo de alcohol a un aumento de la violencia doméstica, especialmente la sufrida por mujeres y perpetrada por sus parejas.[14]

Según los datos arrojados por la base del 911, en muchos casos (1.404 llamados) el agresor se encuentra en estado de ebriedad y en menor medida intoxicado (156) al momento de ejercer la violencia. El factor "dinero" (Tabla 4) es otro elemento que estaría estrechamente vinculado al tipo de violencia económica y patrimonial, como sostiene el informe de la ONU, el tiempo de aislamiento social exacerba las tensiones, y el dinero es un factor más que incide decisivamente en la "explosion" de hecho violento, ya que las limitadas oportunidades económicas es un factor agravante para la existencia hombres desempleados o subempleados, asociado con la perpetuación de la violencia; y es un factor de riesgo para mujeres y niñas, de abuso doméstico, matrimonios forzados, matrimonios precoces, la explotación sexual y trata. La presencia de disparidades económicas, educativas y laborales entre hombres y mujeres al interior de una relación íntima, son factores que deben tenerse en cuenta al momento de diseñar políticas preventivas.

CLASIFICADOR	TOTAL
ebriedad	1404
intoxicado	156
dinero	20

Tabla 4: Clasificador de los Factores que influyen en la violencia

Si relacionamos estas primeras aproximaciones con los datos de la incidencia de días y horarios, aparecen como hechos estrechamente vinculados al uso del tiempo libre y formas de esparcimiento de los grupos familiares. Si bien existían medidas restrictivas en eventos sociales, al interior del núcleo familiar se realizaban prácticas sociales que contribuyeron a la reproducción de la violencia contra la mujer (como las reuniones familiares y el consumo de alcohol).

3.2.5 Estados en los que se encuentra la Víctima o Agresor

Para analizar este dato, es necesario diferenciar que se encontraron estados vinculados de salud física y mental. Según la ONU existen otros factores adicionales de riesgo que se encuentran relacionados con la violencia: los deficientes niveles de salud mental relacionadas a una baja autoestima, ira, depresión, inestabilidad emocional y dependencia, rasgos de personalidad antisocial o fronteriza y aislamiento social, son algunos de ellos, se podría incluir estos factores dentro del estado de salud mental en el que se encuentra la víctima o el agresor al momento de la relación violenta. Estos aparecen concretizados en los datos que figuran en la Tabla 5, en ella encontramos diferentes estados emocionales en los que se encuentra la víctima y/o el agresor al momento de la violencia, como ser: violento, crisis, alterada, molesto; siendo "alterada" y "violento" los predominantes.

En cuanto al estado físico encontramos "embarazo", "sangrando" y "encerrada" como las más relevantes, esto nos permite evidenciar las condiciones extremas en que se da la violencia, a pesar del estado físico en que se encuentre la mujer.

CLASIFICADOR	TOTAL
alterada	132
violento	61
embarazo	59
sangrando	23
encerrada	19
herida	14
crisis	6
molestos	6
acosada	3
convulsionando	3
autolesionando	2
grave	2

Tabla 5: Estados en los que se encuentran los protagonistas

3.2.6 Reacciones que generan los hechos de violencia

Para analizar las reacciones que genera el hecho de violencia, es necesario comprender que la violencia trae aparejado un sufrimiento psicológico para la víctima por la forma en cómo se ejerce la misma (física, psicológica, económica, etc.) . En el clasificador aparecen las conductas concretas de la víctima (grita, llora, miedo, nerviosa, desmayada) y del agresor (agresivo), todas ellas vinculadas al sufrimiento emocional que genera en la víctima el acto agresivo del victimario.

CLASIFICADOR	TOTAL
agresivo	1216
grita	645
llora	339
discusión	317
disturbio	206
agresiva	165
rompio	160
miedo	16
nerviosa	15
abandono	8
desmayada	8

Tabla 6: Reacción de los protagonistas en los hechos de violencia

4. CONCLUSIONES

- Las llamadas que ingresan a un Centro de Emergencias como es el 911, requieren de una actuación inmediata por parte de los organismos de seguridad, en estos casos el hecho está ocurriendo y se necesita brindar asistencia a la víctima. En tal sentido la actuación del operador debe ser eficaz, esto hace que no se tenga un registro completo y detallado de lo

que sucede en el lugar del hecho (y la fuente rica en información se encuentra en el relato que hace la víctima), razón por la cual, con los relatos, no se podría trabajar con técnicas propiamente de minería de datos sino con técnicas de minería de texto.

- En el presente trabajo se realizó un análisis por cada palabra nombrada en el relato que permitió clasificar las alguna variables, tales como: "los protagonistas" entre ellos se encuentran las víctimas, agresor, involucrados como son la familia, niños entre otros. Por la descripción que da el que emite la llamada, se pudo conocer el "Lugar del Hecho" que tuvo mayor ocurrencia en el domicilio particular, seguido de los lugares públicos como son la calle, hospitales entre otros.
- En relación al "tipo de violencia" que ocurre se obtiene que predominó el tipo de violencia física y seguido por el tipo de violencia psicológica, la brecha que existe entre ambas se presume a que se debe que la víctima realiza el llamado cuando considera que su vida corre peligro
- Durante el periodo comprendido entre enero a octubre de 2019 y 2020, la línea de emergencia 911 recibió un total de **10.595** llamadas por situaciones de Violencia contra la Mujer. Siendo 6.061 en el año 2020 y 4.534 en el año 2019, por lo que evidencia un incremento del 14,41% respecto al año anterior.
- La violencia contra la mujer, es una problemática que se evidencia en la mayoría de los barrios de los departamentos Dr. Manuel Belgrano y Palpalá, el identificar los barrios con mayor incidencia de llamadas permite comprender que vivimos en un sistema social excluyente que se constituye en un contexto propicio para interiorizar formas de violencia en nuestras relaciones interpersonales y cómo se reproducen inequidades y abusos de poder tanto en el ámbito público como privado de una sociedad.
- La violencia en el ámbito doméstico es la modalidad predominante que provoca las llamadas de emergencia a la línea 911, de esta manera constituye el ámbito privado pasa a convertirse en un asunto público de enorme trascendencia, razón por la cual demanda la necesaria intervención de los órganos del estado para el abordaje inmediato de esta problemática.
- Las estadísticas señalan una abrumadora mayoría de hombres agresores y mujeres agredidas que siguen en relación de pareja. Una mujer "que se va" de una relación debe luchar contra los significados y consecuencias de haber desafiado la prerrogativa masculina y de ponerse a sí misma en primer lugar, situación que se complejiza de acuerdo a los recursos que brinda la sociedad y los organismos estatales.
- Es muy probable que la cantidad de denuncias sea significativamente menor en el periodo de confinamiento (este sería un estudio a futuro que se puede realizar), ya que este contexto fue propicio que la víctima desestime radicar la denuncia, aún más cuando las tensiones dentro del ámbito familiar se exacerban. Esto nos permite revelar que en muchos otros casos la dependencia económica por parte de la mujer, donde tanto ellas como sus hijos quedan desamparados.
- Las situaciones de violencia contra las mujeres, no dejaron de producirse a pesar de la restricción en las actividades recreativas en contexto de aislamiento social, preventivo y obligatorio en tiempos de pandemia, por el contrario, las medidas exsacerbaron las tensiones al interior del nucleo familiar en cuanto a seguridad, la salud y la situacion económica.

Trabajos a Futuro:

El grupo de investigación lleva adelante investigaciones relacionadas a las problemática de delitos y en particular de Violencia contra las Mujeres en la Provincia de Jujuy, con

anterioridad se lograrón predecir patrones utilizando técnicas de mineria de datos. Este trabajo inicia un camino a los estudios de la minería de textos.

- Seguir investigando las técnicas de minería de textos para encontrar la relación de textos relacionados, para encontrar patrones en los relatos de las llamadas que permitan dar intervención a distintos especialistas.

Agradecimientos:

Se brinda un agradecimiento muy especial al Centro de Monitoreo y Emergencias 911 del Ministerio de Seguridad, por brindar la información. Este informe de investigación se realizó con fines estadísticos siempre resguardando los datos sensibles que pudieran perjudicar a los protagonistas.

5. BIBLIOGRAFÍA Y REFERENCIAS

[1] "Ley 26.485. Protección Integral para Prevenir, Sancionar y Erradicar la Violencia contra las Mujeres en los Ámbitos en que Desarrollen sus Relaciones Interpersonales". 2009. Consultado en Septiembre de 2.019.

[2] Rico, N. "Violencia de género: un problema de derechos humanos". Naciones Unidas, CEPAL. Santiago, Chile. 1.996. pp. 5-9.

[3] ONU. Departamento de Comunicación Global. https://www.unwomen.org/es/digital-library/annual-report. Consultado el día 02-12-2020.

[4] Asencios,Violeta "Data Mining y el descubrimiento del conocimiento" Editor Industrial Data, vol. 7, núm. 2, 2004, pp. 83-86

[5] Instituto Nacional de Estadística y Censo, consultado 24 de diciembre de 2019. https://www.indec.gob.ar/

[6] Maimon, O., & Rokac, L. (2010). Data Mining and Knowledge Discovery Handbook. En O. Maimon, & L. Rokac, *Data Mining and Knowledge Discovery Handbook* (págs. 1-18). Nueva York: Springer.

[7] Vallejos, S. J. (2006). *Universidad Nacional del Nordeste.* Obtenido de Minería de Datos: http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf

[8] Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, *17*(5-6), 375-381.

[9] Tutorial de rapidminer, https://academy.rapidminer.com/pages/content-library, visitado el 04-12-2020

[10] "Ley 27499 - Ley Micaela, Capacitación Obligatoria en la Tematica de Genero y Violencia contra las mujeres", 2018. https://www.argentina.gob.ar/normativa/nacional/ley-27499-318666 visitado 03-12-2020

[11] CIAC, 2020. Centro de Información y Análisis Criminal - Ministerio de Seguridad de Jujuy. http: seguridad.jujuy.gob.ar

[12] Rodriguez, Mariela; Laureano, Nazarena; Farfan Jose; "Predicción de Factores de Tipos de Violencia contra las Mujeres" - CONAIISI 2020. http://conaiisievento.sanfrancisco.utn.edu.ar/

[13] Bottinelli, Cristina. Herederos y protagonistas de Relaciones Violentas. https://biblioteca.iidh-jurisprudencia.ac.cr/ visitado: 04-12-2020

[14] OMS, 2013; citado por Mejía, Ochoa, Ríos, Yaulema & Veloz, 2019. http://www.revistaespacios.com/a20v41n22/a20v41n22p17.pdf visitado: 04-12-2020

Aplicación de las técnicas Regresión Lineal Múltiple, Regresión Exponencial, Análisis Exploratorio y Descriptivo de Datos, para Analizar el Comportamiento e Influencia de las Variables: Precio Combustible, Precio Dólar, Casos Covid19y Muertes por Covid19 en la Provincia de Jujuy.

Autor: Octavio Daniel Coro

Institución: Facultad de Ingeniería, Universidad Nacional de Jujuy. San Salvador de Jujuy.

Datos de contacto: *Mail:* odcoro@fi.unju.edu.ar *Celular:+54-388-4081142*

RESUMEN

Este trabajo consiste en demostrar algunas características de las variables y su comportamiento individuales en primer lugar; luego se realizó la Diagnosis y Validación del modelo, para ver si se cumplen las hipótesis básicas para trabajar con el Modelo de Regresión Lineal Múltiple, para demostrar que influencia tuvieron el Precio Dólar, Casos y Muertes por COVID19, sobre la variable Precio del combustible (Nafta Súper en Jujuy), mediante las siguientes técnicas y métodos: Regresión Lineal Múltiple, Regresión Exponencial, Análisis Exploratorio y Descriptivo de datos. Todo esto utilizando Excel y algunos paquetes de Software Estadístico.

Palabras Claves: COVID19 - Muertes por COVID19 - Regresión Lineal Múltiple - Análisis Descriptivo y Exploratorio de Datos – Regresión Exponencial.

INTRODUCCIÓN

Debido a la situación vivida durante este año con la pandemia surgió la necesidad de realizar este trabajo que puede mostrar la influencia del precio del dólar, los casos Covid19 y Muertes por esta causa en la Provincia de Jujuy, sobre el Precio de Combustible. Se eligió la variable Precio de la nafta en Jujuy, porque generalmente si sube su valor suben los precios de todos los productos y servicios, ya que todos los productos llegan a nuestra provincia por los medios de transporte que ya conocemos, como camiones, colectivos y demas.

Esta pandemia influyó en todos los aspectos de nuestras vidas, desde el trabajo cambiando a modalidad virtual, los nuevos cuidados y permisos para salir, hasta privarnos de cosas tan simples como abrazar y demostrar nuestros afectos a la gente más querida, nuestros familiares, en particular los padres y abuelos que tienen la mayor sensibilidad en cuando a los contagios y las mayores complicaciones que pueden tener si se enferman. Fuimos bombardeados y aturdidos por todos los medios hasta lograr cierta conciencia a nivel pueblo, pero tuvimos que pasar muchas pérdidas y situaciones, en muchos casos de hambre al perder el trabajo y no tener con que sostener a su propia familia, hablando de lo básico como ser alimentos, ropas, calzados, etc.

Este fue el motivo principal para realizar este estudio, trabajando con datos a partir de marzo hasta noviembre de este año 2020, tomando los valores al día 30 de cada mes. Obteniendo los datos del COVID19 de Reportes e informes del COE Jujuy, los precios del dólar de la página web Cotización Dólar Histórico Año 2020 - Cotizaciones históricas del dólar en Argentina, también de la página Dólar Hoy; mientras que los datos de combustible en las páginas en Precios de combustible del DataSet de la página www.datos.gob.ar como se muestra en la Bibliografía y links al final del documento.

METODOLOGIA

Para realizar este trabajo en primer lugar se realizó el análisis de las variables en forma individual utilizaremos Regresión Exponencial, Analisis Exploratorio y Descriptivo de datos, tanto para las variables Precio Combustible como para el Precio Dólar, luego para las variables Casos covid19 y Muertes por covid19 en Jujuy, se utilizó la Regresión Exponencial y Análisis Descriptivo de datos, para demostrar estadísticamente como se distribuyen las mismas.

Por último se relacionaron todas y cada una de las variables regresoras con la variable dependiente (Precio del combustible), con el método de Regresión Lineal Simple para ver como influyen por separado estas variables en el modelo. Luego con el el método de Regresión Lineal Múltiple, para ver tanto el comportamiento individual como en forma general de las variables que internienen en este estudio.

En el transcurso de este trabajo se irán mostrando y analizando en forma mas detallada la metodología mencionada.

Este trabajo está organizado de la siguiente manera:

1. Descripción de las variables Dependientes e Independientes.
2. Aplicamos la Regresión Exponencial y Análisis Descriptivo de Datos mediante distintos gráficos para visualizar el comportamiento de las variables relacionadas a Covid19 en este proceso en forma individual.
3. Análisis comparativo del Precio del Dólar y del Combustible mediante BoxPlot, ya que el precio de un dólar generalmente está asociado al precio de un litro de combustible.
4. Utilizamos las variables mencionadas para realizar los siguientes procedimientos:
    i. Realizamos la Diagnosis y Validación del Modelo de Regresión Lineal Múltiple para ver si es aplicable el mismo.
    ii. Analizamos como influye cada una de las variables regresoras, en el comportamiento de la variable dependiente Precio de Combustible, mediante la aplicación del método de Regresión Lineal Simple.
    iii. Realizamos el análisis completo y en conjunto de la variable dependiente con las variables independientes (regresoras) empleando Regresión Lineal Múltiple.
    iv. Calculamos la ecuación de Regresión Lineal Múltiple, con las variables mencionadas anteriormente. Para poder estimar o proyectar para distintos valores que podrían tomar las variables explicativas.
5. Conclusiones: Realizamos un análisis final de todo lo desarrollado en este estudio, con los puntos mas destacables.

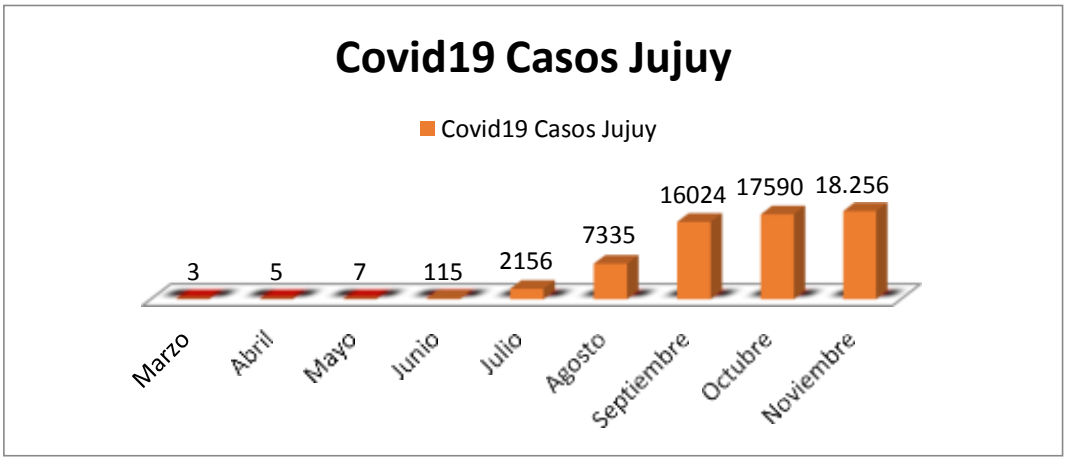
DESARROLLO

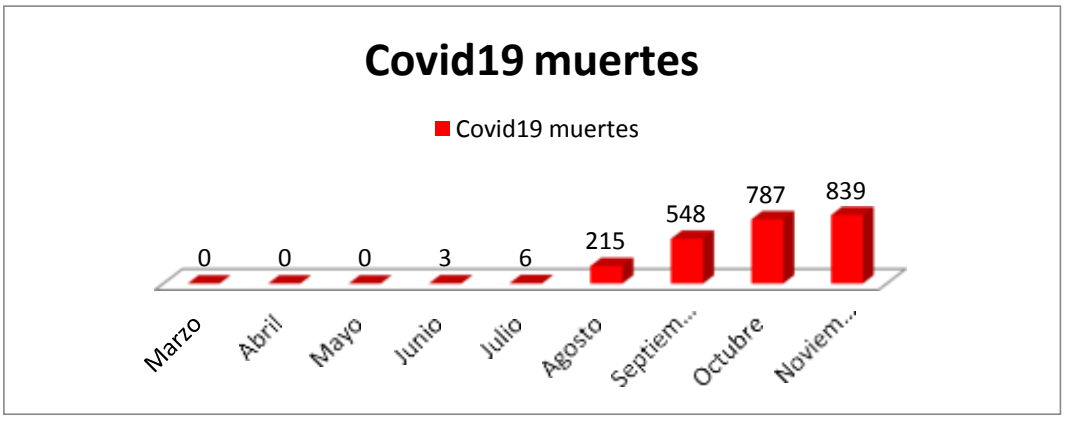
Variables que intervienen en este estudio:

El periodo que se seleccionó para obtener la información es de Marzo a Noviembre de este año, porque son los meses en que comenzó a manifestarse la pandemia en la Provincia de Jujuy.

Mes	Precio Combustible NSJ	Dolar Of. Venta	Covid19 Casos Jujuy	Covid19 muertes	Fecha de obtención
Marzo	60,39	66,67	3	0	30/3/2020
Abril	60,39	69,65	5	0	30/4/2020
Mayo	60,39	71,01	7	0	31/5/2020
Junio	60,39	74,12	115	3	30/6/2020
Julio	60,39	76,02	2156	6	30/7/2020
Agosto	62,99	78,23	7335	215	28/8/2020
Septiembre	65,29	80,45	16024	548	2/10/2020
Octubre	67,59	83,74	17590	787	26/10/2020
Noviembre	69,7	87	18.256	839	30/11/2020

La información de estas variables se obtuvieron de las páginas web, informes y reportes del COE Jujuy, referenciadas al final en la Bibliografía y Links.





Regresión Exponencial: Aplicamos esté método para las variables relacionadas al Covid19 y ver como se comportarían si es que no se hubieran tomado las medidas correspondientes.

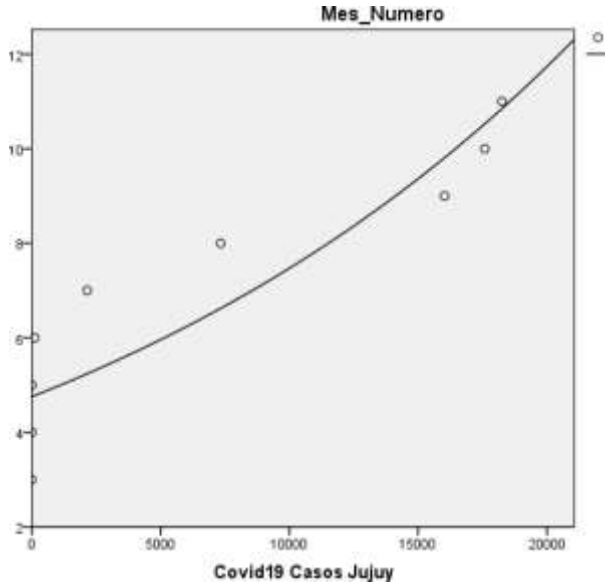
Variable Casos Covid19 en Jujuy

Resumen del modelo y estimaciones de los parámetros

Variable dependiente:Mes_Numero

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Exponencial	,721	18,067	1	7	,004	4,753	4,522E-5

La variable independiente esCovid19 Casos Jujuy.



Ecuación Exponencial para pronosticar Casos de Covid19 en Jujuy:

$$Y = A\ e^{Bx}$$

$$Y = 4{,}753^{\ (4{,}522E-5)}$$

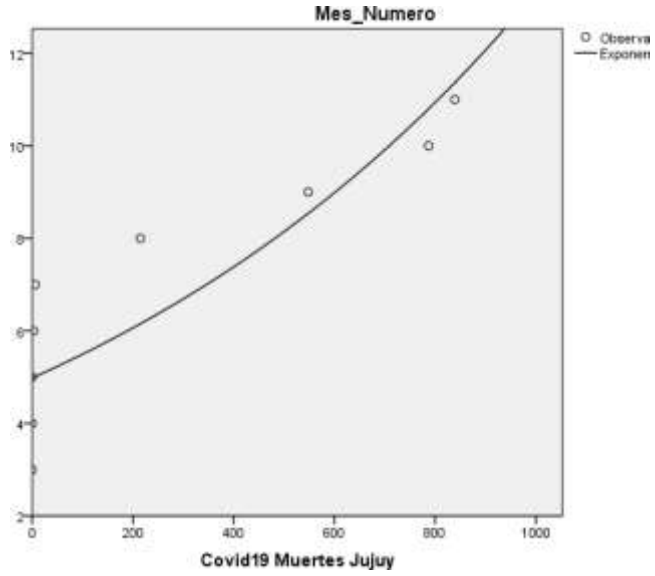
Variable Muertes por Covid19 en Jujuy

Resumen del modelo y estimaciones de los parámetros

Variable dependiente:Mes_Numero

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Exponencial	,653	13,179	1	7	,008	4,982	,001

La variable independiente esCovid19 Muertes Jujuy.



Ecuación Exponencial para pronosticar Muertes por Covid19 en Jujuy:

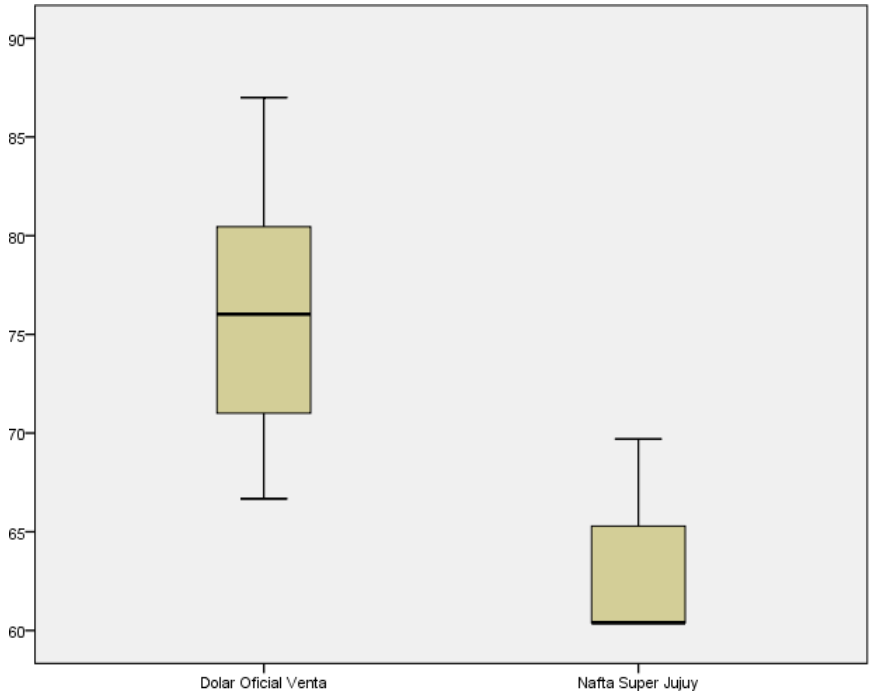
$$Y = A\, e^{Bx}$$

$$Y = 4{,}982\, e^{0{,}001\, X}$$

Diagramas boxplot comparando el precio del Dólar con precio combustible (nafta súper Jujuy)

Nafta Súper Jujuy	
Media	63,0577778
Error típico	1,20895246
Mediana	60,39
Moda	60,39
Desviación estándar	3,62685738
Varianza de la muestra	13,1540944
Curtosis	-0,47451738
Coeficiente de asimetría	1,01847618
Rango	9,31
Mínimo	60,39
Máximo	69,7
Suma	567,52
Cuenta	9

Dólar Oficial Venta	
Media	76,3211111
Error típico	2,23781047
Mediana	76,02
Moda	No tiene
Desviación estándar	6,7134314
Varianza de la muestra	45,0701611
Curtosis	-0,90381917
Coeficiente de asimetría	0,18438843
Rango	20,33
Mínimo	66,67
Máximo	87
Suma	686,89
Cuenta	9



Si analizamos los boxplot vemos que el precio del dólar está más desplazado hacia arriba, cuando con respecto al precio del combustible en Jujuy, esto es debido al congelamiento temporal de los precios del combustible, ya que históricamente, el precio del litro de combustible es prácticamente equivalente al precio de un dólar.

En cuanto a la Asimetría podemos ver que tanto el precio del dólar es positiva con cola derecha, y el precio del combustible también, pero vemos que no tiene bigotes en la parte inferior, debido al congelamiento de los precios y también nos coincide con la Moda (60,39), porque es justo cuando comenzó la pandemia

También podemos ver que estas dos variables con respecto a la Curtosis son Platicúrticas, por los coeficientes menores a cero.

Diagnosis y validación del modelo de Regresión Lineal Múltiple

Ahora vamos a valorar la calidad del ajuste a través, por ejemplo, del coeficiente de determinación no es lo mismo que valorar el cumplimiento de las hipótesis básicas del modelo.

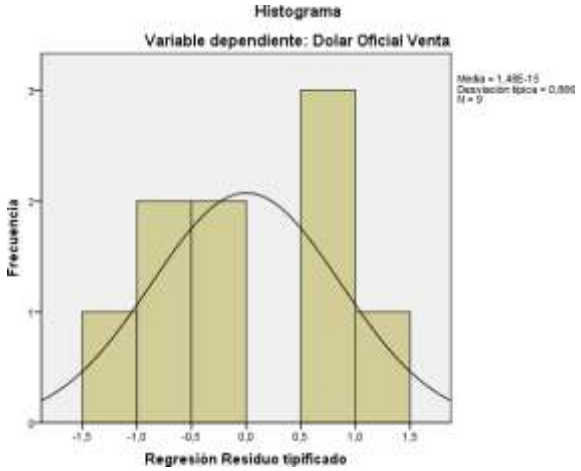
Las labores de inferencia que realizamos bajo el modelo lineal, tienen sentido suponiendo que los datos proceden de tal modelo, tal y como se ha formulado, con todas sus hipótesis básicas

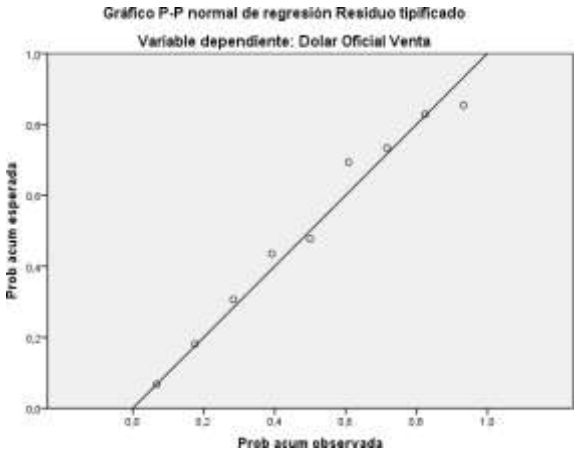
Si las hipótesis no se corresponden con la realidad, se pueden estar cometiendo graves errores en las conclusiones obtenidas de la inferencia basada en el modelo

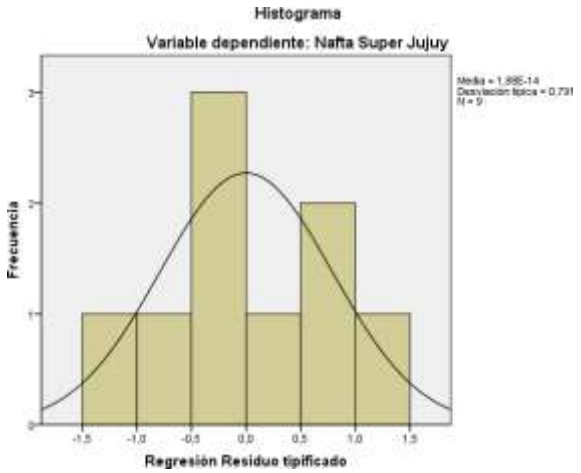
Para llevar a cabo una buena interpretación de un modelo de regresión debemos acompañar siempre nuestro estudio de la diagnosis y validación del modelo. Dicha diagnosis consiste en analizar si se verifican las hipótesis básicas del modelo:

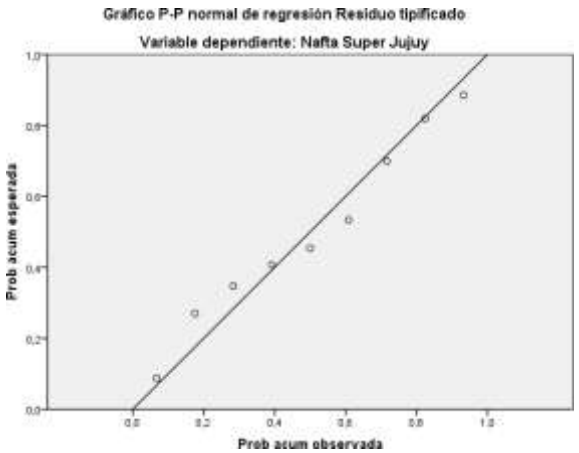
1) Linealidad: los parámetros y su interpretación carecen de sentido si en realidad los datos no proceden de un modelo lineal, situación en la que además las predicciones pueden ser completamente equivocadas. En nuestro caso estos datos provienen de un modelo lineal, ya que analizando por separado cada una de las variables se relacionan linealmente con la variable dependiente seleccionada.

2) Normalidad de los errores: El modelo de regresión lineal asume que la distribución de los errores es Normal, y se lo demuestra con los siguientes gráficos.

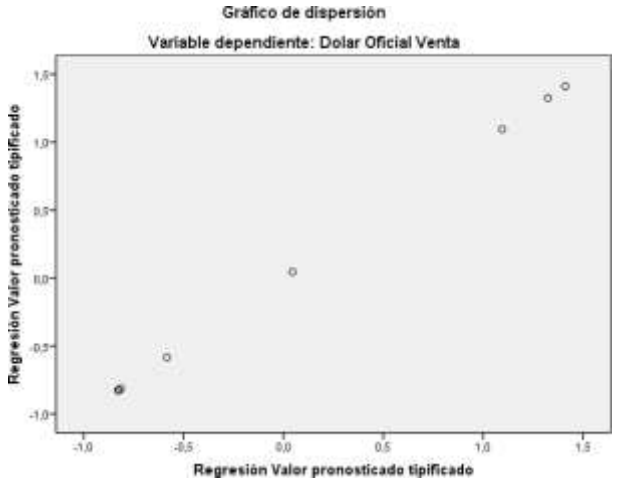


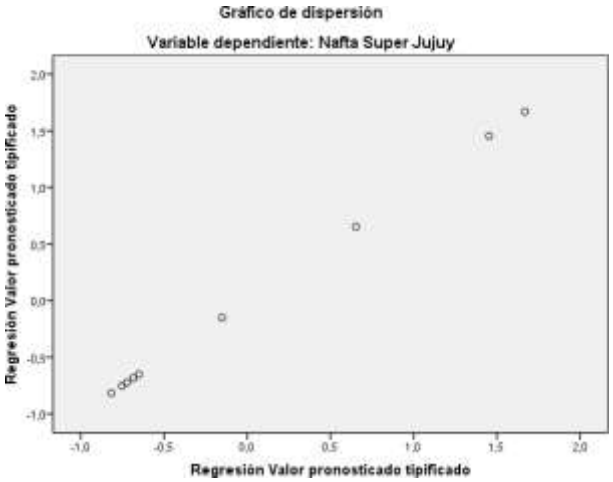






3) Homocedasticidad: La varianza del error es constante. Estos gráficos lo demuestran.





4) Independencia de los errores: Asumimos que las variables aleatorias que representan los errores son mutuamente independientes.
5) Multicolinealidad: Las variables explicativas X1; X2; X3, son linealmente independientes.

El modelo de regresión lineal múltiple asume que las variables explicativas X1; X2; X3, son linealmente independientes, lo que se demuestra en los siguientes resultados, e interpretaciones.

Diagnósticos de colinealidad[a]

Modelo	Dimensión	Autovalores	Índice de condición	Proporciones de la varianza			
				(Constante)	Dólar Oficial Venta	Covid19 Casos Jujuy	Covid19 Muertes Jujuy
1	1	3,330	1,000	,00	,00	,00	,00
	2	,660	2,247	,00	,00	,01	,01
	3	,010	18,312	,00	,00	,86	,99
	4	,001	76,643	1,00	1,00	,13	,00

a. Variable dependiente: Nafta Súper Jujuy

Si analizamos la columna de índice de condición, vemos que tanto el Dólar oficial venta tiene 2,247 menor que 10 es decir que no presenta problemas de Colinealidad, los Casos por Covid19 tiene 18,312 indicando que puede haber un leve problema de colinealidad ya que esta entre 10 y 30, mientras que las Muertes por Covid19 tiene 76,643 lo que indica que presenta un problema importante de Colinealidad, ya que supera el valor 30.

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	55,011	4,960		11,091	,000		
	Dolar Oficial Venta	,077	,070	,142	1,101	,321	,156	6,429
	Covid19 Casos Jujuy	,000	,000	-,328	-1,101	,321	,029	34,389
	Covid19 Muertes Jujuy	,012	,003	1,185	4,273	,008	,034	29,817

a. Variable dependiente: Nafta Super Jujuy

En este último cuadro analizamos el Factor de Inflación de la Varianza (FIV) vemos que la Variable Dólar Oficial Venta no tiene problema de multicolinealidad porque el FIV es menor que 10, mientras que los Casos covid19, seguramente tienen problema de colinealidad debido a que su FIV es mayor que 30, y la Cantidad de Muertes por covid19 es posible que presente Multicolinealidad porque su FIV está entre 15 y 30.

En conclusión pudimos verificar que se cumplen las hipótesis básicas para la Regresión Lineal Múltiple, asi que pasamos ahora a realizar en Análisis de los datos.

ANÁLISIS EN FORMA INDIVIDUAL PARA VER LA INFLUENCIA DE CADA VARIABLE REGRESORA O EXPLICATIVA CON LA VARIABLE DEPENDIENTE.

Ecuación del Modelo de Regresión Lineal Múltiple para el Precio del Combustible

Y	Precio Combustible (Nafta Super en Jujuy)
X1	Precio Oficial Dólar
X2	Casos Covid19
X3	Muertes por Covid19

Y= 55,0107920964138 + 0,0766190475532524 * X1 - 0,00014496911436652 * X2 + 0,0119717720475121 * X3

Pronóstico del valor del Precio del combustible asignando algunos valores particulares a las variables regresoras o explicativas.

Y-Combustible	X1-Dolar	X2-Casos	X3-Muertes
70,5252766	90	19000	950
71,7450866	100	20000	1000
86,8860522	120	50000	2500
112,631789	160	100000	5000

Análisis de las variables mencionadas con el método de RLM

Luego de verificar la Diagnosis y validación del modelo de Regresión Lineal Múltiple, ahora nos abocamos a explicar los aportes de este modelo para estas variables.

En primer lugar analizaremos como incluyen en forma individual cada una de las variables en el modelo, para esto lo hacemos aplicando Regresión lineal simple para cada una de las variables explicativas asociadas a la variable dependiente Precio del combustible.

Estadísticas de la regresión	Precio Dólar	Casos Covid19	Muertes Covid19
Coeficiente de correlación múltiple	0,91396783	0,96747051	0,99122669
Coeficiente de determinación R^2	0,83533719	0,93599918	0,98253034
R^2 ajustado	0,81181393	0,92685621	0,98003468
Error típico	1,57334588	0,98088752	0,51247022
Observaciones	9	9	9

Si analizamos por separado el Coeficiente de Determinación R cuadrado, vemos que más influye sobre la variable precio de combustible (VD) es la variables Muertes covid19 (R^2=0,983), seguida por Casos covid19 (R^2=0,936) y por último Precio dólar (R^2=0,835), esto nos quiere decir que esta última variable si la sacamos del modelo no habría grandes cambios.
Ahora analizaremos coeficiente Beta y T-test

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	55,011	4,960		11,091	,000
	Dólar Oficial Venta	,077	,070	,142	1,101	,321
	Covid19 Casos Jujuy	,000	,000	-,328	-1,101	,321
	Covid19 Muertes Jujuy	,012	,003	1,185	4,273	,008

a. Variable dependiente: Nafta Súper Jujuy

En este cuadro analizaremos la influencia por separado que tiene cada variable regresora en el modelo:

Significación de t-test: si es menor de 0,05 es que esa variable independiente se relaciona de forma significativa con la variable dependiente, por tanto, influye sobre ella, es explicativa

Coeficiente beta (β): indica la intensidad y la dirección de la relación entre esa variable independiente (VI) y la variable dependiente (VD):

- cuanto más se aleja de 0 más fuerte es la relación
- el signo indica la dirección (signo + indica que al aumentar los valores de la VI aumentan los valores de la VD; signo – indica que al aumentar los valores de la VI, los valores de la VD descienden)

Podemos ver en el **Coeficiente T-test** que las Muertes por covid19 aportan en forma muy significativa al modelo por su 0,008 menor que 0,01, mientras que tanto el Precio del dólar como la cantidad de casos no influyen en el modelo de manera significativa.
Si analizamos el coeficiente tipificado **Beta** vemos que el precio del dólar (Beta 0,142) aporta al modelo y mientras sube el precio del dólar sube precio de combustible; si la cantidad de casos covid (Beta=-0,328) aumenta, entonces el precio del combustible tiende a bajar, pero sobre todo por las medidas políticas; si la cantidad de muertes aumenta (Beta=1,185) esto quiere decir que el precio del combustible aumentará en forma notoria, influyendo más en el modelo.

Ahora analizando en forma general el modelo en base a los siguientes cuadros resultados que nos brinda el paquete estadístico:

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	Covid19 Muertes Jujuy, Dólar Oficial Venta, Covid19 Casos Jujuy[a]	.	Introducir

a. Todas las variables solicitadas introducidas.

b. Variable dependiente: Nafta Súper Jujuy

Significación de F-test: si es menor de 0,05 es que el modelo es estadísticamente significativo y por tanto las variables independientes explican "algo" la variable dependiente, cuánto "algo" es la R-cuadrado

En este cuadro siguiente nos muestra que elegimos la variable dependiente Nafta Súper en Jujuy, con las variables regresoras: Precio dólar, Casos covid19 y Muertes covid19 mencionadas anteriormente.

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	103,876	3	34,625	127,602	,000[a]
	Residual	1,357	5	,271		
	Total	105,233	8			

a. Variables predictoras: (Constante), Covid19 Muertes Jujuy, Dólar Oficial Venta, Covid19 Casos Jujuy

b. Variable dependiente: Nafta Súper Jujuy

Si vemos el cuadro anterior del Análisis de la Varianza vemos que el coeficiente arrojado por el F-test es de 0,000 ; esto nos indica que es un muy buen modelo para aplicar este método de RLM.

R cuadrado: En el siguiente cuadro se vé en cuánto las variables independientes explican el comportamiento de la variable dependiente, indica el porcentaje de la varianza de la variable dependiente explicado por el conjunto de variables independientes. Cuanto mayor sea la R-cuadrado más explicativo y mejor es el modelo causal.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,994[a]	,987	,979	,5209162

a. Variables predictoras: (Constante), Covid19 Muertes Jujuy, Dolar Oficial Venta, Covid19 Casos Jujuy

Para finalizar con este cuadro del modelo vemos que el R cuadrado es de 0,987, lo que indica que el 98,7 % de la recta de regresión estimada es explicada por las variables predictoras mencionadas en este cuadro. Esto nos esta demostrando que es un muy buen modelo el de Regresión Lineal Múltiple para explicar el comportamiento de las variables elegidas.

CONCLUSIONES

En primera instancia este trabajo se pensó analizando o teniendo en cuenta dos variables independientes que eran el Precio Dólar y el Precio Combustible, contrastando con las variables predictoras o explicativas: Casos y Muertes por Covid19 en Jujuy. Lo que no daba muy buenos resultados si tomábamos como variable dependiente al Precio del Dólar, por este motivo se decidió trabajar con una variable dependiente única Precio de Combustible, y tomando como variables predictoras al Precio dólar, Casos covid19 y Muertes covid19, y se obtuvieron los resultados mostrados en el Desarrollo de este trabajo.

Luego de seleccionadas las variables que trabajarían para el modelo de Regresión Lineal Múltiple se analizó por separado las variables relacionadas al Covid19, en primer lugar mostrando un gráfico de barras descriptivo, que demostraba su distribución exponencial, por lo que en segundo lugar utilizando la técnica de Regresión Exponencial para esas dos variables, se obtiene el gráfico que representa dicha Regresión y la fórmula para predecir los valores a futuro, si no se hubieran tomado las medidas correctas. Ahora si bien estas variables tienen distribución exponencial, con respecto a las otras variables se demostró que hay una relación lineal múltiple.

También se analizaron en comparación las variables Precio Dólar y Precio de Combustible con BoxPlot aplicando el Análisis Exploratorio de Datos y se demostró que el Precio Dólar tiene valores superiores, debido al congelamiento momentáneo por la pandemia del precio del combustible, porque por lo general el precio de un litro de combustible equivale a un dólar.

Cuando analizamos por separado el Coeficiente de Determinación R cuadrado, vemos que más influye sobre la variable precio de combustible (VD) es la variables Muertes covid19 (R^2=0,983), seguida por Casos covid19 (R^2=0,936) y por último Precio dólar (R^2=0,835), esto nos quiere decir que esta última variable si la sacamos del modelo no habría grandes cambios.

Realizando el T-test también se vio que las Muertes por covid19 aportan en forma muy significativa al modelo por su Coeficiente T-test 0,008 menor que 0,01, mientras que tanto el Precio del dólar como la cantidad de casos no influyen en el modelo de manera significativa.
Vimos que el coeficiente tipificado Beta nos indica que el precio del dólar (Beta 0,142) aporta al modelo y mientras sube el precio del dólar sube precio de combustible; si la cantidad de casos covid (Beta=-0,328) aumenta, entonces el precio del combustible tiende a bajar, pero sobre todo por las medidas políticas; si la cantidad de muertes aumenta (Beta=1,185) esto quiere decir que el precio del combustible aumentará en forma notoria, influyendo más en el modelo.

Por fin analizando el modelo completo, vimos en el cuadro del Análisis de la Varianza que el coeficiente arrojado por el F-test es de 0,000 ; esto nos indica que el método de RLM es un muy buen modelo, para las variables seleccionadas, confirmando esta deducción con el valor del coeficiente de Determinación (R^2) que es de 0,987, lo que demuestra que el 98,7 % del comportamiento del Precio de Combustible es explicado por las variables Precio del dolar, Casos covid19 y Muertes por covid19, utilizando la recta de regresión estimada que obtuvimos en el desarrollo de este trabajo.

Por consiguiente si aumentan los Casos de covid19 y Muertes por covid19 en Jujuy podemos decir que seguramente aumentará el precio de combustible, lo que estallaría en una suba general de todos los productos y servicios de nuestra Provincia de Jujuy.

BIBLIOGRAFIA

- 2009 - Ronald E. Walpole, Raymond H. Myers, Sharon l. Myers y Keying Ye. Probabilidad y estadística para ingeniería y ciencias, 9ª Edición. Ed. Pearson educación
- Estadística. Novena edición. Triola, Mario F. Ed. Pearson educación
- 2008 - Jay L. Devore Probabilidad y Estadística para Ingeniería y Ciencias. Séptima Edición.
- 2008 - Anderson, David R.,Dennis J. Sweeney y Thomas A. Williams. Estadística para administración y economía, Ed. Cengage Learning. 10a. Edición.
- 2004 - Douglas C Montgomery. Diseño y Análisis de Experimentos. Ed. Limusa. 2da Edición
- 2004 – Levin Richard, Rubin David. Estadística para Administración y Economía. Séptima edición. Ed. Pearson.
- 1996 - Douglas C. Montgomery y George C. Runger - Probabilidad y Estadística aplicadas a la Ingeniería. Ed. Mc Graw-Hill.

Regresión Lineal Múltiple - Links

http://networkianos.com/regresion-lineal-multiple/

https://wpd.ugr.es/~bioestad/guia-de-r/practica-3/

https://rpubs.com/Joaquin_AR/226291

https://www.youtube.com/watch?v=n5RnoR9oLlc

Video Ejercicios resuelto de Reg Lineal Multiple

https://www.youtube.com/watch?v=_nXiN03cBjo

https://www.youtube.com/watch?v=zk_75JPEvnw

https://www.youtube.com/watch?v=8d-yYngHa3o

https://www.youtube.com/watch?v=VwjLUKoSoPY

https://www.youtube.com/watch?v=eEsqMNxPN0E

Origen de los datos para este estudio

https://www.cotizacion-dolar.com.ar/dolar-historico-2020.php

https://www.dolarhoy.com/

http://cecha.org.ar/site/index.php/esncuastas-y-consultas/

https://datos.gob.ar/dataset/energia-precios-surtidor---resolucion-3142016/archivo/energia_f8dda0d5-2a9f-4d34-b79b-4e63de3995df

Informes y reportes COE Jujuy

10 y 11 de Diciembre de 2020

Análisis estadístico con R de diferentes métodos de extracción de ácidos nucleicos de *Salmonella* Paratyphi B para el diseño de dispositivos de detección

María del Milagro Said-Adamo[1,2*], Sarita Reyes[1], María Noel Maidana-Kulesza[1], Verónica Rajal[1,3], Ramiro Poma[1], Héctor Cristóbal[1,2]

[1]INIQUI-CONICET – UNSa, [2] Facultad de Ciencias Naturales – UNSa, [3] Facultad de Ingeniería – UNSa,

Salta Argentina.

milagro.said@gmail.com – 0387-154554994

RESUMEN

Existen metodologías oficiales para la investigación y detección de patógenos que se trasmiten a través de alimentos contaminados. Estas implican técnicas microbiológicas convencionales que son laboriosas y consumen tiempo. En los últimos años se intensificó la búsqueda de nuevas tecnologías basadas en ácidos nucleicos para la detección rápida de patógenos, que garanticen la especificidad, sensibilidad y bajo costo en laboratorios miniaturizados que contemplen todas las etapas clásicas de extracción, amplificación y detección. Este trabajo propone evaluar cinco métodos de extracción de ácidos nucleicos a partir de *Salmonella* Paratyphi B aplicando R como herramienta estadística. Se realizaron extracciones de ácidos nucleicos a partir de cultivos puros de la bacteria en estudio y se midió su concentración mediante un espectrofotómetro de microvolúmenes. Los datos se analizaron mediante la prueba estadística no paramétrica de Kruskal-Wallis. Los resultados muestran que los métodos de extracción shock térmico y buffer de lisis permitieron recuperar concentraciones elevadas de ácidos nucleicos; aunque fueron las muestras con mayor dispersión de los datos en comparación con los otros métodos. Las pruebas estadísticas aplicadas de comparación múltiple permiten seleccionar la metodología de extracción óptima y, aportar a la toma de decisiones en cuanto a los intereses de esta línea de investigación.

Palabras Claves: *Salmonella* sp, software R, análisis no paramétrico, prueba de Kruskal-Wallis

INTRODUCCIÓN

Tradicionalmente, los laboratorios de análisis han sido un servicio de apoyo para dar respuesta a diversas necesidades de salud pública, brindando resultados de alta calidad y, de esta manera, colaborando en el diagnóstico, tratamiento y prevención de enfermedades. En los últimos años, se ha expandido el uso de las pruebas *Point-of-care-test* (POCT) o *Lab-on-a-chip* (LOC), lo que ha llevado a introducir cambios en la estructura de los laboratorios convencionales con el fin de contribuir al desarrollo de microtecnologías que sean fáciles de usar, de bajo costo, portátiles, reproducibles, sensibles y que permitan obtener resultados de manera rápida (Berli, 2016; Narayan, 2016; Rodriguez y Shocron, 2019).

Si bien estos laboratorios miniaturizados, cuentan con una serie de ventajas en comparación a las pruebas de laboratorio convencionales, también presentan algunas desventajas no resueltas. Sin embargo, los avances en nuevos desarrollos y tecnologías, han permitido resolver y mejorar los nuevos diseños (Celis-Morales, 2014).

En el marco de la Seguridad Alimentaria (SA), para afrontar la problemática de la contaminación de alimentos con microrganismos patógenos, se emplean de forma tradicional dos grupos de metodologías: las inmunológicas y las génicas. A pesar de que los inmunoensayos son prácticos y efectivos, en general poseen menor sensibilidad que los métodos basados en ácidos nucleicos (Mafra et al., 2008). Adicionalmente, debe considerarse que el procesamiento de alimentos puede modificar la conformación de las proteínas, alterando la especificidad del reconocimiento antígeno-anticuerpo.

Por estas circunstancias, existe una tendencia a desarrollar aplicaciones basadas en la identificación y cuantificación de secuencias específicas de ADN (Asensio et al., 2008). Actualmente se desarrollan numerosos protocolos de qPCR (reacción cuantitativa en cadena de la polimerasa en tiempo real) y la amplificación isotérmica mediada por bucles (LAMP). Además, existe un especial interés por sistemas que permitan analizar muestras sin la necesidad de utilizar operarios especializados, por lo que, el desarrollo de estrategias innovadoras constituye una necesidad.

Estos dispositivos tienen aplicación en diferentes ámbitos, como la industria alimentaria, monitoreo ambiental, defensa, investigación, en el campo médico y cuidado de la salud.

El desarrollo de estos sistemas tiene que contemplar e integrar todas las etapas clásicas de extracción de ácidos nucleicos, amplificación y detección en un solo dispositivo que reúna todas las características esperadas que aumenten la eficiencia y reduzcan los costos (Blair y Corrigan, 2019).

En base a esto, y aplicando R como herramienta estadística, el presente trabajo propone evaluar cinco métodos diferentes de extracción de ADN de *Salmonella* Paratyphi B; un microorganismo patógeno que se transmite a través de una gran variedad de alimentos contaminados. R es un lenguaje de alto nivel y un entorno para el análisis de datos y gráficos más potente y profesional que existe actualmente para realizar tareas estadísticas de todo tipo, desde las más elementales, hasta las más avanzadas. Se trata de un ambiente de programación formado por un conjunto de herramientas muy flexibles que pueden ampliarse fácilmente mediante paquetes, librerías o definiendo nuestras propias funciones. Además es gratuito, de descarga e instalación sencillas y es producto de la filosofía *Open Source* (o código abierto). Esto significa que desde sus inicios, una extensa comunidad de usuarios y programadores de alto nivel contribuye a desarrollar nuevas funciones, paquetes y actualizaciones que son rápidamente accesibles a todo público. Esto

convierte a R en una herramienta estadística estable, confiable y a la vanguardia, ya que está sometida a una actualización permanente (R Core Team, 2016).

El trabajo tiene como propósito determinar la recuperación relativa de distintos métodos de extracción de ácidos nucleicos, tomando como datos las concentraciones obtenidas de los mismos. De este modo, será posible seleccionar el procedimiento de extracción óptimo para formar parte del dispositivo de detección específico para esta *Salmonella* Paratyphi B.

METODOLOGÍA

Se incubó *S.* Paratyphi B en 50 ml de caldo LB (Broth, Miller) a 30°C por 12 h. A partir de este cultivo puro se realizó la extracción de ácidos nucleicos utilizando cinco métodos según describen los autores y fabricantes (Tabla 1). Para cada ensayo se realizaron tres réplicas. Para determinar la concentración de ácidos nucleicos obtenida se utilizó 1 µl de muestra, se midió por triplicado.

Tabla 1. Métodos de extracción de ADN empleados, abreviaturas y referencias

Métodos de extracción de ácidos nucleicos	Abreviatura	Referencia
Fenol-Cloroformo	FC	Sambrook et al., 1989
Kit Comercial 1: PURO Virus RNA Productos Biológicos	KC1	Productos Biologicos
Kit Comercial 2: QIAmp Viral RNA QIAGEN	KC2	QIAGEN
Shock Térmico	ST	Thompson et al., 2006
Buffer de Lisis (proteinasa K y Lisozima)	BL	

Se seleccionaron métodos que requieren el empleo de insumos y equipamiento específicos de laboratorio, como es el caso del método de extracción con fenol-cloroformo y los kits comerciales que, además, son costosos. Y se eligieron métodos más sencillos, rápidos y económicos como la extracción de ácidos nucleicos por shock térmico y por buffer de lisis (Tabla 1).

Las determinaciones de ácidos nucleicos y proteínas se llevaron a cabo utilizando un espectrofotómetro NanoDrop™ (Thermo Scientific 2000) que mide muestras de volúmenes pequeños con alta precisión y reproducibilidad.

Las variables que se midieron fueron: concentración de ácidos nucleicos (ng/ µl), relación de absorbancia a 260 nm y 280 nm (260/280), y la relación de absorbancia a 260 nm y 230 nm (260/230). Los datos fueron registrados por un programa de una computadora asociada al equipo.

Con los datos obtenidos se procedió al análisis estadístico con el programa R (versión 4.0.2 Copyright © 2020 The R Foundation for Statistical Computing: http://cran.r-project.org/bin/windows/base/).

Se calcularon las medidas de resumen para cada método: media, mediana, mínimo, máximo,

desvío estándar. Por otro lado, para comparar estadísticamente los distintos métodos de extracción de ADN, se buscó el procedimiento que se ajustara mejor a los datos obtenidos.

Se realizaron gráficos BoxPlot para completar el análisis y representar el pool de información obtenida.

DESARROLLO

A partir de todos los métodos empleados para la extracción de ácidos nucleicos de cultivos bacterianos de *S.* Paratyphi B, se logró obtener datos cuantitativos de su concentración. Esta cuantificación se realizó mediante espectrofotometría. Las medidas de resumen para cada variable por método se muestran en la Tabla 2. Por su parte, para cada método evaluado se calculó la recuperación relativa de ácidos nucleicos con respecto al método con mayor desempeño. En la Figura 1 están representadas las distribuciones de los conjuntos de datos.

Inicialmente para comparar los métodos de extracción de ácidos nucleicos, se planteó realizar un análisis de la varianza (ANOVA), sin embargo para que estos resultados sean válidos, se debe verificar que se cumplan los supuestos de normalidad de los errores, homogeneidad de varianzas y la no interacción factor-bloque. Al hacer la prueba de normalidad de Shapiro-Wilk, se determinó que las variables estudiadas no tenían distribución normal, por lo que no fue posible aplicar un ANOVA.

Por lo tanto, se realizó el test estadístico no paramétrico Kruskal-Wallis cuyos resultados para cada variable se muestran en la Tabla 2. En principio, este test pone a prueba la hipótesis sobre la igualdad de las medianas de una variable. Esta prueba indica si existen diferencias estadísticamente significativas en las medianas de los rankings de los datos ($p\text{-valor}<0.05$).

En cuanto a la concentración de ácidos nucleicos (ng/µl) y la recuperación relativa, se puede observar que los métodos BL y ST recuperaron mayor concentración (372,50 ± 213,12 ng/µl y 356,60 ± 576,56 ng/µl respectivamente) por sobre los otros métodos (FC 250,00 ±46,50 ng/µl; KC1 139,60 ± 1,45 ng/µl; KC2 118,40 ± 10,25ng/µl). Sin embargo, estos métodos presentaron mayor dispersión de los datos (Figura 1).

La relación de absorbancia a 260/280 se utiliza para evaluar la pureza del ADN y presencia de ARN. Una proporción cercana a 1,8 generalmente se acepta como “pura” para el ADN y una relación cercana a 2,0 para el ARN. Si la relación es sensiblemente menor en cualquier caso, puede indicar la presencia de proteínas, fenol u otros contaminantes que absorben fuertemente a una densidad óptica (DO) de 280 nm o cerca de ellos (NanoDrop 2000). En el presente estudio, los valores de esta relación de absorbancia se asumieron como “puros” para ARN para los métodos KC1 y KC2. Sin embargo. Los metódos FC y ST arrojaron valores cercanos a 1,8 lo que indica cierta “pureza” para ADN. El método BL constituyó las muestras con más proteínas y otros contaminantes (Tabla 2).

Tabla 2. Medidas de resumen de cada variable y recuperación relativa de ácidos nucleicos con respecto al método de mejor desempeño (FC: Fenol-Cloroformo, KC1: Kit Comercial 1, KC2: Kit Comercial 2, ST: Shock Térmico, BL: Buffer de Lisis)

Variable	**Método de extracción de ADN**				
Concentración ng/µl	**FC**	**KC1**	**KC2**	**ST**	**BL**
N	9	9	9	9	9
Media	269,02	140,10	120,83	579,38	444,93
Mediana	250,00	139,60	118,40	356,60	372,50
Desvío Estándar	46,50	1,45	10,25	576,56	213,12
Máximo	328,00	142,90	134,50	2068,00	1002,00
Mínimo	222,60	138,10	108,00	258,80	314,50
Relación de absorbancia 280/260					
N	9	9	9	9	9
Media	1,99	2,87	3,09	1,50	0,95
Mediana	2,01	2,86	3,13	1,53	0,92
Desvío Estándar	0,09	0,03	0,12	0,09	0,06
Máximo	2,09	2,92	3,20	1,58	1,08
Mínimo	1,80	2,84	2,90	1,36	0,91
Relación de absorbancia 260/230					
N	9	9	9	9	9
Media	1,96	3,23	0,92	0,74	0,36
Mediana	2,02	3,23	1,03	0,77	0,34
Desvío Estándar	0,21	0,07	0,20	0,08	0,05
Máximo	2,17	3,36	1,10	0,82	0,48
Mínimo	1,47	3,15	0,64	0,59	0,30
Recuperación relativa	0,46	0,24	0,21	1,00	0,77

La relación de absorbancia a 260/230 constituye por su parte una medida secundaria de la pureza del ADN. Los valores 260/230 para un ácido nucleico "puro" son a menudo más altos que los valores respectivos a 260/280 y se encuentran comúnmente en el rango 1,8-2,2. Si la proporción es sensiblemente menor, puede indicar la presencia de contaminantes copurificados (NanoDrop 2000). Este fue el caso de los métodos BL, ST y KC2. Las extracciones de ADN con FC y KC1 resultaron las más "puras" (Tabla 2).

Finalmente, en el marco de la prueba de Kruskal-Wallis, se realizó una comparación múltiple para determinar si los valores obtenidos en los diferentes grupos eran estadísticamente similares (Figura 2). En cuanto a la concentración de ácidos nucleicos, BL y ST no mostraron diferencias estadísticas entre sí. Sin embargo, con respecto a los otros métodos sí hubieron diferencias.

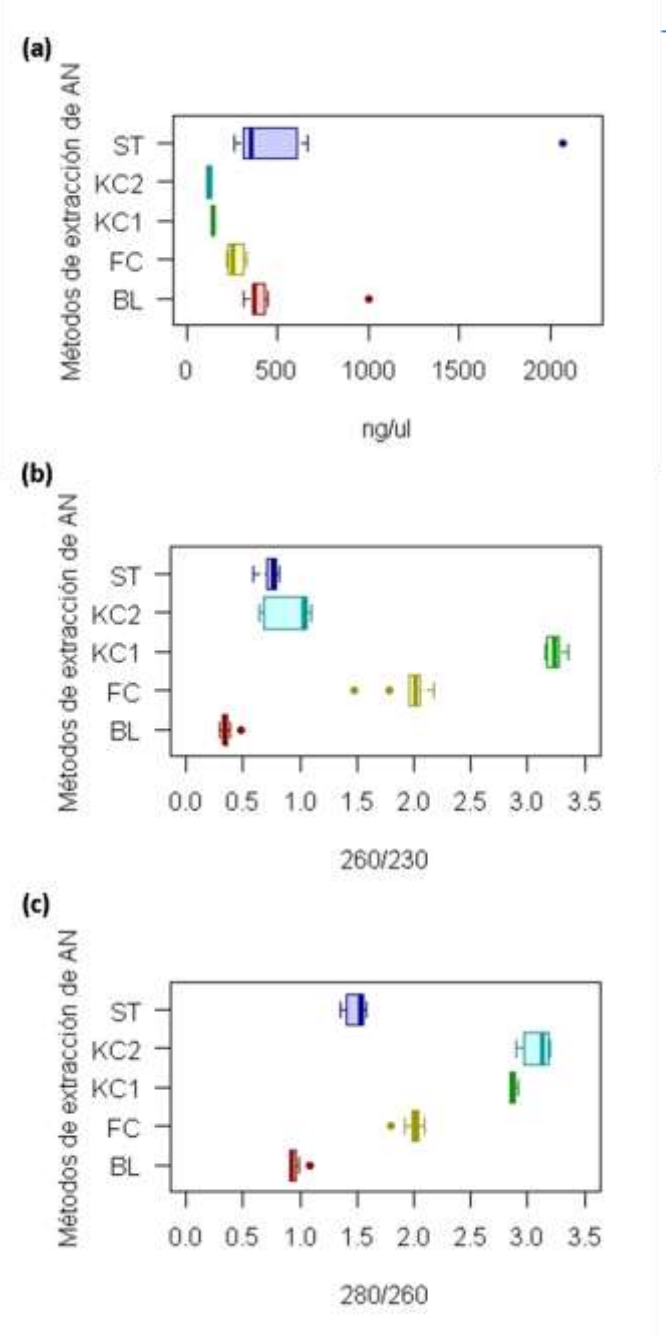


Figura 1. BoxPlot con las medidas de resumen de cada variable para cada método de extracción de ácidos nucleicos (AN) analizado: Mediana (línea central de las cajas), cuartil Q50 (área de las cajas), cuartil Q25 y cuartil Q75 (bigotes de las cajas), valores atípicos (puntos) **(a)** Concentración de ácidos nucleicos (ng/µl) **(b)** Relación de absorbancia 260/280 **(c)** Relación de absorbancia 260/230.

Tabla 3. Test de Comparación Múltiple por Kruskal-Wallis de cada variable estudiada. Las letras de los grupos indican si hay diferencias estadísticas (p-valor<0,05).

Test de Comparación Múltiple		
Concentración ng/µl	Relación 260/280	Relación 260/230
VR1 groups BL 37.55556 a ST 33.77778 a FC 24.66667 b KC1 14.00000 c KC2 5.00000 d	VR4 groups KC2 40.83333 a KC1 32.16667 b FC 23.00000 c ST 14.00000 d BL 5.00000 e	VR5 groups KC1 41.00000 a FC 32.00000 b KC2 20.61111 c ST 16.38889 d BL 5.00000 e

CONCLUSIONES

El presente trabajo permitió caracterizar cinco métodos de extracción de ácidos nucleicos a partir de un cultivo puro de *S.* Paratyphi B.

Mediante espectrofotometría se logró estimar la concentración de ácidos nucleicos de cada muestra y relacionar la absorbancia como indicadores de presencia de posibles contaminantes.

Por su parte, la herramienta R como programa estadístico, ofreció todas las técnicas para el análisis de datos de manera fácil y robusta. No obstante, es necesario conocer y dominar de manera fluida el lenguaje de su programación en línea de comando.

Los métodos evaluados mostraron diferencias estadísticas entre ellos. Sin embargo, en base a las medidas de resumen visualizadas, la recuperación relativa y a los fines de la investigación, fue posible tomar decisiones.

Teniendo en cuenta que el dispositivo que se diseñará posteriormente se basará en la técnica LAMP, es posible seleccionar los métodos de extracción que recuperaron mayor concentración de ácidos nucleicos, Shock Térmico y Buffer de Lisis; aunque estos no hayan logrado las extracciones con mayor pureza, como fue el caso de los métodos de Fenol-Cloroformo y ambos Kits Comerciales. Esta decisión se fundamenta en que el rendimiento de la técnica LAMP es satisfactorio aún en presencia de inhibidores (Nixon y col., 2014), por lo que el interés de este análisis se centra más bien en la concentración de ácidos nucleicos.

De esta manera, al elegir el método de Shock Térmico y el Buffer de Lisis como métodos de extracción de ácidos nucleicos para probar en un POCT o LOC, es posible reducir costos y alcanzar otras características deseadas para este tipo de dispositivos. Esto se probará en ensayos futuros.

BIBLIOGRAFÍA

Asensio L, González I, García T, Martín R (2008) Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA) Food Control 19: 1-8.

Berli CLA (2016) Revista FABICIB 20.

Blair EO, Corrigan DK (2019). A review of microfabricated electrochemical biosensors for DNA detection. Biosensors and Bioelectronics 134: 57-67.

Celis Morales M (2014) Recomendaciones para el uso de pruebas de laboratorio en el lugar de asistencia del paciente (POCT). Departamento de Laboratorio Biomédico Nacional y de R-Referencia. Instituto de Salud Pública. Chile.

Mafra, I, Ferreira, IM, Oliveira, MBP. Food authentication by PCR-based methods. Eur. Food Res. Technol., 2008, 227, 649-665.

Narayan RJ (2016) Medical Biosensors for Point of Care (POC) Applications. Woodhead Publishing.

Nixon G, Garson JA, Grant P, Nastouli E, Foy CA, Huggett JF (2014). Comparative study of sensitivity, linearity, and resistance to inhibition of digital and nondigital polymerase chain reaction and loop mediated isothermal amplification assays for quantification of human cytomegalovirus. *Analytical chemistry*, *86*(9), 4387-4394.

R Development Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rodriguez, EF, & Shocron, RG (2019). Recomendaciones para la implementación de equipos point of care. Revista Argentina de Terapia Intensiva, 36(1).

Thermo Scientific NanoDrop 2000/2000c Spectrophotometer (2009) User Manual.

10 y 11 de Diciembre de 2020

PROBLEMÁTICA DELICTIVA EN LA PROVINCIA DE JUJUY DURANTE EL AISLAMIENTO SOCIAL, PREVENTIVO Y OBLIGATORIO

Autores: Anl. VARGAS, Gerardo José León, Anl. MACHUCA, Matías Alejandro, Ing. RODRIGUEZ, Mariela Ester, Op. Tec. CHAVEZ, Fabio Joel.

Institución: "Ministerio de Seguridad de la Provincia de Jujuy"

Datos de contacto: Gherarleo.GV@gmail.com - (388) 6851701

RESUMEN.

La estadística policial actual es aquella que recoge y compila el Ministerio de Seguridad de la provincia de Jujuy, en el mismo lugar de las actividades delictivas; informan sobre el número de delitos cometidos, personas detenidas y circunstancialmente acusadas. Estas estadísticas se extraen de los registros de organismos oficiales, es decir de las comisarías y dependencias de investigación criminal. Reflejan las fluctuaciones del delito, en determinado tiempo y espacio.

Como prueba de ello es que se desarrolla un informe sobre el Delito contra la Propiedad ya que es el delito con mayor cantidad de registros (ROBOS, HURTOS y ESTAFAS) y son hechos producto de desviaciones biológicas y/o el resultado de la influencia ambiental o social que afecta al "sujeto infractor". De allí es que surge la necesidad de realizar "Estadísticas Criminales" como forma fehaciente de indagar en las causas de este delito, conocer la "Realidad Delictiva" e interpretar el problema del delito, a fin de proponer los cambios necesarios para prevenirlo.

Se pretende estudiar el movimiento general de la criminalidad, tiendo en cuenta sus ritm os, variaciones y relaciones con las condiciones personales (sexo y edad), condiciones geográficas y sociales, todo ello en un marco de conducta adaptado al Aislamiento Social Preventivo y Obligatorio.

Palabras Claves: Covid-19-Delitos-Estafas-Aislamiento-Analisis

INTRODUCCIÓN

En toda sociedad han habido individuos y grupos que se han desenvuelto fuera de los márgenes de la buena convivencia y transgredido las normas y las leyes establecidas, alejándose de los comportamientos aceptables. En la Sociología más clásica se ha utilizado el término "desviación" para explicarlos. Por lo que tiene sentido decir que se refiere a un acto o a alguien que no se ajusta a las normas sociales.

Ahora bien, debemos tener en cuenta que la idea de desviación social abarc a desde un comportamiento descortés (por ejemplo: una murmuración, una calumnia, un insulto, hasta incumplir el confinamiento derivado de la pandemia del COVID-19, o en casos de robo o asesinatos); y es la propia organización social la que establece y controla los niveles y extensión de lo correcto y lo incorrecto, así como el grado y correctivo que de su incumplimiento resulta.

Cada sociedad propone sus propias normas e impone sus sanciones para mantener el orden social, de donde se deriva que define su propia visión de lo alejado del comportamiento aceptable, ligado a su historia. Así pues, resulta de extrema importancia el aprendizaje, interiorización y cumplimiento de las pautas sociales establecidas por cada cultura, porque las normas instauradas en cada ámbito social tienen como finalidad la consecución de una armoniosa convivencia, la supervivencia del individuo y el mantenimiento de la buena convivencia.

Consecuente con lo anterior, lo desviado, las normas y las sanciones son relativas a cada sociedad. Están determinadas por un espacio y un tiempo, cambian, quedan obsoletas, se transforman, se modifican los marcos tolerados y su cumplimiento adopta distintas modalidades, valoraciones y niveles.

En el siguiente trabajo, se abordará la criminalidad en la provincia de Jujuy, R.A., en los tiempos de pandemia, tratando de poner en manifiesto desde un enfoque descriptivo las conductas criminales, comportamientos de víctimas y escenarios adaptados a este nuevo entorno.

METODOLOGÍA

En cuanto a la Metodología utilizada para esta publicación, se utilizaron diferentes métodos de Análisis:

- ***Filtración***: Se eliminaron datos sobrantes para centrarse solo en lo que es importante.
- ***Categorización***: Se agruparon y clasificaron los datos en grupos lógicos: ROBOS, HURTOS y ESTAFAS. Esto permite identificar, analizar y comunicar la información.
- ***Agrupación***: Se contaron, resumieron, promediaron y agruparon los datos en categorías.
- ***Correlación***: La correlación es una técnica estadística que determina si un grupo de datos se relacionan con otro [1].

Luego, para la presentación de los datos se utilizaron distintos gráficos estadísticos con el fin de facilitar su interpretación para el usuario de la información.

- ***Diagrama de Rectángulos***: Se construye sobre un sistema de Ejes cartesianos, situación en uno de los ejes de las distintas modalidades del carácter y el otro los valores de las frecuencias.
- ***Diagrama de Sectores***: Una distribución que se construye trazando una circunferencia de radio arbitrario y dividiendo su círculo en sectores, cada sector se asocia a una de las modalidades de

carácter de modo que el ángulo central de cada sector sea proporcional a la frecuencia de la correspondiente modalidad

- ***Histograma***: Está fundamentada en la proporcionalidad de las áreas de rectángulos a las frecuencias de cada modalidad. [2].

Al final se utilizó un proceso de mapeo delictivo mediante un Sistema de Información Geográfico (GIS), a través del cual se procedió, en primera instancia a la confección de Cartogramas (representación sobre un mapa de las diversas variables).

DESARROLLO

Medidas Sanitarias ante el COVID-19 en la Provincia de Jujuy

Teniendo en cuenta la Declaración de Pandemia de la Organización Mundial de la Salud sobre el brote de COVID-19 (coronavirus), importando un estado de emergencia de la salud pública internacional. El 12 de marzo el Gobernador de Jujuy CPN Gerardo Morales, decretó mediante Decreto 696-S/2020 la Emergencia Sanitaria y Epidemiológica en la provincia de Jujuy. Posteriormente el 15 del mismo mes la medida fue adoptada a Nivel Nacional quedando instaurado el aislamiento social preventivo y obligatorio.

En este sentido, se detallan algunas de las medidas adoptadas hasta la fecha:

El 10 de abril se implementa el uso de barbijo obligatorio. Posteriormente el día 15 de abril se activan sectores comerciales, como ser: Supermercados mayoristas y minoristas, farmacias, ferreterías, veterinarias, restaurantes, locales de comidas rápida (solo reparto a domicilio), servicio de construcción privada y afines, ópticas y actividades oftalmológicas, corralones, oficios varios, rubros vinculados con la actividad forestal y minera y actividades vinculadas al comercio exterior.

El 22 de abril inició en la provincia la “cuarentena administrada”, con circulación según la terminación del DNI par/impar de lunes a sábados, excluyendo los días domingos; como así también se permitió la actividad física como la caminata y el trote de no más de (02) personas, con distanciamiento social y uso de barbijo.

El 29 de abril el gobernador Gerardo Morales firma un decreto en el cual quedan habilitadas las actividades esenciales de la administración pública que tengan injerencia con la emergencia sanitaria, como así también se habilitan las visitas a internos privados de su libertad según terminación de DNI. Por último, se incorpora el ciclismo como nueva actividad física permitida. [3]

El 09 de mayo se habilita la actividad gastronómica en restaurantes y confiterías, como así también la habilitación de shopping y galerías. El 22 de mayo Jujuy es la primera provincia en habilitar el Turismo Interno. El 25 de mayo se produce la habilitación de gimnasios de lunes a sábados de 08:00 a 19:00 hs.

Ante el surgimiento de contagios covid-19, el 28 de mayo se dispuso la cuarentena obligatoria de las localidades de Fraile Pintado y Calilegua.

El 12 de junio se produce el primer deceso a causa del covid-19, en la Provincia. El 16 de junio ante la suba de casos se vuelven a Fase 1 Capital, Palpalá, Yala y Perico.

El día 15 de junio La Quiaca vuelve transitoriamente a Fase 1 de aislamiento obligatorio por el

término de 7 días, atento al preocupante escenario sanitario y epidemiológico instalado en la ciudad boliviana de Villazón, donde aumentaron significativamente los casos de COVID-19.

El 20 de junio la Provincia en su totalidad regresa a fase 1, a raíz de que se registraron (11) casos nuevos, permaneciendo en Fase 1 hasta el 06 de Julio donde se habilitan actividades recreativas en zonas amarillas; siendo estas: caminar, correr, ciclismo y deportes individuales como el golf, tenis y paddle.

El 22 de Julio producto del aumento de casos de Covid-19, se regresa a Fase 1 en toda la Provincia, mientras que el día 29 de Julio se habilitan actividades y servicios esenciales autorizados en las localidades de Libertador General San Martín, Fraile Pintado, Calilegua y Perico.

En el mes de agosto, la provincia se dividió en zonas según la cantidad de contagios y nivel de riesgo, llevándose a cabo a partir del día 09 de agosto.

Con motivo de la Fiesta Nacional de los Estudiantes, el día 21 de septiembre se toma la medida de suspender en todos los niveles y en toda la provincia las clases virtuales.

Debido al descenso de los casos positivos, el día 4 de octubre la provincia en su totalidad queda en "zona verde". El 21 de octubre se habilita el ingreso a la provincia en colectivos, aviones y vehículos particulares, como así también vuelve a funcionar el aeropuerto. Como incentivo en la industria comercial y de todos los sectores en general, el 26 de octubre se habilitó el turismo interno.

Recientemente el gobernador anunció el nuevo decreto, quedando sin efecto las salidas transitorias según terminación par/impar, estableciendo salida libre a partir del 09 de noviembre, finalizando la etapa de aislamiento obligatorio y entrando a la nueva etapa de distanciamiento social. Con el inicio de las clases presenciales con los debidos protocolos de bioseguridad.

Por último, el 14 de noviembre se dictaminó la "HABILITACIÓN DE DISCIPLINAS DEPORTIVAS CON PROTOCOLOS ESTRICTOS" [4]

Toma de Datos

Los datos de la Policía de la Provincia se centralizan en el Centro de Información y Análisis Criminal del Ministerio de Seguridad de Jujuy, el cual es el órgano homologado por la Dirección Nacional de Estadística Criminal del Ministerio de Seguridad de la Nación, para la reunión, procesamiento y difusión de Información Criminal.

Teniendo en cuenta la base de Datos Policiales que maneja el C.I.A.C., discriminamos los mismos mediante la tipología delictiva, teniendo que un 33,78% pertenecen a hechos Informativos , que no configuran delitos en sí (problemas vecinales, accidentes de tránsitos, incendios, suicidios entre otros), siendo un 32.80% Delitos contra la Propiedad, se eligió trabajar sobre estos tipos de delitos, porque son los que incluyen mayor relevancia en el contexto social y presentan mayor problemática urbana. Dentro de los Delitos contra la Propiedad, los que presentan mayor cantidad son Robo con un 12,08%, Hurto con un 11, 68% y Estafa con un 2.30%, eligiendo estos 3 delitos para este trabajo.

Selección de tema a Analizar

Para el presente, se eligió trabajar con los delitos contra la propiedad de mayor ocurrencia (Robos, Hurtos y Estafas), durante los meses en que duró la denominada socialmente "cuarentena" en la Provincia de Jujuy, tomando como periodo a analizar desde la segunda quincena de marzo hasta la primera quincena de noviembre del 2020, e igual periodo del año 2019 con los datos seleccionados de la base de datos del Centro de Información Criminal y Análisis Criminal.

¿Qué entendemos por Robo, Hurto y Estafa?

- ***Robo***: Hechos registrados como el apoderamiento de una cosa por violencia sobre la persona o el objeto, generalmente calificado los términos del Artículo 164 y subsiguiente del Código Penal, sin considerar la imputabilidad o culpabilidad del autor.
- ***Hurto***: Hechos registrados como el apoderamiento de una cosa sin violencia sobre la persona o el objeto, generalmente calificado los términos del Artículo 162 y subsiguientes del Código Penal, sin considerar la imputabilidad o culpabilidad del autor
- ***Estafa***: Hechos registrados como el que defraudare a otro con nombre supuesto, calidad simulada, falsos títulos, influencia mentida, abuso de confianza o aparentando bienes, crédito, comisión, empresa o negociación o valiéndose de cualquier otro ardid o engaño. Calificado en los términos del Artículo 172 y subsiguientes del Código Penal, sin considerar la imputabilidad o culpabilidad del autor. [5]

Aspectos a tener en cuenta

Para los datos de personas contagiadas por Covid-19, se utilizaron los datos disponibles en la página del Ministerio de Salud de la Nación, de la cual se extrajeron únicamente los datos de la provincia de Jujuy. A su vez también se efectuó una explotación digital de medidas sanitarias dispuestas por el Gobernador de la Provincia. [6]

La Organización Mundial de la Salud, define a los coronavirus como una extensa familia de virus que pueden causar enfermedades tanto en animales como en humanos. En los humanos, se sabe que varios coronavirus causan infecciones respiratorias que pueden ir desde el resfriado común hasta enfermedades más graves como el síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SRAS). El coronavirus que se ha descubierto más recientemente causa la enfermedad por coronavirus COVID-19.

El COVID-19 es la enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente. Tanto este nuevo virus como la enfermedad que provoca eran desconocidos antes de que estallara el brote en Wuhan (China) en diciembre de 2019. Actualmente COVID-19 es una pandemia que afecta a muchos países de todo el mundo.

Se define como "cuarentena", a la separación y restricción de movimientos de personas que estuvieron expuestas a una enfermedad infecciosa, pero que no tienen síntomas, para observar si desarrollan la enfermedad.

Esto se diferencia del "aislamiento", que es la separación de personas que padecen una enfermedad contagiosa, de aquellas que están sanas. Ambas medidas son estrategias de salud pública que tienen como objetivo prevenir la propagación de enfermedades contagiosas. Política que fue implementada tanto por el Gobierno Nacional como Provincial.

Análisis General de los fenómenos Criminológicos

En todo hecho criminal, existen tres factores que convergen para la concreción de un delito sea cual fuere. Estos son: una víctima, un victimario y un lugar.

Entiéndase como víctima a cualquier persona objeto o institución que se vea afectada por la comisión de un hecho delictivo. Victimario comprende a la persona o grupo de personas que cometen el delito, seleccionando una o varias víctimas según su modalidad de actuar y operar. En cuanto al lugar, nos referimos a todo medio ambiental físico o no, que permita la concreción entre la víctima y el victimario para producirse el delito.

Durante el periodo del Aislamiento Social, Preventivo y Obligatorio, la circulación libre de las personas se encontraba restringida, autorizada únicamente la salida de las personas consideradas con funciones esenciales para el gobierno.

Para poder trabajar los datos se estructuraron en una sola base dependiendo del lugar donde sucedieron los hechos, por lo cual se analizó la información de 75 Comisarías, tomando todos aquellos robos, hurtos y estafas consumados en cada uno de ellas, durante el periodo del 15 de marzo al 15 de noviembre, generando en el año 2020 un total de 5.846 observaciones y para el año 2019 un total de 7.968 observaciones:

	15Mar al 15Nov 2019	15Mar al 15Nov 2020	2019 vs 2020
Total, Registros Robos, Hurtos, Estafas	**7968**	**5846**	**-27%**

Tabla 1: Comparativos de delitos entre los años 2019 y 2020

En la tabla que precede, se denota que en el periodo en el cual duró el aislamiento social, estos tipos de delitos han disminuido un 27%, en comparación al año 2019, teniendo en cuenta esto, se comparó cada uno de los delitos.

	15Mar al 15Nov 2019	15Mar al 15Nov 2020	2019 vs 2020
Robo	**4196**	**2650**	**-36,8%**
Hurto	**3442**	**2565**	**-25,5%**
Estafa	**330**	**631**	**+91,2%**

Tabla 2: Comparativos de discriminados por tipo de Delitos entre los años 2019 y 2020

Como se observa, los delitos de Robo y Hurto han disminuido un 36.8% y un 25,5% respectivamente, en relación anual. Ello se explica debido a la restricción de circulación de personas, los victimarios no han podido consumar los diferentes hechos ante la ausencia de víctimas.

El delito de Estafas es el que ha aumentado en un 91,2%, en comparación al año anterior. El comportamiento de este tipo de delito, que en años anteriores es relativamente similar, en el año 2020, la pandemia ha generado favorecedores delictivos. Considerando que fue uno de los delitos más relevantes durante este tiempo, se decide centralizar más el análisis en esta tipología.

Para una mejor observación de los datos, se optó por unificar estas variables en semanas, desde el 15 de marzo al 15 de noviembre, obteniendo 35 semanas, de igual forma se trabajaron los datos de cada delito y los datos de contagios de Covid-19.

Desde una apreciación lógica, podemos decir que, al haber menos circulación de personas en las calles, el delito ocurrido en la vía pública debe de disminuirse, lo que se sustenta en el siguiente gráfico encontramos un diagrama lineal de los hechos, por el lugar de ocurrencia, comparados entre años 2019 y 2020.

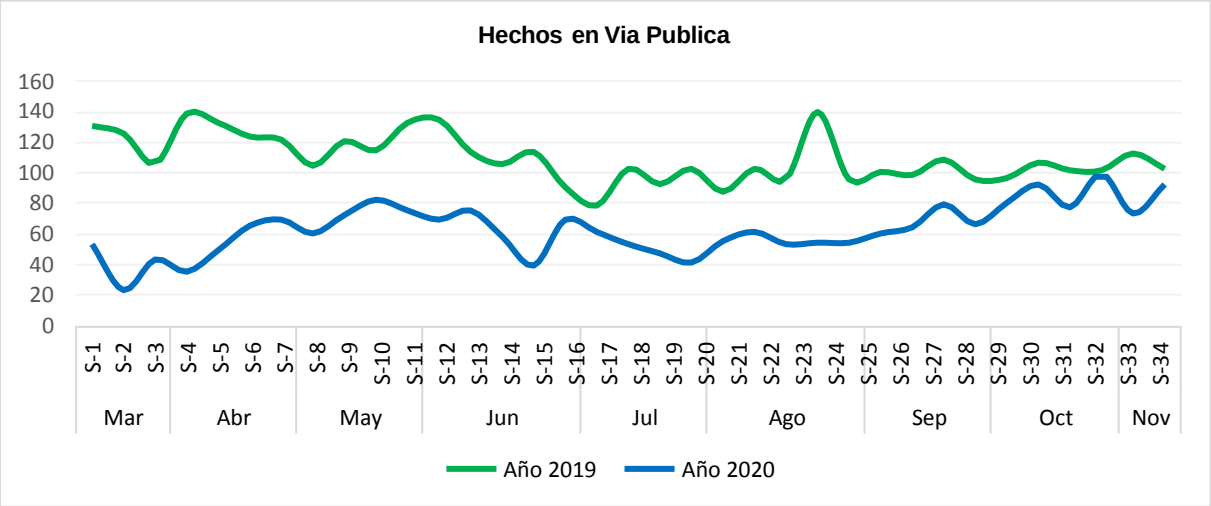


Figura 1: Comparativo Cronológico del comportamiento del delito en la Vía Publica durante el año 2019 y 2020

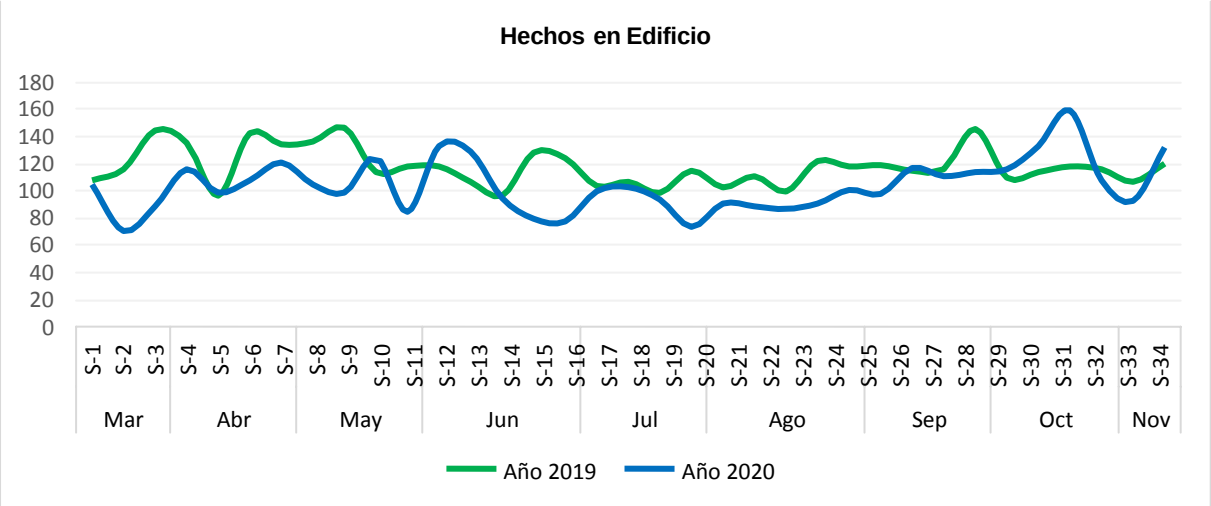


Figura 2: Comparativo Cronológico del comportamiento del delito en Lugares Cerrados durante el año 2019 y 2020

El comportamiento de los hechos varía de acuerdo al lugar de ocurrencia, observamos que los hechos en lugares cerrados “Edificios” (domicilio particular, edificios públicos, privados, etc.) se comportan relativamente similar al año anterior. Mientras que aquellos hechos ocurridos en la Vía Pública, se observa que esta clase de delitos disminuyó desde el 15 de marzo, los delitos vienen en baja, esto puede darse por el resguardo que tomaron los ciudadanos de no salir a la calle, lo que conllevó a la reducción del delito previo a lo anunciado por el gobernador, lo que se conoce como “beneficio anticipado”, en la prevención del delito.

En cuanto a los contagios registrados en nuestra Provincia, los datos suministrados por la página del Ministerio de Salud de Nación [7], tenemos que la masividad de contagios se dio a partir del 20 de Julio donde se empezó a registrar el alza en los contagios de Covid19, teniendo su pico máximo el 3 de septiembre.

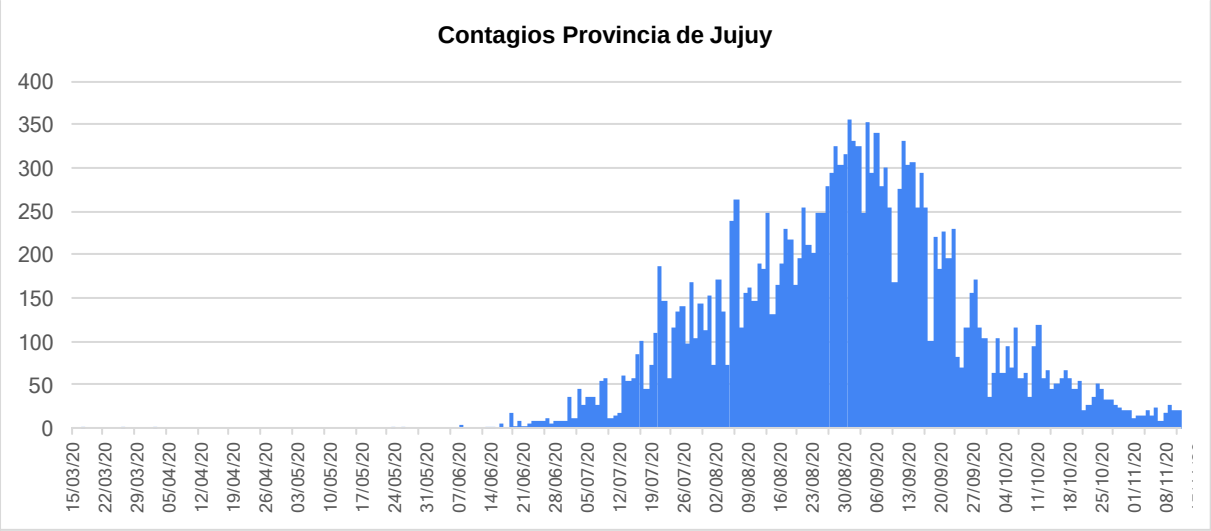


Figura 3: Grafico histórico de contagiados de COVID19 en la Provincia de Jujuy

Vemos que el delito de Estafa ha ido aumentando a gran escala a partir del mes de Mayo, en el mismo periodo en que los picos de contagios se fueron dando. Analizando la base de datos vemos que las Estafas que predominan son las estafas telefónicas, teniendo en cuenta el anuncio que hizo el presidente Alberto Fernández el 23 de Marzo, aludiendo que se abonará 10.000 pesos en carácter de ingreso familiar de emergencia (IFE), el cual apuntó a sostener la economía de los hogares que hayan sufrido un cese de ingreso debido a la emergencia nacional por la pandemia del coronavirus, a pagarse en el mes de Abril, para personas entre 18 a 65 años

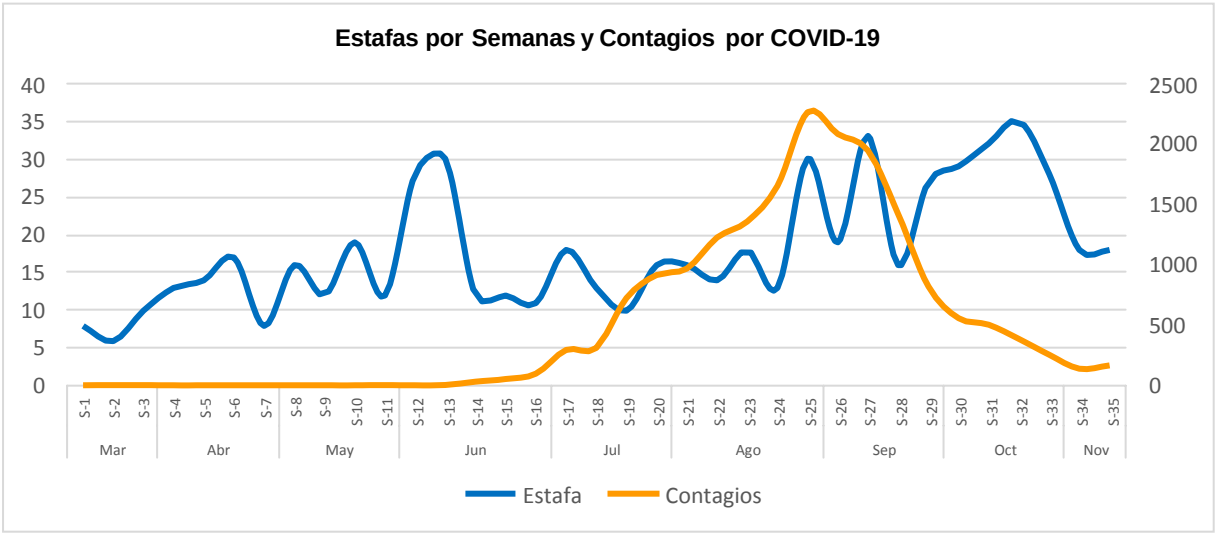


Figura 4: Grafico comparativo semanal entre Delitos de Estafas registrados y cantidad de contagios por Covid19

Este anuncio fue utilizado por los delincuentes como “carnada”, para efectuar estafas telefónicas. En el que, según el análisis efectuado con la base de datos, el modo de operar de los infractores

consistía en la selección de víctimas de entre 18 a 65 años (edad de las personas que gozan del beneficio del IFE), con mayor selección en víctimas de avanzada edad.

Al comparar los delitos de estafa con igual periodo del año 2019, identificamos que los primeros meses el comportamiento es similar en ambos años, empezando a notar un crecimiento en el mes de mayo, aumentando exponencialmente, llegando a su máxima expresión en el mes de octubre, decayendo luego considerablemente en el mes de noviembre. De continuar analizándose los meses subsiguientes, podría suponerse que las estafas se comportarían de igual forma que al inicio del año 2020 y 2019

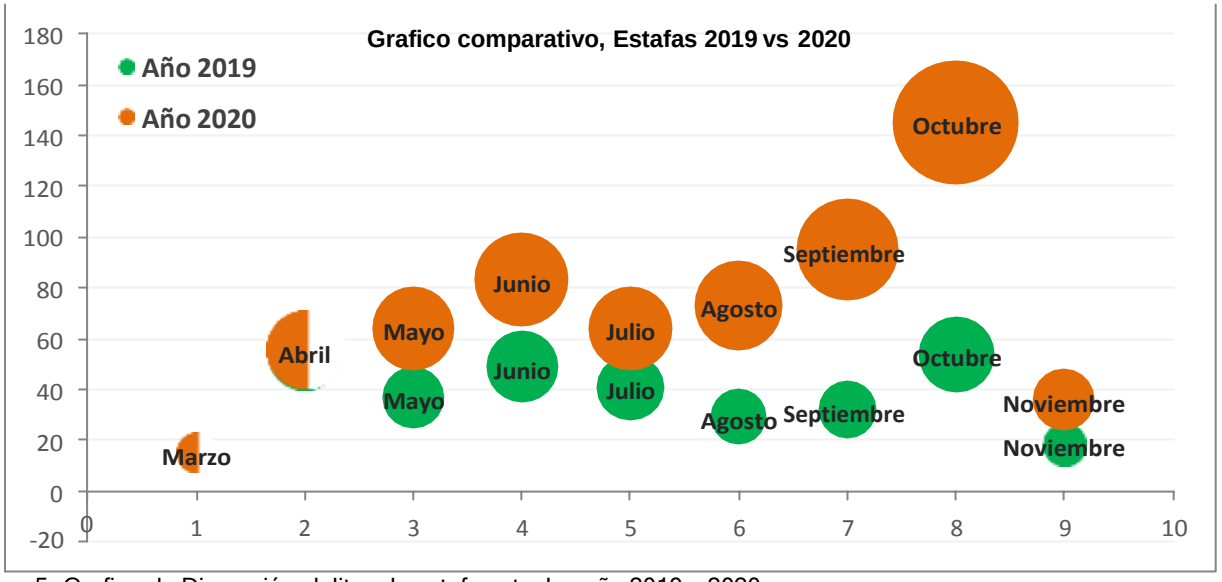


Figura 5: Grafico de Dispersión, delitos de estafa entre los año 2019 y 2020

Efectuado un cuadro con las variaciones relativas y absolutas de las Estafas, nos dio como resultado que la variación relativa de agosto a noviembre presenta un aumento superior al 100%.

Periodo	Año 2019	Año 2020	Variación Absoluta	Variación Relativa
Marzo	15	15	0	-
Abril	55	56	1	**+2%**
Mayo	37	64	27	**+73%**
Junio	49	83	34	**+69%**
Julio	41	64	23	**+56%**
Agosto	29	73	44	**+152%**
Septiembre	32	95	63	**+197%**
Octubre	54	145	91	**+169%**
Noviembre	18	36	18	**+100%**

Tabla 3: Comparativo Mensual de delitos de Estafa entre el año 2019 y 2020

Al igual que la industria del comercio, el marketing, la recreación, etc.; fueron adaptándose al nuevo escenario creado a partir del Aislamiento Social, la comisión de hechos delictivos, tuvo que adaptarse a nuevas modalidades y tipos. Esto según la elección racional de cada persona enfocada a cometer

un crimen.

En cuanto a las estafas, podemos decir que, al no necesitar de un espacio físico para consumarse, se produjo un notable crecimiento de las mismas durante el contexto de encierro, al no tener lugar físico se quiso verificar si las estafas por departamento se modificaron en relación al año anterior.

Departamento	Año 2019	Frecuencia Relativa	Año 2020	Frecuencia Relativa
Cochinoca	7	2%	7	1%
D M Belgrano	188	57%	303	48%
El Carmen	42	13%	65	10%
Humahuaca	8	2%	24	4%
Ledesma	24	7%	54	9%
Palpalá	18	5%	72	11%
Rinconada	-	-	-	-
San Antonio	-	-	3	-
San Pedro	24	7%	73	12%
Santa Barbara	1	0%	4	1%
Santa Catalina	1	0%	-	-
Suques	1	0%	2	0%
Tilcara	3	1%	7	1%
Tumbaya	3	1%	5	1%
Valle Grande	-	0%	-	0%
Yavi	10	3%	12	2%

Tabla 4: Comparativo por Departamento, delitos de Estafa años 2019 y 2020

Para determinar se calculó la frecuencia relativa de las estafas por departamento para cada año, observándose que la mayoría de los departamentos no presentaban variaciones considerables, el aumento si se dio efectivamente en los departamento de Palpalá y San Pedro, donde si se comparan las frecuencias relativas, presentan un aumento de un 50%.

Otras variables que se analizaron fueron los días y horarios en que se perpetraron los ilícitos

Cuadro de doble entrada: días y horarios 2020

	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
LUNES	1			1				2	3	3	5	7	11	5	2	4	6	2	4	2	2	3		
MARTES									1	3	6	3	9	5	2	2	7	6	2	6	2		1	2
MIERCOLES							1		2	5	4	7	6	7	7	5	4	4	2	1	3		1	2
JUEVES	1						1	2	5	1	5	8	7	6	3	4	5	6	2	3	5	2		2
VIERNES			1						5	8	8	4	8	3	7	4	4	2	4	6	1	1		2
SABADO		1							1	2	3	3	6	2	3	1		2	3	1	5	2		
DOMINGO	1								1	1		2	4	1	1	2		1	1	1	3	1		

Figura 6: Grafico de Doble entrada, días y horarios afectados en los delitos de Estafa en el año 2020

Cuadro de doble entrada: días y horarios 2019

	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
LUNES		1	1	1				1	1	3	8	6	4	1	6	4	3	3	6	1	1	1	1	1
MARTES		1								3	7	6	2	2	1	4	3	5	1	1		1		
MIERCOLES	1				1				2	2	2	12	6	2		4	3	1	2	3	1	3	1	
JUEVES				1			2		1	6	4	4	4	2	1	5	1	1	2	5	3	1	1	2
VIERNES			1							3	1		4	2	3	2	3	2	2	3	1	1		
SABADO	1										3	2	4	1		1	2	1	2	3	3	2		
DOMINGO	2					1					1	1	4	2			1	1	1		1			

Figura 6: Grafico de Doble entrada, días y horarios afectados en los delitos de Estafa en el año 2019

Al usar un formato condicional, vemos la agrupación de hechos, en el año 2019 la agrupación de hechos se dio principalmente los días Lunes a Jueves entre los horarios de 09:00' a 12:00', reduciéndose considerablemente hacia el fin de semana. Mientras que durante la pandemia los hechos por este delito se han hecho más extensos, tanto en días como en horarios, abarcando los días Lunes a Sábados, en el horario de 08:00' a 13:00'.

Ahora bien, los tipos de estafas y víctimas seleccionadas fueron estudiadas cuidadosamente por cada victimario. Siendo estas Estafas telefónicas a personas de sexo femenino un 48% y un 52% al sexo masculino durante el 2020. En cuanto a las edades, vemos que los hombres más afectados son aquellos que se encuentran entre en los rangos de 40 a 49 años, mientras que en las mujeres se distribuye de forma casi similar en los rangos de edad que van de 30 a 59 años. Por otro lado, este tipo de delitos afecta mucho menos a los jóvenes adultos, esto debido a la facilidad para el uso de las tecnologías y el acceso a la información, lo que permite que sean menos propensos a caer en estafas, y posean más herramientas al momento de identificar una estafa.

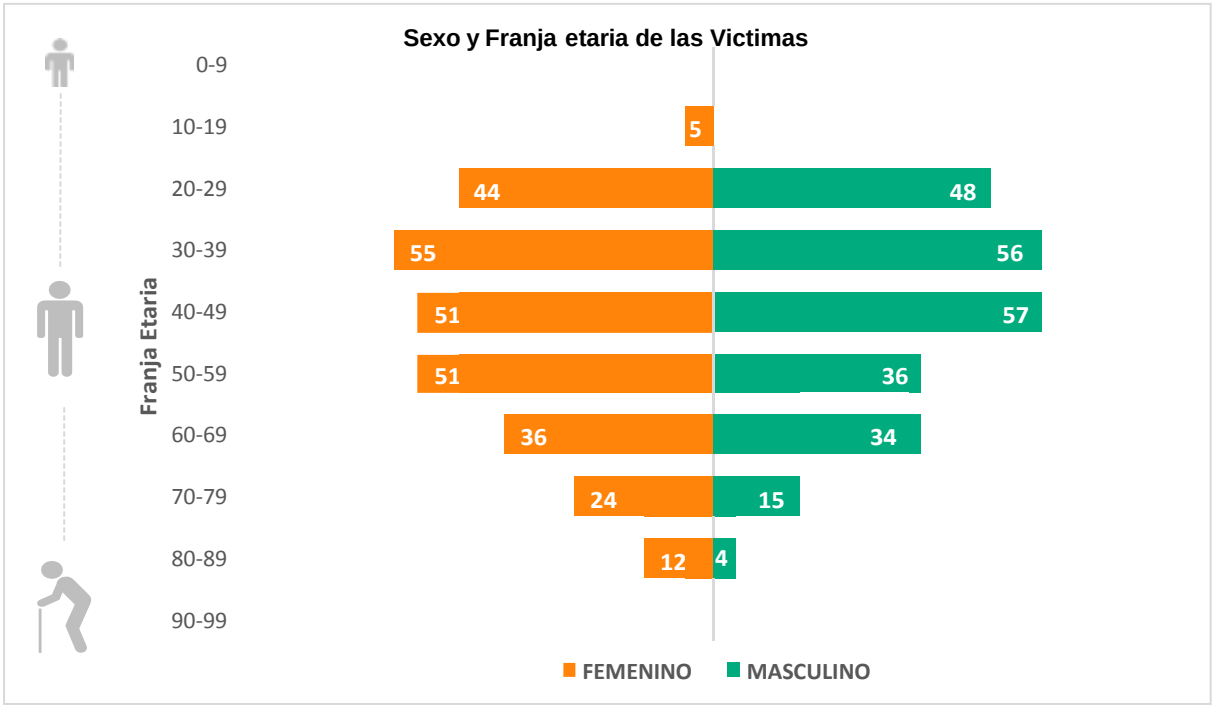
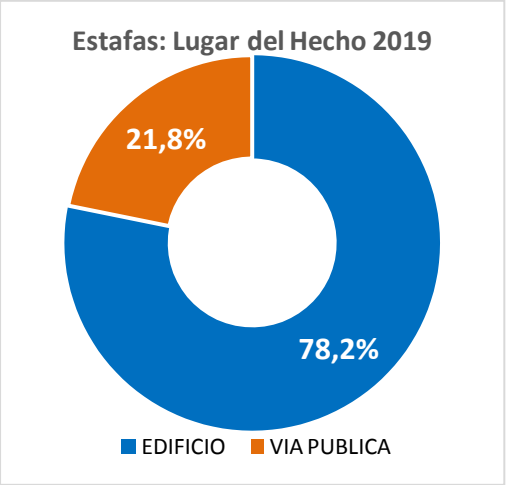


Figura 7: Grafico de barras agrupadas, granja etaria y sexo de las víctimas de Estafas 2020

En cuento al lugar del hecho en las estafas en comparación al año anterior, si bien la mayoría de las estafas ocurren en lugares cerrados y muy poca cantidad en lugares abiertos, pero durante el año 2020, esta diferencia se ha incrementado a aun más.



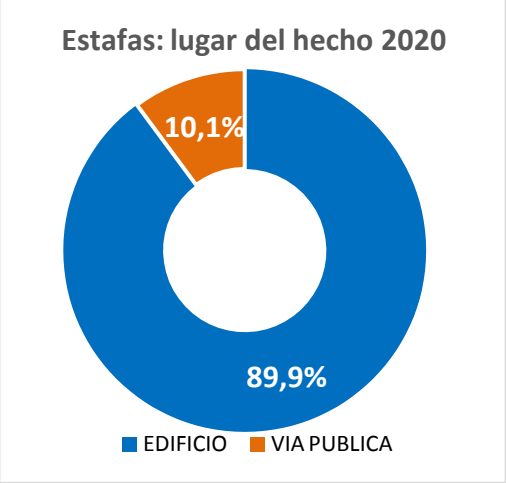


Figura 8a y 8b: Grafico de Anillos, de lugares donde ocurren los hechos de estafa en los años 2019 y 2020

CONCLUSIONES

Se comprobó una correlación delictual entre el espacio-tiempo trabajado en el presente análisis, dado que durante el confinamiento obligatorio el Delito de Estafa, lideró el Rankin de los hechos criminales, siendo a través del uso de la tecnología de la comunicación. Esto se refleja en la adaptabilidad de la delincuencia ante los nuevos entornos que el mundo impone.

Por lo que, a partir de la identificación de las nuevas conductas criminales, se propone paralelamente trabajar sobre las formas de Prevención Policial, orientadas a los nuevos escenarios digitales que tienen repercusión durante el confinamiento, y de esta manera estar preparados ante otros delitos que pudieren surgir en una segunda etapa de Aislamiento Social a raíz de un rebrote de Covid-19.

El análisis desagregado de datos y el entrecruzamiento con otras bases de información, permite identificar patrones y tendencias que datos generales no permiten identificar. Pudiendo generar medidas preventivas, específicas para las personas afectadas y los lugares con mayor incidencia.

Cada sociedad tiene una problemática en particular, una distinta de otra según su conformación y realidad, la cual debe ser abordada desde su interpretación más exacta en base a los datos obtenidos de las fuentes confiables, que permitirán posteriormente la difusión de información a raíz de su procesamiento, para la implementación de Políticas Públicas en materia de Seguridad.

BIBLIOGRAFÍA

[1] Patricio Tudela Poblete “Análisis delictual: enfoque y metodología para la reducción del delito”, Fundación Paz 2010, Pag 47 y 48.

[2] Antonio Bargas Sabadia “Estadística Descriptiva e Inferencial”, Universidad de Castilla 1996, Pag. 45, 46 y 48.

[3] Jujuy, C, 2020, “Ultimas Noticias”, Obtenido de COE Jujuy: http://coe.jujuy.gob.ar/noticias/.

[4] Jujuy, P.D. 2020, “Jujuy Energía Viva”, Obtenido de Decreto y Resoluciones: http://jujuy.gob.ar/home/

[5] Ley 11.179 “Código Penal de la Nación Argentina”. Buenos Aires Argentina. 1984.

[6] Organización Mundial de la Salud, 2019, “Organización Mundial de la Salud”. Obtenido de Preguntasy Respuestas sobre la enfermedad por coronavirus (COVID-19): https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses.

[7] Ministerio de Salud Argentina 2020 “Datos Abiertos del Ministerio de Sadu”. Obtenido de COvid-19. Casos registrados en la República Argentina: http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina.

10 y 11 de Diciembre de 2020

Procedimiento para detección de casos de COVID19 en poblaciones finitas

Autores: Rodriguez, Héctor Iván / Jakulica, Ricardo / Mautino, Gisella

Instituciones:
Facultad de Ingeniería, Universidad Nacional de Salta.
IIDISA, Instituto de Ingeniería y Desarrollo Industrial de Salta.
DI3 Doctorado Interinstitucional de Ingeniería Industrial.

Datos de contacto:
ivan@ing.unsa.edu.ar - Cel +5493874129731
rjakulica@gmail.com – Cel +54 93875080758
gmautino@ing.unsa.edu.ar – Cel +5493884440566

RESUMEN.

En este documento se describe un procedimiento de muestreo para poblaciones finitas, constituidas por personal de empresas o instituciones, estableciendo la cantidad de personas a ser hisopadas, y realizar las inferencias estadísticas sobre el porcentaje de positivos COVID 19, estimado por intervalo de confianza mediante aproximaciones sucesivas, con el fin de tomar las medidas que resulten necesarias en función de los resultados obtenidos.

Palabras Claves: Muestreo. Hisopado. COVID 19. Distribución Hipergeométrica. Inferencia estadística por intervalo de confianza.

INTRODUCCION:

Las empresas e instituciones que manejan un volumen importante de personas en relación de dependencia, directa o indirecta, se vieron en la necesidad de tomar decisiones acordes a la gestión en pandemia COVID 19, a efectos de evitar el contagio del personal y la propagación del

virus hacia la comunidad. El hisopado al personal constituyó una forma de conocer si había contagiados a efectos de tomar las medidas pertinentes, pero esta actividad de hisopado constituye no solo un recurso escaso sino también costoso. Es entonces donde se requiere un método de muestreo que permita la inferencia reduciendo cantidades de personas a ser hisopadas. Se presenta en este documento dicho procedimiento de muestreo e inferencia estadística para poblaciones finitas constituidas por personal de empresas o instituciones.

METODOLOGIA:
Describir el método de muestreo económico sobre una población finita y el procedimiento para realizar inferencias sobre el porcentaje de positivos, estimado por intervalo de confianza mediante aproximaciones sucesivas.

DESARROLLO:

Elección de la distribución de muestreo
Dado que se trata de una población finita, se recurre a la distribución Hipergeométrica para calcular probabilidades y poder determinar el tamaño muestral.
***N**: Es el tamaño de la población finita.*
***C**: Es la cantidad de casos positivos Covid19 en la población.*
***n**: Es el tamaño de la muestra elegida en la población finita.*
***x**: Es la variable aleatoria que representa la cantidad de casos positivos en la muestra n.*
Condiciones:
N ∈ {1, 2, 3, 4,.....} Pertenece al campo de los números enteros y es cantidad finita.
C ∈ {1, 2, 3, 4,.....N}
n ∈ {1, 2, 3, 4, ... N}
x ∈ max{(0, n+C-N),…,min(C,n)}. Esto significa que el valor de x no puede superar la cantidad real de casos positivos en la población (C), ni tampoco ser mayor que la población misma N. El caso más extremo sería que x puede tomar valores desde 0 hasta N cuando toda la población este infectada, o sea N=C.

Función de Probabilidad:

$$P(x) = \frac{\binom{N-C}{n-x}\binom{C}{x}}{\binom{N}{n}}$$

Media:

$$E(x) = \frac{nC}{N}$$

Desvío Estándar:

$$\sigma(x) = \sqrt{\frac{(C/N)(1-(C/N))(N-n)(N}{-1)}}$$

Procedimiento para el cálculo del tamaño de muestra
1) Se construye una tabla de distribución de probabilidades p(X≥1) para los distintos valores de n y C.

2) Se grafica dicha tabla para construcción del Abaco.

3) Se selecciona el valor C: para ello se fija un criterio razonable del número de contagios que se desea detectar, este valor debe ser fijado de acuerdo con el grado de agresividad en la velocidad de contagio del virus, información que es aportada por los centros de estudios respectivos. Por ejemplo, en el caso de un grupo perteneciente a una empresa, si hay contagiados en el grupo, es más probable que haya 5 o más que uno solo o menos de 5. Este es un criterio importante para reducir el tamaño de muestra.

4) Se fija la probabilidad deseada. Es un valor que está en el orden de los porcentajes de falsos positivos o negativos.

5) Con los valores de C y Probabilidad de detección se extrae del Abaco el tamaño de muestra.

Ejemplo: Sobre una población de 195 con un solo caso positivo, se desea calcular la probabilidad de detectarlo con una muestra de 20 casos.

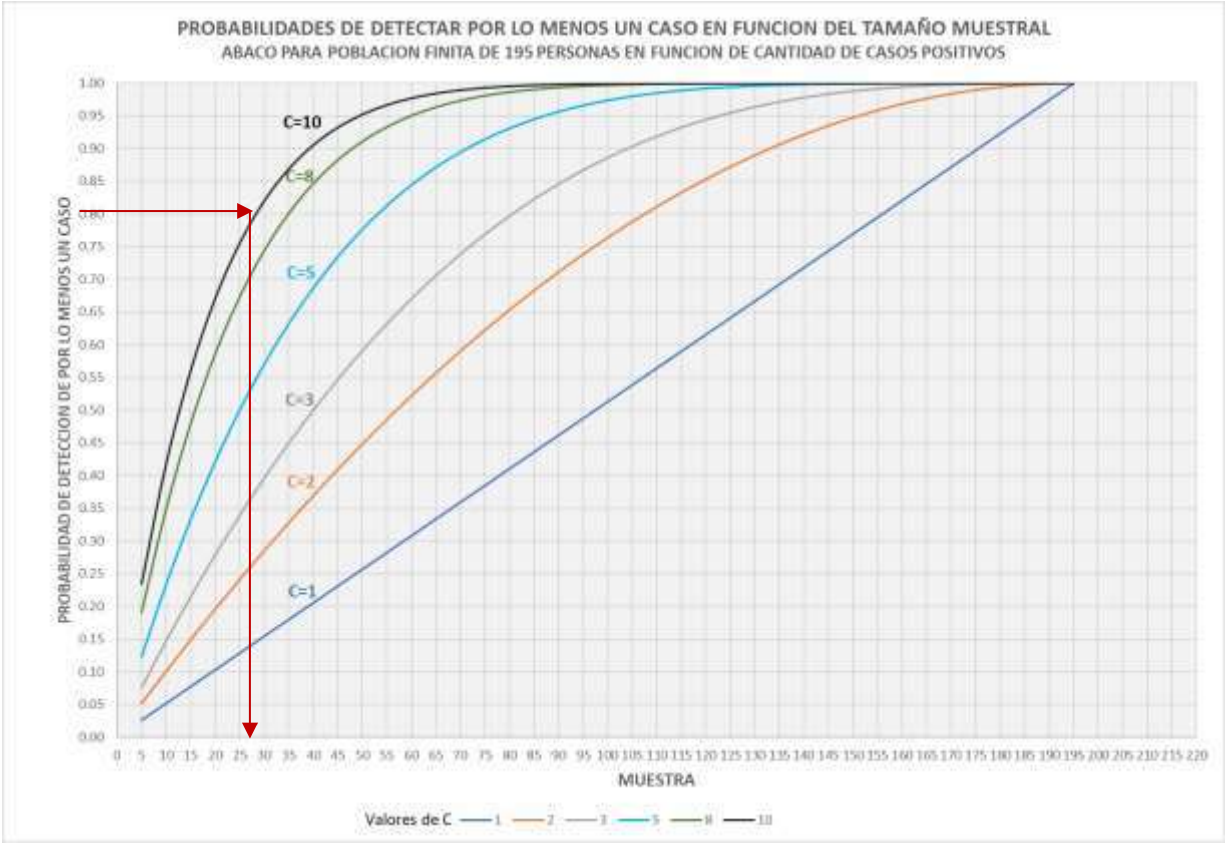
N	195
C	1
X	0
n	20

Calculamos con la distribución hipergeométrica la probabilidad de no detectarlo, luego por el complemento encontramos la probabilidad de detectarlo.

p(k=0)	0.897
p(k>=1)	**0.103**

De esta manera construimos la tabla de distribución de probabilidades **p(k>=1)** para todos los casos y la graficamos.

p(X≥1)						
n	C					
5	0.026	0.051	0.075	0.123	0.191	0.234
10	0.051	0.100	0.147	0.234	0.349	0.417
15	0.077	0.148	0.214	0.333	0.479	0.560
20	0.103	0.195	0.278	0.421	0.586	0.670
25	0.128	0.241	0.339	0.500	0.673	0.755
30	0.154	0.285	0.396	0.570	0.744	0.820
35	0.179	0.328	0.449	0.632	0.801	0.869
40	0.205	0.369	0.500	0.687	0.847	0.905
45	0.231	0.409	0.547	0.735	0.883	0.932
50	0.256	0.448	0.591	0.777	0.911	0.952
55	0.282	0.486	0.632	0.813	0.933	0.967
60	0.308	0.522	0.670	0.845	0.951	0.977
65	0.333	0.557	0.706	0.872	0.964	0.985
70	0.359	0.590	0.739	0.895	0.974	0.990
75	0.385	0.623	0.769	0.915	0.981	0.993
80	0.410	0.653	0.797	0.931	0.987	0.996
85	0.436	0.683	0.823	0.945	0.991	0.997
90	0.462	0.711	0.846	0.957	0.994	0.998
95	0.487	0.738	0.867	0.966	0.996	0.999
100	0.513	0.764	0.886	0.974	0.997	0.999
105	0.538	0.788	0.903	0.980	0.998	1.000
110	0.564	0.811	0.919	0.985	0.999	1.000
115	0.590	0.833	0.932	0.989	0.999	1.000
120	0.615	0.853	0.945	0.992	1.000	1.000
125	0.641	0.872	0.955	0.995	1.000	1.000
130	0.667	0.890	0.964	0.996	1.000	1.000
135	0.692	0.906	0.972	0.998	1.000	1.000
140	0.718	0.921	0.978	0.998	1.000	1.000
145	0.744	0.935	0.984	0.999	1.000	1.000
150	0.769	0.948	0.988	0.999	1.000	1.000
155	0.795	0.959	0.992	1.000	1.000	1.000
160	0.821	0.969	0.995	1.000	1.000	1.000
165	0.846	0.977	0.997	1.000	1.000	1.000
170	0.872	0.984	0.998	1.000	1.000	1.000
175	0.897	0.990	0.999	1.000	1.000	1.000
180	0.923	0.994	1.000	1.000	1.000	1.000
185	0.949	0.998	1.000	1.000	1.000	1.000
190	0.974	0.999	1.000	1.000	1.000	1.000
195	1.000	1.000	1.000	1.000	1.000	1.000



Los criterios fijados fueron de C=8 con una probabilidad de 0.80. lo que determina un tamaño muestral de n=35. Además, este tamaño de muestra cubre con 0.90 de probabilidad si hubiere 10 casos, que es lo sospechado podría haber.

INFERENCIA ESTADISTICA POR INTERVALO DE CONFIANZA

El segundo paso, después de tomada la muestra, es realizar en base a los casos positivos observados, una inferencia sobre la cantidad de casos positivos en la población.

Para eso se realiza dos tipos de estimaciones, una aplicando el teorema de la desigualdad de Chebyshev y otro por aproximación de la hipergeométrica a la Binomial. Luego se revisa la consistencia entre ambos para una estimación puntual.

Teorema de Chebyshev

Establece que la probabilidad que cualquier variable aleatoria X tome un valor dentro de *k* desviaciones estándar de la media es al menos *1-1/k*^{2}.

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Probabilidad y estadística para ingeniería y ciencias, Ronald E. Walpole, Raymond H. Myers, Sharon l. Myers y Keying Ye. (2012). Probabilidad y estadística para ingeniería y ciencias, 8ª Edición. Ed. Pearson educación. Pág. 132

Inferencia aplicando el teorema de Chebyshev
Procedimiento:

1. De la muestra obtener los "x" números de casos favorables.
2. Construir una tabla de valores de "C" y su desvío estándar calculado con la distribución hipergeométrica
3. Asignar un valor de "k" coincidente con la probabilidad de confianza deseada.
4. Completar la tabla para cada valor de "C" el intervalo de Chebyshev.
5. El intervalo buscado es aquel formado por el mínimo y máximo de todos los valores de "C" que generan intervalos que contienen a "x" del punto 1)

En nuestro ejemplo:
N=195
n=35 (valor seleccionado en el cálculo de tamaño muestral)
x=7 (son los casos positivos en la muestra n)

k=	2
Probabilidad de Chebyshev	0.75

Frecuencia relativa o proporción de casos positivos en la muestra n
p = x/n
p = 0.20

Intervalos de Chebyshev				
C	**Desvío**	**LI**	**LS**	**contiene x=7**
5	0.849	-0.80	2.60	no
10	1.185	-0.58	4.17	no
15	1.432	-0.17	5.56	no
20	1.630	0.33	6.85	no
25	1.796	0.89	8.08	si
30	1.938	1.51	9.26	si
35	2.062	2.16	10.41	si
40	2.169	2.84	11.52	si
45	2.264	3.55	12.60	si
50	2.346	4.28	13.67	si
55	2.418	5.04	14.71	si
60	2.480	5.81	15.73	si
65	2.533	6.60	16.73	si
70	2.577	7.41	17.72	no
75	2.614	8.23	18.69	no
80	2.643	9.07	19.64	no

Con una probabilidad superior a 0.75 el valor de casos positivos en la población finita esta entre 25 y 65 casos.

Inferencia por intervalos de confianza aplicando la aproximación de la distribución hipergeométrica a la binomial.

Como la muestra es mayor de 30 podemos aplicar el teorema de límite central para aproximar la varianza poblacional con la muestral y calcular los valores de z con la distribución normal, caso contrario usaríamos la distribución t de student.

Intervalo de confianza:

$$x - S.z_{\frac{a}{2}} \leq C \leq x + S.z_{a/2}$$

S^2=n.(x/n)(1-x/n).(N-n)/(N-1)

S	5.07266782
alfa	0.05
z=	1.96

Intervalo por aproximación Binomial		
29	≤ C ≤	49

Este intervalo está contenido en lo encontrado con Chebyshev.

Estimación Puntual

c= N.(x/n)

c = 195.(7/35)

c = 39

También verificamos que este valor está contenido en el intervalo de confianza.

Las estimaciones mediante este método se verifican con la metodología utilizada.

CONCLUSIONES:

Con el procedimiento desarrollado, se trabaja con una población finita, por ello se recurre a la distribución Hipergeométrica para calcular probabilidades. Luego se describe el procedimiento para el cálculo del tamaño de muestra, teniendo en cuenta la cantidad de casos positivos C. Este valor debe ser fijado de acuerdo con el grado de agresividad en la velocidad de contagio del virus, información que es aportada por los centros de estudios respectivos.

Se fija la probabilidad deseada, que es un valor que está en el orden de los porcentajes de falsos positivos o negativos. Con los valores de C y probabilidad de detección, se extrae del Abaco el tamaño de muestra.

El siguiente paso, después de tomada la muestra, es realizar en base a los casos positivos observados, una inferencia sobre la cantidad de casos positivos en la población.

Para eso se realizan dos tipos de estimaciones, una aplicando el teorema de la desigualdad de Chebyshev y otro por aproximación de la hipergeométrica a la Binomial. Finalmente, revisamos la consistencia entre ambos mediante una estimación puntual. Verificando que las estimaciones realizadas mediante esta metodología, son consistentes.

BIBLIOGRAFIA:

Probabilidad y estadística para ingeniería y ciencias, Ronald E. Walpole, Raymond H. Myers, Sharon l. Myers y Keying Ye. (2012). Probabilidad y estadística para ingeniería y ciencias, 9ª Edición. Ed. Pearson educación

10 y 11 de Diciembre de 2020

Perfil y actitud de los estudiantes de posgrado hacia el curso de estadística: Caso Universidad Nacional de San Antonio Abad del Cusco

Yheni Farfán Machaca[1], Elba Vega Durand[2]

[1]Facultad de Ciencias, Universidad Nacional de San Antonio Abad del Cusco. Cusco.
[2]Escuela de Posgrado, Universidad Nacional de San Antonio Abad del Cusco. Cusco

yheniosterreich@gmail.com. Teléfono 0051967492060

RESUMEN

Esta investigación es de naturaleza aplicada, con enfoque cuantitativo, diseño no experimental, transversal, descriptivo correlacional a nivel multivariado, tuvo como unidad de análisis al estudiante de posgrado de la Universidad nacional de San Antonio Abad del Cusco que llevó el curso de Estadística, a quienes se les aplicó un encuesta en forma virtual mediante el Formulario de Google, cuyos objetivos fueron determinar fueron determinar el perfil de los estudiantes y los factores latentes de la variable actitud; para tal efecto se seleccionó una muestra de tamaño 160, la confiabilidad del instrumento previamente validado, se midió con el estadístico Alfa de Cronbach, resultando 0,802, conteniendo 14 ítems para la actitud y cuatro preguntas socio demográficas, habiéndose encontrado que la mayoría de los estudiantes eran del género femenino (55,6%), de entre las edades de 25 a 29 años(45%), con título profesional (71,9%) y procedían del Departamento del Cusco(44,4%). Los datos de la variable actitud fueron procesados mediante el Análisis Factorial por componentes principales, determinándose 4 factores que explican el 65,08% de variabilidad total de la variable actitud: estos son: un factor inseguridad del estudiante, un factor curso de alto nivel, un factor curso no aplicable y un factor curso no útil.

Palabras Claves: estudiantes, estadística, actitud, análisis multivariado

INTRODUCCIÓN

Siendo la Estadística, ciencia que tiene aplicación en todos los campos del desarrollo profesional es que se ve la necesidad de que todos los estudiantes de las maestrías de la Escuela de posgrado debieran manejar en forma eficaz las herramientas y /o técnicas estadísticas que permiten analizar un conjunto de datos, pero ese dominio de la Estadística depende de muchos factores, siendo uno de ellos el tratado en esta investigación que es, la actitud que muestra el profesional hacia la aplicación de las herramientas estadísticas en las investigaciones que tienen que realizar llámese: tesis, artículos, etc.

Entre las investigaciones se tiene a Tarazona, Bazán, & Aparicio (2013) que en su investigación señalan que la actitud hacia la estadística se muestra en forma diferenciada por especialidad del estudiante, es decir los estudiantes de ingeniería industrial muestran actitudes más positivas hacia la Estadística que los de Ingeniería de Telecomunicaciones y redes los cuales muestran actitudes menos positivas. Así como los autores Comas, Martins, Nascimento, & Estrada (2017), indica que las actitudes hacia la Estadística en alumnos de Psicología, en general fueron moderadas o positivas, que las mujeres muestran peores actitudes y sobre todo que la actitud global hacia la estadística se empeora con los años de estudio de la misma es decir en estudiantes de posgrado, posiblemente porque encuentran mayores dificultades con el tema. En esta misma línea de investigación Darias (2000), utlizando una muestra de estudiantes de la Universidad de La Laguna, trató de validar la Escala de Actitudes hacia la Estadística (EAE); así como aplicando una de las técnicas del análisis multivariado, el análisis factorial de componentes principales con rotación varimax encontraron cuatro factores fundamentales, éstos son: un factor seguridad, un factor importancia, un factor utilidad y un factor deseo de saber. Otro trabajo de investigación realizado por Vilá & Rubio (2014), quienes midieron las actitudes de los estudiantes de Pedagogía de la Universidad de Barcelona hacia la Estadística, a través de la escala EAE, cuyos resultados mostraron niveles neutros – bajos de actitud general hacia la estadística, identificaron grupos de estudiantes con perfiles diferenciados: un grupo con una actitud desfavorable, pero no especialmente ansioso, con apenas conocimientos previos de Estadística y muy pocas habilidades numéricas; otro grupo caracterizado por una actitud positiva, al que le gustaba la Estadística, pero que muestra preocupacióny ansiedad; un tercer grupo que destacaba por la ansiedad ante la Estadística, es asi que dada la potencial relación entre la actitud y logro académico, concluyeron que existía la necesidad de explorar estrategias de enseñanza adaptativa para los diferentes grupos de la Universidad.

En este sentido, esta investigación se modela un constructo a fin de investigar el perfil y la actitud de los estudiantes de la Escuela de posgrado hacia la Estadística.

METODOLOGÍA

Esta investigación se realizó en la Universidad Nacional de San Antonio Abad del Cusco, con los estudiantes de la escuela de posgrado en el mes de noviembre del presente año 2020 para conocer su perfil y los factores latentes de la variable actitud, se utilizó un cuestionario con preguntas sociodemográficas y de actitud hacia el curso de Estadística, determinándose el nivel más frecuente de las características de información socio demográficas, así como para determinar los factores latentes o fundamentales de la variable actitud, se aplicó el Análisis factorial de componentes principales conteniendo la prueba KMO de adecuación de la muestra y la prueba de Esfericidad de Bartlett de la matriz de correlaciones de los 14 ítems; la extracción de 4 factores latentes con el método de rotación varimax.

La técnica utilizada fue por medio de una encuesta virtual, utilizando el formulario de Google, se encuestó a los estudiantes que llevaron el curso de Estadística en la escuela de posgrado. El instrumento utilizado fue validado y posteriormente se realizó su correspondiente confiabilidad, cuyo resultado fue de 0,802 de acuerdo con el Estadístico Alfa de Cronbach, el cual es confiable. El Instrumento de la actitud fue adecuado de Santabárbara & López (2019), con modalidades tipo likert, estos son: 1 es totalmente en desacuerdo, 2 es en desacuerdo, 3 es ni en desacuerdo, ni de acuerdo, 4 es de acuerdo y 5 es totalmente de acuerdo.

DESARROLLO

Se elaboró una matriz de datos de 160 estudiantes de posgrado que llevaron el curso de estadística, cuyo procesamiento se realizó con el software estadístico SPSS v.25, versión prueba con disponibilidad para el procesamiento de las tablas de frecuencia de los datos socio demográficos, así como para el modelo multivariante análisis factorial de componentes principales obteniéndose el perfil de los mencionados estudiantes, así como los factores latentes de la variable actitud.

a. Perfil de los estudiantes de posgrado

Tabla 1: Variables socio demográficas de los estudiantes de posgrado

Variables socio demográficas	Modalidad más frecuente	Frecuencia	Porcentaje
Sexo	Femenino	89	55,6
Edad	De 25 a 29 años	72	45,0
Grado de instrucción	Título profesional	115	71,9
Lugar de residencia	Cusco	71	44,4

En la tabla 1 se muestra que los estudiantes de posgrado de la Universidad Nacional de San Antonio Abad del Cusco, mayoritariamente fueron del sexo femenino (55,6%), de entre las edades de 25 a 29 años (45%), con título profesional (71,9%) y que residían en la ciudad del Cusco.

b. Factores fundamentales de la variable actitud hacia el curso de Estadística

Tabla 2: Prueba KMO y de esfericidad de Bartlett de la actitud

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0,709
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	940,364
	gl	91
	Sig.	0,000

La tabla 2 nos muestra la prueba KMO = 0.709, nos indica que la muestra es adecuada para el estudio. La prueba de esfericidad de Bartlett con significación = 0.000 < 0.01, nos indica que existe suficiente evidencia para concluir que la matriz de correlación R ≠ I, es decir contiene correlaciones de pares de variables altamente significativas y por tanto es conducente, el análisis factorial.

Tabla 3: Matriz de varianza explicada de la actitud de los estudiantes

Varianza total explicada									
	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
Componente	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,678	33,416	33,416	4,678	33,416	33,416	3,745	26,753	26,753
2	1,693	12,096	45,513	1,693	12,096	45,513	2,497	17,837	44,590
3	1,498	10,699	56,212	1,498	10,699	56,212	1,527	10,909	55,499
4	1,241	8,866	65,078	1,241	8,866	65,078	1,341	9,579	65,078
5	0,908	6,485	71,563						
6	0,850	6,072	77,635						
7	0,709	5,065	82,699						
8	0,581	4,150	86,849						
9	0,526	3,760	90,609						
10	0,422	3,013	93,622						
11	0,301	2,153	95,775						
12	0,263	1,875	97,651						
13	0,187	1,333	98,984						
14	0,142	1,016	100,000						

Nota: Método de extracción: análisis de componentes principales. Fuente: elaboración propia

En la tabla 3 se observa los porcentajes de varianza total explicada por los 4 factores a retener, es del 65,078 %, es decir en promedio, el 16,27% por cada factor; mientras que los 10 factores restantes, explicarían el 34,922 %, es decir, en promedio 3,49 % cada uno, por lo que se extrajo los 5 factores para sintetizar la variabilidad total de la variable actitud hacia el curso de estadística de los estudiantes de posgrado, lo que se confirma con la figura 1 de sedimentación de los autovalores de la variable actitud.

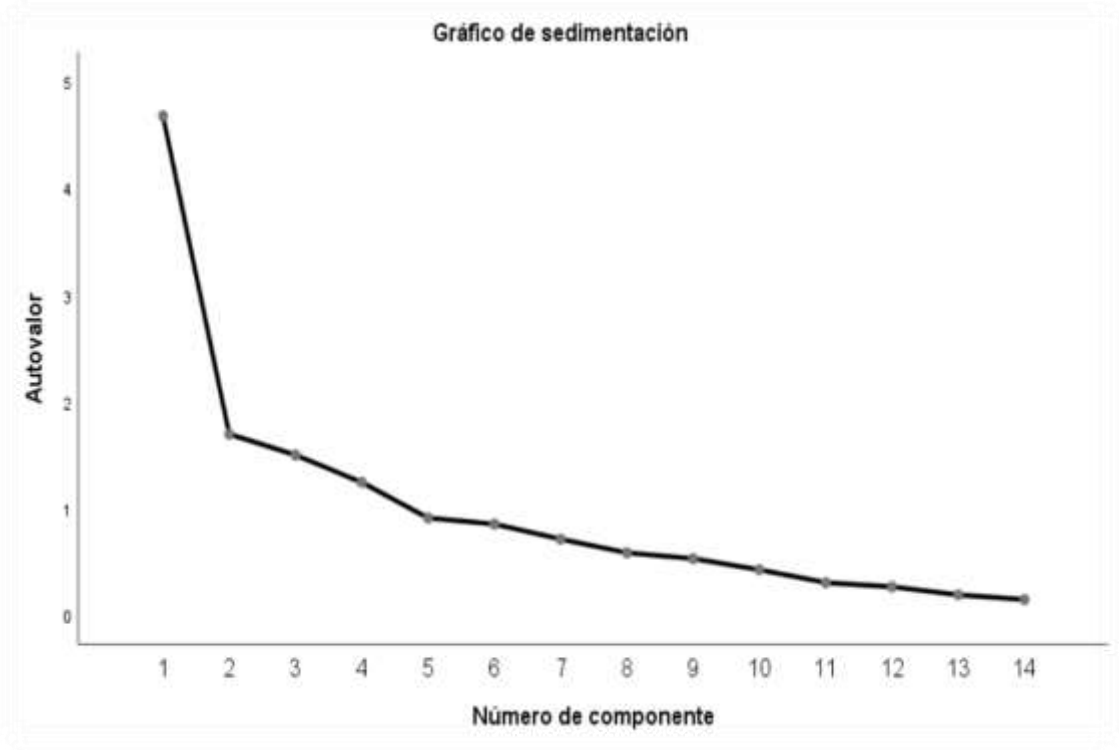


Figura 1: Gráfico de sedimentación de la variable actitud

En la tabla 4, se muestra los 4 factores latentes extraídos: El primer factor altamente correlacionado con sentirse inseguro cuando se realiza problemas en el curso de estadística, las fórmulas de estadística son difíciles de entender, el curso de estadística es complicado, sentirse frustrado al realizar las pruebas estadísticas y en las clases de estadística estoy con tensión, por lo que a este factor se ha denominado *Factor inseguridad del estudiante*.

Por otro lado, el segundo factor altamente correlacionado con, no entiendo mucho el curso de estadística debido a mi forma de pensar, las conclusiones estadísticas raramente se dan en la vida, aprender estadística requiere de mucho tiempo y disciplina, cometo muchos errores al realizar las interpretaciones de los resultados estadísticos, por lo que se le denominó Factor *Curso de alto nivel.*

Así mismo, el tercer factor me gusta el curso de estadística y la ciencia estadística raramente se aplica en la vida diaria; por lo que se le llamó *Factor curso no aplicable*.
Finalmente, el cuarto factor altamente correlacionado con, la mayoría de los estudiantes aprenden estadística rápidamente y en el ejercicio profesional no utilizaré estadística a este factor se le ha denominado *Factor curso no útil*.

Tabla 4: Matriz factorial rotada de correlaciones de la variable actitud

	Componente			
	1	2	3	4
Me gusta el curso de estadística			**0,713**	
Me siento inseguro cuando realizo problemas en el curso de estadística	**0,758**			
No entiendo mucho el curso de estadística debido a mi forma de pensar		**0,563**		
Las fórmulas estadísticas son difíciles de entender	**0,834**			
El curso de estadística es complicado	**0,771**			
Me siento frustrado al realizar las pruebas de hipótesis estadísticas	**0,642**			
La ciencia estadística raramente se aplica en la vida diaria.			**0,774**	
En las clases de estadística estoy con tensión	**0,626**			
Las conclusiones estadísticas raramente se dan en la vida.		**0,507**		
La mayoría de los estudiantes aprenden estadística rápidamente				**0,862**
Aprender estadística requiere de mucho tiempo y disciplina		**0,758**		
En el ejercicio profesional, no utilizaré estadística				**0,520**
Cometo muchos errores al realizar las interpretaciones de los resultados estadísticos		**0,660**		
Me resulta difícil aprender los conceptos estadísticos	**0,787**			

Nota: Método de extracción: análisis de componentes principales. Método de rotación: Varimax con normalización Kaiser. Fuente: Elaboración propia

CONCLUSIONES

1. Los estudiantes de posgrado de la Universidad Nacional de San Antonio Abad del Cusco, mayoritariamente fueron del sexo femenino (55,6%), de entre las edades de 25 a 29 años (45%), con título profesional (71,9%) y que residían en la ciudad del Cusco (44,4%)
2. Los cuatro factores latentes que sintetizan el 65,078 %, de la variabilidad total de la variable actitud hacia el curso de estadística de los estudiantes de posgrado son: un factor altamente correlacionado con sentirse inseguro cuando se realiza problemas en el curso de estadística, las fórmulas de estadística son difíciles de entender, el curso de estadística es complicado, sentirse frustrado al realizar las pruebas estadísticas y en las clases de estadística estoy con tensión, denominado *Factor inseguridad del estudiante*, el segundo factor altamente correlacionado con, no entiendo mucho el curso de estadística debido a mi forma de pensar, las conclusiones estadísticas raramente se dan en la vida, aprender estadística requiere de mucho tiempo y disciplina, cometo muchos errores al realizar las interpretaciones de los resultados estadísticos, denominado Factor *Curso de alto nivel;* un tercer factor me gusta el curso de estadística y la ciencia estadística raramente se aplica en la vida diaria; denominado *Factor curso no aplicable* y un cuarto factor altamente correlacionado con, la mayoría de los estudiantes aprenden estadística rápidamente y en el ejercicio profesional no utilizaré estadística a este factor se le ha denominado *Factor curso no útil*.

BIBLIOGRAFÍA

Comas, C., Martins, J., Nascimento, M., & Estrada, A. (2017). Estudio de las actitudes hacia la Estadística en estudiantes de Psicología. *BOLEMA*, 479 - 496.

Darias, E. (2000). Escala de actitudes hacia la Estadística. *Psicothema*, 175 - 178.

Santabárbara, J., & López, R. (2019). Actitudes hacia la estadística en residentes de medicina que cursan un posgrado de investigación. *Fundación educación Medica(FEM)*, 79 - 83.

Tarazona, E., Bazán, J., & Aparicio, A. (2013). Actitudes hacia la estadística en universitarios peruanos de mediana edad. *Revista Digital de Investigación en Docencia Universitaria (RIDU)*, 57 - 76.

Vilá, R., & Rubio, M. (2014). Actitudes hacia la Estadística en el alumnado del grado de Pedagogía de la Universidad de Barcelona. *Revista de docencia universitaria (REDU)*, 131 - 149.

Evolución del número de pacientes detectados con COVID-19 en Salta Capitalfrente al resto de Argentina.

Autores: Jorge A. Silvera, Angel R. Barberis

Catedra Métodos Estadísticos Y Análisis Cuantitativo - Facultad Escuela de Negocios - UNIVERSIDADCATOLICA DE SALTA.

Datos de contacto: jsilvera@unsa.edu.ar, +54 9 387 4028837.

RESUMEN

En el presente trabajo se estudiara la evolución del número de pacientes detectados con COVID-19 en Salta Capital frente al resto de Argentina, utilizando la técnica de regresión lineal, con la herramienta Microsoft Excel con el módulo XLSTAT by addinsoft.

Palabras Claves: Covid-19, Regresión Lineal, Estadística Descriptiva.

INTRODUCCION

El objetivo del presente Trabajo final es aplicar los conocimientos adquiridos en la materia Métodos Estadísticos y Análisis Cuantitativo, de la Maestría en Administración de Negocios dela Escuela de Negocios de la Universidad Católica de Salta, a un conjunto muestral de datos relacionados a la epidemia de Covid-19, en particular en nuestra Provincia de Salta Capital.

METODOLOGIA

En una etapa previa se trabajó en la obtención de información, utilizando una metodología de recolección, análisis y vinculación de datos cuantitativos propuesta por Hernández Sampieri[5], Tomando datos provienen de diferentes fuentes, en distintos formatos y calidades.

Los datos utilizados fueron obtenidos del Ministerio de Salud de la Provincia de Salta, en formato excel, y del sitio https://ourworldindata.org/coronavirus-data, Our World in Data es unaorganización de investigación y datos para el progreso de los problemas mundiales.

Toda esta información es Open Source, pudiendo obtenerse actualizada del Github https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/ecdc/full_data.csv.

En este trabajo, una vez de tener los datos seleccionados, depurados y formateados, se procedió a aplicar el método de regresión lineal, obtener resultados y sacar conclusiones.

DESARROLLO

Se intentara de dar respuesta a la pregunta si ha evolucionado de igual forma el número de pacientes detectados con COVID-19 en Salta capital que en el resto de Argentina.

Utilizando la técnica de la regresión simple, si la evolución en Salta Capital fuera similar a la del resto de Argentina, los datos deberían guardar una cierta proporcionalidad debido a la diferencia de tamaño de las poblaciones, además, posiblemente haya algún desplazamiento debido a posibles diferencias en el comienzo de la epidemia en cada Provincia. En decir, salvo diferencias en traslación y escala, los datos correspondientes deberían ser similares, es decir deberían situarse alrededor de una recta, esta sería nuestra hipótesis.

Procedemos entonces a ajustar el modelo de regresión simple empírico, definamos las siguientes variables aleatorias X (Independiente) e Y (Dependiente):

Xi = FA Salta Capital = {Número acumulado de pacientes detectados en Salta Capital hasta el día i}

Yi = Diferencia Acumulado = {Diferencia entre el acumulado en Argentina y el correspondiente dato en Salta capital}

Los datos obtenidos corresponden a las fechas comprendidas entre 11/04/2020 y el 30/08/2020).

Como primera etapa, se procesó los datos relevados, los cuales fueron obtenidos en formato xlsx y cvs, unificado formatos, procesando información, calculando frecuencias F y acumulados FA, para recién poder aplicar los estadísticos descriptivos a las dos variables.

Vamos a utilizar la herramienta Excel con el módulo XLSTAT *by addinsoft,* es un complemento de análisis de datos para Excel potente a la vez que flexible que permite a los usuarios analizar, personalizar y compartir resultados en Microsoft Excel. Gracias a sus más de 240 funciones estadísticas estándar y avanzadas, XLSTAT es la herramienta preferente de análisis estadístico en empresas y universidades.

Estadísticos descriptivos X Salta Capital:

Variable	Observaciones	Obs. con datos perdidos	Obs. sin datos perdidos	Mínimo	Máximo	Media	Desv. típica
F	113	0	113	1,000	27,000	5,611	5,054

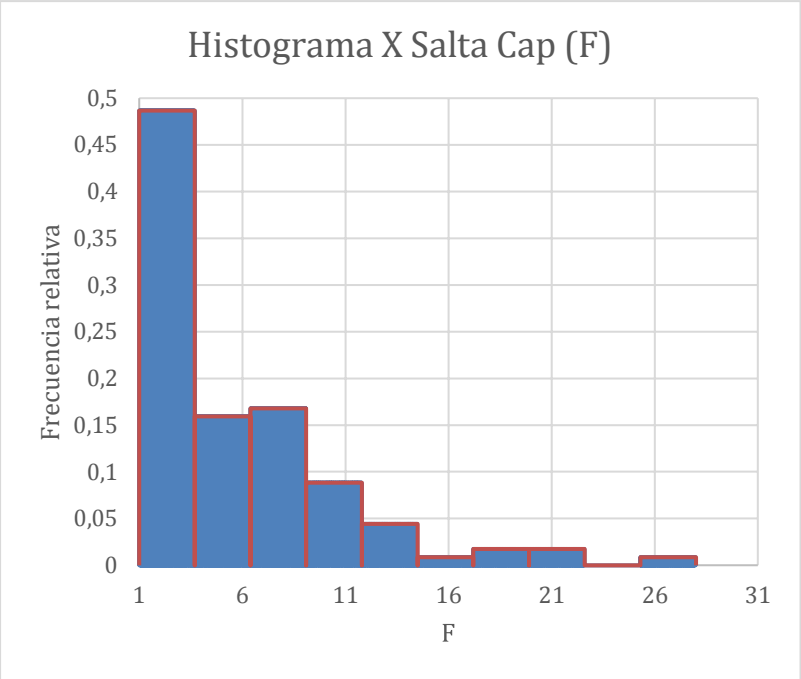


Ilustración 1 - Histograma X Salta Capital (F).

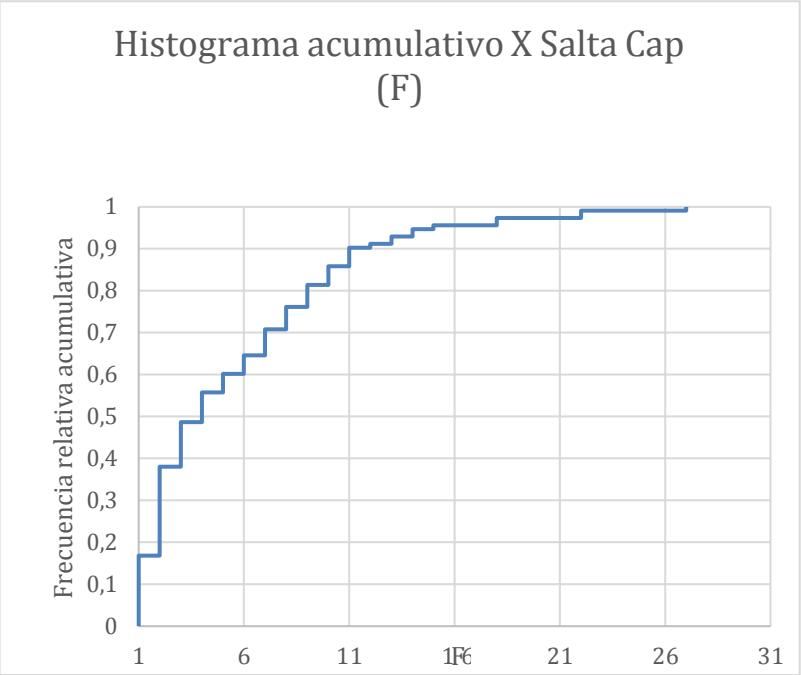


Ilustración 2 - Histograma acumulativo X Salta Capital (F).

Estadísticos descriptivos para los intervalos X Salta Capital (F):

Límite inferior	Límite superior	Frecuencia	Frecuencia relativa	Densidad
1	3,7	55	0,487	0,180
3,7	6,4	18	0,159	0,059
6,4	9,1	19	0,168	0,062
9,1	11,8	10	0,088	0,033
11,8	14,5	5	0,044	0,016
14,5	17,2	1	0,009	0,003
17,2	19,9	2	0,018	0,007
19,9	22,6	2	0,018	0,007
22,6	25,3	0	0,000	0,000
25,3	28	1	0,009	0,003

Estadísticos descriptivos Y Argentina:

Variable	Observaciones	Obs. con datos perdidos	Obs. sin datos perdidos	Mínimo	Máximo	Media	Desv. típica
F	113	0	113	0,000	11730,000	3284,133	2976,285

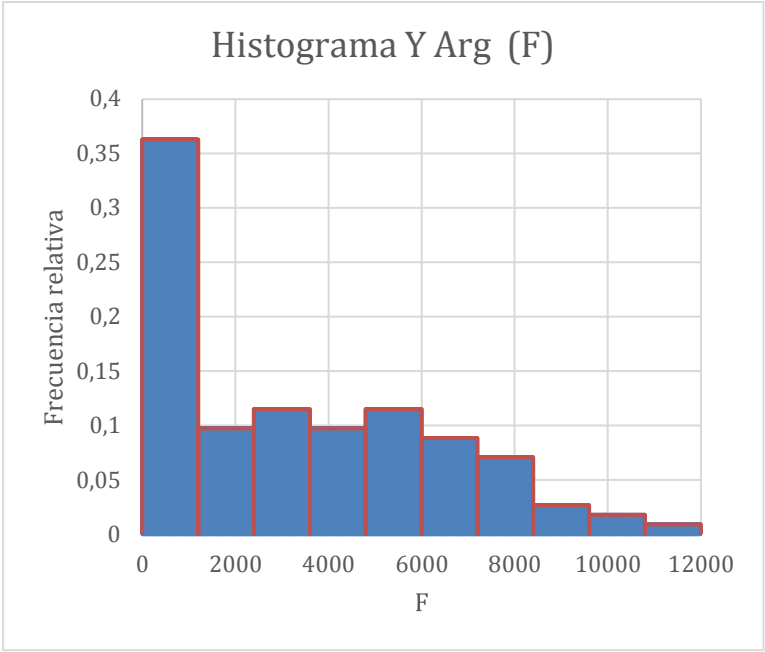


Ilustración 3 - Histograma Y Argentina (F).

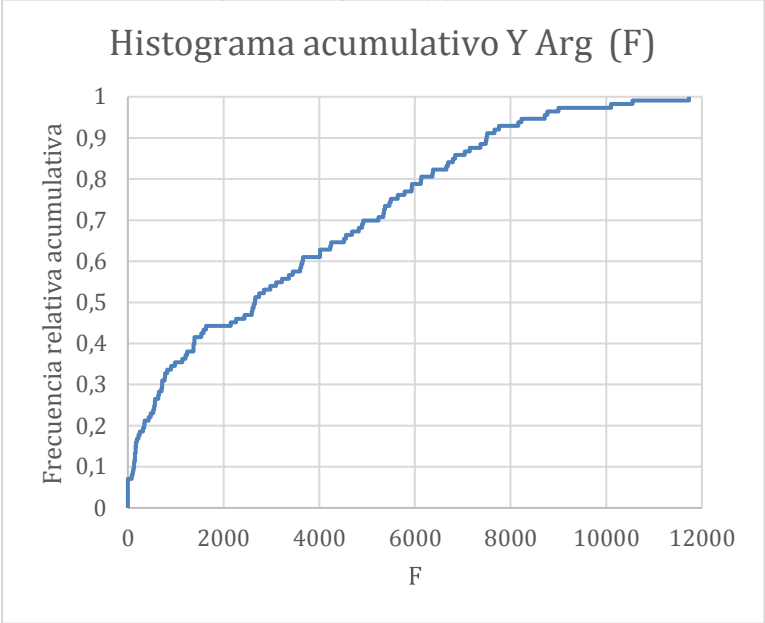


Ilustración 4 - Histograma acumulativo Y Argentina (F).

Estadísticos descriptivos para los intervalos Y Argentina (F):

Límite inferior	Límite superior	Frecuencia	Frecuencia relativa	Densidad
0	1200	41	0,363	0,000
1200	2400	11	0,097	0,000
2400	3600	13	0,115	0,000
3600	4800	11	0,097	0,000
4800	6000	13	0,115	0,000
6000	7200	10	0,088	0,000
7200	8400	8	0,071	0,000
8400	9600	3	0,027	0,000
9600	10800	2	0,018	0,000
10800	12000	1	0,009	0,000

Utilizando MS Excel, obtenemos los siguientes resultados:

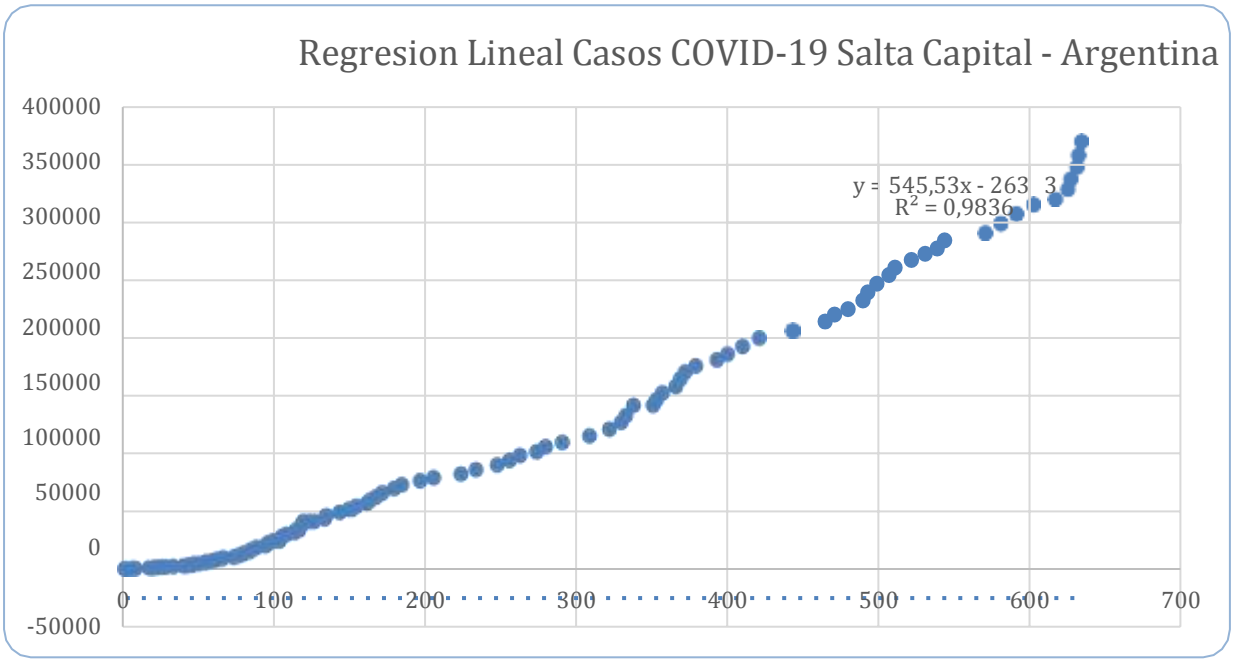


Ilustración 5 - Regresión Lineal casos COVID-19 Salta Capital vs. Argentina.

Utilizando la función ESTIMACION.LINEAL() obtenemos:

Completamos el análisis utilizando XLSTAT *by addinsoft,* obteniendo los siguientes resultados:

Estadísticos descriptivos X:

Variable	Observaciones	Obs. con datos perdidos	Obs. sin datos perdidos	Mínimo	Máximo	Media	Desv. típica
FA X Salta Capital	113	0	113	2,000	634,000	230,637	196,117
Diferencia Acumulado XY	113	0	113	79,000	370473,000	99516,301	107878,766

Matriz de correlaciones:

	Diferencia Acumulado	FA Salta Capital
Diferencia Acumulado XY	**1**	0,992
FA X Salta Capital	0,992	**1**

Estadísticos de bondad del ajuste (FA Salta Capital X):

El siguiente cuadro de resultados proporciona los coeficientes de ajuste del modelo. El R'² (coeficiente de determinación) proporciona una idea del % de variabilidad de la variable a modelizar, explicado por las variables explicativas. Mientras más cerca está de 1 este coeficiente, mejor es el modelo.

Observaciones	113
Suma de los pesos	113
GL	111
R²	**0,984**
R² ajustado	**0,983**
MEC	638,290
RMSE	25,264
MAPE	75,893
DW	0,048
Cp	2,000
AIC	731,826
SBC	737,280
PC	0,017

Análisis de varianza (FA Salta Capital X):

Fuente	GL	Suma de cuadrados	Cuadrados medios	F	Pr > F
Modelo	1	4236901,951	4236901,951	6637,896	**<0,0001**
Error	111	70850,173	638,290		
Total corregido	112	4307752,124			

Calculado contra el modelo Y=Media(Y)

El cuadro de análisis de la varianza es un resultado que debe ser atentamente analizado, vemos que la prueba del F de Fisher es utilizada. Dado que la probabilidad asociada al F, en este caso, es inferior de 0.0001, significa que nos arriesgamos de menos del 0.01% concluyendo que la variable explicativa origina una cantidad de información significativa al modelo.

Análisis Suma de Cuadrados Tipo I (FA Salta Capital X):

Fuente	GL	Suma de cuadrados	Cuadrados medios	F	Pr > F
Diferencia Acumulado	1	4236901,951	4236901,951	6637,896	**< 0,0001**

Análisis Suma de Cuadrados Tipo III (FA Salta Capital X):

Fuente	GL	Suma de cuadrados	Cuadrados medios	F	Pr > F
Diferencia Acumulado	1	4236901,951	4236901,951	6637,896	**< 0,0001**

Parámetros del modelo (FA Salta Capital X):

Fuente	Valor	Error estándar	t	Pr > |t|	Límite inferior (95%)	Límite superior (95%)
Intercepción	51,216	3,240	15,807	**<0,0001**	44,796	57,637
Diferencia Acumulado	0,002	0,000	81,473	**<0,0001**	0,002	0,002

Ecuación del modelo (FA Salta Capital X):

FA Salta Capital X = 51,2160942115741+1,80293150302517E-03*Diferencia Acumulado X Y

Coeficientes estandarizados (FA Salta Capital X Y):

Fuente	Valor	Error estándar	t	Pr > |t|	Límite inferior (95%)	Límite superior (95%)
Diferencia Acumulado XY	0,992	0,012	81,473	**<0,0001**	0,968	1,016

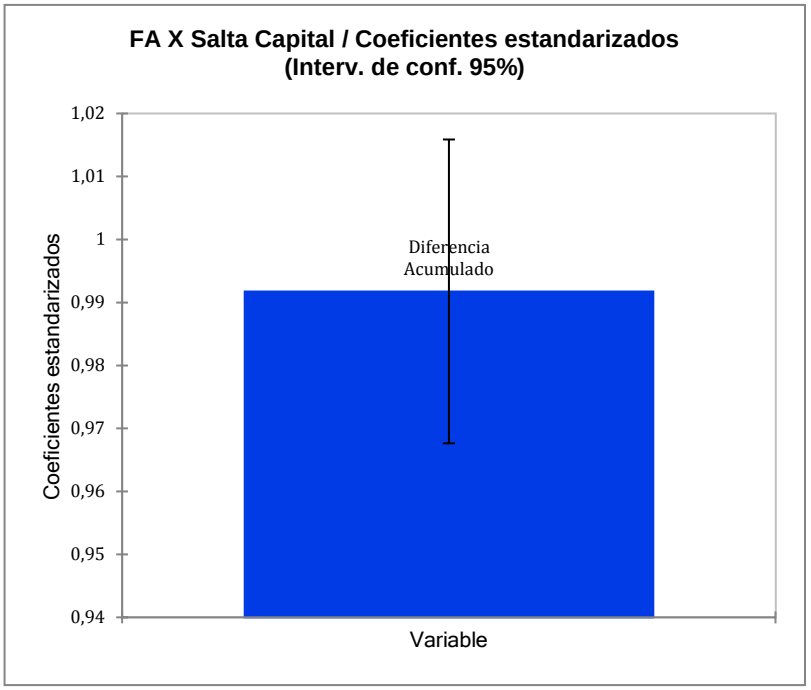


Ilustración 6 - FA Salta Capital X/ Coeficientes estandarizados (Interv. de conf. 95%).

Predicciones y residuos (FA Salta Capital X):

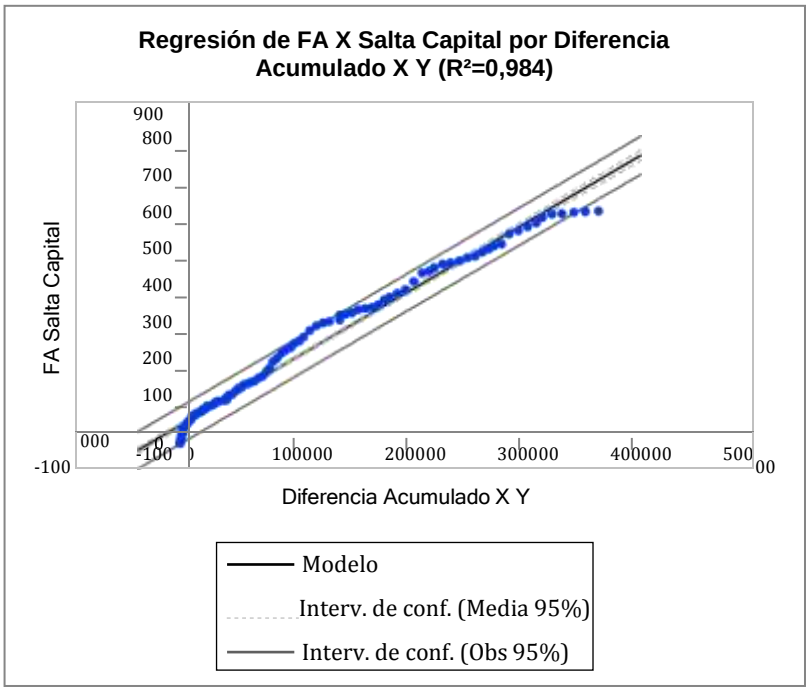


Ilustración 7 - Regresión de FA Salta Capital X por Diferencia Acumulado X Y (R²=0,984).

En este gráfico podemos visualizar los datos, la recta de regresión, y los dos intervalos de confianza, el intervalo alrededor de la media del estimador es lo más cerca de la curva, el segundo es el intervalo alrededor de la estimación puntual. Vemos claramente una tendencia lineal.

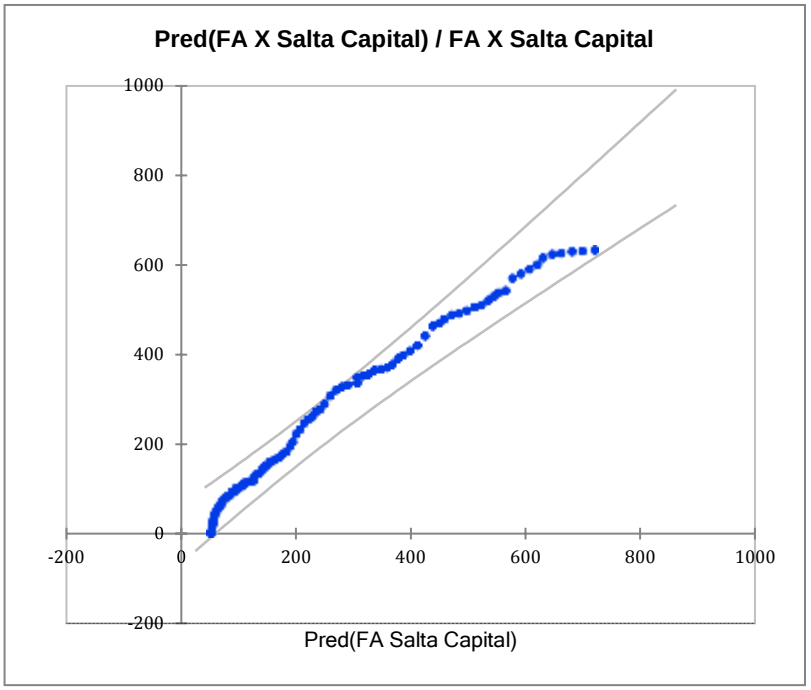


Ilustración 8 - Pred(FA Salta Capital X) - FA X Salta Capital.

Este grafico nos permite comparar las predicciones con las observaciones.

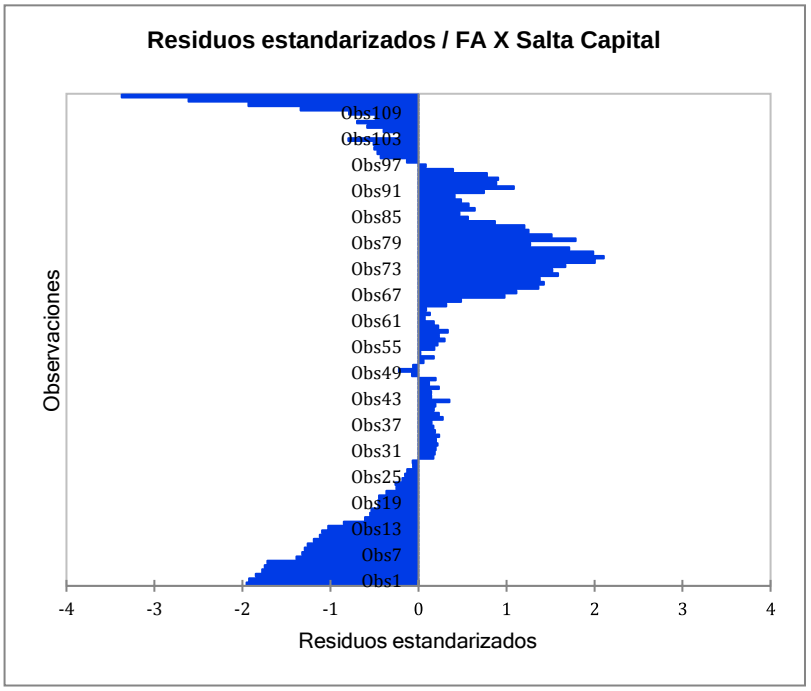


Ilustración 9 - Residuos estandarizados / FA Salta Capital X.

El histograma de los residuos estandarizados permite señalar rápidamente y visualmente la presencia de valores fuera del intervalo.

Interpretación (FA Salta Capital X) XLSTAT by addinsoft:
Dado el valor R^2, la variable explicativa explica el 98% de la variabilidad de la variable dependiente FA Salta Capital.

Dado el valor p asociado al estadístico F calculado en la tabla ANOVA, y dado el nivel de significación del 5%, la información aportada por las variables explicativas es significativamentemejor que la que podría aportar únicamente la media.

CONCLUSIONES

Luego de aplicar el análisis de regresión lineal simple, para poder realizar una inferencia estadística, obtuvimos una ecuación de regresión que describe el comportamiento lineal entre dos variables, que nos permitió observar como los puntos se alinean en torno a la recta, con un valor del coeficiente de correlación de Pearson, r^2=0.992 alto, lo cual, según la idea intuitiva expresada anteriormente indicaría que no se aprecian diferencias en la evolución del número de pacientes detectados entre Salta Capital y el resto de Argentina.

Pero hay que ser precavidos al realizar este tipo de afirmaciones, no hemos realizado un proceso de inferencia, tan solo estamos describiendo el comportamiento de los datos observados, explicando una relación, no necesariamente la causa, para lo cual se debería hacer otros estudios.

Finalmente, vemos que en todo el mundo ha quedado patente que existen multitud de problemas asociados a la recopilación de los datos diarios relativos a la pandemia del COVID-19, especialmente en Salta capital fue muy difícil conseguir datos, sobre todos oficiales, las causas son múltiples, desde cambios metodológicos a problemas administrativos y de gestiónen la Provincia o la falta de medios. Como consecuencia, a menudo se producen importantes distorsiones en las series temporales que describen la epidemia en una zona geográfica concreta, haciendo bastante complicado encontrar un modelo.

Además recientemente Edouard Mathieu, analista de datos en Our World in Data, especializadoen estadística en tiempo real con base en la Universidad de Oxford, el sitio web desde donde tomados los datos de Argentina para realizar este trabajo final, comunicó vía Twitter que Argentina dejará de formar parte de su mapa de testeos porque las cifras no tendrían la calidadsuficiente. “Para asegurar la calidad y confiabilidad de los datos de tests de COVID-19 de @OurWorldinData, hemos decidido eliminar a la Argentina de nuestro conjunto de datos por elmomento. Las cifras oficiales recopiladas por el Gobierno no tienen la calidad suficiente para reflejar correctamente el alcance de las pruebas” publicó en la red social.

Por todo esto, nos queda claro que tratar de realizar un estudio más profundo del COVID-19, como encontrar un modelo determinista para la previsión de casos confirmados de contagio, basado en una modificación de la curva de Gompertz o el modelo logístico por ejemplo, requiere la cooperación entre distintas ramas científicas, con un equipo multidisciplinario altamente experimentado compuesto por epidemiólogos, estadísticos y científicos.

BIBLIOGRAFIA

[1] Material de la catedra MÉTODOS ESTADÍSTICOS Y ANÁLISIS CUANTITATIVO-MAESTRÍA EN ADMINISTRACIÓN DE NEGOCIOS – UCASAL.

[2] Sitio Web https://ourworldindata.org/coronavirus-data

[3] Sitio Web https://covid19.inverence.com/ Un modelo epidemiológico de series temporalespara la COVID-19.

[4] Sitio Web https://help.xlstat.com/s/article/regresion-lineal-simple-tutorial-en-excel
Hernández Sampieri, Metodología de la investigación, 6ta edición, Ed. McGraw-Hill, 2014.

ANÁLISIS SOBRE CASOS DETECTADOS CON COVID 19 EN ARGENTINA

Mamaní A. Ismael, José H. Farfán, Mariela Rodríguez

Facultad de Ingeniería – Universidad Nacional deJujuy - Jujuy[1]

ismael.mamani.ar@gmail.com[1], jhfarfan@gmail.com[1],mariela.rodriguez@fi.unju.edu.ar[1]

Resumen

El presente trabajo realiza un análisis exploratorio de los datos a fin de detectar patrones destacables sobre los casos que padecieron las personas que tuvieron COVID-19 en la República Argentina, además pretende mostrar las características que presentan los casos de personas fallecidas, aplicando para ello técnicas de Minería de Datos, tales como Reglas de Asociación y por último se trata de describir los casos de personas infectadas ocurridos en el país agrupadas por edades.

Palabras Claves: COVID-19, Argentina, Análisis Exploratorio, Minería de Datos, Reglas deAsociación**.**

INTRODUCCIÓN

En estos últimos tiempos uno de los temas que ha repercutido en todo el mundo fue la aparición de la Pandemia del COVID-19 la cual continúa siendo estudiada y con la cual grandes equipos de científicos se vieron en la necesidad de dar una respuesta al mundo. Si bien grandes comunidades científicas en todo el mundo han trabajado arduamente en investigar al virus causante de tal pandemia, aún se desconoce por completo el comportamiento del mismo; y cómo ciertos casos particulares toman resultados de falsos positivos o falsos negativos lo que es mucho más peligroso, las repercusiones tras un recontagio o un periodo de contagio entre otros.

Por tanto, este trabajo pretende encontrar patrones que den indicios o describan el comportamiento del virus en personas de distintas edades, en ciudadanos de la República Argentina, residentes en dicho país. El periodo de tratamiento que tuvieron las personas fallecidas y como la ola de contagio se produjo en el país.

A continuación se procederá a realizar una descripción de los atributos del dataset mantenido por la Dirección Nacional de Epidemiología y Análisis de Situación de Salud, confecha de publicación del 15 de mayo de 2020.

datos.gob.ar

Cantidad total de registros
2.078.267

Atributos	Tipos	Descripción
clasificacion_resumen	Texto (string)	Clasificación del caso
id_evento_caso	Número entero (integer	Numero de caso
sexo	Texto (string)	Sexo
edad	Número entero (integer	Edad
edad_años_meses	Texto (string)	Edad indicada en Años o meses
residencia_pais_nombre	Texto (string)	Pais de residencia
residencia_provincia_nombre	Texto (string)	Provincia de residencia
residencia_departamento_nomb	Texto (string)	Departamento de residencia
carga_provincia_nombre	Texto (string)	Provincia de establecimiento de carga
fecha_inicio_sintomas	Fecha ISO-8601 (date)	Fecha de inicio de síntomas
fecha_apertura	Fecha ISO-8601 (date)	Fecha de apertura del caso
sepi_apertura	Número entero (integer	Semana Epidemiológica de fecha de apertura
fecha_internacion	Fecha ISO-8601 (date)	Fecha de internación
cuidado_intensivo	Texto (string)	Indicación si estuvo en cuidado intensivo
fecha_cui_intensivo	Fecha ISO-8601 (date)	Fecha de ingreso a cuidado intensivo en el caso de corresponder
fallecido	Texto (string)	Indicación de fallecido
fecha_fallecimiento	Fecha ISO-8601 (date)	Fecha de fallecimiento en el caso de corresponder
asistencia_respiratoria_mecanic	Texto (string)	Indicación si requirió asistencia respiratoria mecánica
carga_provincia_id	Número entero (integer	Código de Provincia de carga
origen_financiamiento	Texto (string)	Origen de financiamiento
clasificacion	Texto (string)	Clasificación manual del registro
residencia_provincia_id	Número entero (integer	Código de Provincia de residencia
fecha_diagnostico	Fecha ISO-8601 (date)	Fecha de diagnóstico
residencia_departamento_id	Número entero (integer	Código de Departamento de residencia
ultima_actualizacion	Fecha ISO-8601 (date)	Última actualización

Tabla Nº1: descripción de los atributos del dataset.

METODOLOGÍA

Tabla Nº1: descripción de los atributos del datasetPara llevar a cabo el cumplimiento de los objetivos propuesto se procede a utilizar la metodología KDD junto con la herramienta Rapidminer, un software Open-Source para el análisis y minería de datos, en su versión 9.7 con Licencia Estudiantil. Cabe recalcar que el dataset utilizado fue extraído de la página oficial de dataset de la República Argentina y que el mismo fue utilizado en la actualización de datos del 06/10/2020 ya que este mismo recibe actualizaciones cada cierto tiempo.

KDD es un proceso complejo que incluye no sólo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de los mismos, tal y como se refleja en la Figura 1.1

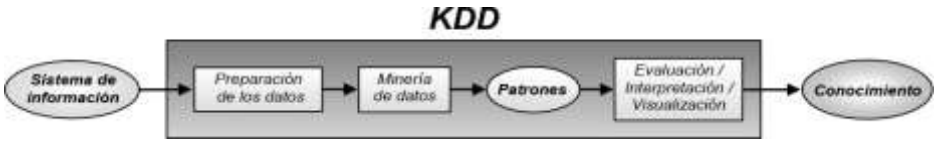


Grafico N°1 - Proceso de KDD

El proceso de KDD se organiza en torno a cinco fases como se ilustra en la Figura 2.1. En la fase de integración y recopilación de datos se determinan las fuentes de información quepueden ser útiles y dónde conseguirlas. A continuación, se transforman todos los datos a un formato común. En la fase de selección, limpieza y transformación se tratan los valoreserróneos o faltantes, debido a que en ocasiones los datos provienen de diferentes fuentes. Además, se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma sean más útiles. En la fase de minería de datos, se decide cuál es la tarea a realizar (clasificar, agrupar, etc.) y se elige el método que se va autilizar. En la fase de evaluación e interpretación se evalúan los patrones y se analizan porlos expertos, y si es necesario se vuelve a las fases anteriores para una nueva iteración. Finalmente, en la fase de difusión se hace uso del nuevo conocimiento y se hace partícipede él a todos los posibles usuarios.

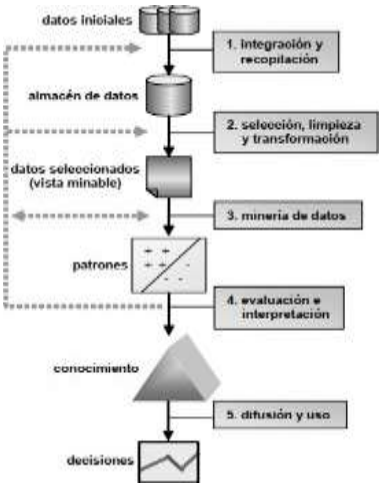


Grafico N°2 Fases de la Metodología KDD

DESARROLLO

Para la etapa de desarrollo se tienen en cuenta los procesos realizados según cada uno de los objetivos propuestos.

Compresión del dominio y establecimiento de los objetivos

Para esto se realiza una investigación sobre la temática, abordada a un problema de dominio público, en el periodo comprendido hasta principios de Octubre del año 2.020. En cuanto a su origen es de una fuente oficial, del mismo país a ser investigado cuya fuente primaria es del Ministerio de Salud, la Dirección Nacional de Epidemiología y Análisis de la Situación de Salud. Área de Vigilancia. El formato de trabajo del DataSet en cuestión es Microsoft Excel (xlsx).

Objetivos: Para la realización del presente trabajo se tienen en cuenta los siguientes objetivos

- Hallar la distribución de casos de coronavirus en Argentina por provincia por mes
- Determinar un modelo de clasificación de las personas infectadas.
- Obtener un patrón de los pacientes fallecidos y categorizarlos por edades.

Para todo trabajo de Minería de Datos es importante definir un label u objetivo, para este caso en particular se ha seleccionado el atributo “clasificación_resumen”

En la etapa de Selección se realiza una elección de los datos excluyendo registros tales como datos redundantes, faltantes o erróneos que no contribuyan a los objetivos a los que se pretende llegar. Contribuyendo al rendimiento del modelo generado, a través de la reducción de las dimensiones de los datos y al nivel de procesamiento generado en el equipo informático.

Para la visualización de los datos se procede a utilizar la codificación de archivo UTF-8, talcomo se muestra en el gráfico Nº3.

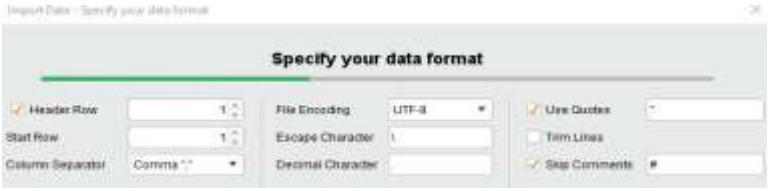


Grafico Nº3: especificación del formato del DataSet

Debido a la gran cantidad de datos y atributos se procede en un primer momento a seleccionar los atributos considerados necesarios, bajo distintos criterios, para el cumplimiento de los objetivos planteados:

- Hallar la distribución en el país por periodo de tiempo

Se considera necesario que, para el cumplimiento de este objetivo, el uso del atributo “fecha_diagnostico”, como parámetro de tiempo, por la relevancia en el contagio de un paciente, y por ser el atributo con menor cantidad de datos con missing values, de todos los atributos de tipo fecha.

Para el atributo “edad_meses”, el cual contiene edades tanto en meses como en años y es de tipo polinomial, se lo convierte en un atributo del tipo entero y como posee una cantidad del 0,096%, se reemplazan los valores faltantes por el promedio. De la misma forma el mes con el valor 12,03%, en este caso reemplazando por la fecha mayor.

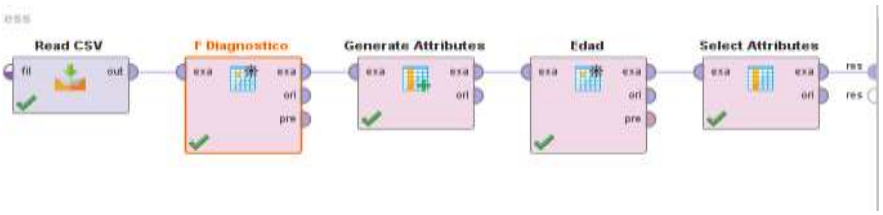


Grafico 4: operadores empleados para la selección de atributos.

Posterior a ello se presenta un análisis exploratorio de los datos donde podemos observar que los datos presentados, se concentran en su mayoría en los meses de Agosto, Septiembre, un poco Abril y Marzo. Por otro lado, podemos ver que la mayoría de los casos registrados en el DataSet, se concentran entre las edades de 20 a 60 años, centrándose más entre 28 y 36 años, tal como se muestra en el gráfico Nº4.

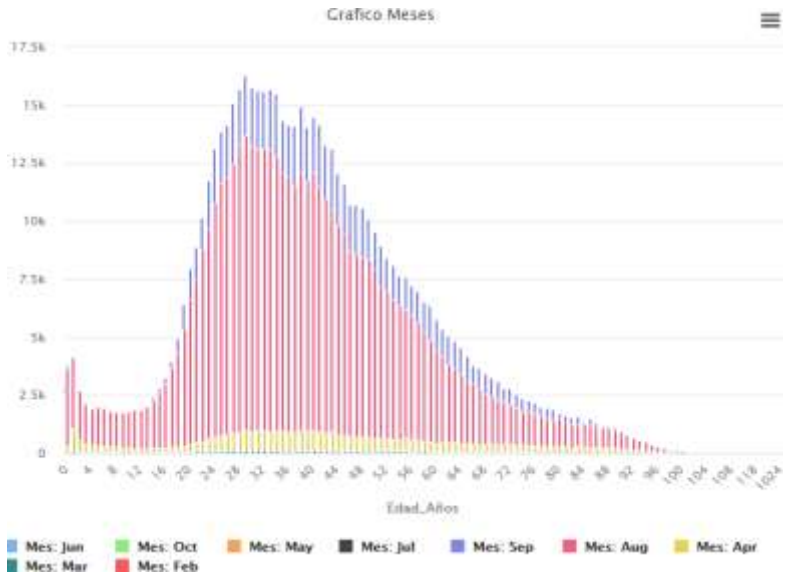


Gráfico Nº5: distribución por meses.

A continuación en el Gráfico Nº5 se observa que, las provincias con mayores casos registrados de contagios confirmados en el DataSet fueron la provincia de Buenos Aires, Buenos Aires, Ciudad autónoma de Buenos Aires, Santa Fe, Mendoza y Córdoba, Salta, Tucumán y Jujuy por nombrar los principales.

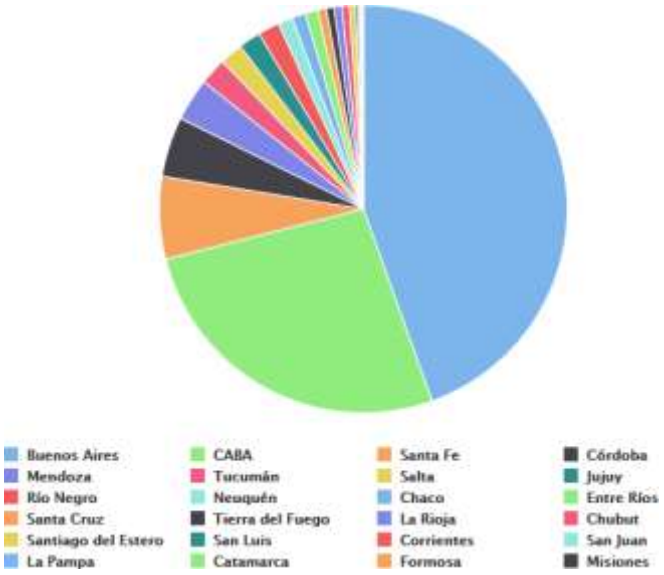


Gráfico Nº6: distribución de casos por provincias.

A continuación, en el Gráfico Nº6 se muestran las provincias en relación a la cantidad personas fallecidas de las mismas, destacándose Buenos Aires y CABA .

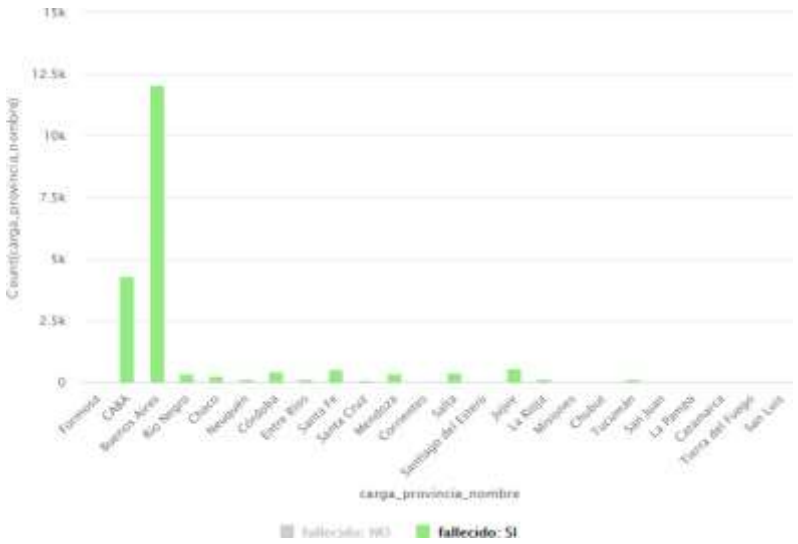


Gráfico Nº7: distribución de fallecimientos por provincias.

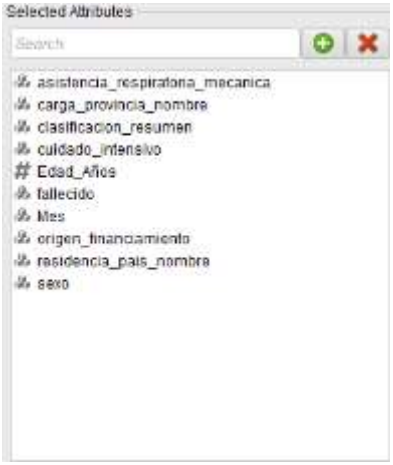
Se analiza la población con el fin de obtener una tasa poblacional, obteniendo los datos proporcionados por el INDEC en su último censo registrado en la página oficial, tal como se muestra en la Tabla Nº1 sumándole los datos del DataSet del presente trabajo.

Provincia	Población					
	Total Hab	Contagiados Confirmados	IP(100 mil hab)	Fallecidos	Tasa de letalidad (7 meses)	Tasa de mortalidad
Total del país	**40.117.096**	**779.932**	**1.944**	**20.211**	**2,6%**	**0,05%**
Ciudad Autónoma de Buenos Aires	2.890.151	206.638	7.150	4.310	2,1%	0,15%
Buenos Aires	127.205	347.415	273.114	12.067	3,5%	9,49%
Catamarca	15.625.084	252	2	0	0,0%	0,00%
Chaco	638.645	8.400	1.315	269	3,2%	0,04%
Chubut	673.307	4.615	685	64	1,4%	0,01%
Córdoba	273.964	37.153	13.561	439	1,2%	0,16%
Corrientes	551.266	1.230	223	23	1,9%	0,00%
Entre Ríos	3.194.537	7.697	241	137	1,8%	0,00%
Formosa	1.738.929	109	6	0	0,0%	0,00%
Jujuy	333.642	13.743	4.119	558	4,1%	0,17%
La Pampa	1.214.441	760	63	9	1,2%	0,00%
La Rioja	1.448.188	4.777	330	118	2,5%	0,01%
Mendoza	3.308.876	26.425	799	337	1,3%	0,01%
Misiones	509.108	81	16	3	3,7%	0,00%
Neuquén	1.055.259	9.050	858	145	1,6%	0,01%
Río Negro	1.235.994	13.238	1.071	369	2,8%	0,03%
Salta	874.006	14.320	1.638	410	2,9%	0,05%
San Juan	432.310	872	202	39	4,5%	0,01%
San Luis	318.951	1.586	497	26	1,6%	0,01%
Santa Cruz	681.055	5.424	796	70	1,3%	0,01%
Santa Fe	992.595	50.904	5.128	547	1,1%	0,06%
Santiago del Estero	367.828	3.682	1.001	65	1,8%	0,02%
Tierra del Fuego, Antártida e Islas del Atlántico Sur	530.162	4.794	904	64	1,3%	0,01%
Tucumán	1.101.593	16.767	1.522	142	0,8%	0,01%

Tabla Nº2: distribución de fallecimientos por provincias.

- Determinar las características de las personas infectadas

Para este apartado se seleccionan los atributos acorde al label **“**clasificación_resumen”, tal como se muestra en la gráfica Nº8.



Grafica Nº8: selección de atributos para el label “clasificacion_resumen”.

Los mismos son considerados teniendo en cuenta el país actual de residencia, “Argentina”y los casos clasificados como “Confirmados” a través de un filtro. Se procederá a generar atributos: Edad_años, que indica edad en años, Mes, que índica el mes de fecha de diagnóstico, y se generará una diferencia entre fecha diagnóstico y fecha fallecido llamado periodo_internacion usando la función date_diff.

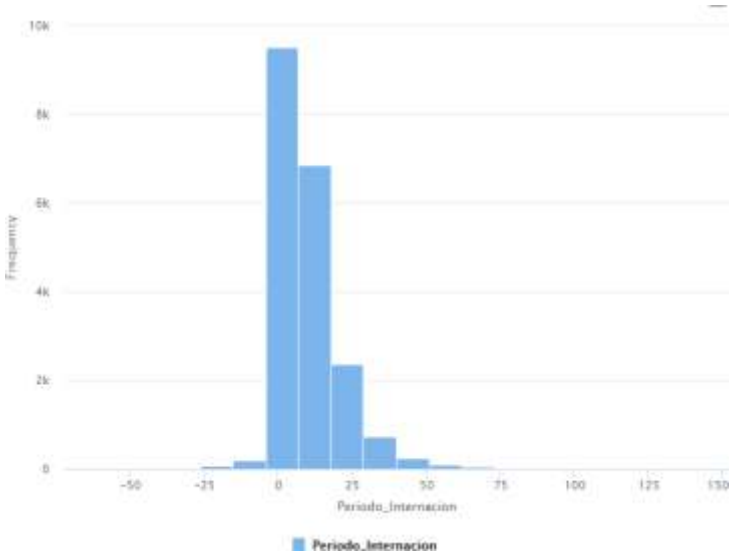


Gráfico Nº9: períodos (fecha_diagnostico-fecha_internacion)

Como se puede observar en el Gráfico Nº9 hay fechas en que el diagnóstico de positivo se determinó o el mismo día de su muerte o días después del mismo(valores negativos), Para este caso se centrara en los días de anticipación, es decir los periodos mayores a cero. Loscuales rondan hasta los 73 días antes de su fallecimiento. Se puede observar que la mayoría de los datos se centran entre 0 a 15 días del diagnóstico de la enfermedad aproximadamente.

Por otro lado, se analizará el sexo en función a las edades de personas contagiadas

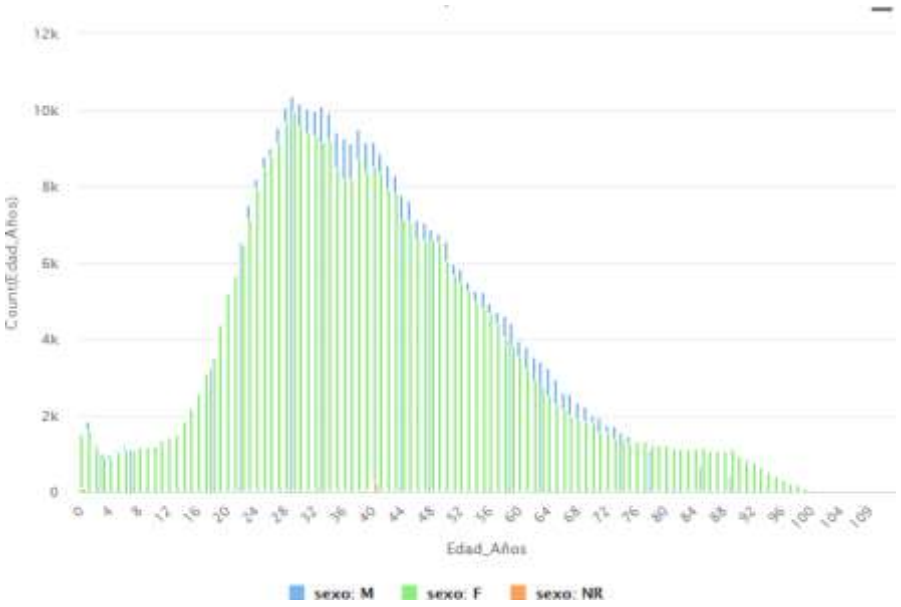


Gráfico Nº10: distribución de contagiados por edad y sexo.

Donde se puede observar (Gráfico N° 10) que hay cierta simetría entre los datos tanto masculino como femenino siendo un poco superior los del sexo masculino.

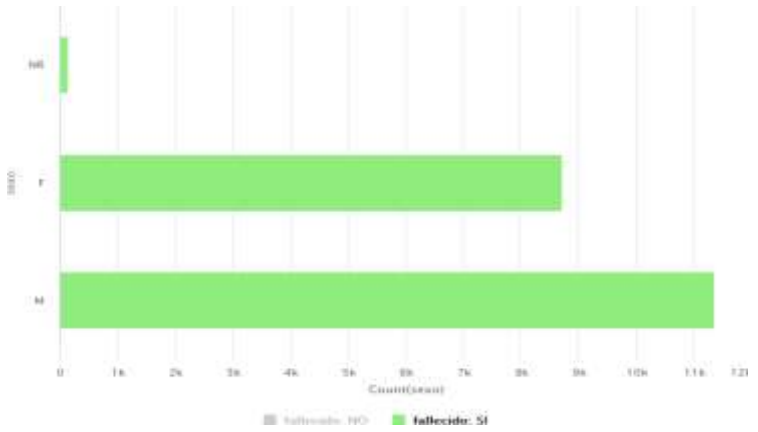


Gráfico Nº11: distribución de contagiados por sexo.

En el Gráfico N°11 se observa una diferencia entra cantidad de fallecidos comparado por sexo.

Para determinar los atributos fuertemente relacionados se utilizará un operador de matriz de correlación y definiendo al label al atributo “clasificación_resumen” con el cual se obtienelos siguientes resultados

attribute	weight
Edad_Años	0.910
cuidado_intensivo	0.521
fallecido	0
asistencia_respiratoria_mecanica	0.538
origen_financiamiento	1
Periodo_Internacion	0.219

Grafico Nº12: cálculo de peso de atributos en relación al label.

De este modo se obtienen como atributos fuertemente relacionados a:

Edad_Años, Cuidado_intensivo, asistencia_respiratoria_mecanica y origen_financiamento

A partir de estos atributos se buscará encontrar las características con operadores de árbol de decisión y redes neuronales profundas. Debido al Volumen de datos se tomó una muestra estratificada del 0,1 relativo y se filtraron los casos que fueron categorizados en el dataset como casos descartados y se tomó a pacientes residentes en Argentina

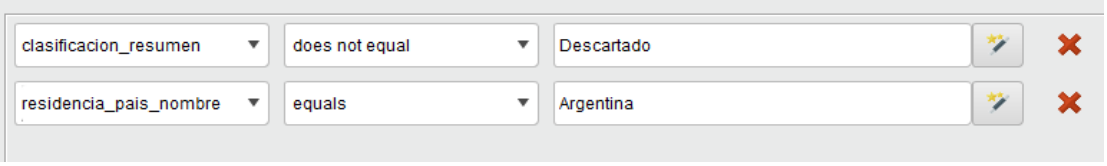


Gráfico Nº13: aplicación de filtro.

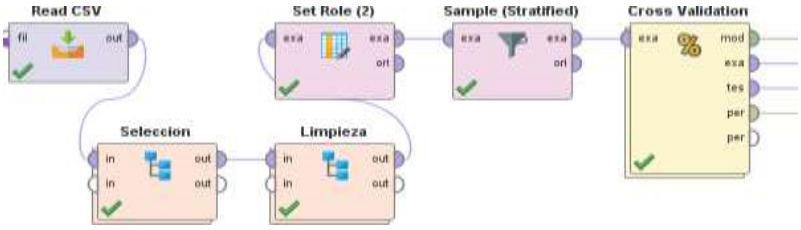


Gráfico Nº14: diagrama de modelo - Primer Nivel.

Se probaron diversos operadores y se obtuvieron los siguientes resultados:

	Operadores		
	Vote(Random Forest-Deep learning)	Deep Learning	WJ48
Performance	84,22%	100%	100%

Tabla N°3 – Performance de operador

De los cuales se optará por elegir el operador wj48.

- Obtener un patrón de los pacientes fallecidos y categorizarlos por edades

Para este punto se basará en armar reglas de asociación teniendo en cuenta el valor del label en solo los casos confirmados.

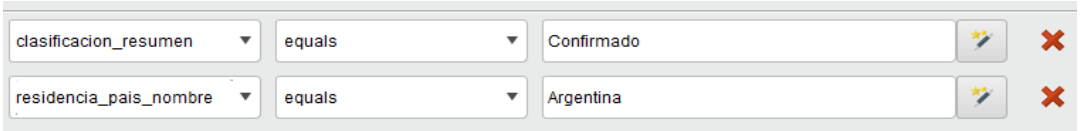


Gráfico Nº15: filtrado de atributos.

Para la determinación de patrones se utilizó los operadores de reglas de asociación con unsoporte mínimo del 0,95 y además se convirtieron los datos a binominales para este análisis(requisito del operador)

Gráfico Nº16: diagrama de modelo de reglas de asociación.

Obteniendo las siguientes reglas

AssociationRules

```
Association Rules
[origen_financiamiento] --> [clasificacion_resumen = Confirmado] (confidence: 1.000)
[Edad_Años = adulto mayor] --> [clasificacion_resumen = Confirmado] (confidence: 1.000)
[Edad_Años = joven] --> [clasificacion_resumen = Confirmado] (confidence: 1.000)
[Edad_Años = adulto] --> [clasificacion_resumen = Confirmado] (confidence: 1.000)
[origen_financiamiento, Edad_Años = adulto mayor] --> [clasificacion_resumen = Confirmado] (confidence: 1.000)
```

Gráfico Nº17: reglas de asociación

Gráfico Nº18: resultado del operador de reglas de asociación.

Donde se puede observar que hay una fuerte asociación en los datos registrados, ya que la mayoría de los casos confirmados dependió del origen del financiamiento, es decir si erapúblico o privado, también una fuerte relación entre adulto_mayor, joven y adultos como potenciales casos confirmados. Otra asociación está entre el origen de financiamiento y losadultos mayores.

Otra forma de determinar patrones es utilizando operadores de clasificación tal como lo es el operador W-J48 el cual pertenece a Weeka, comentado anteriormente para este caso se

tendrán en cuenta todos los valores posibles del atributo “clasificacion_resumen”, se agregarán atributos para arrojar más indicadores y se mantendrá la muestra estratificada al0,1. Se obtuvieron las siguientes reglas de las cuales se enfocará en analizar los casos de los pacientes que fueron casos confirmados.

W-J48

```
J48 pruned tree
------------------

clasificacion = Caso Descartado: Descartado (106821.0)
clasificacion = Caso sospechoso - Con muestra sin resultado concluyente: Sospechoso (9302.0)
clasificacion = Caso confirmado por laboratorio - No Activo por criterio de laboratorio: Confirmado (3325.0)
clasificacion = Caso confirmado por laboratorio - No activo (por tiempo de evolución): Confirmado (56278.0)
clasificacion = Caso confirmado por laboratorio - Activo Internado: Confirmado (6489.0)
clasificacion = Caso Sospechoso - Sin muestra: Sospechoso (5089.0)
clasificacion = Caso confirmado por laboratorio - Fallecido: Confirmado (1997.0)
clasificacion = Caso Sospechoso - Muestra no apta: Sospechoso (264.0)
clasificacion = Caso confirmado por criterio clinico-epidemiológico  - No activo (por tiempo de evolución): Confirmado
clasificacion = Caso confirmado por laboratorio - Activo: Confirmado (6210.0)
clasificacion = Otro diagnostico: Descartado (59.0)
clasificacion = Caso Invalidado Epidemiologicamente: Descartado (59.0)
clasificacion = Caso confirmado por criterio clinico - epidemiológico -  Activo internado: Confirmado (104.0)
clasificacion = Caso confirmado por criterio clinico-epidemiológico - Activo: Confirmado (596.0)
clasificacion = Sin clasificar: Sin Clasificar (6.0)
clasificacion = Caso confirmado por criterio clinico-epidemiologico - Fallecido: Confirmado (23.0)

Number of Leaves  :     16

Size of the tree :      17
```

Gráfico Nº19: resultado de reglas de decisión.

Las personas contagiadas, fueron reportados por los laboratorios, por tiempo de evolución del paciente, es decir sin evolución medica realizada con un total de 56278 casos

Un total de 6489 casos fueron confirmados contagiados estando en activa

internación.

De los confirmados por laboratorio 1997 casos terminaron en fallecimiento.

De los confirmados por laboratorio 2971 casos fueron realizados por criterio clínico-epidemiológico, por tiempo de evolución, es decir justificada por evolución medica ya sea clínico o especialista.

De los confirmados por el laboratorio 6210 casos fueron de personas que no teníaninternación activa, es decir no estaban internados.

Otros aportes a partir de un análisis exploratorio

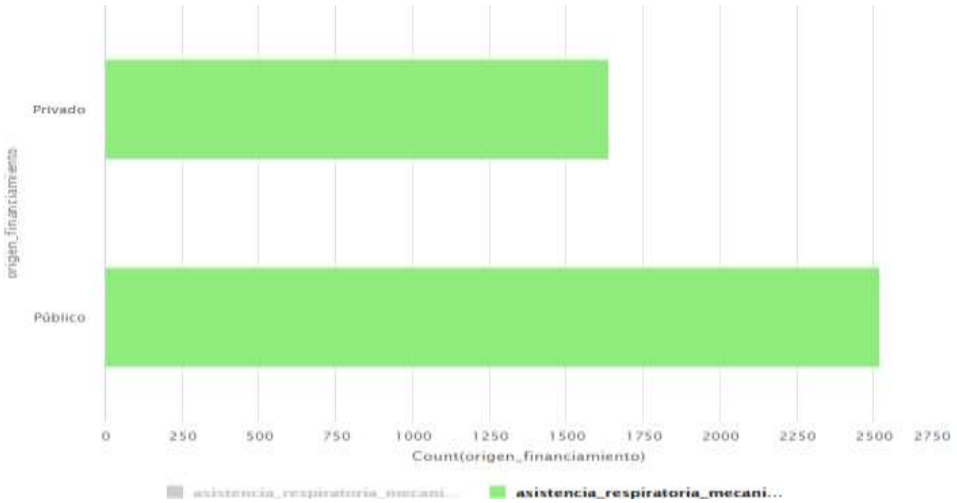


Gráfico Nº20: relación origen de financiamiento y asistencia respiratoria mecánica

De los datos analizados recibieron asistencia_respiratoria_mecanica tanto pacientes de origen de financiamiento público como privado, aunque los privados en menor medida comparado a los públicos. Mientras que una gran mayoría no lo recibió

Se realizó un análisis frente a las edades

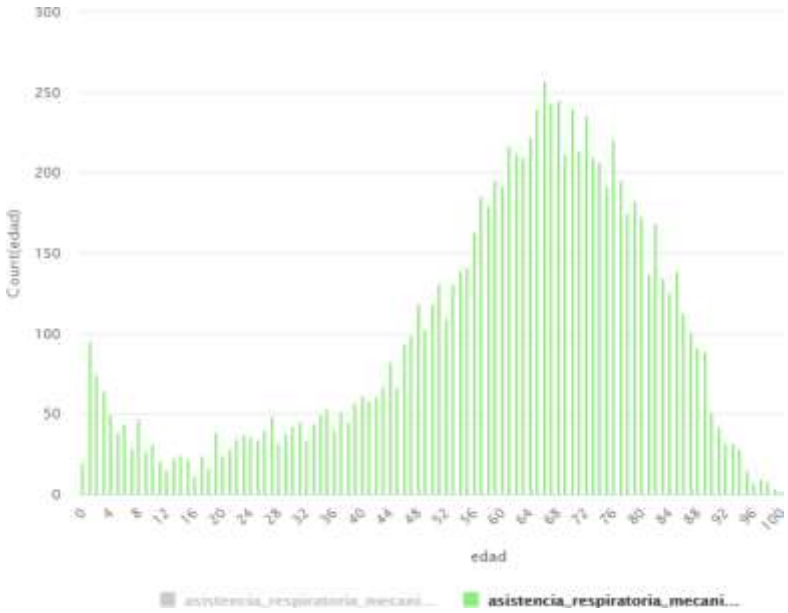


Gráfico Nº21: asistencia respiratoria por edad.

Observando que la gente entre 64 y 72 aprox recibió más asistencia respiratoria mecánica.

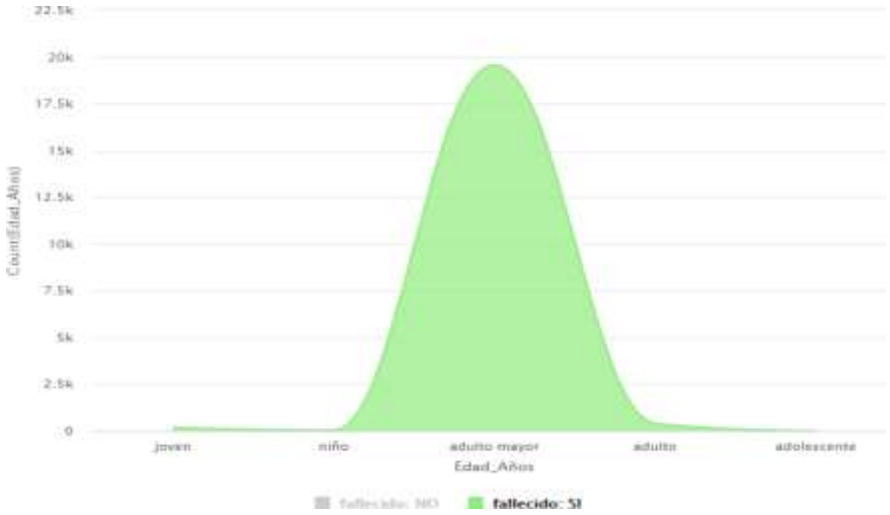


Gráfico Nº22: relación de fallecidos por edad agrupada.

Este grafico visualizamos que las personas fallecidas tienen un alta índice en personas adultas mayores.

En términos generales se tiene la siguiente grafica de contagiados

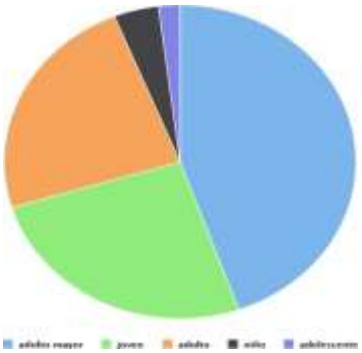


Gráfico Nº23: casos confirmados de contagiados por edad.

Por categorías de casos confirmados siendo los adultos, adultos mayores y jóvenes las categorías con mayores casos

CONCLUSIONES

Mediante este trabajo realizado se pretende exponer cierta información del comportamientodel virus en nuestro país, relevante para tener en cuenta a que personas fueron las infectadas en su mayoría, a que personas tratar frente a un segundo rebrote en caso de no tener una vacuna alternativa en ese periodo o bien para análisis de la tendencia del mismoen los periodos especificados en el análisis. Se define a este trabajo como un aporte más sujeto a otros para la afianzacion de la información ya que se considera que aún no se puede definir el comportamiento total del virus sosteniendo solo este análisis, sino que la misma debe ser respaldada por más estudios. Se debe tener en cuenta también que hay casos que quizás nunca se registraron y por tanto difiera con otros datos, aclaramos por tanto que debe realizarse el juicio dependiendo del dataset que se esté analizando y no como una imposición oficial. Sin embargo, es interesante analizar tales patrones para determinar cómo fue el comportamiento y como tomar decisiones en cuanto a medidas preventivas, en cuanto a personas expuestas, prioridades en asistencia, confirmaciones por parte de laboratorios y médicos (personal de salud).

En cuanto al trabajo futuro se puede extraer más información de otros dataset publicados y buscar justificar tales comportamientos con personal activo de la salud ya que la información dada quizás sea más enriquecedora acompañada de argumentos médicos y de personal de salud en general. Además, consideraremos el trazado de una línea de resumen de información extraída de este estudio.

BIBLIOGRAFÍA

[Hernandez Orallo, 2005] - Introducción a la Minería de Datos, España: Editorial Pearson EducaciónS.A.

[Documentación – Rapidminer, 2020], Recuperado de: https://docs.rapidminer.com/ Fecha de consulta: 25/11/2020.

[Estimaciones – Estimacionescovid19, 2020], Recuperado de: https://institucional.us.es/blogimus/2020/03/como-estimar-el-numero-de-infectados-reales- por-covid-19-el-caso-de-andalucia-e-italia/ Fecha de consulta: 27/11/2020.

[INDEC – Indicadores, 2020], Recuperado de: https://www.indec.gob.ar/indec/web/Nivel4- Tema-2-18-77

Fecha de consulta: 27/11/2020.

[Tasas – Indices,2020], Recuperado de:

https://as.com/diarioas/2020/04/21/actualidad/1587464905_473162.html Fecha de consulta: 27/11/2020.

[Dataset – Covid19Casos, 2020], Recuperado de: https://datos.gob.ar/dataset/salud-covid-19-casos-registrados-republica-argentina Fecha de consulta: 27/10/2020

10 y 11 de Diciembre de 2020

Herramientas para el manejo de la incertidumbre del Diseño de procesos en Ingeniería Química

Orlando José Domínguez y Julieta Martínez

Facultad de Ingeniería, CIUNSa, Universidad Nacional de Salta. Salta

odominguez@ing.unsa.edu.ar; jmartinez@ing.unsa.edu.ar

RESUMEN

El presente trabajo pretende contribuir al Diseño de Procesos, de la carrera de Ingeniería Química, como una alternativa a los diferentes problemas que conlleva el tema del manejo de la incertidumbre. Se propone, como novedad, resolver estos problemas, a través de la integración de conceptos económicos y tecnológicos, tratando el problema de la incertidumbre con la toma de decisión, considerando que la función objetivo, los parámetros, y los atributos están sujetos a la incertidumbre. Se formula un problema global donde se incorporan ya sea valores de probabilidad o bien distribuciones de probabilidad para salvar el desconocimiento. Este tipo de problemas se resuelve mediante herramientas disponibles actualmente, y software accesible que permite obtener un resultado más ajustado a la realidad, presentando los resultados semejantes a un pronóstico de ocurrencias en el futuro. Se muestran ejemplos donde se aplican los métodos y técnicas que se enseñan y aplican en la Cátedra de Diseño de Procesos. Finalmente, se resuelve un problema específico aplicando la herramienta al análisis de la sensibilidad de los parámetros a fin de observar la variación en los indicadores del proyecto para dar un pronóstico sobre el VAN y la TIR.

Palabras claves: diseño de proceso, diseño bajo incertidumbre, análisis sensibilidad, diseño integrado

INTRODUCCIÓN

La característica de los problemas y las operaciones en Ingeniería química, están sujeto a incertidumbre. Esta se manifiesta de múltiples formas, se la puede resumir en que la incertidumbre está presente en el entorno del proceso, debido a la ausencia de cierta información para definir completamente el análisis, como precios, demandas, etc. Se pueden presentar los siguientes ejemplos:

Incertidumbre en los parámetros, acerca del valor verdadero de los parámetros de tipo tecnológico, usados en el análisis.

Incertidumbre en el modelo, que no son apropiado para representar la realidad.

Incertidumbre en el tiempo, ya que los resultados de las determinaciones se materializarán en el futuro, futuro que se desconoce.

Incertidumbre en los componentes climáticos y temporales, ya que un fenómeno climático puede influir y cambiar un proyecto.

El enfoque conocido como diseño en condiciones de incertidumbre, ha sido estudiado por más de cuatro décadas. En las décadas de los 80 y 90, se comenzaron a enfocar en su estudio, antes de que se desarrollaran las herramientas sistemáticas, el enfoque era utilizar los valores nominales para el diseño básico y luego aplicar factores de sobre diseño empíricos, como se aplica para tamaños de equipos para considerar las incertidumbres implicadas, Maroto, et al., (2001)

La utilización excesiva de factores de diseño y el uso de valores nominales, ignora otros valores posibles de las incertidumbres, además de que el empleo de factores de sobre diseño no garantiza el funcionamiento viable en toda la incertidumbre y puede variar si no se tiene conocimiento sobre el grado de flexibilidad del diseño en cuestión, lo que puede resultar en costos adicionales innecesarios.

Sobre la base de los objetivos de diseño, González Cortes et al., (2012), mencionan que los enfoques sistemáticos han sido por lo general agrupados en dos categorías. El primero, referido al diseño óptimo para un grado fijo de flexibilidad en las que el diseño debe ser factible en todos los valores inciertos en un conjunto discreto de escenarios factibles que varían con el tiempo (problema de diseño multiperíodo), y el segundo en el que el diseño debe ser factible en rangos especificados de un conjunto seminfinito de escenarios (problema general de diseño bajo incertidumbre).

La característica, que debe considerarse en el modelado matemático de estos sistemas, es el hecho de que las variaciones de las variables tienen un comportamiento aleatorio, Scenna, (2000).

Esto es, debido a características inherentes al proceso, factores climáticos o de mercado, etc., las variables de operación no tienen valores únicos y/o fijos, sino que pueden fluctuar en torno a un valor estable, normal o nominal, admitiendo cualquier valor comprendido en un determinado rango de incertidumbre. Un ejemplo de esta problemática es la formulación de los problemas de síntesis de redes de intercambio calórico flexibles y de trenes de destilación integrados flexibles, tal como se cita en Scenna y Benz (2000).

En Ingeniería Química, la función objetivo, desde un punto de vista económico, son normalmente muy sensibles a los precios utilizados a las variaciones de las variables de entradas, tales como lo son a la materia prima, energía, y también a las estimaciones del costo del capital del proyecto. Estos costos y precios se pronostican o estiman, para situaciones futuras, por lo que normalmente están sujetos a un error considerable. La estimación de costos y la predicción de los precios son inciertos, desconocidos al instante de usar las mismas en la función objetivo. También existe incertidumbre en las variables de decisión, ya sea por variación en las condiciones de las entradas de la planta, por variaciones climáticas, por variaciones introducidas por operación inestable de la planta, o por la imprecisión en los datos del diseño y las ecuaciones de restricción (Sinnott y Towler, 2012).

METODOLOGÍA

Aplicaciones en Ingeniería Química

En el contenido mínimo del programa de la materia Diseño de Procesos de la carrera de Ingeniería Química de la Facultad de Ingeniería de la Universidad Nacional de Salta, figura el diseño en condiciones de incertidumbre. En esta sección se presentan los contenidos teóricos mínimos para abordar el tema los cuales son tres: Árbol de decisión, Diseño bajo incertidumbre y el Análisis de sensibilidad. Para este último tema, se enfoca el desarrollo de este estudio.
En esta sección se presentan y explican las diferentes herramientas utilizadas para cada uno de los ejemplos desarrollados en la catedra, que permiten obtener mejoras en los cálculos y análisis de estos.

Árbol de decisión
Desde el punto de vista práctico se tocan problemas típicos, uno es resolver el árbol de toma de decisión, mediante su optimización a través de programación lineal y programación lineal mixta, con búsqueda de extremo.
Esta técnica permite analizar decisiones de tipo secuenciales, basada en el uso de resultados y probabilidades asociadas. Los árboles de decisión se pueden utilizar para generar sistemas expertos, búsquedas binarias y árboles de juegos. Mediante este tipo de estructuras se permite visualizar todas las diferentes alternativas que pueden ocurrir con su correspondiente valoración económica o valor esperado de cada alternativa.
Para este tipo de problemas, se utiliza para resolverlo una aplicación que trabaja sobre planilla de cálculo, con el cual se realiza un seudo análisis de sensibilidad. Seudo en el sentido que los cambios de se realizan manualmente, de forma muy rudimentaria. Para incorporar y facilitar este tipo de análisis, se puede aplicar una herramienta, un complemento que se incorpora sobre Excel, denominado en inglés como *Simple Decision Tree*, que permite construir de forma progresiva, automatizada y muy simple, árboles de decisión complejos y elaborados (Slashdot Media, 2012).
Este tipo de complemento facilita la construcción del árbol de decisión, e incorpora en las celdas las fórmulas automáticamente, por lo que se debe ingresar de forma manual solo algunos valores, los demás son calculados inmediatamente. Este tipo de automatización permite disponer de los resultados de forma más rápida, y destinar de ese ahorro de tiempo en proponer algunos cambios en los parámetros, como también realizar un informe de los resultados aún más detallado, que para los casos realizados en papel.

Diseño bajo condiciones de Incertidumbre
Es otro aspecto que se desarrolla dentro del marco de Diseño de Procesos bajo condiciones de incertidumbre. En este tópico se presentan cuatro conceptos nuevos que son: criterios de diseños, los que también se denominan atributos del diseño, alternativas de diseño, escenarios y los resultados.
Una vez definidos los atributos del diseño, generadas las alternativas e identificados los escenarios, queda planteada una matriz de resultados, para cada una de las alternativas. Dando un problema de decisión multicriterio, por lo que se debe reducir cada matriz a un vector, ya sea por algún método, tal como el uso de la teoría de juegos con algunas de las estrategias como el de Maximin denominado criterio de Wald con una visión pesimista, Maximax con una visión optimista, el criterio de Minimax del costo de oportunidad o de Savage, el criterio Hurwicz intermedio entre la visión pesimista y la optimista o alguna otra como programación lineal, PL, (Taha, 2012).
Posteriormente para cada alternativa se debe reducir los vectores a un escalar, mediante el uso de la función de utilidad mediante la cual se realiza la transformación del vector a un solo valor escalar. Este escalar involucra la contribución de todos los atributos para los diferentes escenarios para cada alternativa, Gallardo Ku, (2018). De tal manera que al elegir la mejor de todas las alternativas, también se está eligiendo la mejor combinación de los atributos de diseños involucrados en la función

de bondad (Varian, 2012).
De los cuatro conceptos mencionados anteriormente, el conjunto de escenarios son los que están sujetos a la incertidumbre, a través de la asignación de probabilidad de ocurrencia, de materialización de dicho escenario. Los escenarios están sujetos al siguiente conjunto de propiedades:

a. Influyen significativamente sobre el resultado del sistema.
b. No dependen de la voluntad de quien toma la decisión.
c. Su valor es incierto en el momento de tomar la decisión.

Este subconjunto puede subdividirse en aquellos que:

i. No dependen de ninguna voluntad. En estos casos se estudia mediante la teoría de las decisiones individuales. Ejemplo de estos escenarios, son las situaciones climáticas.
ii. dependen de otras voluntades, con intereses distintitos. Estos se estudian mediante la teoría de juegos y del planeamiento estratégico. Un ejemplo de escenario de este tipo es cuando en el proyecto influye la posible incorporación de una nueva competencia.

La incertidumbre en este tipo de problemas está incorporada en la probabilidad de ocurrencia del escenario propuesto.
El mayor desafío e inconveniente se presenta en el método de reducción de la matriz de resultados de los criterios de diseños para cada alternativa a un valor escalar por alternativa. La tabla 1 representan los resultados de los criterios de diseños y_i para cada escenario s_k, para una determinada alternativa a_j.

$[a_j]$	s_1	s_2	s_3	...	....	s_k
y_1						
y_2						
y_3			$y_i(s_k, a_j)$			
:						
:						
:						
y_l						

Tabla 1. Matriz de resultados para la alternativa *j*.

Donde y_i (s_k, a_j) es la probabilidad de que el criterio de diseño tome el valor y_i dada la ocurrencia del escenario s_k, la alternativa de diseño a_j, y las demás hipótesis (H) impuestas al diseño, esto se expresa como p (y_i | s_k, a_j, H).
Como ejemplos de criterios de diseño se pueden mencionar algunos ejemplos dependiendo de los diferentes enfoques o puntos de vista, tales como: con un enfoque económico, el valor actual neto (VAN) de cada proyecto, desde un punto de vista social, la cantidad de mano de obra (MO) necesaria para un proyecto, desde un punto de vista ambiental podría ser la cantidad de ppm de contaminante en los efluentes (PPM) de las diferentes alternativas para el proceso, entre otros.
Con respecto a las alternativas, se podrían mencionar por ejemplo citar las diferentes alternativas para producir ácido sulfúrico: alternativa 1, proceso de cámara de plomo, alternativa 2, por el método de contacto, y así sucesivamente. Otro ejemplo también podría ser un mismo proceso y las alternativas serían las diferentes posibilidades de localización.
Finalmente, como ejemplos de escenarios desde la visión climática podrían ser: la posibilidad de que ocurra un tsunami durante un proyecto en Japón, la probabilidad de que se produzcan lluvias o sequias prolongadas para un determinado proyecto que depende de una materia prima que es un

cultivo.
Los problemas de diseño bajo incertidumbre se pueden representar también como árbol de decisión, utilizando la herramienta aplicada en el punto anterior.
Para la asignación de probabilidades, se aplica las funciones predeterminadas disponibles en los paquetes convencionales de planillas de cálculos, tales como las funciones ALEATORIO() que devuelve un numero aleatorio mayor o igual que 0 (cero) y menor que 1 (uno), por el cual el valor de la celda cambia al actualizarse o dar "Intro" (Enter) en cualquiera de las celda. También la función ALEATORIO.ENTRE (inferior; superior) que regresa un numero aleatorio diferente entre los limites especificados.
Estas funciones se pueden combinar con la función ENTERO, para generar un numero entero aleatorio. Con esta primera aproximación, se pasa desde un típico problema determinístico a un problema estocástico que cambia al azar.
Por ejemplo, para la resolución de un problema necesitamos saber el valor de la probabilidad de ocurrencia de un escenario, al especificar el valor del parámetro como un valor constante e igual a 0,3, él problema queda resulto, por lo que el problema se denomina determinístico. Este tiene un solo resultado, mientras se usen los mismos valores de las variables, se tendrá el mismo resultado.
En cambio, al usar las funciones ALEATORIO () y ALEATORIO.ENTRE(), con los que se ingresa el valor de probabilidad, estos dejan de ser fijo, haciendo que el resultado cambie cada vez que se actualiza cualquier celda, una vez asignado este valor el problema es estocástico. Por ejemplo se puede disponer de información de una fuente, donde le adjudica al valor de la probabilidad del escenario un intervalo, es decir que puede tomar un valor de 0,3 ±10%, por lo que la probabilidad se encuentra entre 0,27 y 0,33, estos valores representan los límites de la función, por lo que los valores se pueden generar de la función ALEATORIO.ENTRE(0,27;0,33), regresando valores diferentes entre esos límites, teniendo diferentes resultados para cada vez que se actualice una celda. Este problema estocástico, brindan resultados diferentes para cada valor de probabilidad, con un resultado para cada corrida o determinación.
Un aporte para el ejercicio es incorporar una tabla de datos, recurso de Excel, donde una celda toma valores de probabilidad de una tabla, por lo que la misma cambia por cada uno de los valores de la tabla, regresando una tabla de resultados diferentes. Se puede proponer o generar una tabla de datos de 100 o 1.000,….. resultados, creando con ello una simulación de tipo Monte Carlo. Con los valores de los resultados se logra elaborar una gráfica de tipo de distribución de probabilidades.
Un análisis de la salida de los resultados de este último problema nos brinda más información que en el original problema determinístico, ya que se obtiene una tabla de múltiples resultados, la que permite disponer de una distribución de probabilidad de estos.

Análisis de sensibilidad
El análisis de sensibilidad es una herramienta de gestión que permite a las organizaciones predecir los resultados de un proyecto, colaborando en la comprensión de la incertidumbre.
Desde el punto de vista formal el análisis de sensibilidad es una técnica que estudias el impacto que tiene sobre una variable dependiente de un modelo de valor (función objetivo) las variaciones en una de las variables independiente que lo conforman.
En definitiva, es observar las variaciones del proyecto ante el aumento o disminución en alguna de sus variables o parámetros claves, manteniendo el valor de las demás constante. Es decir, este análisis se ejecuta de a una variable a la vez y se supone independencia entre las distintas variables que sí o sí influyen en el valor del proyecto.
Una contribución a la mejora de este tratamiento es la incorporación de complementos de análisis de riesgo y manejo de la incertidumbre, tales como Cristal Ball, @Risk o bien Risk Simulator, todos ellos operan sobre planillas de cálculos, como Excel.

Ambas empresas cuentan con este tipo de complementos similares, presentan versiones académicas y/o trial de prueba que permite acceder a ellos por un corto periodo de tiempo. Se basan en el modelado predictivo, previsión, simulación y optimización del sistema abordado. En general los problemas de análisis de sensibilidad trabajan sobre la simulación Monte Carlo, que consiste en generar un cuadro de pronósticos que muestra rango entero de posibles valores y la posibilidad de alcanzar cualquiera de ellos, León Sánchez et al. (2004), Del Carpio Gallegos, (2007).
Permite describir un rango de posibles valores para cada celda de incertidumbre dentro de la planilla de cálculo, que corresponden a las celdas donde se ubican las variables independientes y que tienen valores inciertos. De modo que todos los supuestos que se ingresen al modelo son expresados al mismo tiempo. De acuerdo con el complemento utilizado, este rango de posibles valores puede estar en 1.000, 10.000, 100.000, valores diferentes que se toman como valores de entrada, obteniendo la misma cantidad de valores de salida, uno para cada valor ingresado. El o los resultados, valores de salida, dependen fundamentalmente de los variables independientes, por lo que, al ingresar una distribución de valores a la entrada, se obtiene una distribución de valores para la variable de salida. Para cada variable independiente se puede incorporar su rango de variación, mediante diferentes distribuciones de probabilidad. Esta simulación realiza todas las combinaciones posibles entre las variables. Por lo que brinda una distribución de salida que contempla todos los posibles resultados que se pueden dar.
Esta distribución de la variable de salida permite realizar un análisis más profundo de la situación, además se dispone de numerosa información que proporciona una conclusión más profunda, y un comportamiento más real del sistema estudiado.
En general se puede proponer diferentes tipos de distribuciones, para el valor de la variable de entrada, dependiendo de la información disponible por el tomador de decisión.
Entre las distribuciones que se disponen para asignar, a la variable de entrada o independiente, se encuentra la distribución Normal, Triangular, Uniforme, Logarítmico, Uniforme discreta, Binomial, Exponencial, Pert, Poisson, Gamma, Weibull, además de incorporar una tabla de valores.

DESARROLLO

Introducción al análisis de sensibilidad

El ejemplo típico de análisis de sensibilidad, en las carreras de Ingeniería, es el estudio de los índices de rentabilidad de un proyecto variando los diferentes factores que influyen en un proyecto de inversión. Los indicadores de rentabilidad más utilizados en el estudio de factibilidad económica de un proyecto son: el valor actual neto (VAN), también expresado como valor presente neto (VPN), la tasa interna de retorno (TIR), la relación beneficio costo (B/C), el periodo de recuperación de la inversión y el índice de rentabilidad (Towler, Sinnott, 2012), también Peters (2003).
Variando las variables independientes de las cuales depende estos indicadores, tales como el costo de la materia prima (MP), costo de la mano de obra (MO), inversión fija (IF), capital de trabajo (CW), capacidad de producción (CP), precio de ventas (PV), etc.
La selección de estas variables se realiza de acuerdo con la forma que se detalla a continuación. Primeramente, se realiza el cálculo del valor del costo total de inversión (CT), y se detallan todos los ítems, variables independientes, y en qué porcentaje contribuyen al valor del costo total de inversión. Finalmente se seleccionan aquellas variables que contribuyen e influyen en mayor medida, tomando estas como variables inciertas de entrada.
Towler y Sinnott (2012), presentan una tabla con diferentes parámetros característicos y con rangos de variaciones típicos. En general estas variaciones son simétricas e iguales, variando levemente entre variable y variable. Estos deltas de variación pueden estar entre 5%, 10% o 20% tanto para su incremento como para su disminución. Esta simetría es formal ya que la experiencia en nuestro país

muestra que es difícil que los valores por ejemplo de costos de materia prima o costos de mano de obra disminuyan de forma simétrica a su aumento, estos valores nunca decrecen.

Aplicación al ejemplo de sensibilidad

El análisis de sensibilidad se aplica al estudio de factibilidad económica, para el proyecto de producción de pentaborato de sodio con déficit de ácido, investigado por Domínguez, (2019).

Los valores característicos de los parámetros del proyecto considerado como caso base se presentan a continuación:

- ✓ Valor de la Inversión Total: 4,12 Millones U$S
- ✓ Tasa de oportunidad: 13%

Las variables independientes o, de entrada:

- ✓ Costo Materia Prima: 386,1 U$S/t
- ✓ Precio Producto: 1.200 U$S/t
- ✓ Inversión en Capital Fijo: 3,42 millones U$S/t
- ✓ Capacidad de producción: 10.000 t/a

Mientras que las variables de Salida a analizar:

- ✓ Valor Actual Neto (VAN): 10,89 millones U$S
- ✓ Tasa Interna de Retorno (TIR): 38% a

Se considera una variación de ±20% simétrica al valor base, para proceder con el análisis de sensibilidad.

Primeramente, a cada una de las variables de entrada indicadas como variables inciertas se las considera que presentan una distribución de probabilidades, para este caso se asigna una distribución triangular, ya que aparte del valor nominal se conoce un valor inferior y uno superior de un posible futuro incremento, tal como se muestra en la Fig. 1.

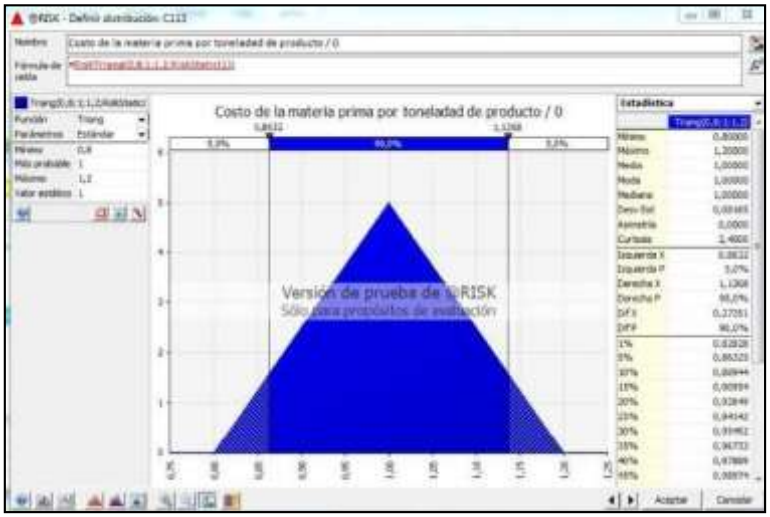


Figura 1. Distribución de probabilidad triangular para la variación de la materia prima.

De igual manera se le asigna una distribución de probabilidades a cada una de las demás variables independientes, dependiendo del conocimiento o información disponibles de las mismas. En general cuando se desconoce, de su forma de variación se toma una distribución del tipo Normal (gaussiana). Posteriormente, para la variable de salida, como el VAN, de acuerdo al complemento utilizado, sea esté Cristral Ball, @Risk o Risk Simulator, se selecciona el botón *definir prevención, Añadir salida* o bien *propiedades de pronóstico*, con dicha acción se define a la celda que presenta un

comportamiento incierto a la salida, la misma se representa por una distribución de probabilidad del VAN, de igual manera se procede para la TIR, o cualquier otro indicador que se quiera estudiar.
Finalmente, definidas las entradas, las salidas, se inicia la simulación tipo Monte Carlo, se presiona el botón Iniciar, Iniciar simulación o Correr, comienza la simulación. Este proceso realiza tantas simulaciones como se quiera y hasta 100.000 iteraciones, de todas las combinaciones de las variables de entrada posibles para generar una distribución de la variable de salida como se observa en le Fig. 2.

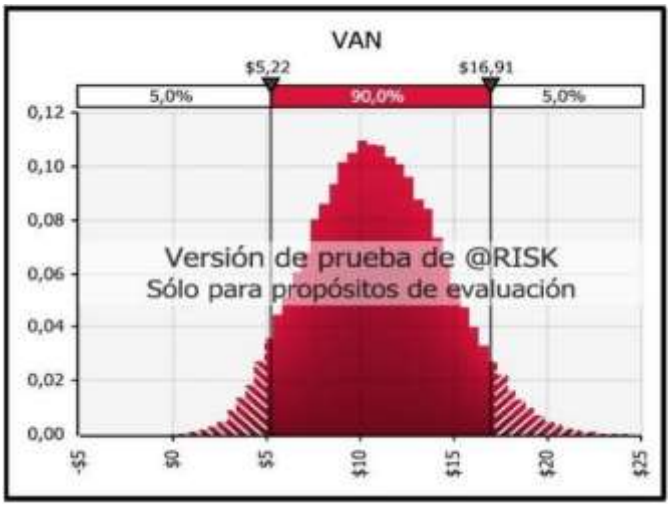


Figura 2. Distribución de probabilidad de la Respuesta VAN

En la Fig. 2 se presenta el resultado de las 100.000 iteraciones llevadas a cabo, obteniendo una distribución de probabilidades del VAN.
De igual manera, se obtiene distribuciones de probabilidad para las demás salidas requeridas.
En la Fig. 2 se observa fácilmente y a simple vista, que la probabilidad de que el VAN sea mayor a cero (considerado proyecto factible) es de 97,5%.

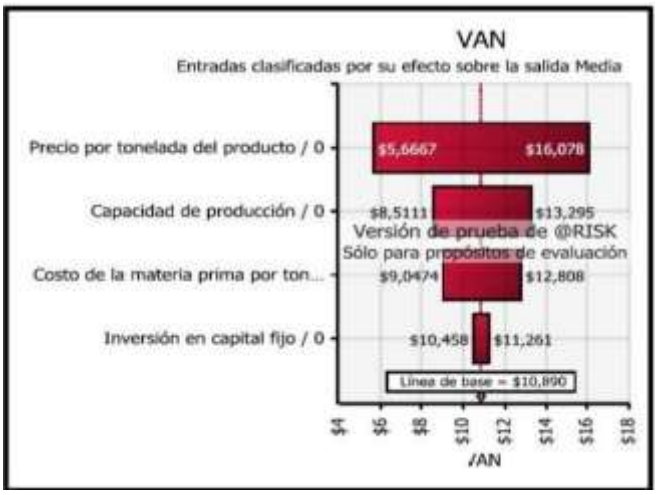


Figura 3. Curva de tornado de la respuesta VAN

Estos softwares, confeccionan para las variables de salidas, diferentes tipos de gráficos, como la

curva acumulada o el grafico de tornado que muestra en que cantidad y en orden jerárquico la contribución de las variables de entrada a la variación de la salida. En la Fig. 3 se visualiza el grafico de tornado del VAN, en él se muestra la importancia relativa de cada variable de entrada.
En la Fig. 3, se observa la importante contribución del precio del producto a la variación del VAN. También se observa la relevancia del precio del producto sobre las demás variables, mostrando una jerarquía sobre la capacidad de producción y a su vez está sobre el costo de materia prima, y finalmente sobre la inversión en capital fijo que influye muy poco en la variación del VAN. Se concluye además que una variación de ±20% de la inversión en capital fijo produce variaciones menores en el VAN.
El análisis de las gráficas de la salida de la tasa interna de retorno proporciona quizás más información, tal como se presenta en la Fig. 4.
Se evidencia en la Fig. 4 la curva de distribución de probabilidad de la TIR lo que representa la contribución de las 100.000 iteraciones de la combinación de las variables de entrada.

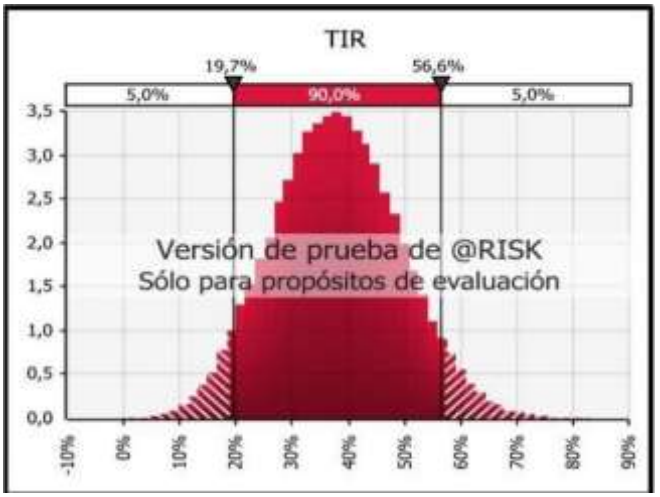


Figura 4. Distribución de probabilidad de la TIR.

El complemento, por defecto representa en la Fig. 4 la situación que se produce para el rango de probabilidad de ocurrencia del 90%, que supone que la TIR se encuentra entre un valor mínimo de 19,7% y un valor máximo de 56,6%. También se expresa que con un 95% de probabilidades el proyecto presenta valores de la TIR mayores a un 19,5%, valor éste superior a la tasa de oportunidad considerada del 13%.
Con probabilidades mayores al 95% de las veces es más rentable y supera a la tasa de oportunidad.

CONCLUSIONES

Con la utilización de estos recursos, se consiguen realizar mejores análisis y con mayor profundidad, con lo que se procesan conclusiones más amplias sobre los resultados, con métodos rigurosos, obteniendo como resultado una disminución considerable de la incertidumbre inicial.
Al emplear estos complementos, de forma independiente, o mediante la incorporación de fórmulas de planillas de cálculos, es un principio una mejora en el análisis de incertidumbre.
Para los problemas del tipo de análisis de sensibilidad de proyecto, se tiene la posibilidad de disponer de distribuciones de probabilidad de las variables de salida, debida a variaciones aleatorias en las estimaciones de los parámetros del proyecto, que también lo hacen mediante una distribución de

probabilidades, mejorando así el abordaje clásico, de variaciones individuales, con el que se obtiene un solo valor.
Al incorporar escenarios futuros e inciertos en los estudios mejoran los pronósticos de análisis.
Se ha potenciado el análisis de los resultados, en particular en los problemas abordados en la cátedra de Diseño de Procesos de la carrera de Ingeniería Química, de la Universidad Nacional de Salta.

BIBLIOGRAFÍA

Del Carpio Gallegos, J., Análisis de riesgo en la evaluación de alternativas de inversión utilizando Cristal Ball, *Gestión y Producción, Ind. Data* 10(1), p55-58, 2007.

Domínguez, O. J., *Desarrollo de Tecnologías para la obtención de boratos refinados*, 1er Edición. O.J. Domínguez, Buenos Aires, Argentina. 2019.

Gallardo Ku, J. D., *Notas en teoría de la incertidumbre*. 1a ed., Pontificia Universidad Católica del Perú, Fondo Editorial, Lima, Perú, 2018. ISBN 978-612-317-433-0

González Cortés, M., Pedraza Gárciga, J., Clavelo Sierra, D., González Suárez, E., Incertidumbre en la Integración de Procesos para el desarrollo de Biorefinerías. *Rev. Centro Azúcar* Vol 42, No. 3, Julio-Septiembre 2015 (pp. 30-38).

León Sánchez, D. P., Quintero Rodríguez, I. M., Zuñiga Muñoz, W., *Crystal Ball. Software de Análisis y Simulación de Riesgo*. Unidad de informática y Comunicaciones, Facultad de Ciencias Económicas, Universidad Nacional de Colombia, Bogotá, Colombia, PDF, 2004. http://www.fce.unal.edu.co/media/files/UIFCE/Finanzas/Crystal_Ball_1.pdf.

Maroto, A., Boqué, R., Riu, J., y F. Xavier Rius *Incertidumbre y precisión*. 2001. Técnicas de Laboratorio, 266 (2001) 834-837. http://www.quimica.urv.es/quimio/general/incert.pdf

Peters, M., Timmerhaus, K. y West, R. E., *Plant Design and Economics for Chemical Engineers*, 5th Edition, s.l.: McGraw-Hill, 2003. ISBN 0-07-119872-5.

Scenna, N., *Modelado, simulación y optimización de procesos químicos*. Ciudad Autónoma de Buenos Aires, Argentina: Universidad Tecnológica Nacional, 2000.

Scenna, N. J., Benz, S. J., Introducción al diseño de procesos químicos, Breves nociones, En Nicolás J. Scenna. (Ed.), *Modelado, simulación y optimización de procesos químicos*. Universidad Tecnológica Nacional, (pp 29-82), CABA, Argentina, 2000.

Sinnott, R., Towler, G. *Diseño en Ingeniería química*, 5ta edición, Ed. Reverte, Barcelona, España, 2012. ISBN: 978-84-291-7199-0.

Slashdot Media, Source Force: Simple Decision Tree, Sacramento, California, EEUU, (2012). https://sourceforge.net/projects/decisiontree/files/decisiontree/1.4/

Taha, H. A., *Investigación de Operaciones*, Person Education, 9na edición, Naucalpan de Juárez, Mexico, 2012.

Towler G. P., Sinnott, R., *Chemical engineering design: principles, practice, and economics of plant and process design*, Ed. Reverte, 2da Ed., United Stades of America, 2012.

Varian, Hal R., Microeconomía Intermedia, octava edición, Antoni Bosch editor, Barcelona, España, 1996.

10 y 11 de Diciembre de 2020

Breve análisis sobre los primeros brotes históricos por virus Zika en Salta, Argentina.

Juan Carlos Rosales, Américo Acosta, Celeste Herrera, Pablo Quintana, Emanuel Osedo, Diego Zerpa y Betina Abad.

Institución: Departamento de Matemática. Facultad de Ciencias Exactas Facultad de Ciencias Naturales. Universidad Nacional de Salta.

Datos de contacto: jcrmodeling@gmail.com 378-4255385

RESUMEN

La situación originada por el virus Zika, visualiza con hechos reales, algunos de los aspectos descuidados por la medicina de enfermedades tropicales y las instituciones gubernamentales y farmacéuticas. Similar a la situación del Dengue, Chikingunya, Leishmaniasis, entre otras enfermedades neglicenciadas, esta desatención permitiría, que las nuevas situaciones epidemiológicas cambien a una dinámica inusitada y peligrosa; como actualmente se observa con el nuevo corona virus, SARS-CoV-19 que, por la gravedad de la enfermedad y el impacto en la sociedad, superó a las potenciales enfermedades candidatas a originar pandemias, como, por ejemplo, la influenza.

El ingreso del Zika a Salta por primera vez, lo hizo con una intensidad mayor en el año 2017, mientras que, en el año 2018, la fuerza de infección disminuyó aproximadamente en un 34%. Al mejorar los modelos, eliminando residuos atípicos, las tasas estimadas fueron α=0,34 y α=0,23, para 2017 y 2018 respectivamente.

Los valores numéricos simulados del número reproductivo básico, R_o, en función de los tiempos de generación, para el primer brote de Zika en Salta, según los modelos de Anderson & May y de Begon *et al*., fueron R_o =1,63 (95% IC: [1,52-1,74]) y R_o =2,68 (95% CI: [2,25 - 3,11]).

Palabras Claves: Modelos Teóricos. Simulación. Fuerza de infección. ZikaV.

INTRODUCCIÓN

En poco tiempo el virus del Zika (ZikaV) adquirió un potencial explosivo para diseminarse. Su capacidad de propagación parece haber aumentado en los últimos tiempos, especialmente después de su llegada a Brasil, donde, según estimaciones del gobierno, ya ha infectado entre 440.00 y 1.300.000 personas. El ZikaV se transmite a las personas principalmente por la picadura de un mosquito de la especie Aedes que esté infectado (*Ae. aegypti y Ae. albopictus*). Estos mosquitos son los mismos que propagan los virus del dengue y del chikungunya. [1].

La situación actual originada por el ZikaV, visualiza con hechos reales, algunos de los aspectos descuidados por la medicina de enfermedades tropicales y las instituciones gubernamentales y farmacéuticas; similares a los casos de Dengue, Chikingunya, Leishmaniasis, entre otras enfermedades neglicenciadas. Esta desatención permite que, las nuevas situaciones epidemiológicas cambien a una dinámica inusitada y peligrosa, como actualmente se observa con la situación del nuevo corona virus, SARS-CoV-19; que por la gravedad de la enfermedad y el impacto que tiene en toda la sociedad, superó a las potenciales enfermedades candidatas a originar pandemias, como, la Influenza. En este caso, la preparación de reasortantes de alto crecimiento para la producción de una vacuna para la cepa H5N1 demostró ser dificultosa y lenta debido a los problemas técnicos encontrados durante los procesos de selección (p.ej., toxicidad para los huevos embrionados de gallina) [2].

Luego de haber ingresado a América del Sur, en mayo de 2015 por el Noreste de Brasil; en enero de 2016 el ZikaV ingresa a la Argentina, y en el siguiente año continua su expansión llegando a Salta. En el caso de Salta, la situación del zika es nueva, reportándose los primeros casos a principios del año 2017entre abril y junio de 2016, y proporciona paradójicamente, oportunidades de investigación de vital importancia en diferentes aspectos. Por ejemplo, el concepto teórico básico en epidemiología matemática, el número reproductivo básico ($\boldsymbol{R_0}$) [3], podría ser relacionado con el crecimiento inicial de un brote epidémico. Las hipótesis de tasas de crecimiento exponencial y logística juegan un rol fundamental [4]. Surgen inmediatamente varios interrogantes, por ejemplo, ¿Cuál será la tasa de crecimiento intrínseco de los casos de zika? ¿Es posible obtener primeras estimaciones para $\boldsymbol{R_0}$? ¿Existen diferencias entre los brotes epidémicos ocurridos en 2017 y 2018? ¿Existen diferencias entre la situación presentada en la provincia de Salta y el resto del país? ¿Cuáles fueron las rutas hipotéticas de ingreso a Salta? Establecer las tendencias de la enfermedad, identificar las áreas geográficas que requieren medidas de control estacional-patrón regular de variación entre estaciones del año, etc. precisan estudios para complementar el conocimiento del impacto del zika en Salta.

Los parámetros que se podrían estimar, conjuntamente con los modelos que se pueden analizar, resultan importantes para un registro retrospectivo, el cual permitirá la descripción y en el futuro la comparación de datos con relación a determinadas características en el tiempo. Las estimaciones y sus conocimientos permiten, tomar mejores decisiones y optimizar el diseño de planes para el abastecimiento y suministro de remedios para tratamientos, medidas de prevención y control, por parte de las autoridades de Salud Pública para las regiones afectadas [5].

METODOLOGÍA

Los brotes de enfermedades infecciosas se observan a menudo en forma de series de tiempo univariada ó multivariada, por ejemplo, el número de casos notificados acumulados durante un período de tiempo [6], número de casos acumulados durante días, semanas, meses o años. Para estimar la relación entre estas variables se estudia del primer brote histórico de Zika en la provincia de Salta, Argentina, ocurrido en el año 2017. El marco que se utiliza está clasificado como el clásico para investigar variabilidad espacio-temporal aplicada a nivel provincial. Se utiliza la información de casos de Zika por localidades georreferenciadas para los Departamentos General San Martín, Orán y Rivadavia. La modelización espacial se realiza por el diseño de un mapa de calor. El mapa de calor se construyó con archivos tipos “shapefile”, del Instituto Nacional de Geografía [7]. El software QGIS, del Equipo de desarrollo de software (QGIS 3.12) [8], fue utilizado, en particular, las herramientas para este tipo de análisis espacial. Por medio de la técnica de ventanas de promedio móviles de 4SE, se obtendrá la serie de tiempo suavizada. Se intenta construir un modelo explicativo de la evolución temporal de la variable, número acumulado de casos de Zika al comienzo del brote epidémico, con el fin de cuantificar los posibles efectos de la fuerza de infección y el número reproductivo básico $\boldsymbol{R_0}$ en la población, donde efectivamente se registraron casos.

Los datos de los casos humanos de zika notificados sospechosos y/o confirmados, corresponden a los publicados en los boletines integrados de vigilancia epidemiológica, (BIV), del Ministerio de Salud de la Argentina [9] o en los reportes de informes epidemiológicos de la OMS /PAHO, según la definición de caso correspondiente [10].

Se asume como hipótesis que, los datos acumulados de casos de Zika al comienzo del brote epidémico puede describirse mediante el modelo exponencial [11]. Una vez verificada esta situación, ya que existen expresiones qué, relacionan la tasa de crecimiento intrínseca con la fuerza de infección y el número reproductivo básico [3], [12], se obtienen primeras estimaciones de estos parámetros para Salta.

El número reproductivo básico, proporciona información sobre el potencial de propagación de la enfermedad y la dificultad de controlar la misma. La idea más simple es, considerarlo como un número único: un promedio o un tipo apropiado de promedio ponderado. Pero el número reproductivo también puede interpretarse como una distribución de probabilidades en la población de posibles infectores; distintos hospedadores que pueden tener diferentes tendencias a transmitir enfermedades.

El número reproductivo y la tasa de propagación de una enfermedad, están vinculados por el intervalo de generación $\boldsymbol{T_g}$: el intervalo entre el momento en que un individuo es infectado por un infector y el momento en que el infectado se convierte en infector. Al igual que el número reproductivo, el intervalo de generación puede pensarse en un solo número (generalmente su valor medio) o como una distribución. En este caso, se estimará $\boldsymbol{R_o}$ utilizando expresiones básicas, propuestas por Anderson & May [3] y Begon *et al*. [12], además considerando $\boldsymbol{T_g}$, como un número promedio, que considere el tiempo de infección y el tiempo de latencia del individuo infectado, simulados y distribuidos uniformemente en esos periodos, $\boldsymbol{T}\sim U$ (a, b).

El tiempo de latencia e infección influye en la estimación $\boldsymbol{R_o}$. El periodo latencia e infección del zika, en el ser humano y en los mosquitos mencionado anteriormente se podrían relacionar con los tiempos de generación que resulta en un nuevo infectado.

Se realizan simulaciones numéricas tipo Monte-Carlo, de casos con distintos tipos de tiempo de incubación (latente) y de infección, clasificados como tiempo bajo (TB), tiempo medio (TM) y tiempo alto (TA), según los tiempos de latencia e infección que podría tener el infector, se realizan para analizar la influencia de este factor en R_o.
Se simularon corridas de 50 casos para las estimaciones de R_o y de 15 casos para analizar la influencia del factor tiempo en R_o, considerando 5 casos de tiempo bajo, 5 casos de tiempo medio y 5 casos de tiempo alto, dado la incerteza de esos parámetros, en base a los tiempos de latencia e infección en seres humanos y tiempo de infección en los mosquitos. Se estimaron los R_o, con los modelos obtenidos y se realizó el análisis para determinar si existe influencia del factor tiempo, en los valores obtenidos de R_o. Las simulaciones y suavizaciones se realizaron con implementaciones en entorno Matlab y los análisis estadísticos con el software Statgraphics Centurión (16.1.03).

DESARROLLO

Distribución espacial y evolución temporal

La **Figura 1** muestra la ubicación geográfica de la región de estudio en Salta, Argentina, más concretamente, el norte y noreste de la provincia de Salta. Las localidades de los Departamentos de Orán, General San Martín y Rivadavia se indican con puntos de color rojo. Nuevamente, como con otras enfermedades negligenciadas, las zonas de ocurrencias de casos corresponden a las regiones con mayores problemáticas en el sistema de salud y en el aspecto económico de la provincia de Salta.

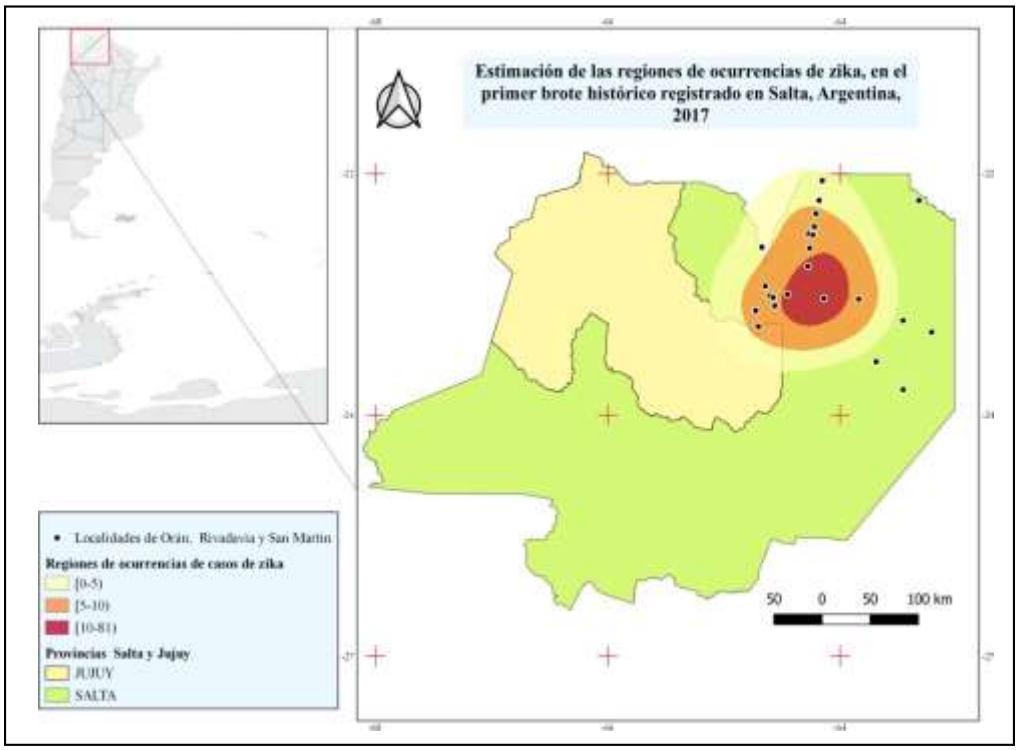


Figura 1 *Estimación espacial de regiones de mayor ocurrencia de casos de Zika del primer brote en Salta, Argentina, año 2017.*

La **Figura 2** Representa la curva suavizada de casos humanos infectados por ZikaV, durante el primer y segundo brote epidemiológico registrado en la provincia de Salta, Argentina, durante las semanas epidemiológicas comprendidas entre SE1-SE22, año 2017 y SE8- SE28. Las semanas fueron renumeradas según el origen de coordenadas. La misma fue obtenida usando la técnica de ventanas de promedio móviles de 4SE.

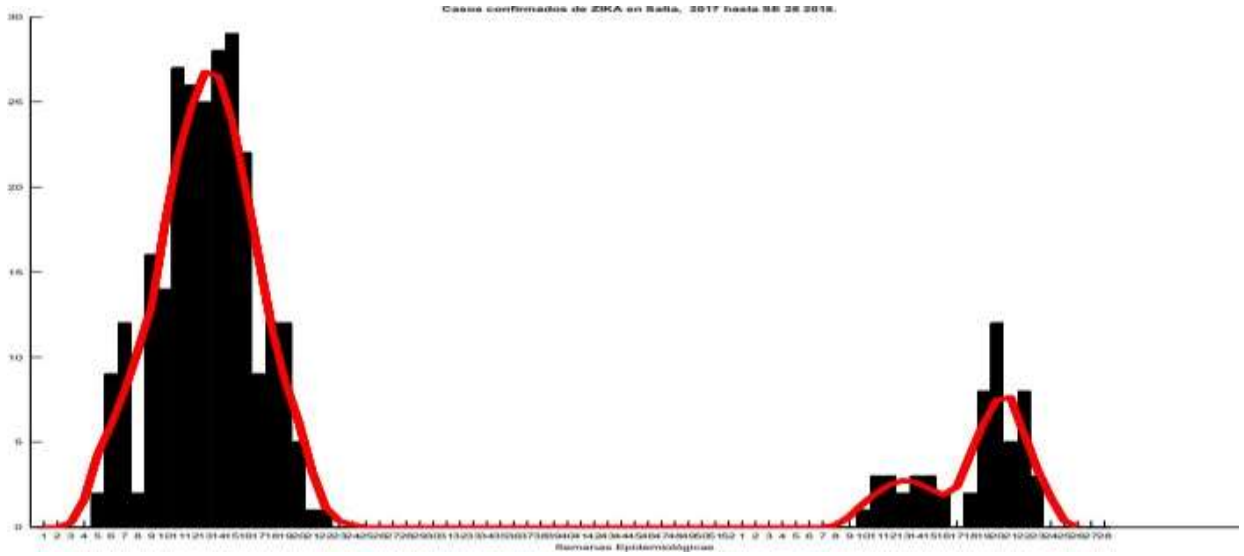


Figura 2 *Arriba: Serie de tiempo suavizada con promedios móviles de 4SE, para los casos de zika ocurridos por primera vez en la provincia de Salta, Argentina (línea roja), luego de haber ingresado a América del Sur, en mayo de 2015 por el Noreste de Brasil. El primer pico histórico de zika registrado en la provincia de Salta, ocurrió en el año 2017, entre las semanas epidemiológicas SE4-SE22. Izquierda: Segundo pico histórico de zika registrado en la provincia de Salta, ocurrido en el año 2018, entre las semanas epidemiológicas SE10-SE23. Gráficos en base a los datos publicados en BIV N°392.*

La **Tabla 1**, muestra el resumen estadístico para el primer y segundo brote histórico de zika en Salta. Incluye medidas de tendencia central, medidas de variabilidad y medidas de forma. De particular interés aquí son el sesgo estandarizado y la curtosis estandarizada, las cuales pueden utilizarse para determinar si la muestra proviene de una distribución normal. Valores de estos estadísticos fuera del rango de -2 a +2 indican desviaciones significativas de la normalidad, lo que tendería a invalidar cualquier prueba estadística con referencia a la desviación estándar. En el caso de primer brote, año 2017, el valor del sesgo estandarizado se encuentra dentro del rango esperado para datos provenientes de una distribución normal. El valor de curtosis estandarizada se encuentra dentro del rango esperado para datos provenientes de una distribución normal. Mientras que, en el caso del segundo brote histórico año 2018, el valor de sesgo estandarizado no se encuentra dentro del rango esperado para datos provenientes de una distribución normal. En este caso, el valor de curtosis estandarizada se encuentra dentro del rango esperado para datos provenientes de una distribución normal. Si se elimina el dato atípico 11, que resultó 12 casos de zika y corresponde a la SE18, se obtienen sesgo y curtosis estandarizada indicados en la Tabla 1, columna 2018sAtip, ahora dentro del rango esperado para datos provenientes de una distribución normal.

Resumen	Año	2017	2018	2018_{sAtip}
Recuento		18	13	12
Promedio		14,0	3,84615	3,16667
Desviación Estándar		9,97644	3,38738	2,4433
Coeficiente de Variación		71,2603%	88,0719%	77,1567%
Mínimo		1,0	0	0
Máximo		29,0	12,0	8,0
Rango		28,0	12,0	8,0
Sesgo Estandarizado		0,351153	2,17664	1,89451
Curtosis Estandarizada		-1,19723	1,22348	0,961304

Tabla 1 *Resumen Estadístico para casos zika en Salta ocurridos el primer y segundo brote histórico en el año 2017, SE4-SE22, 2018, SE10-SE23 y 2018 sin atípico.*

La **Figura 3** muestra la ACF para el brote de zika de la **Figura 2**. Esta gráfica muestra las autocorrelaciones estimadas entre los valores ajustados de casos de zika a diferentes retrasos. El coeficiente de autocorrelación con retardo *k* mide la correlación entre los valores ajustados de casos zika al tiempo *t* y al tiempo *t-k*. También se muestran los límites de probabilidad del 95,0% alrededor de 0 (línea roja). Si los límites de probabilidad a un retraso particular no contienen el coeficiente estimado, hay una correlación estadísticamente significativa a ese retraso al nivel de confianza del 95,0%. En este caso, uno de los 24 coeficientes de autocorrelación (se muestran 6), retraso *k*=1, au=0,657 (es= 0,236), es estadísticamente significativo al nivel de confianza del 95,0%, implicando que la serie de tiempo puede no ser completamente aleatoria (ruido blanco). Esta estimación podría sugerir posibles modelos autoregresivos para representar la dependencia encontrada en los datos, por ejemplo, del tipo AR (1).
En el caso del segundo brote 2018, ver **Figura 4**, ninguno de los 24 coeficientes de autocorrelación son estadísticamente significativos, implicando que la serie de tiempo puede ser completamente aleatoria.

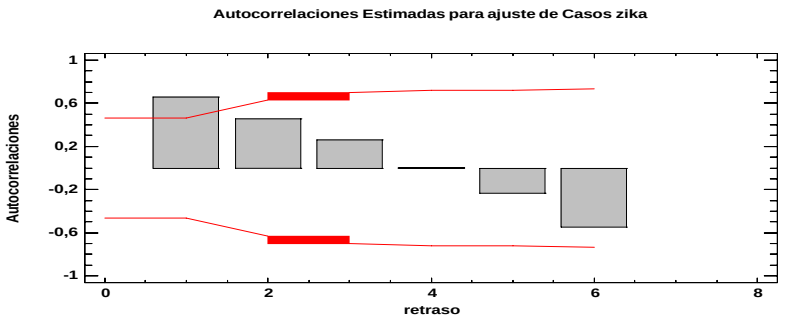


Figura 3 *Función de autocorrelación muestral (ACF) para el primer brote histórico de Zika en Salta, ocurrido en el año 2017 entre las semanas epidemiológicas SE4-SE22.*

Estimación de tasas de crecimiento

Se consideran las SE del primer brote histórico de Salta, ocurrido en el año 2017, entre SE4-SE22. El número de casos acumulados de infectados por ZikaV (N_{czac}), se ajustaron según el modelo exponencial $N_{czac}=\beta \cdot exp\ (\alpha \cdot SE)$, se estimaron los parámetros por regresión lineal. Para aplicar la regresión a los datos de la variable repuesta se les aplicó la función logaritmo natural (LN_{czac}).

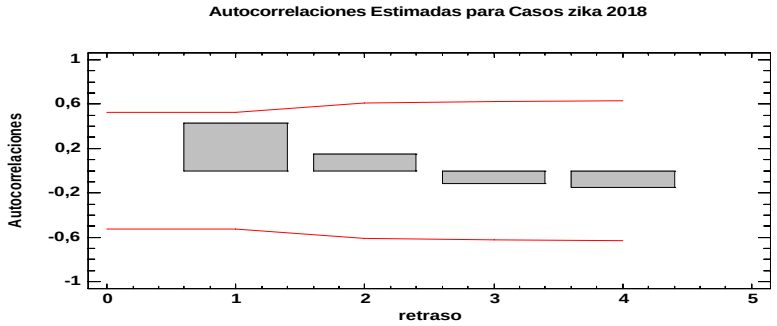


Figura 4 *Función de autocorrelación muestral (ACF) para el segundo brote histórico de Zika en Salta, ocurrido en el año 2018 entre las semanas epidemiológicas SE10-SE23.*

Analizando en base a estadísticos de bondad de ajuste del modelo lineal y al valor del Criterio de información Akaike, AIC, se determinó el modelo adecuado, en función de la variable repuesta casos acumulados de zika y la variable explicativa SE.
En la **Tabla 2** se muestran las estimaciones para 5, 6, 7, 8, 9, y 10 SE, tanto para el primer brote 2017 y el segundo brote en 2018.

SE	Función	Modelo	R^2	R^2_{Aj}	sd	MAE	AIC
10	$LNcz_{ac}$= 1,34 + 0,42 · SE	Ncz_{ac}= 3,82 · $e^{(0,42 \cdot SE)}$	88,34	86,88	0,48	0,33	-1,05
9	$LNcz_{ac}$= 1,21 + 0,45 · SE	Ncz_{ac}= 3,36 · $e^{(0,45 \cdot SE)}$	88,20	86,52	0,48	0,32	-1,02
8	$LNcz_{ac}$= 1,08 + 0,49 · SE	Ncz_{ac}= 2,95 · $e^{(0,49 \cdot SE)}$	87,74	85,70	0,48	0,32	-0,95
7	$LNcz_{ac}$= 0,94 + 0,53 · SE	Ncz_{ac}= 2,58 · $e^{(0,53 \cdot SE)}$	86,56	83,87	0,50	0,34	-0,82
6	$LNcz_{ac}$= 0,80 + 0,53 · SE	Ncz_{ac}= 2,23 · $e^{(0,59 \cdot SE)}$	84,70	80,87	0,52	0,35	-0,37
5	$LNcz_{ac}$= 0,57 + 0,69 · SE	Ncz_{ac}= 1,77 · $e^{(0,69 \cdot SE)}$	84,35	79,14	0,54	0,38	-0,43

SE	Función	Modelo	R^2	R^2_{Aj}	sd	MAE	AIC
10	$LNcz_{ac}$= 0,69 + 0,28 · SE	Ncz_{ac}= 2,00 · $e^{(0,28 \cdot SE)}$	79,93	77,42	0,46	0,34	-1,15
9	$LNcz_{ac}$= 0,61 + 0,31 · SE	Ncz_{ac}= 3,36 · $e^{(0,31 \cdot SE)}$	77,99	74,85	0,48	0,35	-1,03
8	$LNcz_{ac}$= 0,46 + 0,35 · SE	Ncz_{ac}= 1,58 · $e^{(0,35 \cdot SE)}$	80,15	76,85	0,46	0,34	-1,03
7	$LNcz_{ac}$= 0,26 + 0,42 · SE	Ncz_{ac}= 1,31 · $e^{(0,42 \cdot SE)}$	83,73	83,48	0,43	0,31	-1,09
6	$LNcz_{ac}$= 0,08 + 0,48 · SE	Ncz_{ac}= 1,08 · $e^{(0,48 \cdot SE)}$	85,42	81,78	0,42	0,30	-1,06
5	$LNcz_{ac}$= -0,13+ 0,58 · SE	Ncz_{ac}= 2,59 · $e^{(0,58 \cdot SE)}$	86,46	81,95	0,42	0,29	-0,95

Tabla 2 *Estimaciones del modelo exponencial por regresión lineal según semanas epidemiológicas para los casos acumulados de Zika en Salta ocurridos en el primer y segundo brote histórico en el año 2017, SE4-SE22 y 2018, SE10-SE23, respectivamente.*

Los coeficientes de determinación varían entre 84,35 y 88,34 para el brote de 2017, mientras que para el brote de 2018 varían desde 77,99 hasta 86,46. Para el primer pico histórico ocurrido en 2017, el modelo adecuado según el coeficiente de determinación y el criterio de información de AIC, corresponde a la semana epidemiológica 10, SE10. Mientras que, para el caso, del brote de 2018, según el mayor coeficiente de determinación se tendría el modelo para la semana SE5, mientras que por el criterio de información AIC, se tendría la SE 10, el mejor modelo que minimiza el criterio de información.
Las tasas estimadas proporcionan información de la fuerza de infección de la enfermedad en los brotes en sus primeras semanas de ocurrencia. Así para el primer brote histórico del ingreso del

zika en Salta, la estimación de la fuerza de infección ($\hat{\alpha}= \alpha$), resultó, α_{2017} =0,42; mientras que, en el año 2018, fue α_{2018} =0,28 si se considera AIC o bien α_{2018} =0,58 si se considera el coeficiente de determinación, no obstante, habría que analizar la situación de la interpretación del intercepto negativo.

Refinamiento de modelos estimados

La mejor estimación para las tasas de crecimiento intrínseco del primer y segundo brote histórico, en la fase inicial, según el análisis anterior, corresponden a las primeras 10 semanas epidemiológicas. Puede observarse, en la curva epidémica, que posteriormente a la SE 10, ver **Figura 2** en los brotes a derecha e izquierda, la curva comienza a descender en las próximas semanas.

En la **Tabla 3**, se observa que, tanto para 2017 y 2018_{sAtip}, los valores del sesgo, curtosis estandarizada, correspondientes a la variable LN_{czac}, se encuentran dentro del rango esperado para datos provenientes de una distribución normal.

	2017	2018	2018_{sAtip}
Recuento	10	10	9
Promedio	3,62	2,26	2,51
Desviación Estándar	1,34	0,98	0,58
Coeficiente de Variación	36,93%	42,75%	2324%
Mínimo	0,69	0	1,38
Máximo	5,08	3,29	3,29
Rango	4,39	3,29	1,91
Sesgo Estandarizado	-1,52	-2,06	-1,00
Curtosis Estandarizada	0,92	1,77	0,24

Tabla *3* *Resumen estadístico para $LNcz_{ac}$ de casos zika en Salta, ocurridos en las 10 primeras SE en los brotes del año 2017 y 2018.*

Para verificar la distribución normal de las variables correspondiente a LN_{czac}, de los brotes de zika, se observaron además las descripciones estadísticas proporcionadas por el método gráfico de probabilidad normal.

Los parámetros se estiman por regresión lineal, para el primer brote año 2017, dado que el valor-P (P=0,0001), en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre $LNcz_{ac}$ y SE con un nivel de confianza del 95,0%. En la **Tabla 2** se puede ver que el estadístico R^2, coeficiente de determinación, indica que el modelo ajustado explica 88,34% de la variabilidad en $LNcz_{ac}$. El coeficiente de correlación es igual a 0,94, indicando una relación relativamente fuerte entre las variables. El error estándar del estimado indica que la desviación estándar de los residuos es 0,48. Este valor puede usarse para construir límites de predicción para nuevas observaciones. El error absoluto medio (MAE) de 0,33 es el valor promedio de los residuos. Mientras que para el segundo brote histórico se obtuvieron un coeficiente de determinación, R^2 =79,93%, sd=0,46 y MAE=0,34, respectivamente.

Para intentar mejorar las estimaciones, se elimina el dato con mayor residuo atípico. La **Tabla 4**, y la **Figura 5** muestran los modelos seleccionados para la 10 SE, eliminando los residuos atípicos. Se observa que, si se elimina el valor con residuo atípico, mayores que 2, SE1, en el brote 2017, se mejoraría el coeficiente de determinación R^2=96,99, sd=0,18 y MAE=0,14. Si se

eliminan valores con residuos atípicos, mayores que 2, SE1 y SE4, en el brote 2018, se mejoraría R^2=89,90, sd=0,22 y MAE=0,17.

Año	Función	Modelo	R^2	sd	MAE
2017	$LNcz_{ac}$= 1,81 + 0,34 · SE	Ncz_{ac}= 3,82 · $e^{(0,34 \cdot SE)}$	96,99	0,18	0,14
2018	$LNcz_{ac}$= 1,04 + 0,23 · SE	Ncz_{ac}= 2,83 · $e^{(0,23 \cdot SE)}$	89,90	0,22	0,17

Tabla 4 Estimación de parámetros del modelo exponencial, por regresión lineal, eliminando residuos atípicos, para los primeros brotes históricos de zika, 2017-2018, ocurridos en Salta, Argentina.

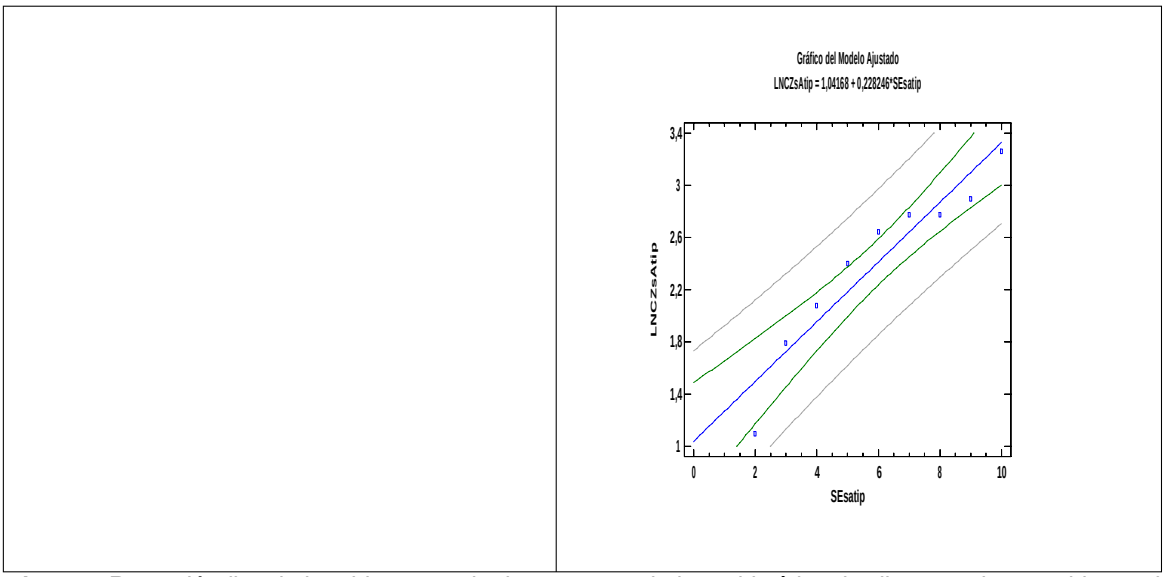


Figura 5 Regresión *lineal obtenidas para el primer y segundo brote histórico de zika en Salta ocurrido en el año 2017y 2018 eliminando residuos atípicos*.

El análisis y los parámetros estimados, permiten seleccionar los modelos exponenciales explicitados en la **Tabla 4**.

Las correspondientes tasas de crecimiento intrínseco *a*=0,42 y *a*=0,28 para los brotes históricos de zika ocurridos en Salta, **Tabla 2**, indican que el ingreso del zika en Salta por primera vez, lo hizo con una intensidad mayor en el año 2017, mientras que, en el año 2018, la fuerza de infección disminuyó aproximadamente en un 34%. Similarmente, con los modelos refinados Tabla 4, las tasas estimadas fueron a=0,34 y a=0,23, para 2017 y 2018 respectivamente, sin embargo, la disminución de la fuerza de infección resultó también del mismo orden, 34%.

Estimaciones de R_0

En los seres humanos el período de incubación del zika es de 3 a 12 días. El período de infección en los seres humanos dura entre 5 y 7 días, después de la aparición de los síntomas. El período de incubación extrínseca del virus en los mosquitos es de alrededor de 10 días. En *Aedes aegypti*, se podrían hallar altos niveles de virus en los mosquitos entre 20 y 60 días después de la infección, aunque el promedio de vida de la hembra *Aedes aegypti* adulta, es más

corto en las zonas tropicales [9], [13]. Estos valores determinan intervalos que permiten simular los tiempos de generación aleatoriamente.
Los modelos encontrados proporcionan estimaciones del número reproductivo básico. Los valores de las estimaciones para el número reproductivo básico R_0, según el tiempo de generación T_g, utilizando las tasas estimadas, en orden creciente, se muestran en la **Figura 6**.

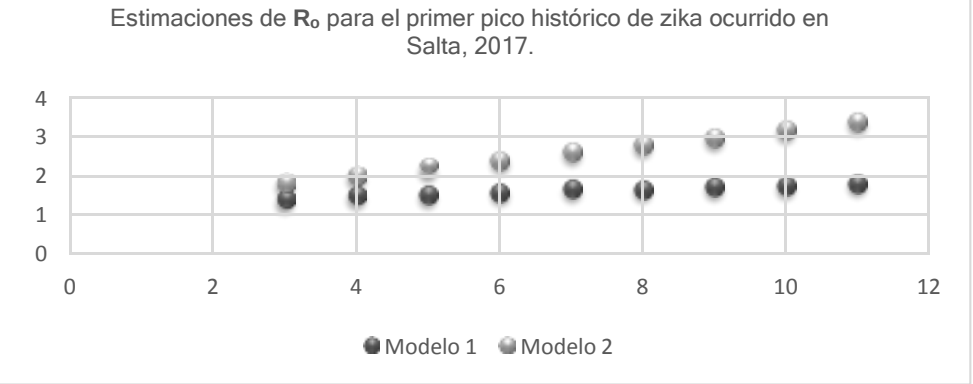


Figura 6 *Estimaciones para R_0 obtenidas a partir de los Modelos de Begon et. al. y Anderson May, según diferentes tiempos de generación T_g, para el primer brote histórico de zika en Salta, Argentina ocurrido en 2017.*

El resumen estadístico de los valores estimados en la **Figura 6**, se presenta en la **Tabla 5**, donde α ($\alpha = 0{,}34$), es la tasa de crecimiento intrínseco, T_g el tiempo de generación y R_0 el número reproductivo básico. Los autores mencionados, presentan las relaciones entre R_0, α, T_g, en sus modelos [3], [12].

Investigadores	Begon *et al.* [12]	Anderson May [3]
Modelo	$R_0 = exp(\alpha T_g)$	$R_0 = 1 + \alpha T_g$
Recuento	10	10
Promedio	2,68	1,63
Desviación Estándar	0,60	0,15
Coef. de Variación	22,42%	9,2%
Mínimo	0,69	1,34
Máximo	1,78	1,84
Rango	4,39	0,47
Sesgo Estandarizado	-5,85E-10	-0,25
Curtosis Estandarizada	-0,77	-0,54
Intervalo de Confianza	[2,25 - 3,11]	[1,52 - 1,74]

Tabla 5 *Descripción estadística de las estimaciones de R_0 usando la tasa estimada para el primer brote histórico de zika, ocurridos en Salta, Argentina, en al año 2017, periodo SE4-SE22, según los modelos especificados.*

Así, según los modelos lineal de Anderson & May [3] y exponencial de Begon *et al.* [12], se obtuvo para esta simulación, un valor promedio de R_0 =1,63 (95% IC: [1,52-1,74]) y R_0 =2,68 (95% CI: [2,25 - 3,11]), respectivamente.

La tabla ANOVA descompone la variabilidad de ***Ro*** estimado por el modelo lineal en contribución

debido al factor latencia-infección. El valor-P prueba la significancia estadística del factor analizado. Puesto que el valor-P es menor que 0,05, el factor en el caso analizado, no tiene un efecto estadísticamente significativo sobre las estimaciones del ***Ro*** proporcionadas tanto para el Modelo lineal como para el exponencial, con un de nivel de confianza del 95,00%.

En la **Figura 7** se muestran los intervalos LSD (Least Significant Difference) correspondientes a los tipos de tiempo de latencia e infección, baja, media y alta (relacionado con la viremia). Se observa que los valores estimados del número de reproducción básico, Ro, tanto para el modelo lineal, como para el exponencial, estudiados, no difieren significativamente uno del otro, los intervalos se solapan. Se observa también que a factor tiempo más alto mayor valor de Ro.

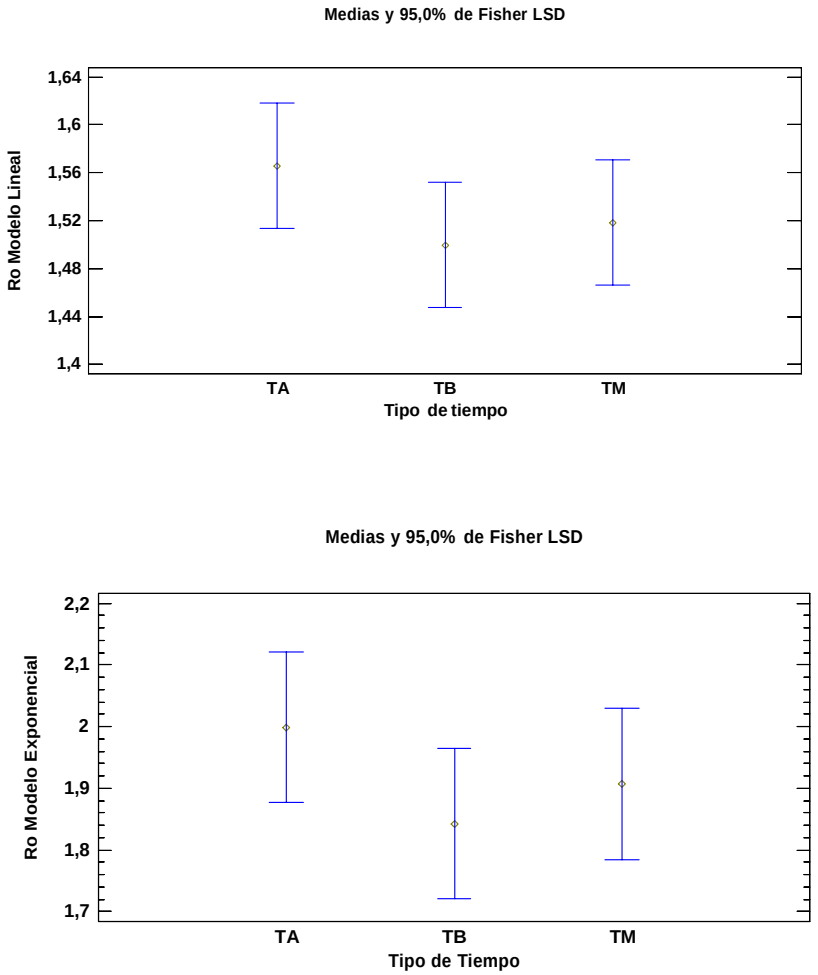


Figura 7 ***Arriba****: Influencia simulada del tipo de tiempo (latencia e infección) (TA: tiempo alto, TB: tiempo bajo y TM: tiempo medio) en la estimación del* R_0 *(número reproductivo básico), según el modelo lineal.* ***Abajo****: Influencia del tipo de tiempo (latencia e infección) en* R_0 *según el modelo exponencial, en el primer brote histórico de zika ocurrido en Salta, año 2017, en el periodo SE4-SE22.*

CONCLUSIONES

Se presentó un breve análisis de los primeros brotes históricos que han ocurrido en Salta por infecciones de Zika. La distribución espacial de casos obtenida señala regiones de los departamentos General San Martín, Orán y Rivadavia, la infección afectó más a localidades de los departamentos en el orden mencionado. Embarcación localidad del Departamento Gral. San Martín, registró el mayor número de casos en el ingreso a Salta del ZikaV por primera vez. Estas regiones con grandes problemáticas en la provincia de Salta, en lo que respecta al sistema sanitario, social y económico, se caracterizan además por ser afectadas por una serie de enfermedades que la OMS considera negligenciadas o desatendidas. Las inversiones gubernamentales, debido a las bajas tasas de incidencia y mortalidad de las mismas se direccionan en otro sentido. Paradójicamente, hasta la irrupción, del SARS-CoV-19, enfermedad con dinámica peligrosa, incontrolable y letal, donde la muerte en sí misma, no pudo disimularse y pone en dudas cualquier tipo de registros estadísticos e inversiones realizadas las instituciones gubernamentales. Esto no debería ocurrir, porque las bases de datos primarias fidedignas son fundamentales para cualquier tipo de análisis y de gran importancia para la toma de decisiones, beneficiando al gobierno en sí mismo.
El gobierno de la provincia de Salta, por medio del Ministerio de Salud Pública, podría mejorar las acciones de concientización y prevención con la provisión no solo de repelentes, sino proveer telas mosquiteras para puertas y ventanas en estas regiones subtropicales, como así también mosquiteros para las camas, destinados a aquellos pobladores con necesidades básicas insatisfechas que viven en esas regiones subtropicales. Todo esto, beneficiaría el control no solo del Zika, sino también, de enfermedades como, Dengue, Chikingunya y Leismaniasis.

La aparición del ZikaV en Salta por primera vez, ha presentado características y patrones generales con las cuales las enfermedades emergentes se inician, con relativamente pocos casos, picos cortos, alternados por fases de silencio, en este caso de tipo estacional.

Por otro lado, se estimaron las tasas de crecimiento intrínseco para los brotes históricos de zika registrados por primera vez en Salta, Argentina, en el año 2017 y un brote menor como segundo pico en 2018. Las tasas de crecimiento intrínseco estimadas fueron α=0,42 y α=0,28 para los años 2017 y 2018 respectivamente. El ingreso del zika a Salta por primera vez, lo hizo con una intensidad mayor en el año 2017, mientras que, en el año 2018, la fuerza de infección disminuyó aproximadamente en un 34%. Se mejoraron (refinaron) los modelos eliminando los residuos atípicos, las tasas estimadas fueron α=0,34 y α=0,23, para 2017 y 2018 respectivamente, sin embargo, la disminución de la fuerza de infección resultó también del mismo orden, 34%.

Con estas tasas se estimaron valores numéricos del número reproductivo básico, $\boldsymbol{R_o}$ en función de los tiempos de generación, para el primer brote de zika en Salta. Según los modelos lineal Anderson & May [3] y exponencial de Begon *et al*. [12], se obtuvo un valor promedio de $\boldsymbol{R_o}$ =1,63 (95% IC: [1,52-1,74]) y $\boldsymbol{R_o}$ =2,68 (95% CI: [2,25 - 3,11]), respectivamente. Se considera, más realista para la región la estimación proporcionada por el modelo de Anderson & May, por las descripciones observadas en la serie temporal de casos.

Simulaciones numéricas de casos con distintos tipos de tiempo de generación, clasificadas como alto, medio y bajo, según el tiempo de latencia e infección del infector y sus $\boldsymbol{R_o}$ relacionados

fueron analizados. El factor mencionado permitió observar, que el tipo de tiempo, en el caso de zika, no tiene un efecto estadísticamente significativo sobre las estimaciones del número reproductivo básico, R_o, proporcionadas tanto por el Modelo lineal como el exponencial, encontrados a un de nivel de confianza del 95,00%.

Este tipo de análisis proporciona parámetros, que podrían utilizarse en estudios retrospectivos y también como estimaciones que pueden ser introducidos en sub-modelos para describir la dinámica de los ciclos tantos urbanos como silvestres de esta zoonosis que afectan diferentes lugares del mundo, pero precisan el conocimiento de parámetros regionales para modelizaciones realistas [14].

BIBLIOGRAFÍA

[1] Schuler-Faccini L, Ribeiro E, Feitosa I, Cavalcanti D, Pessoa A, Doriqui M, Schuler-Faccini J, et al. Possible association between Zika virus infection and microcephaly – Brazil, 2015. Morbidity and Mortality Weekly Report. 22 January. 2016.
[2] OMS Plan de Preparación para la Pandemia de Influenza. El Rol de la Organización Mundial de la Salud y Guías para la Planificación Nacional y Regional. 2000.
[3] Anderson R, May R. Infectious Diseases of Human. Dynamics and Control. Oxford UK, Oxford Sciences Publications. 1991.
[4] Villela DAM, Bastos LS, Carvalho LM, Cruz OG, Gomes MFC, Durovni B, Lemos MC, Saraceni V, Coelho C and Codeco. Zika in Rio de Janeiro: Assessment of basic reproduction number and comparison with dengue outbreaks. Epidemiology Infectious. 2017; 1-9.
[5] MOPECE. Módulo de Principios de Epidemiología para el Control de Enfermedades, Tema3. Organización Mundial de la Salud. Organización Panamericana de la Salud, 2011.
[6] Stocks T, Britton T, Hohle M. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. Biostatistics. 2020;21(3):400-416
[7] Instituto Geográfico Nacional (IGN). Argentina. Available:http://www.ign.gob.ar/.Geodatos
[8] QGIS Software Development Team (QGIS 3.1.12). Disponible: https://qgis.org/es/site/
[9] Boletín Integrado de Vigilancia. Dirección Nacional de Epidemiología. Análisis de la situación de salud. Ministerio Salud de la Argentina. ISSN 2422-698X
[10] The year Zika evolved from an emergency into a long-term public health challenge. Pan American Health Organization PAHO; 2016.
[11] Banks RB. Growth and diffusion phenomena. Mathematical Frameworks and Applications. Text in Applied Mathematics 14. Springer; 1994.
[12] Begon M, Townsend CR, Harper JL. Ecology, From individuals to ecosystems. Blackwell Publising, Oxford UK; 2006.
[13] Guía para la vigilancia integrada de la infección por zika y recomendaciones para el equipo de salud. Ministerio de Salud. Presidencia de la Nación Argentina, 2016.
[14] Rosales JC, Avila Blas OJ and Yang HM. Monte Carlo Simulation of American Tegumentary Leishmaniasis Epidemic. The Case in Orán, Salta, Argentina, 1985-2007. Applied Mathematical Sciences, Vol. 11, 2017, (1), 1 – 14. https://doi.org/10.12988/ams.2017.610255

III Jornadas Internacionales de Estadística Aplicada
10 y 11 de diciembre de 2020

Utilización de la estadística en la toma de decisiones en Medicina

Dr. Matías Torres Jiménez; Dr. Facundo Tomás; Dr. Iván Rollan; Dr. Jorge Floridia; Dr. Gustavo Herrera

Instituto Neurológico Salta

drgustavoherrera.gah@gmail.com – 0387-683-8017

Asesor: Ingeniero Héctor Iván Rodríguez; profesor de las cátedras de probabilidades y estadísticas y diseño experimental de la facultad de Ingeniería de la Universidad Nacional de Salta

Resumen. La toma de decisiones en la práctica clínica constituye uno de los problemas más complejos que enfrenta el médico a diario en su desempeño. Varios son los conceptos que se tienen al respecto, el más sencillo plantea que es un "proceso para elegir entre una o más opciones", otro más explicativo lo define como un "proceso de identificación y selección de la acción más adecuadapara la solución de un proceso específico".

Palabras Claves: Medicina Basada en la evidencia, riesgo relativo, el odds ratio, la reducción absoluta de riesgo y el número de pacientes que será necesario tratar (NNT), número de pacientes necesario para dañar (NND)

INTRODUCCIÓN

La toma de decisiones en la práctica clínica constituye uno de los problemas más complejos que enfrenta el médico a diario en su desempeño. Varios son los conceptos que se tienen al respecto, el más sencillo plantea que es un "proceso para elegir entre una o más opciones", otro más explicativo lo define como un "proceso de identificación y selección de la acción más adecuada para la solución de un proceso específico". Se trata de una decisión probabilística.

METODOLOGIA

Se realizó una revisión bibliográfica acerca de la aplicación de la estadística en la toma de decisiones en la práctica clínica. Se aprecian las dificultades en la comprensión de

la estadística en los diferentes niveles de atención médica, al no comprender su utilidad en la solución de los problemas de salud de la población. Se plantea la interrogante: ¿Por qué la aplicación de la estadística constituye un pilar en el buen desempeño del profesional de la salud?

DESARROLLO

Dos son las situaciones fundamentales en la toma de decisiones: la primera, el proceso del diagnóstico y la segunda, la aplicación terapéutica.

MEDICINA BASADA EN LA EVIDENCIA:

La MBE, al igual que los meta-análisis, las múltiples publicaciones aleatorizadas y los bancos de datos procedentes de investigaciones permiten acceder a una amplia información médica que ayuda en la toma de decisiones, sin embargo, la magnitud de la información dificulta su aplicación y, en ocasiones, brinda una información que a posteriori resulta falsa o de poca utilidad.

Existe sin embargo controversia entre la MBE y la medicina basada en la experiencia, debemos recordar que la MBE comenzó a utilizarse hace apenas 30 años y se basa en la utilización de las mejores evidencias disponibles en la atención al paciente, los que defienden la medicina basada en la experiencia argumentan que "la mejor evidencia científica no puede eliminar el arte milenario de la medicina". (Revista Cubana de Medicina. 2014;53(2): 114-115)

Lo más habitual es que para cuantificar el efecto de un tratamiento, si el resultado se expresa por una variable binaria, se utilicen el riesgo relativo, el odds ratio, la reducción absoluta de riesgo y el número de pacientes que será necesario tratar (NNT).

Siempre que queramos saber algo sobre una variable biológica (la diferencia de presión arterial según sexo, el efecto de un fármaco, etc.) nos encontraremos con dos problemas de difícil solución. El primero, que a nosotros nos interesa saber el valor de esa variable en la población, pero la población es inaccesible en su totalidad, motivo por el que tenemos que seleccionar una muestra representativa de esa población y trabajar sobre ella.

El segundo problema, que nos encontramos con el azar.

Es imposible librarnos del azar. Su efecto siempre estará presente en nuestros estudios. La buena noticia es que podemos intentar reducirlo (aumentando el tamaño de la muestra, seleccionando con cuidado los participantes, etc.) y, sobre todo, podemos medir cuál es su efecto. Para esto se han diseñado los contrastes de hipótesis. (hipótesis nula – hipótesis alternativa)

Aquí es donde entran en juego los diferentes test estadísticos. Para ello, a partir de los resultados, calculamos un estadístico que siga una distribución de probabilidad conocida como, por ejemplo, una t de Student. Esto nos permite saber cuál es la probabilidad de obtener un valor como el obtenido o más alejado de la nulidad

Si la probabilidad es alta, diremos que la diferencia se debe al azar y que no es probable que se cumpla en la población.

Pero si la probabilidad de obtener este valor por azar es muy baja, podremos decir que, probablemente, sí existe una diferencia real. Dicho de otro modo, rechazaremos la hipótesis nula y abrazaremos la alternativa.

Y este es el valor de p: la probabilidad de obtener, por azar, una diferencia tan grande o mayor de la observada,

Así, por convenio suele establecerse que si este valor de probabilidad es menor del 5% (0,05) es lo suficientemente improbable que se deba al azar como para rechazar con una seguridad razonable la H0 y afirmar que la diferencia es real. Si es mayor del 5%, no tendremos la confianza necesaria como para poder negar que la diferencia observada sea obra del azar.

Este es el significado de la ansiada $p < 0,05$ que muchas veces buscamos con determinación al leer los trabajos de las revistas científicas. De todo lo dicho hasta ahora, parece claro que cuando planificamos un estudio desearemos que nuestra p nos salga significativa, para poder rechazar la hipótesis nula y quedarnos con la hipótesis alternativa.

El valor de p tiene relación con la fiabilidad del estudio, cuyo resultado será más fiable cuanto menor sea la p, pero hay muchos factores que pueden intervenir además del hecho de que exista o no diferencia real: el tamaño de la muestra, la varianza de la variable medida, el tamaño del efecto, la distribución de probabilidad empleada, etc.

(Rev Pediatr Aten Primaria vol.19 no.76 Madrid oct./dic. 2017)

RIESGO RELATIVO:

El riesgo relativo (RR) es un cociente de riesgos, y su valor es 1, si su valor es mayor de 1 es más frecuente o menos frecuente si su valor es menor de 1.

ODDS RATIO:

El concepto de Odds se maneja habitualmente en el mundo anglosajón, en especial en las apuestas y corresponde a la razón entre la probabilidad de que un evento ocurra y la probabilidad de que no ocurra. Supongamos que una muestra de 100 pacientes que han recibido un tratamiento médico se ha alcanzado el éxito en 75 de ellos. Si se divide la probabilidad de curación (p= 75/100 = 0,75) por la probabilidad de no curación (25/100 = 25), se obtendrá la Odds de curación para ese tratamiento. Que valdría 3, que es el resultado de dividir 75% entre 25% (Odds= 0,75/0,25 = 3), o bien simplemente dividir 75 entre 25. ¿Cómo se interpretaría una Odds de 3 en el ejemplo? Se entendería que por cada paciente en que no se alcanzó el éxito terapéutico hay 3 en que si se logró, es decir, con ese tratamiento la probabilidad de éxito es 3 veces mayor que la de fracaso.

REDUCCION ABSOLUTA DE RIESGO:

También se le denomina Reducción Atribuible del Riesgo y Riesgo Atribuible. Es una medida útil para expresar la efectividad de un tratamiento o de una intervención. Corresponde a la diferencia entre el riesgo en el grupo sin el FR en estudio y el riesgo en el grupo con el FR en estudio. Dicho de otra forma, expresa la reducción del riesgo de aparición del EI en el grupo de sujetos con la intervención en estudio respecto de los sujetos que no reciben esta intervención, que reciben un placebo o que reciben una intervención diferente. Se calcula entonces como la diferencia del RA o incidencia en el grupo no expuesto y el RA o incidencia en el grupo expuesto. El valor del RAR puede variar entre -1 y 1; por lo que se debe interpretar de la siguiente forma: si es igual a 0, significa que no hay asociación entre el FR y el EI; si el valor es menor a 0, significa que la asociación es positiva, es decir que la presencia del FR se asocia a mayor ocurrencia del EI; y si el valor es mayor a 0, significa que la asociación es negativa, es decir que la presencia del FR se asocia a menor ocurrencia del EI (Blume & Peipert; Faulkner et al.; GENESIS, 2014).

NNT (Número Necesario a Tratar):

Medida útil para evaluar el impacto de un tratamiento o de una intervención. Se define como el número de individuos que hay que tratar con la terapia experimental para producir, o evitar, un evento adicional respecto a los que se producirían con la terapia estándar o el placebo (Cook & Sackett, 1995; GENESIS, 2014). Permite expresar la magnitud del efecto de un tratamiento en términos comprensibles, lo que facilita la toma de decisiones en salud (Cook & Sackett, 1995); expresa el efecto del tratamiento en términos que permiten comparar sus ventajas con sus inconvenientes (efectos adversos, costes, etc.). Se calcula como el cociente entre 1 y el RAR.

NND (Número Necesario para Dañar):

Este índice se puede usar para evaluar efectos adversos de una intervención. Se puede definir como el número de sujetos que deberían recibir el tratamiento experimental en lugar del estándar o el placebo, para que un sujeto adicional obtenga un perjuicio. Su cálculo tiene sentido cuando el riesgo del evento perjudicial es mayor en el grupo sometido al tratamiento experimental que en el grupo con tratamiento estándar.

Representa que el tratamiento experimental consigue menos beneficio que el estándar o un placebo; o que los efectos adversos inherentes al tratamiento son mayores en el grupo experimental. Existen calculadoras online que permiten obtener fácilmente el NNT y el NND con sus respectivos IC 95% (Sierra, 2005). El NND se calcula con respecto a la "exposición" y "no exposición", y puede determinarse para los datos brutos o corregidos de factores de confusión; de tal modo que fármacos con NND bajo pueden estar indicados en situaciones en las que el NNT, es menor que el NND.

CONCLUSIONES

La toma de decisiones en la práctica clínica constituye uno de los problemas más complejos que enfrenta el médico a diario en su desempeño; dado el creciente aumento sostenido de publicaciones científicas nos vemos en la tarea de conocer las distintas variables de asociación e interpretarlas para una correcta atención y tratamiento general de nuestros pacientes.

Por tal motivos se torna indispensable contar con los conocimientos de las Estadísticas aplicadas a la toma de decisión en medicina.

BIBLIOGRAFIA

Revista Cubana de Medicina. 2014;53(2): 114-115

(Rev Pediatr Aten Primaria vol.19 no.76 Madrid oct./dic. 2017)

Abraira, V. Medidas del efecto de un tratamiento (I): reducción absoluta del riesgo, reducción relativa del riesgo y riesgo relativo. Semergen,26(11):535-6, 2000.

Akobeng, A. K. Understanding measures of treatment effect in clinical trials. Arch. Dis. Child., 90(1):54-6, 2005.

Blume, J. & Peipert, J. F. What your statistician never told you about P-values. J. Am. Assoc. Gynecol. Laparosc., 10(4):439-44, 2003.

Cook, R. J. & Sackett, D. L. The number needed to treat: a clinically useful measure of treatment effect. B. M. J., 310(6977):452-4, 1995.

Sierra, F. Evidence-Based Medicine (EBM) in practice: applying number needed to treat and number needed to harm. Am. J. Gastroenterol., 100(8):1661-3, 2005.

MANTENIMIENTO DE LA SALUD EN ADULTOS CON DIAGNÓSTICO DE OBESIDAD QUE ASISTEN AL CENTRO DE SALUD Nº55 DEL BARRIO 17 DE OCTUBRE, AÑO 2020. (PROYECTO CIUNSA N° 2587)

Autores: Carlos Ariel Ramos Díaz - Angélica Beatriz Farfán – Mónica Millán

Facultad de Ciencias de la Salud, Sede Central.
Universidad Nacional de Salta. Salta Capital

Datos de contacto: Carlos Ariel Ramos Díaz. Mail: ariel.carlos09@hotmail.com
teléfono: 387 - 4628726

RESUMEN. La obesidad constituye un factor de riesgo y es un problema de salud multifactorial y prevenible. La investigación tuvo como objetivo determinar el Mantenimiento de la salud que presentan los adultos con diagnóstico de obesidad que asisten al Centro de Salud Nº 55. Fue un estudio observacional, de tipo descriptivo correlacional y corte transversal que se realizó sobre una muestra probabilística aleatoria (n=50). Los datos se recolectaron a través de visitas domiciliarias, utilizando un cuestionario de preguntas abiertas y cerradas. Se realizó un análisis univariado y bivariado de las características sociodemográficas, factor relacionado y características definitorias. Se concluyó que los pacientes son mujeres, obesas, adultas medias sin cobertura social, nivel de formación básico que presentan Mantenimiento Ineficaz de la Salud, y el factor relacionado deterioro en la toma de decisiones independientes que se evidencia por la presencia de las características definitorias Patrón de falta de conducta de búsqueda de salud: alimentación y actividad física. Hay relación estadística entre el Deterioro en la toma de decisiones independientes y el Patrón de falta de conducta de búsqueda de salud: actividad física, por lo que se deben plantear estrategias desde el centro de salud para disminuir el sedentarismo en este grupo de riesgo.

Palabras Claves: obesidad – etiqueta diagnostica – factor relacionado – características definitorias

INTRODUCCIÓN

La obesidad es considerada pandemia y epidemia en la Argentina, ya que su ocurrencia se presenta tanto entre países desarrollados como en aquellos que se encuentran en vías de desarrollo. Tampoco existen distinciones significativas respecto de la clase social a la que pertenece la población afectada por obesidad. El exceso de peso constituye un factor de riesgo para desarrollar enfermedades cardiovasculares, y se asocia a otros factores de riesgo como la Hipertensión, Dislipemia y Diabetes. También se relaciona con el desarrollo de apnea del sueño, depresión, empeoramiento de la osteoartritis y deterioro de la calidad de vida (Bray, 2009).

En Argentina se llevó a cabo en los años 2005, 2009, 2013 y en el presente año la Encuesta Nacional de Factores de Riesgo (ENFR) que presenta la situación de los principales determinantes del riesgo de las Enfermedades Crónicas No Transmisibles (ECNT), siguiendo con el propósito establecido en el Plan Federal de Salud 2004-2007de elaborar un estudio para superar "la ausencia de información para establecer líneas de base en algunos de los principales problemas de salud y factores de riesgo"
La obesidad es un problema de salud prevenible, su carácter multifactorial y su cronicidad requieren del compromiso tanto del equipo de salud como del mismo paciente para alcanzar los objetivos que se planteen en su tratamiento. La adquisición de comportamientos saludables que permitan preservar estilos de vidas positivos que permanezcan a largo plazo, necesita de la adherencia del paciente al tratamiento. El estudio realizado por Campos Paniagua (2015) en adolescentes de México, comprueba que la adherencia a los programas de promoción de hábitos saludables por parte del mismo paciente logra no sólo alcanzar los objetivos propuestos, sino también a mantenerlos a lo largo del tiempo. Sin embargo, el artículo realizado por Bronsens (2009) "EOPs: Barreras en la adherencia al tratamiento de la obesidad" menciona que pese a la implementación de estrategias para bajar de peso, se sabe que aproximadamente dos tercios de los pacientes que realizan un tratamiento efectivo y logran bajar de peso, recuperan su peso inicial luego de un año, y casi todos lo hacen luego de cinco años.

Las políticas públicas sanitarias de Argentina cuenta con medidas costo-efectivas para el abordaje y control de esta enfermedad y sus factores de riesgo. Algunas de estas medidas son poblacionales, por ejemplo las que se llevan a cabo bajo la Estrategia Nacional para la Prevención y Control de las Enfermedades Crónicas no Transmisibles, y otras son de carácter individual al realizar la consulta de salud en los centros asistenciales de atención primaria. La intervención del personal de enfermería que forma parte del equipo de salud de los centros asistenciales, juega un papel fundamental en su relación terapéutica con el paciente, porque trabaja a través de la educación sanitaria para que el usuario alcance un nivel adecuado de capacitación en su autocuidado. Sin embargo a pesar de todo el esfuerzo realizado, el paciente presenta dificultades para lograr este objetivo.

La realización de planes de cuidados es una herramienta fundamental de enfermería, que permite identificar los problemas de salud del paciente que acude a la consulta, realizar las intervenciones de enfermería y establecer los objetivos de trabajo. En primer lugar el personal de enfermería debe formular los diagnósticos enfermeros, para ello debe identificar al sujeto, familia o comunidad que pueden presentar signos y síntomas que se consignarán como características definitorias que den cuenta de la presencia de un problema de salud, pueden ser obtenidas mediante la observación, entrevista, exploración física o inspección. Seguidamente el enfermero deberá establecer los factores causantes o etiológicos (factores causales) que guarden relación con el problema de salud identificado. Sin embargo los diagnósticos enfermeros son individualizados, y obedece a las manifestaciones y particularidades de un sujeto, familia o comunidad de atención determinado.

La etiqueta diagnóstica "Mantenimiento ineficaz de la salud" (00099), que se define como la "incapacidad para identificar, manejar o buscar ayuda para mantener la salud" (NANDA 2015- 2017) figura en investigaciones de enfermería en pacientes con obesidad, realizados por Vargas Olegario y cols. (2017) "Plan de cuidados de enfermería en el adolescente con obesidad". También el Manual realizado por Piñar Oya y Pérez dirigido a enfermeras de las unidades de hospitalización polivalente en los Hospitales de Alta Resolución en España, ofrece entre otros, planes de cuidados para pacientes con diagnóstico de obesidad.

La mencionada etiqueta, presenta factores relacionados que son un componente integral de todos los diagnósticos de enfermería enfocados en el problema, son etiologías, circunstancias, hechos o influencias que tienen algún tipo de relación con el diagnóstico enfermero (por ejemplo, causa, factor contribuyente). Para la siguiente investigación se utilizó el factor relacionado Deterioro en la toma de decisiones independientes, el cual se entiende como el proceso de toma de decisiones del individuo en relación con los cuidados sanitarios que no incluye el conocimiento del afectado ni tiene en cuenta las normas sociales. También se tuvo en cuenta las características definitorias Patrón de falta de conducta de búsqueda de salud y Desinterés por mejorar las conductas de salud. Estas son señales o inferencias observables que se agrupan como manifestaciones de un diagnóstico (signos o síntomas) de carácter objetivo o subjetivo. Una evaluación que identifica la presencia de una serie de características definitorias presta apoyo a la precisión del diagnóstico enfermero y permite abordar de manera específica dichas manifestaciones.
Por todo lo expuesto, el problema a investigar se delimita de la siguiente manera:

¿Cuál es Mantenimiento de la Salud en adultos con diagnóstico de obesidad que asisten al Centro de Salud Nº 55 del Barrio 17 de octubre, Salta Capital, año 2019?

METODOLOGIA

Objetivo General

- Determinar el Mantenimiento de la salud (00099) que presentan los adultos con diagnóstico de obesidad que asisten al Centro de Salud Nº 55 del Barrio 17 de octubre, agosto a diciembre del año 2019.

Objetivos Específicos

- Caracterizar sociodemograficamente, según edad, sexo, presencia de obra social, nivel de estudio, ocupación y el estado nutricional de los pacientes objeto de estudio.

- Identificar el factor relacionado Deterioro en la toma de decisiones independientes (00243 dominio 10 clase 3) de asistencia al centro de salud.

- Describir las características definitorias, Patrón de falta de conducta de búsqueda de salud: alimentación, actividad y ejercicio; Desinterés por mejorar las conductas de salud: consumo de tabaco y alcohol.

- Determinar la relación entre el Deterioro en la toma de decisiones y la presencia de una característica definitoria.

Hipótesis estadística

H0: el **factor relacionado** es independiente de las **características definitorias.**
H1: el **factor relacionado** no es dependiente de las **características definitorias.**

El estudio es observacional, de tipo descriptivo correlacional y corte transversal, se trabajó con una muestra seleccionada mediante el método de muestreo probabilístico aleatorio con reposición de unidades de análisis, sobre la población con diagnóstico de Obesidad

(N=185), que pertenece al área de responsabilidad del Centro de Salud Nº 55 de Barrio 17 de Octubre en el período de agosto a diciembre del año 2019. El tamaño muestral para la investigación (n = 50) fue calculada a través de una hoja de cálculo del programa Excel con un 95% de confianza, un p = 0,05 y un error muestral de 0,04. Se consideró como criterio exclusión para el presente estudio, a los pacientes que habían sido seleccionados pero que cambiaron de domicilio y aquellos que presentarán un problema biológico que le impida realizar actividad física.

El acceso a los datos de la muestra poblacional de interés se vio facilitada por contar con la disponibilidad de los contactos telefónicos de los pacientes, junto con la planilla de seguimiento y control suministrado por el Centro de Salud Nº 55,con previa autorización al Proyecto CIUNSa Tipo B, titulado "Disposición para mejorar la gestión en salud de patologías metabólicas: hipotiroidismo y obesidad", siempre manteniendo la confidencialidad de la información obtenida de las unidades de observación. Asimismo los participantes otorgaron su consentimiento para participar del estudio, pudiéndose retirar del mismo cuando lo soliciten.

Para la recolección de datos se realizó una entrevista como instrumentos un cuestionario, una balanza Omron de bioimpedancia y un tallimetro, con estos últimos se controló peso y talla a los pacientes. El cuestionario utilizado fue estructurado y permitió relevar información sobre: 1) características sociodemográficas (edad, sexo, presencia de obra social, nivel de estudio y ocupación); 2) estado nutricional a través del Índice de Masa Corporal; 3) los factores relacionados como la asistencia al centro de salud y 4) las características definitorias como alimentación, actividad física, consumo de tabaco y consumo de alcohol, para la elaboración de un Diagnóstico Enfermero focalizado en el problema, construido en función de un modificador y el foco diagnóstico. Para lo que se consideró:

- La Etiqueta Diagnóstica: Mantenimiento ineficaz de la Salud. Partiendo de la variable Mantenimiento de la salud que se categorizó en:
  1. Eficaz, cuando no presenta ninguna característica definitoria,
  2. Ineficaz, cuando presenta una o más características definitorias

- El Factor relacionado: Deterioro en la Toma de Decisiones Independientes que se consigna de acuerdo a la identificación de la variable Asistencia al Centro de Salud categorizada de la siguiente manera
  1. Asiste
  2. No asiste

- Las Características definitorias:
  1. Patrón de Falta de Conducta de Búsqueda de Salud: Alimentación. Se consideró las variables Alimentación (consumo de cuatro comidas diarias, consumo de menos de 4 comidas diarias).
  2. Patrón de Falta de Conducta de Búsqueda de Salud: actividad y ejercicio. Se consideró la Variable Actividad Física (realiza – no realiza).
  3. Desinterés por Mejorar las Conductas de Salud: Consumo de Tabaco. Se consideró la variable Consumo de Tabaco (fuma – no fuma).
  4. Desinterés por Mejorar las Conductas de Salud: Consumo de Alcohol. Se consideró la variable Consumo de Alcohol (consume – no consume).

La aplicación de los instrumentos se realizó en un horario programado y acordado previamente con los pacientes que conformaron la muestra, obteniendo el consentimiento

de los mismos de forma verbal o por vía telefónica. El cuestionario estuvo dirigido por el investigador y no se estableció un tiempo límite para ser respondido.

Se llevó a cabo una prueba piloto del cuestionario en una muestra con características similares, para comprobar su confiabilidad y validez.

Para el procesamiento, sistematización y análisis de los datos se utilizó el programa Excel, SPSS 20 y EPIDAT 3.1. Se realizó un análisis univariado y bivariado de las características sociodemográficas, factor relacionado y características definitorias. Parala prueba de hipótesis estadísticas se utilizó la Prueba de Chi Cuadrado con Corrección por Continuidad de Yates.

DESARROLLO

CUADRO N° 1: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNÓSTICO DE OBESIDAD, SEGUN CARACTERISTICAS SOCIODEMOGRAFICAS. CENTRO DE SALUD Nº55 DEL BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE AÑO 2019.

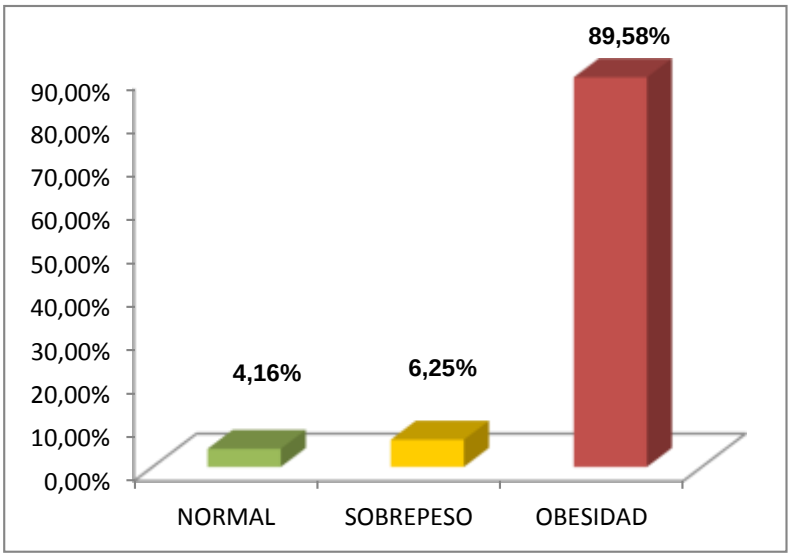
EDAD	N	%
ADOLESCENTE	2	4
ADULTO JOVEN	24	38
ADULTO MEDIO	19	48
ADULTO MAYOR	5	10
TOTAL	50	100
SEXO	**N**	**%**
MASCULINO	5	10
FEMENINO	45	90
TOTAL	50	100
COBERTURA SOCIAL	**N**	**%**
SI	15	30,6
NO	35	69,4
TOTAL	50	100
NIVEL DE FORMACION	**N**	**%**
BASICO	43	86
AMPLIADO	7	14
TOTAL	50	100

En el CUADRO N° 1 se observa que del total de pacientes con diagnóstico de obesidad que asisten al Centro de salud N° 55, el 48% es adulto medio (edad de 41 a 65 años) y el 38% adulto joven (edad de 20 a 40 años). Esta distribución de datos presenta un promedio de edad de 41 años con un desvío estándar de 15 años, es decir que la mayor concentración de los pacientes presenta edades comprendidas entre los 26 y los 56 años, siendo la edad mínima de 16 años y la máxima de 73 años. Según las medidas de forma, se trata de una distribución levemente asimétrica (AS = 0,26) y platicurtica (curtosis = -1, 04).

Se evidencia además, que el 90% son mujeres y del total, el 69,4% no presentan cobertura social (sin obra social) razón por la cual asisten al centro de salud a realizar controles de enfermería y consultas médicas.
Por último, el 86% de los pacientes tiene un nivel de formación básico (primario y secundario), conocer estas características sociodemográficas, permiten al personal de salud programar, y ejecutar estrategias de salud para la promoción de hábitos saludables

y la prevención de enfermedades y/o complicaciones principalmente en base a la edad y el nivel de formación.

GRAFICO N°1: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN ESTADO NUTRICIONAL. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.

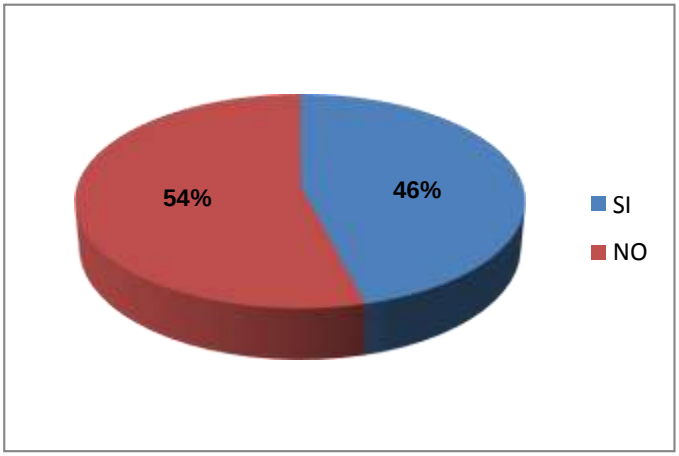


Se observa que el 89,58% de la muestra de estudio presenta obesidad ya que el IMC calculado es mayor a 30 kg/cm2.

El Índice de Masa Corporal es una medida calculada a partir del cociente entre peso expresado en Kg y la talla al cuadrado en cm, lo que permite determinar el estado nutricional de un individuo a partir del valor de dicho cociente que permite establecer lo siguiente: a) Normal, IMC < 25 kg/cm2; b) Sobrepeso, IMC entre 25,1 – 29,9 kg/ cm2 y, c) Obesidad, IMC > 30 kg/cm2. De esta manera se puede brindar educación para la salud y pautas de estilos de vida saludables (alimentación saludable, realización de ejercicio físico diario, eliminación de hábitos nocivos como el tabaquismo y el consumo de alcohol, entre otros).

FACTOR RELACIONADO

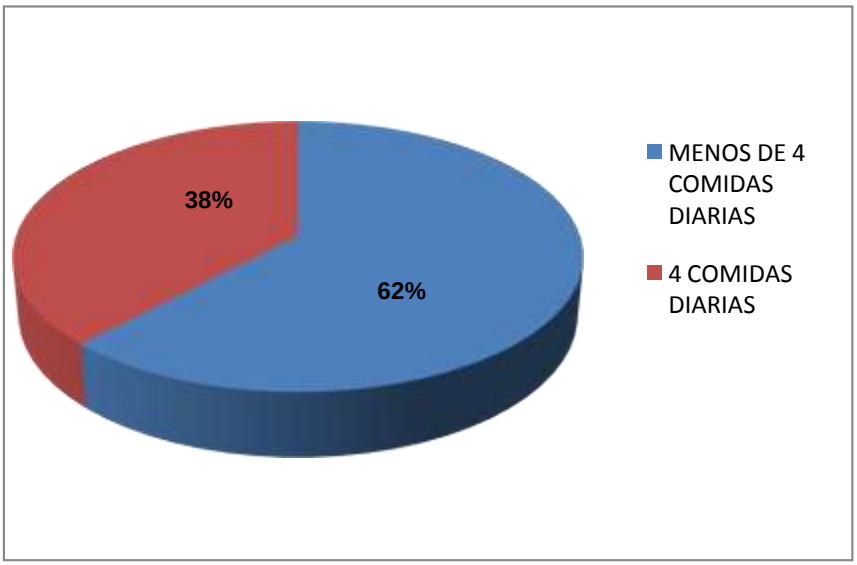
GRAFICO N° 2: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN DETERIORO EN LA TOMA DE DECISIONES INDEPENDIENTES. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019



Con respecto al factor relacional para la variable Mantenimiento de la Salud, se observa que el 46% presenta deterioro en la toma de decisiones independientes, evidenciado por la no concurrencia al Centro de Salud para la realización de los controles de rutina (peso, talla, control de tensión arterial) y consulta con los distintos profesionales de la institución: médicos, nutricionistas, psicólogos, etc.
Conocer este factor, permite al profesional de enfermería abordar estrategias para garantizar que el individuo en cuestión utilice los servicios del sistema de salud como una herramienta de prevención y no como una solución rápida a su enfermedad o complicaciones. Por este motivo, se utiliza de manera periódica y sostenida la visita domiciliaria a este grupo de riesgo.

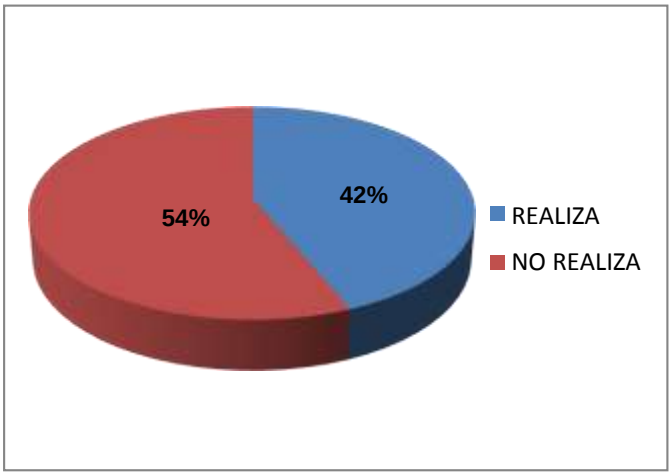
CARACTERISTICAS DEFINITORIAS

GRAFICO N° 3: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN PATRON DE FALTA DE CONDUCTA DE BUSQUEDA DE SALUD: ALIMENTACION. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.



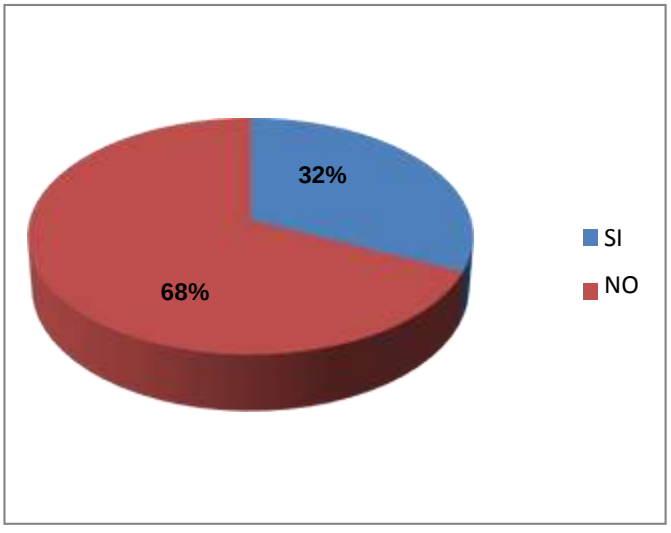
En este grafico se observa, que el 62% de la muestra de estudio presenta el Patrón de falta de conducta de búsqueda de salud: alimentación.
Una alimentación saludable y equilibrada consiste en que el individuo debe respetar el número máximo de comidas diarias y el intervalo entre las mismas (desayuno, almuerzo, merienda y cena). Esta característica definitoria se manifiesta a través de que los pacientes con diagnóstico de obesidad consumen menos de cuatro comidas al día, donde el 8% omite el desayuno y el 54% la cena.

GRAFICO N° 4: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN PATRON DE FALTA DE CONDUCTA DE BUSQUEDA DE SALUD: ACTIVIDAD Y EJERCICIO. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.



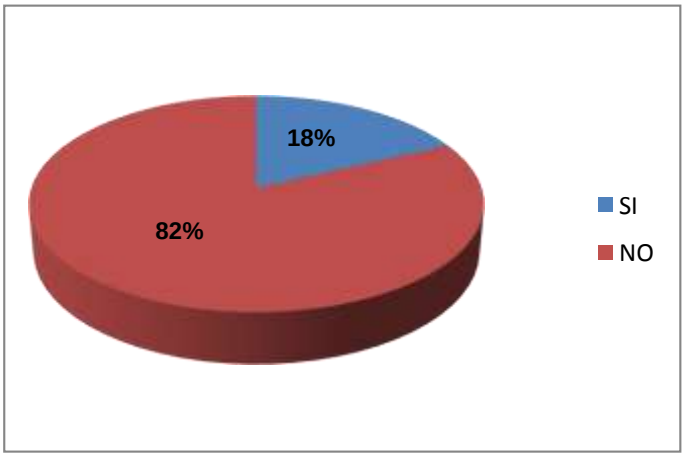
En este gráfico, se evidencia que 54% de los pacientes presenta el patrón de falta de conducta de búsqueda de salud que se manifiesta a través de la no realización deactividad física de 30 minutos de duración recomendable para evitar el sedentarismo y sus consecuencias.

GRAFICO N°5: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN DESINTERES POR MEJORAR CONDUCTAS DE SALUD: CONSUMO DE TABACO. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.



desinterés por mejorar conductas de salud, lo que se manifiesta a través del no consumo de cigarrillos, lo que provoca daños irreversibles a largo plazo a nivel pulmonar y arterial.

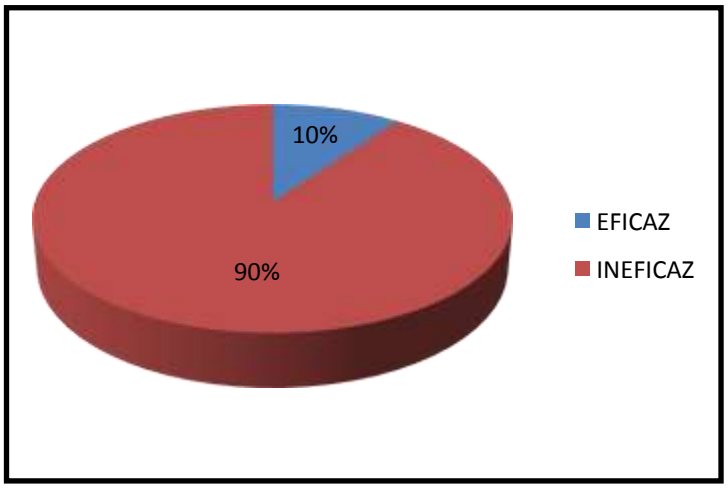
GRAFICO N° 6: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN DESINTERES POR MEJORAR CONDUCTAS DE SALUD: CONSUMO DE ALCOHOL. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.



La característica definitoria Desinterés por mejorar conductas de salud no se presenta en el 82% de los pacientes con diagnóstico de obesidad y se manifiesta a través del no consumo de bebidas alcohólicas. El consumo de alcohol a lo largo de la vida produce en estas personas la acumulación de grasa abdominal y, como consecuencia, un aumento del perímetro de la cintura y, especialmente en los hombres aumenta el índice de masa corporal (IMC).

ETIQUETA DIAGNOSTICA

GRAFICO N°7: DISTRIBUCION PORCENTUAL DE ADULTOS CON DIAGNOSTICO DE OBESIDAD, SEGÚN MANTENIMIENTO DE LA SALUD. CENTRO DE SALUD N° 55 BARRIO 17 DE OCTUBRE, AGOSTO A DICIEMBRE DEL AÑO 2019.



pacientes con diagnóstico de obesidad que asisten al Centro de Salud N° 55, presentan Mantenimiento Ineficaz de la Salud, lo que evidencia que tienen al menos una o más de las siguientes características definitorias:

1. Patrón de Falta de Conducta de Búsqueda de Salud: Alimentación. Consumen menos de 4 comidas al día.
2. Patrón de Falta de Conducta de Búsqueda de Salud: actividad y ejercicio. No realizan actividad física diaria durante 30 minutos.
3. Desinterés por Mejorar las Conductas de Salud: Consumo de Tabaco. Consumen al menos un cigarrillo a la semana.
4. Desinterés por Mejorar las Conductas de Salud: Consumo de Alcohol. Consumen alcohol al menos una vez a la semana.

HIPOTESIS ESTADISTICAS

Hipótesis estadística 1

H0: El Deterioro en la toma de decisiones independientes no se relaciona con el Patrón de falta de conducta de búsqueda de salud: alimentación.
H1: El Deterioro en la toma de decisiones independientes se relaciona con el Patrón de falta de conducta de búsqueda de salud: alimentación.

Con un $p > 0,05$ ($p = 1,0$), para un grado de libertad y un 95% de confianza se acepta la hipótesis de independencia entre las variables, por lo que se puede afirmar que no existe relación estadística entre el deterioro en la toma de decisiones independientes y el patrón de falta de conducta de búsqueda de salud: alimentación.

Hipótesis estadística 2

H0: El Deterioro en la toma de decisiones independientes no se relaciona con el Patrón de falta de conducta de búsqueda de salud: actividad física.
H1: El Deterioro en la toma de decisiones independientes se relaciona con el Patrón de falta de conducta de búsqueda de salud: actividad física.

Con un $p < 0,05$ ($p = 0,03$), para un grado de libertad y un 95% de confianza se rechaza la hipótesis de independencia entre las variables, por lo que se puede afirmar que existe relación estadística entre el Deterioro en la toma de decisiones independientes y el Patrón de falta de conducta de búsqueda de salud: actividad física.

Hipótesis estadística 3

H0: Deterioro en la toma de decisiones independientes no se relaciona con el Desinterés por mejorar conductas de salud: consumo de tabaco.
H1: Deterioro en la toma de decisiones independientes se relaciona con el Desinterés por mejorar conductas de salud: consumo de tabaco.
Con un p > 0,05 (p = 0,1), para un grado de libertad y un 95% de confianza se acepta la hipótesis de independencia entre las variables, por lo que se puede afirmar que no existe relación estadística entre el Deterioro en la toma de decisiones independientes y el Desinterés por mejorar conductas de salud: consumo de tabaco.

Hipótesis estadística 4
H0: Deterioro en la toma de decisiones independientes no se relaciona con el Desinterés por mejorar conductas de salud: consumo de alcohol.
H1: Deterioro en la toma de decisiones independientes se relaciona con el Desinterés por mejorar conductas de salud: consumo de alcohol.

Con un p > 0,05 (p = 0,4), para un grado de libertad y un 95% de confianza se acepta la hipótesis de independencia entre las variables, por lo que se puede afirmar que no existe relación estadística entre el Deterioro en la toma de decisiones independientes y el Desinterés por mejorar conductas de salud: consumo de alcohol.

CONCLUSIONES

Luego de la organización, análisis e interpretación de los datos, se puede concluir que los pacientes que asisten al centro de salud n° 55 son mujeres, obesas (IMC > 30 kg/cm2), adultas medias (41 a 65 años) con un promedio de edad de 41 años y un desvío estándar de 15 años, sin cobertura social (obra social) y nivel de formación básico (estudios primarios y secundarios). Con respecto a la **Etiqueta Diagnóstica**, presentan Mantenimiento Ineficaz de la Salud, como **Factor Relacionado**, Deterioro en la toma de decisiones independientes (no concurrencia al centro de Salud), evidenciado por la presencia de las **Características Definitorias**, 1) Patrón de falta de conducta de búsqueda de salud: alimentación porque consumen menos de cuatro comidas al día, omitiendo el desayuno y/o la cena y 2) Patrón de falta de conducta de búsqueda de salud: actividad física ya que no realizan la actividad física diaria recomendada. Por último, hay relación estadística entre el factor relacionado Deterioro en la toma de decisiones independientes y la característica definitoria Patrón de falta de conducta de búsqueda de salud: actividad física, por lo que se deben plantear estrategias desde el centro de salud para disminuir el sedentarismo en este grupo de riesgo.

BIBLIOGRAFÍA

- American Psychological Association Herdman, TH (2012) (Ed) NANDA International Diagnósticos Enfermeros. Definiciones y clasificaciones 2012- 2014. Barcelona Elsevier.
- Evidentia Revista Enfermera basada en la evidencia (2005) Plan de cuidados estandarizado de enfermería en obesidad. Disponible en htttps://www.areasaludplacensia.es (acceso 25/11/19).
- Vargas Olegario, A. (2017) Plan de cuidados de enfermería en el adolescente con obesidad. [Internet] Revista Médica Electrónica Portales Médicos. Disponible en https://www.revista-portalesmedicos.com/revista-medica/plan-cuidados-enfermeria-adolescente-obesidad/ (acceso 25/11/19).
- World Health Organization. Obesity. Preventing and managing the global epidemic. Report of a WHO consultation of obesity. Geneva: WHO; 1998.
- Silva, P y cols. (2019). Factors associated with metabolic syndrome in older adults: a population-based study. *Revista Brasileira de Enfermagem*, *72*(Suppl. 2), 221-228. E pub December 05, 2019.https://doi.org/10.1590/0034-7167-2018-0620
- Rodrigo-Cano, S. y cols. (2017) Causas y tratamiento de la obesidad. Nutr. Clín. Diet. hosp. 2017; 37(4):87-92 DOI: 10.12873/374 https://revista.nutricion.org/PDF/RCANO.pdf
- Quiroz, C y cols. (2018) Factores asociados con la adherencia a la actividad física en pacientes con enfermedades crónicas no transmisibles Rev. Salud Pública. 20 (4): 460-464, 2018 DOI: https://doi.org/10.15446/rsap.V20n4.62959 https://scielosp.org/pdf/rsap/2018.v20n4/460-464/es
- Quiroga-de Michelena, Maria Isabel. (2017). Obesidad y genética. *Anales de la Facultad de Medicina*, *78*(2), 192-195. https://dx.doi.org/10.15381/anales.v78i2.13216.

SAFETY AND EFFICACY OF THE COMBINED USE OF IVERMECTIN, DEXAMETHASONE, ENOXAPARIN AND ASPIRIN AGAINST COVID 19

Authors

Carvallo Héctor[1], Hirsch Roberto[2], Farinella María Eugenia[3]

[1] Professor of Internal Medicine, Universidad Abierta Interamericana, Argentina; Academic Coordinator at Eurnekian Public Hospital, Argentina

[2] Director of Postgraduate degree program in Infectious Diseases, Universidad de Buenos Aires, Argentina; Head of Department of Infectious Diseases at Hospital of Infectious Diseases Francisco Javier Muñiz, Buenos Aires, Argentina

[3] Head of Service of Clinical Medicine at Eurnekian Public Hospital, Argentina

NOTE: This Protocol and its trial were duly submitted to: ClinicalTrials.gov Identifier: NCT04425863 Eurnekian Public Hospital Protocol Record IDEA, Ivermectin, Dexametasone, Enoxaparin and Aspirin as treatment for Covid 19.

Abstract

From the first outbreak in Wuhan (China) in December 2019, until today the number of deaths worldwide due to the coronavirus pandemic exceeds eight hundred thousand people and the number of infected people arises to more than 25 million.
No treatment tested worldwide has shown unquestionable efficacy in the fight against COVID 19, according to NICE reports.
We have designed an experimental treatment called IDEA based on four affordable drugs already available on the market in Argentina, based on the following rationale:

- **I**vermectin solution at a relatively high dose to lower the viral load in all stages of COVID 19
- **D**examethasone 4-mg injection, as anti-inflammatory drug to treat hyperinflammatory reaction to COVID-infection
- **E**noxaparin injection as anticoagulant to treat hypercoagulation in severe cases.
- **A**spirin 250-mg tablets to prevent hypercoagulation in mild and moderate cases

Except for Ivermection oral solution, which was used in a higher dose than approved for parasitosis, all other drugs were used in the already approved dose and indication. Regarding Ivermectin safety, several oral studies have shown it to be safe even when used at daily doses much higher than those approved already.
A clinical study has been conducted on COVID-19 patients at Eurnekian Hospital in the Province of Buenos Aires, Argentina. The study protocol and its final outcomes are described in this article. Results were compared with published data and data from patients admitted to the hospital receiving other treatments.
None of the patient presenting mild symptoms needed to be hospitalized. Only one patient died (0.59 % of all included patients vs. 2.1 % overall mortality for the disease in Argentina today; 3.1 % of hospitalized patients vs. 26.8 % mortality in published data). IDEA protocol appears to be a useful alternative to prevent disease progression of COVID-19 when applied to mild cases and to decrease mortality in patients at all stages of the disease with a favorable risk-benefit ratio.

Introduction

In late December 2019, the incidence of atypical pneumonia cases of unknown cause was reported in the Chinese city of Wuhan. PCR (Polymerase Chain Reaction) studies found a new coronavirus, named SARS-CoV-2. The disease caused by the virus has been named COVID-19.

To design a treatment protocol using affordable, already marketed drugs, we have considered the information already available about SARS-CoV-2 and COVID-19.

SARS-CoV 2 proved to be remarkably similar to SARS-CoV, the only significant different bein a furin-binding domain in the SARS-CoV-2 protein S, which may expand tropism or increase virus transmission (1). Studies on several RNA viruses have revealed a potential role for IMPα / β1 during infection (2). Moreover, entry of SARS-CoV-2 into cells is facilitated by its optimized binding to ACE2 (angiotensin conversion enzyme 2) (3). This enzyme acts a receptor for SARS-CoV-2 and is found in multiple tissues (4, 5, 6), including alveolar epithelium of the lung, arterial and venous endothelium, smooth muscle, renal tubular epithelium, oropharyngeal mucosa and epithelium of the small intestine, largely explaining the clinical presentation of patients with COVID-19. Ivermectin, an antiparasitic drug, already known to inhibit RNA viruses by interfering with IMP IMPα / β1 (2 7prima) has been shown to inhibit SARS-CoV-2 in vitro (8). Ivermectin is already available as oral tablets and drops on the market in several countries and may be useful to treat COVID-19. Therefore, we have included Ivermectin oral solution as an antiviral drug in our treatment protocol. We selected an oral dose much higher than that being used for parasitosis but within safety margins according to previous clinical trials published in the literature (9, 10). Guzzo et al. (9) administered up to three doses of 60 mg to adults and Levy reports doses higher than 500 μg/kg in children and 400 μg/kg in children below 5 years of age with good safety profile. IDEA protocol includes doses of 24, 36 and 48 mg on days 0 and 7 of treatment equivalent to doses of ca. 300, 450 and 600 g/kg for mild, moderate and severe cases of 80-kg adults, which are lower than the maximum doses already tested in adult humans. The reason of this decision is that,in spite of pharmacokinetic data available show that the maximum plasma concentration are far below the concentration needed to achieve inhibition of SARS-CoV-2 in-vitro (11), we consider ivermectin to be potentially effective to fight COVID-19 because it tends to be distributed in different organs due to its lipophilicity and could reach higher levels in

high ACE2 expressing organs than in bloodstream (12) . Posology consisted of two doses of Ivermectin on day 0 (start of treatment) and 7. This time interval avoids accumulation of drug according to plasma half life reported in literature (9, 10). Ivermectin oral solution was used because previous studies allowed the hypothesis of a better bioavailability (higher plasma concentrations) (13).

A significant percentage of patients, ca. 30 – 50 % of total, suffering from COVID-19 may not present any symptoms (14). Most patients develop mild symptoms. However, a substantial percentage of patients develop moderate to severe forms of the disease, needing hospital care and even ICU treatment. Mortality rate is approximately 26.8 % in these patients (15).

The most frequent symptoms of COVID-19 are fever, cough, dyspnea, myalgias or fatigue. Other reported symptoms are bilateral conjunctival injection without associated secretions, hypogeusia, skin rash and hyposmia (16). A group of patients with severe forms of COVID-19 develop a cytokine storm syndrome (16). Several authors link this cytokine storm syndrome to secondary hemophagocytic lymphohistiocytosis (abridged sHLH), a poorly recognized hyperinflammatory syndrome leading to multiple organ failure and death, triggered by viral infections (17, 18**)**. The main features of sHLH include unremitting fever, cytopenias, and hyperferritinaemia. Pulmonary involvement is present in approximately 50% of patients. A cytokine profile that resembles sHLH is associated with the severity of COVID-19 disease (19). Mortality predictors from a recent multicenter retrospective study of 150 confirmed cases of COVID-19 in Wuhan, China included elevated ferritin (mean 1297.6 ng / ml in non-survivors versus 614.0 ng / ml in survivors; $p < 0.001$) and IL-6 ($p < 0.0001$), suggesting that mortality could be due to viral hyperinflammation (19). Therefore, we have included Dexamethasone injection in our treatment protocol for moderate and severe cases to treat hyperinflammation (20).
Even though COVID-19 is primarily a respiratory disease, accumulating data suggests it to be profoundly prothrombotic (21, 22, 23, 24, 25, 26, 27, 28). In fact, microthrombosis in different locations has been repeatedly reported in patients with COVID 19. This symptom is due to a hypercoagulable state (21, 22, 24) caused by this disease. Therefore, we have included aspirin tablets as antithrombotic agent for mild and moderate cases (29) and enoxaparin injection as anticoagulant for severe cases in IDEA treatment protocol (30, 31, 32).
Thus, this four-drug treatment protocol may allow to control viral load, as well as the most serious symptoms leading patients to ICU treatment and death. In moderate to severe cases we added ventilation and standard supportive care. We present in this article the final outcomes of a clinical trial on COVID-19 patients treated with this protocol.

Materials and Methods

Materials

Ivermectin 0.6 mg/ml solution, Dexamethasone 4-mg injection, Enoxaparin injection and Aspirin 250-mg tablets purchased from the Argentinean market were used.

Study design

The study was a single-center, prospective clinical trial run at Eurnekian Hospital in the Province of Buenos Aires, Argentina from May 2020 to July 2020. For ethical reasons, all patients received treatment and supportive care.

Patients

Patients included in this study protocol were male and female persons not less than 5 years old, with a positive rt-PCR diagnosis of COVID-19 performed on nasal swab specimens, able to provide informed consent and not participating in any other clinical study. polymerase chain reaction (PCR), no participation in other clinical trials during the study period, and able to provide informed consent. Pregnant women and persons with previous reports of allergy to any of the drugs included in the treatment were excluded.

Treatment

Symptoms were classified as mild and severe according to the following table:

MILD SYMPTOMS	SEVERE SYMPTOMS
Fever not above 38.5 °C Isolated diarrheal episodes Hyposmia or Hypogeusia Mild desaturation (93 – 96 %) Dyspnea without matter Polymyoarthralgias, Persistent headache, Abdominal pain	Fever above 38.5 °C Diarrhea (more than 3 daily depositions) Flictenular conjunctivitis Strong desaturation (92% or less) Tachypnea (FR> 25 / minute)

Disease stages were classified according to the following table:

Mild stage	Moderate stage	Severe stage
Only mild symptoms and no clinical sign of viral pneumonia	3 severe symptoms or 2 severe and 2 mild symptoms. Clinical signs of viral pneumonia	4 severe symptoms or 3 severe symptoms and not less than 2 mild symptoms. Clinical signs of bilateral viral pneumonia

The following treatmen protocol was used on each case according to the following table:

DISEASE SEVERITY	IVERMECTIN ORAL SOLUTION	DEXAMETHASONE	ANTITHROMBOTIC /ANTICOAGULANT	VENTILATION
Mild stage	24 mg on days 0 and 7	None	Aspirin 250-mg tablet once daily for at least 30 days	No
Moderate stage	36 mg on days 0 and 7	Dexamethasone 4- mg injection daily	Aspirin 250-mg tablet once daily for at least 30 days	Low Flow Washed Oxygen or Oxygen Concentrator
Severe stage	48 mg via gastric cannulae on days 0 and 7	Dexamethasone 4- mg injection on day daily.	Enoxaparin 100 UI/kg (ca. 1 mg/kg) daily	Mechanical Ventilation

Patients at mild stage of COVID-19 were treated as outpatients. They came to the hospital to receive drugs and remote follow-up via mobile phone was implemented. Patients at moderate stage of disease were immediately admitted to ward care. Patients at severe stage of COVID-19 were immediately admitted to ICU.

The day on which each patient starts treatment is numbered as "day 0". All other days are numbered in relation to this one.
Inpatients were discharged from ICU to hospital ward care when their symptoms were compatible with moderate stage disease.
Inpatients were discharged home from hospital when they tested negative for COVID-19 as determined by rt-PCR on specimens obtained from nasal swabbing or when their symptoms were compatible with mild stage of disease.
Outpatients were considered cured after one negative COVID-19 determination by rt-PCR on specimens or 10 days without any symptoms.

Ethics Committee aproval

The study was approved both by the ethics committee of the hospital and the County Ethics Committee, in accordance with the Declaration of Helsinski and its ammendments. All included patients signed an informed consent before inclusion. This study was registered in ClinicalTrials.gov website under the identifier number: NCT04425863**.**

Outcomes

The primary outcomes were:

- Percentage of patients progressing from mild to moderate or severe stages of disease
- Mortality rate by day 30

The secondary outcomes were safety outcomes related to treatment adverse events and dose adjustments for any of the drugs used.

Statistical analysis
Data were collected and recorded in MS-Excel spreadsheets and processed. Demographic data, prevalence of different comorbidities and outcomes were calculated. The outcomes of the study were compared with data from the literature and, in the case of moderate to severe cases, with a group of patients admitted to the hospital in the same period of time who did not join the study protocol and received other treatments.

Results

Population characteristics
A total of 167 patients were included. All of them had confirmed COVID19 infection by the rtPCR method. Average age was 55.7 years, 48.5 % were female and 51.5 % were male. The stages of disease of all included patients when they join the study were:

Disease stage at inclusion	Patients (n)	Average age (years)	Sex (female/male %)
Mild	135	55.7	48.5 / 51.5
Moderate to severe	32	59.7	57.5 / 42.5
Total	167	56.5	50,2 / 49,8

From the moderate to severe cases included, 23 (71.9 %) presented at list one risk factor.

Disease progression and mortality

All the135 patients who joined the study at a mild stage of COVID-19 did not worsen illness and had no need of hospitalization of any kind.

Regarding the remaining 32 patients, only one of them died. This patient had been included already at a severe stage of disease. The remaining 31 patients did not worsen during treatment.

Comparison with other treatments
Overall mortality rate of patients treated according to IDEA protocol was 0.59 % (1 death in 167 treated cases). As a comparison, estimated overall mortality rate in Argentina is approximately 2.1 % (official data by September 2nd, 2020).
Regarding moderate to severe cases, i.e. patients needing hospitalization, only 1 patient out of 32 receiving IDEA treatment died (3.1 %), whereas mortality rate published in articles from Spain, Italy and Spain is ca. 25 %. Moreover, a group of 12 patients were hospitalized in Eurnekian hospital in the same period but did not receive IDEA treatment. Three of them died, thus presenting a mortality rate of 25 %, i.e. significantly higher than that of those receiving IDEA treatment.

Adverse events
Only one patient suffered from a serious adverse event. It was a patient who had previous history of gastric ulcer and contracted one during treatment, probably caused by Dexamethasone injection. The problem was immediately solved with iced water through oropharyngeal cannulae and omeprazole, without discontinuation of treatment or dose adjustment.

Dose adjustments
No dose adjustment was necessary.

Discussion

IDEA treatment has proved to be efficacious in preventing worsening of the symptoms of COVID-19 in practically all treated patients.
Overall mortality rate in patients treated according to IDEA protocol is significantly lower (0.59 %) than that of the infected population in Argentina (2.1 % by September 2nd, 2020) and other countries, like Brazil (3.1 % by September 2nd, 2020) today. Quite strikingly, only 1 out of 32 hospitalized patients died of COVID-19 when treated according to the IDEA protocol, whereas the published data show a 26.8 % mortality for inpatients (15), i.e. 8 deaths instead of 1. A similar mortality rate of 25 % (3 deaths) was observed in a group of 12 patients admitted to Eurnekian hospital in the same period, who received other treatments.
Regarding disease progression, no patient with mild COVID-19 progressed to moderate or severe disease after treatment according to IDEA protocol. Moreover, even though we did not have a quantitative follow-up of time to absence of symptoms, we have observed it to be less than 3 days in many mild cases.
We consider that our data, even though not being placebo-controlled due to ethical reasons, allow us to conclude that IDEA protocol may be of use to help stop COVID-19 progression and reduce hospitalization and mortality.
Based on the outcomes of this study, a possible preventive strategy for COVID-19 in communities of high viral circulation might consist of an oral dose of ivermectin lower than 24 mg (proposed 12 mg) regularly administered once a week (approximately one incubation period) to low-risk people for a limited period of time, while high risk population remains isolated. This dose might be enough to reduce viral load at a low level to keep COVID-19 at a mild stage, without eliminating SARS-CoV-2 completely, so that immunity against SARS-CoV-2 is developed individually to finally reach herd immunity ("immunizing effect"). This hypothesis is worth further exploration for the prevention of transmission in healthcare workers and close contacts, and, if successful, may be further applied for prevention in the community. In the present situation in some American countries like Brazil and Argentina, this could help reduce overall mortality in the absence of a vaccine.

Conclusions

Given the rapid advance of COVID-19 pandemic in Argentina and several other countries and the lack of evidence of efficacy of any other treatments, we consider IDEA protocol to be a scientifically based, low-cost treatment and sustainable healthcare strategy with a favorable

benefit-risk ratio, based on the results reported in this article: reduction in overall mortality to one fourth, reduction in mortality of hospitalized patients to one eighth and no progression of mild cases to moderate or severe cases. It should be emphasized that the earlier the treatment starts, the better. Mild cases early treated according to our protocol did not progress to more severe stages. Furthermore, even though not exactly quantitated, we noticed the best results in the treatment of mild cases, most of them recovering in less than 3 days. The application of early treatment is consistent with the principles of medicine: even the most efficient therapeutic measures will lose efficacy if they are applied too late. In conclusion, IDEA seems to be an adequate treatment strategy for pandemic COVID-19 disease at all stages of disease, but specially so, if applied at the earliest possible stage thereof.

References

1) Andersen KG, Rambout A, Lipkin WI, Holmes EC, Garry RF (2020), The Proximal Origin of SARS-CoV-2. Nat Med. 2020 Apr; 26 (4):450-452. doi: 10.1038/s41591-020-0820-9
2) Sharun K, Dharma K, Patel SK, Pathak M, Tiwari R, Singh BR et al. (2020), Ann Clin Microbiol Antimicrob, 19:23, doi: 10.1186/s12941-020-00368-w
3) Gebhlawi M, Wang K, Viveiros A, Nguyen O, Zhong JC, Turner AJ (2020), Circulation Research. 2020; 126:1457–1475, doi: 10.1161/CIRCRESAHA.120.317015
4) Carly G K Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, Cao Y, et al. (2020), SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets Across Tissues. Cell. 2020 May 28;181(5):1016-1035.e19. doi: 10.1016/j.cell.2020.04.035.
5) Hamming I, Timens W, Bulthuis MLC, Lely AT, Navis GJ, van Goor H (2004), Tissue Distribution of ACE2 Protein, the Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. J Pathol. 2004 Jun;203(2):631-7. doi: 10.1002/path.1570.
6) Hao Xu, Liang Z, Deng J, Peng J, Dan H, Zeng X et al. (2020, High Expression of ACE2 Receptor of 2019-nCoV on the Epithelial Cells of Oral Mucosa. Int J Oral Sci. 2020 Feb 24;12(1):8. doi: 10.1038/s41368-020-0074-x
7) Heidary F, Gharebaghi R (2020), Ivermectin: a systematic review from antiviral effects to COVID-19 complementary regimen, J Antibiot (Tokyo). Jun 12:1-10. doi: 10.1038/s41429-020-0336-z
8) Caly L, Druce JD, Catton MG, Jans DA, Wagstaff KM (2020), The FDA-approved Drug Ivermectin Inhibits the Replication of SARS-CoV-2 in Vitro, Antiviral Res. Jun; 178: 104787. doi: 10.1016/j.antiviral.2020.104787
9) Guzzo CA, Furtek CI, Porras AG, Chen C, Tipping R, Clineschmidt CM et al. (2002), Safety, Tolerability, and Pharmacokinetics of Escalating High Doses of Ivermectin in Healthy Adult Subjects. J Clin Pharmacol. 2002 Oct;42(10):1122-33. doi: 10.1177/009127002401382731
10) Levy M, Martin L, Bursztein AC, Chiaverini C, Miquel J, Mahé E et al. (2019), Ivermectin safety in infants and children under 15 kg treated for scabies: a multicentric observational study British Journal of Dermatology, doi: 10.1111/bjd.18369
11) Schmith VD, Zhou JJ, Lohmer LRL. (2020), The Approved Dose of Ivermectin Alone is not the Ideal Dose for the Treatment of COVID-19. Clin Pharmacol Ther. 2020 May 7:10.1002/cpt.1889, doi: 10.1002/cpt.1889.
12) Baraka OZ, Mahmoud BM, Marschke CK, Geary TG, Homeida MMA, Williams JF (1996), Ivermectin Distribution in the Plasma and Tissues of Patients Infected with Onchocerca Volvulus, Eur J Clin Pharmacol (1996) 50: 407–410, doi: 10.1007/s002280050131
13) González Canga A, Sahagún Prieto AM, Diez Liébana MJ, Fernández Martínez N, Sierra Vega M, García Vieitez JJ (2008), The Pharmacokinetics and Interactions of Ivermectin in

Humans—A Mini-review, The AAPS Journal, Vol. 10, No. 1, March, doi: 10.1208/s12248-007-9000-9

14) Nishiura H, Jung SM, Kinoshita R, Yuan B (2020), Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19), International Journal of Infectious Diseases · March 2020, doi: 10.1016/j.ijid.2020.03.020
15) Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B et al. (2020), COVID-19 Mortality Risk Assessment: An International Multi-Center Study doi: 10.1101/2020.07.07.20148304
16) Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H et al. (2020), Clinical and Immunological Features of Severe and Moderate Coronavirus Disease 2019. J Clin Invest. 2020;130(5):2620-2629. https://doi.org/10.1172/JCI137244
17) Greene AG, Saleh M, Roseman E, Sinert R (2020), Toxic shock-like syndrome and COVID-19: A case report of multisystem inflammatory syndrome in children (MIS-C), American Journal of Emergency Medicine, doi: 10.1016/j.ajem.2020.05.117
18) Winiarska VO, Grywalska E, Rolinski J (2020), Could hemophagocytic lymphohistiocytosis be the core issue of severe COVID-19 cases? BMC Medicine, 18: 214, doi: 10.1186/s12916-020-01682-y
19) Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ (2020), COVID-19: Consider Cytokine Storm Syndromes and Immunosuppression HLH Across Speciality Collaboration, UK Lancet, Mar 28; 395(10229): 1033-1034, doi: 10.1016/S0140-6736(20)30628-0.
20) Horby P, Lim WS, Emberson JR, Haynes R, Landray MJ et al. (2020), Dexamethasone in Hospitalized Patients with Covid-19 – Preliminary Report, N Engl J Med, doi: 10.1056/NEJMoa2021436
21) Abou Ismail MY, Diamond A, Kapoor S, Arafah Y, Nayak L (2020), The hypercoagulable state in COVID-19: Incidence, pathophysiology, and management, doi: 10.1016/j.thromres.2020.06.029
22) Connors JM, Levy JH (2020), Thromboinflammation and the Hypercoagulability of COVID-19. J Thromb Haemost., Apr 17, doi: 10.1111/jth.14849
23) Ranucci M, Ballotta A, Di Dedda U, Bayshnikova E, Dei Poli M, Resta M et al. (2020) The Procoagulant Pattern of Patients With COVID-19 Acute Respiratory Distress SyndromeJ Thromb Haemost. Apr 17, doi: 10.1111/jth.14854
24) Panigada M, Bottino N, Tagliabue P, Grasselli G, Novembrino C, Chantarangkul V et al. (2020) Hypercoagulability of COVID-19 patients in Intensive Care Unit. A Report of Thromboelastography Findings and other Parameters of Hemostasis, J Thromb Haemost. Apr 17, doi: 10.1111/jth.14850
25) Wichmann D, Sperhake JP, Lütgehetmann M, Steurer S, Edler C, Heinemann A, et al. (2020) Autopsy Findings and Venous Thromboembolism in Patients With COVID-19. Ann Intern Med., May 6:M20-2003, doi: 10.7326/M20-2003
26) Porfidia A, Pola R (2020) Venous thromboembolism in COVID-19 patients. J Thromb Haemost, Jun; 18 (6):1516-1517, doi: 10.1111/jth.14842
27) Tang N, Li D, Wang X, Sun Z (2020), Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost, Apr;18 (4):844-847, doi: 10.1111/jth.14768.
28) Valdes Valderrama E, Humbert K, Lord A, Frontera J, Yaghi S (2020), Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Ischemic Stroke, Stroke. 2020;51: e124–e127, doi: 10.1161/STROKEAHA.120.030153
29) M R Buchanan, J Hirsh: Effect of Aspirin on Hemostasis and Thrombosis. N Engl Reg Allergy Proc. Jan-Feb 1986;7(1):26-31. doi: 10.2500/108854186779045539.
30) Bikdeli B, Madhavan MV, Jimenez D, Chuich T, Dreyfus I, Driggin E et al. (2020) COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention,

Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review. J Am Coll Cardiol. 2020 Jun 16;75(23):2950-2973. doi: 10.1016/j.jacc.2020.04.031

31) Menezes-Rodrigues FS, Padrão Tavares JG, Pires de Oliveira M, Guzella de Carvalho R, Ruggero Errante P, Omar Taha M et al. (2020) Anticoagulant and antiarrhythmic effects of heparin in the treatment of COVID-19 patients. J Thromb Haemost, May 14:10.1111/jth.14902. doi: 10.1111/jth.14902
32) Tang N, Bai H, Chen X, Gong J, Li D, Sun Z (2020), Anticoagulant Treatment Is Associated with Decreased Mortality in Severe Coronavirus Disease 2019 Patients with Coagulopathy. J Thromb Haemost, May;18(5):1094-1099, doi: 10.1111/jth.14817

10 y 11 de Diciembre de 2020

Encuesta Virtual de Victimización y Percepción Social del Temor al Delito. Ciudad de Córdoba 2020.

Roberto González
Observatorio de Estudios sobre Convivencia y Seguridad Ciudadana
Universidad Nacional de Villa María
Ciudad de Córdoba

Olga Puente de Camaño
Universidad Nacional de Córdoba
Ciudad de Córdoba

Matías Caro
CONICET
Universidad Católica de Córdoba
Ciudad de Córdoba

Datos de contacto: inecseg@gmail.com

RESUMEN

El presente trabajo presenta, analiza y discute la metodología desarrollada y aplicada por el Observatorio de Estudios sobre Convivencia y Seguridad Ciudadana para la realización de la Encuesta Virtual de Victimización 2020 en la ciudad de Córdoba. Las medidas tomadas en la Argentina en el marco del Aislamiento Social Preventivo y Obligatorio (ASPO) primero y Distanciamiento Social Preventivo y Obligatorio (DISPO) vedaron la posibilidad de realizar principalmente la tradicional encuesta de victimización del Observatorio en la ciudad de Córdoba.

Es frente a esto que se desarrolló una estrategia que permitió realizar la encuesta de manera virtual, velando por salvaguardar la aleatoriedad, representatividad y anonimato de la misma. Así en base a formularios con códigos de verificación se recolectó y seleccionó una muestra de 1822 casos totales y 1009 efectivos.

El análisis de los resultados obtenidos con los de años anteriores permite dar cuenta de tendencias similares entre las encuestas presenciales y la virtual, por lo que se valoriza la metodología propuesta para la encuesta virtual y los resultados obtenidos mediante la misma. Sin embargo, consideramos que la comparación interanual no es prueba suficiente y que se deben llevar adelante otros estudios a los fines de obtener certezas sobre la misma

Palabras Claves: Encuesta de victimización. Criminología. Delitos.

INTRODUCCIÓN

El crimen es uno de los fenómenos sociales más difíciles de medir, está en la propia naturaleza del delito la intención de quedar oculto. Por otro lado, el delito una vez consumado puede por diversas circunstancias no ser informados por las víctimas a las instituciones públicas, de allí que la criminología se valga de diferentes instrumentos para la recolección de la información criminal (Caro, M. 2021)

La principal estrategia de recolección de estadísticas criminales es la información recolectada a partir de registros oficiales, de esta manera a partir de los datos relevados por el Ministerio Público Fiscal y la Dirección de Estadísticas y Censos de la Provincia de Córdoba, que desde el Observatorio de Estudios sobre Convivencia y Seguridad Ciudadana elaboramos y publicamos anualmente para su análisis las tasas delictivas de la provincia (González, R. et al 2020).

Sin embargo, cómo ya hemos mencionado, no todos los delitos que acontecen en la provincia son efectivamente denunciados y de allí que desde hace 5 años el Observatorio ha llevado adelante diversas Encuestas de Victimización en la Provincia. En este sentido las Encuestas de Victimización consultan a una muestra aleatoria y representativa acerca de si el sujeto a término individual o cómo miembro de su hogar ha sufrido algún delito en un período de tiempo determinado, por lo general 12 meses.

Estas indagaciones no sólo permiten estimar las prevalencias delictivas en una población, sino que, a su vez, brindan información sobre las características de los hechos, las medidas de seguridad adoptadas, la sensación de seguridad, la evaluación del desempeño de las instituciones del sistema de seguridad y por supuesto si la victima realizó o no la denuncia y de corresponder que causas la llevaron a desistir de denuncias.

La valiosa información recolectada por la Encuesta anual de Victimización en Córdoba, luego de varios años de realización debió enfrentar en 2020 las restricciones en el marco del Aislamiento Social Preventivo y Obligatorio (ASPO) y luego el Distanciamiento Social Preventivo y Obligatorio. Este conjunto de medidas adoptadas por el gobierno nacional puso en jaque la modalidad tradicionalmente presencial de recolección de la muestra de la Encuesta.

Ante esto el equipo del Observatorio, en alianza con las Universidades de la provincia y con la especial colaboración de la Policía de la Provincia de Córdoba, diseño e implementó una encuesta de victimización digital, tendiente a asegurar una muestra aleatoria y representativa que permitiese la replicación de los resultados a escala poblacional.

El diseño, implementación y evaluación de la denominada “Encuesta Virtual Córdoba de Victimización y Percepción Social del Temor al Delito” es lo que pretendemos analizar en el presente trabajo.

METODOLOGÍA

El contexto especial de Aislamiento Social Preventivo y Obligatorio (ASPO) dispuesto por el gobierno nacional a los fines de paliar la pandemia de Covid-19, condujo a la necesidad de realizar la Encuesta Córdoba de Victimización y Percepción Social del Temor al Delito de manera no presencial.

De esta manera se acordó con la Policía Barrial el envío de formularios digitales de encuestas a los miembros de los distintos grupos de seguridad de la Ciudad de Córdoba que la Policía Barrial administra. Con una cantidad de 3500 grupos y 120.000 miembros, de los cuales muchos participan en calidad de jefes de hogar y por tanto representantes de su núcleo familiar, la policía brinda asistencia de emergencias a un porcentaje importante de la población de la ciudad capital.

Dividiendo a la ciudad en los 7 segmentos con los que tradicionalmente la encuesta trabaja, se procedió a geo-localizar cada grupo en los distintos segmentos. Luego se sortearon grupos y participantes dentro de los grupos en proporción a la cantidad de población del segmento representado y de la cantidad de miembros del grupo, puesto que los grupos con mayor cantidad de miembros representan zonas más amplias o más densamente pobladas.

A partir de los grupos y sujetos sorteados, se envió en una primera instancia un folleto de sensibilización al grupo, explicando que se trataba de una encuesta de universidades provinciales, que era anónima y que sus resultados serían muy importantes para la formulación de políticas públicas en la materia.

Luego la policía procedió a enviar links específicos de respuestas a cada sujeto sorteado, cada sujeto tenía un link particular puesto que en el mismo iba incorporado un código único de verificación, esto a los fines de controlar que hubiera una respuesta por sujeto y que no se compartiera el link. Sin embargo, la confidencialidad se garantizó puesto que la policía sabía que código de verificación correspondía a cada participante, pero no podía conocer las respuestas, en tanto que un solo miembro del equipo de las universidades tenía acceso a las respuestas con su correspondiente código de verificación, pero no conocía a que personas correspondían.

En caso de encontrarse respuestas con el código de verificación duplicado las mismas se suprimían. Luego de enviados se insistió mediante el grupo de la importancia de responder y al no alcanzar la cantidad de respuestas que se buscaban (1000) se sortearon nuevos sujetos dentro de los grupos y se enviaron nuevos links.

Continuando sin alcanzar los 1200 casos efectivos, y fruto de un análisis de la mesa de trabajo, se llegó a la conclusión de que el envío individual del link generaba desconfianza en la ciudadanía, por lo que se decidió enviar un mensaje por grupo con el link con código de verificación grupal. Esto aumentó considerablemente las respuestas llegando a 1822, de las cuales se seleccionaron 1009 de manera aleatoria en función de la población del segmento y la cantidad de miembros del grupo.

Población

La población sobre la que se aplicó la encuesta abarcó a personas mayores de 18 años residentes en hogares de la ciudad de Córdoba en el año 2020 y que participaran de los grupos de seguridad de la Policía Barrial. El tamaño de la muestra fue de 1009 casos efectivos.

Procedimiento de relevamiento

El procedimiento consistió en dos etapas: una primera etapa de sensibilización que tuvo como objetivo brindar información a los vecinos sobre la encuesta de victimización y su utilidad, promoviendo una participación más comprometida de parte del vecino en la futura encuesta a administrar; y la segunda etapa donde se administró el cuestionario de relevamiento.

Etapa de Sensibilización

El procedimiento de la sensibilización se realizó a través del envío de folletos informativos, los mismos comenzaron siendo imágenes o PDF, pero al ser advertidos por la policía de que no todos tenían los programas requeridos para abrir un PDF o suficiente cantidad de datos para descargar imágenes, los contenidos fueron vertidos en formato de mensaje.

La importancia de la sensibilización radica en que está orientada a que el vecino comprenda que la victimización y el temor al delito son problema de todos, por lo que no deben permanecer silenciados, en tanto los datos aportados constituyen insumos necesarios para la formulación de políticas públicas de prevención.

Se brinda entonces, a los vecinos, la oportunidad al responder sobre los delitos sufridos, la victimización, directa o indirecta de la que fue objeto, aportar a la mejora de las condiciones de vida de la comunidad, en la medida que sus repuestas serán utilizadas para la elaboración y aplicación de políticas públicas preventivas que resultarán en que haya menor número de víctimas y disminuya, de esta manera, el temor al delito.

Etapa de Relevamiento de datos

El relevamiento de datos se realizó a través de una encuesta virtual mediante un formulario auto-administrado. Se procuró utilizar descripciones de los hechos delictivos para favorecer el entendimiento de los encuestados y se procuró presentar una encuesta breve a los fines de favorecer la respuesta de la misma.

El cuestionario utilizado para la Encuesta de Victimización y Percepción Social del Temor al Delito en la Ciudad de Córdoba se divide en siete bloques:

I. Datos sociodemográficos.

II. Sensación de seguridad.

III. Medidas de prevención.

IV. Desempeño de las instituciones.

V. Delitos.

VI. Delitos virtuales.

VII. Características del hecho y denuncia.

RESULTADOS

Victimización

El robo en la vía pública es el delito más común en la ciudad de Córdoba, habiendo afectado al 39.98% de los hogares. Le siguen el robo de casa que afectó al 37.50% de los hogares y el de objetos en el vehículo (34.82%). Estos resultados se condicen con los de encuestas anteriores (González, R. et al 2021) y con las tasas delictivas de la provincia (González, R. et al 2020), en las que los delitos contra la propiedad en la vía pública son los más comunes.

Po otra parte, un 23.12% de los hogares fueron afectados por delitos mediante teléfono, internet y/o redes sociales. Los delitos mediante internet y redes sociales fueron medidos por primera vez en la Encuesta Virtual 2020, a partir de la hipótesis que en el marco del ASPO el delito callejero se estaba necesariamente reconvirtiendo para volcarse en estos ámbitos. Al no haber punto de comparación deberá ser incluido en próximas encuestas para su análisis.

Características de la criminalidad

La violencia física estuvo presente en el 74.10% de los robos con violencia en la vía pública, siendo este el delito de mayor porcentaje de agresiones físicas. Por el contrario, en el robo de partes del auto la violencia solo estuvo presente en el 4.3% de los casos siendo el delito de menor incidencia de violencia de los registrados.

En el 55% de los casos de agresiones físicas y el 32.60% de los casos de robo con violencia en la vía pública se produjeron lesiones, siendo estos los delitos más violentos relevados.

Uso de armas

El robo de motocicleta es ampliamente el delito con mayor utilización de armas de fuego, estando presente en el 47.8% de los casos. En el caso de la utilización de armas blancas o impropias, prevalecen los delitos de agresión física y robo de autos, ambos con una utilización de estos medios en el 25% de los casos.

No hay datos que nos permitan comparar con años anteriores dados que es la primera oportunidad en que se desagrega motocicletas de la categoría de automotores utilizada anteriormente.

Denuncias

Con respecto a las denuncias, el delito con mayor porcentaje de denuncias, de acuerdo a la tendencia histórica y mundial es el robo de autos con una tasa de denuncia del 86.4%. El robo sin violencia o hurto en la vía pública presenta la menor tasa de denuncia, con un 32.4% de los casos.

Un 41.1% de los encuestados manifestó estar disconforme con la respuesta obtenida luego de la denuncia, frente a 24.2% que se manifestaron en algún grado de conformidad. El principal motivo para la no denuncia es la desconfianza en la actuación de las autoridades competentes, siendo el 27.24% de los casos de no denuncia.

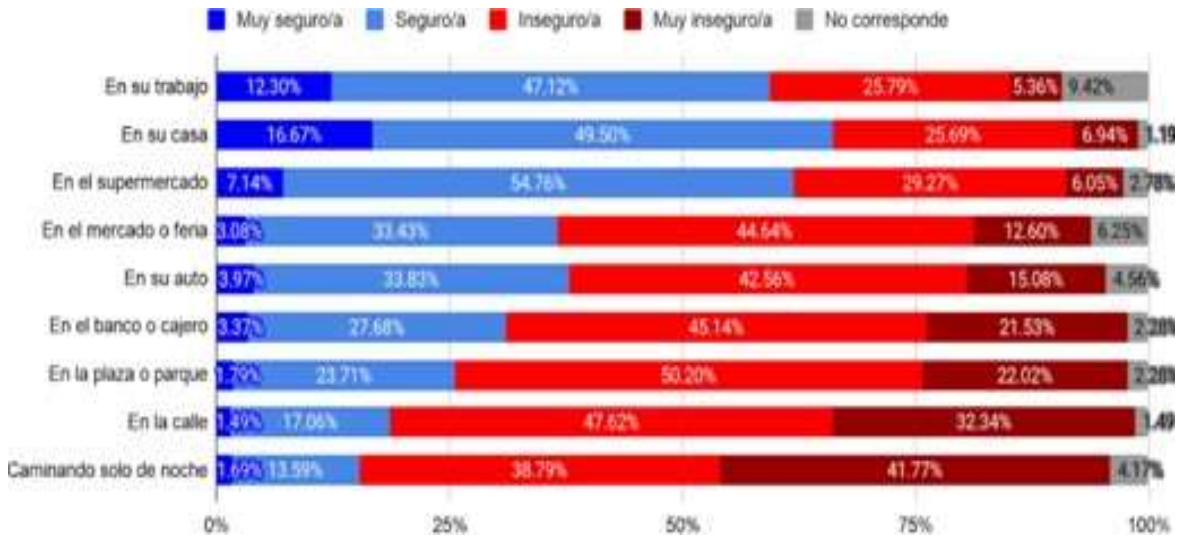
Todas estas tendencias a la denuncia del robo de autos, a la no denuncia de hurtos callejeros, a la desconformidad y descredito de las instituciones de seguridad son también elementos presentes en encuestas presenciales anteriores.

Percepción de seguridad ciudadana

La casa y el trabajo son los espacios donde más seguros se siente los ciudadanos con tasa del 66.17% y 59.42% respectivamente. En la calle y caminando solo de noche fueron las situaciones de menor sensación de seguridad, dado que solo el 18.55% y 15.28% de los ciudadanos manifestaron sentirse seguros respectivamente.

Ambas tendencias están presentes en la encuesta presencial de 2019 (González, R. et al 2021) en la que ocuparon exactamente los mismos lugares, siendo también la casa el lugar más seguro y la calle durante la noche el más inseguro, cómo se puede observar en el gráfico 1.

Gráfico 1: Sensación de seguridad por lugares.

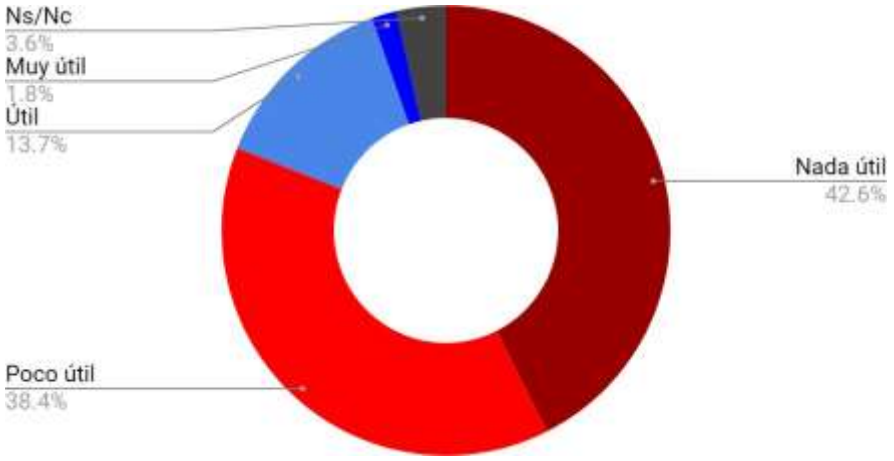


Fuente: Elaboración propia en base a los resultados de la Encuesta Virtual de Victimización y Percepción Social del Temor al Delito. Ciudad de Córdoba, 2020.

Percepción del sistema de seguridad pública

Con respecto a la evaluación del sistema de seguridad pública podemos observar que la opinión relevada en 2019 de manera presencial y en 2020 de manera virtual sobre el accionar de la justicia es muy similar. Así se las puede comprar en los gráficos 2 y 3.

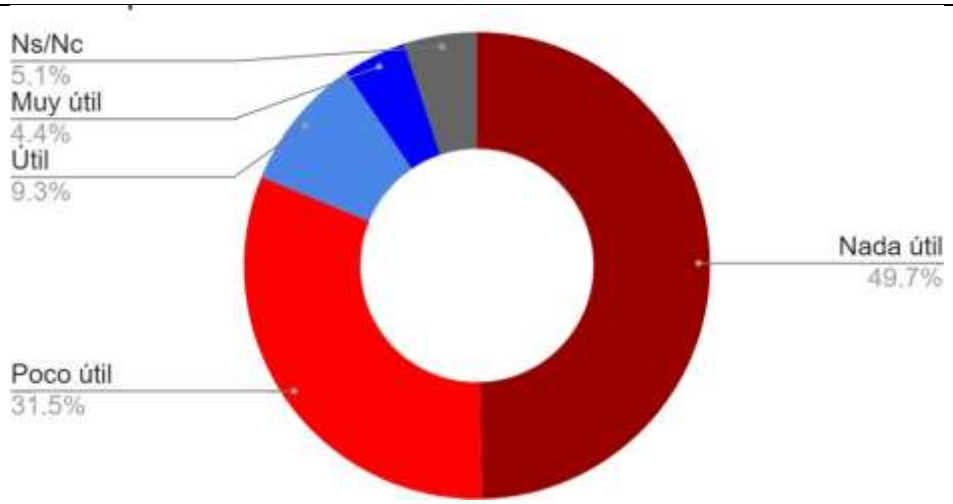
Gráfico 2: Accionar de la Justicia en la resolución de los delitos que se comenten.



Fuente: González, R. et al (2021). Encuesta Córdoba 2019 de Victimización y Percepción Social

del Temor al Delito. Editorial: Matías Alejandro Caro.

Gráfico 3: Utilidad del trabajo de la Justicia para resolver de los delitos que se comenten.

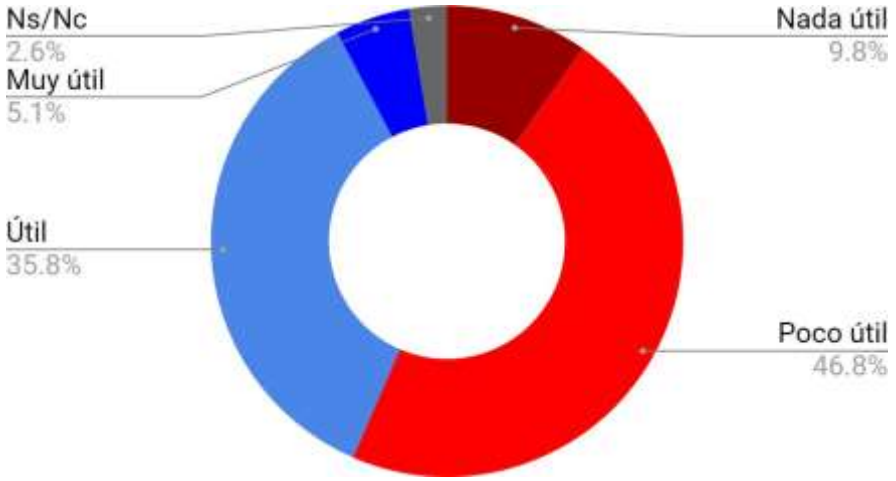


Fuente: Elaboración propia en base a los resultados de la Encuesta Virtual de Victimización y Percepción Social del Temor al Delito. Ciudad de Córdoba, 2020.

Sin embargo, por su parte los resultados de la evaluación del trabajo policial tendieron a ser mucho más positivos en la medición de 2020, que en la de 2019, como puede observarse en los gráficos 4 y 5. Esto quizás pueda deberse a que los miembros de los grupos de WhatsApp están más en contacto con la policía, especialmente con un cuerpo específico de la policía, la Policía Barrial que reviste el carácter de policía de proximidad, estando entrenada para tener una relación de cercanía con el vecino.

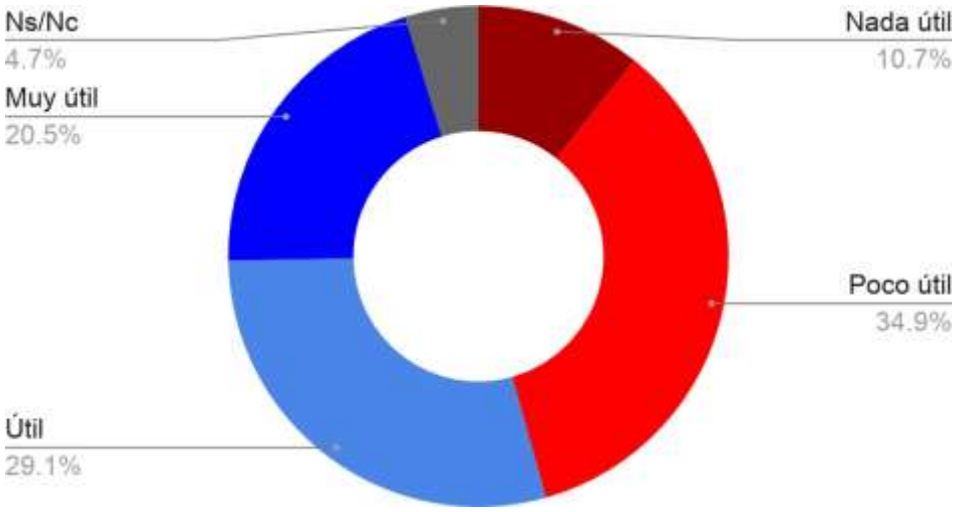
Si bien esto se plantea a modo de hipótesis, la importante discrepancia en la evaluación podría brindar sustento a la efectividad de la implementación de policías de proximidad desde el punto de vista ciudadano.

Gráfico 4: Accionar de la Policía en la prevención del delito.



Fuente: González, R. et al (2021). Encuesta Córdoba 2019 de Victimización y Percepción Social del Temor al Delito. Editorial: Matías Alejandro Caro.

Gráfico 5: Utilidad del trabajo de la Policía para prevenir el delito.



Fuente: Elaboración propia en base a los resultados de la Encuesta Virtual de Victimización y Percepción Social del Temor al Delito. Ciudad de Córdoba, 2020.

CONCLUSIONES

En conclusión, la Encuesta Virtual de Victimización y Percepción Social del Temor al Delito ha demostrado ser un mecanismo económico y rápido para tener una perspectiva de la situación de la seguridad en la ciudad de Córdoba. A su vez las estrategias tecnológicas utilizadas, a los fines de garantizar aleatoriedad, representatividad y anonimato parecen haber dado resultados.

Estos resultados estarían dados por las similitudes interanuales de las tendencias observadas. Sin embargo, las diferencias temporales y de población de las encuestas llevan a ser precavidos sobre su utilidad y a la necesidad de realizar más estudios. Más allá de esto la utilización de tecnologías que implique la reducción de tiempos y la economización de procesos en la investigación y administración pública son motivos suficientes para prestar atención a las potencialidades de la estrategia aquí presentada.

BIBLIOGRAFÍA

Caro, M. (2021) Información criminal para la gestión política y judicial de la seguridad: herramientas para su recolección y análisis. En Granja, M. C. (2021) Hacer justicia en la justicia. Ediciones Lerner SRL.

González, R. et al (2021). Encuesta Córdoba 2019 de Victimización y Percepción Social del Temor al Delito. Editorial: Matías Alejandro Caro.

González, R. et al (2020) Tasas delictivas en la provincia de Córdoba 2019. Editorial: Matías Alejandro Caro.