

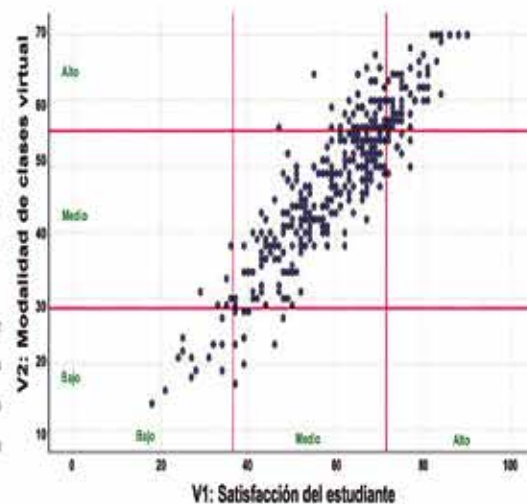
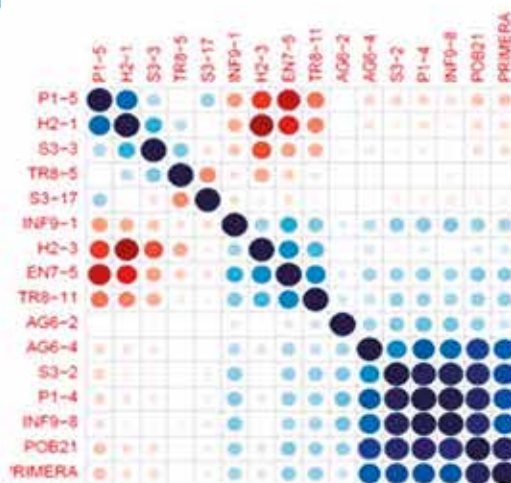
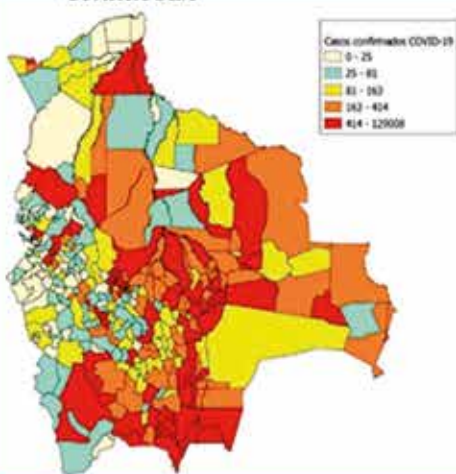


Universidad Mayor de San Andrés
Facultad de Ciencias Puras y Naturales
Carrera de Estadísticas

Varianza

Revista del Instituto de Estadística Teórica y Aplicada

CUARTAOLA



$$\begin{aligned}
 &= E[(x - \mu)^2] \\
 &= E[x^2 - 2x\mu + \mu^2] \\
 &= E[x^2] - 2\mu E[x] + \mu^2 \\
 &= E[x^2] - 2\mu^2 + \mu^2 \\
 &= E[x^2] - \mu^2 \\
 &= E[x^2] - E[x]^2
 \end{aligned}$$

UMSA
FCPN
CARRERA
ESTADÍSTICA

IETA
Instituto de Estadística
Teórica y Aplicada



Varianza

Revista de Investigación de la Carrera de Estadística

Publicación del Instituto de Estadística Teórica y Aplicada

Número 19
Abril, 2022
La Paz - Bolivia

Universidad Mayor de San Andrés
Facultad de Ciencias Puras y Naturales
Carrera de Estadística
Instituto de Estadística Teórica y Aplicada (I.E.T.A.)

ISSN 2789-3510 VERSIÓN IMPRESA
ISSN 2789-3529 VERSIÓN EN LÍNEA

DEPÓSITO LEGAL
4-1-285-2021 P.O.

REVISTA VARIANZA
Nº 19 - Abril, 2022

DIRECTOR CARRERA DE ESTADÍSTICA
M. Sc. Juan Carlos Flores López

ASISTENTE
A. Rosmery Blanco Mamani

DIAGRAMACIÓN Y DISEÑO
M. Zulema Vargas Cerrudo

Los artículos presentados son entera responsabilidad de los autores

VISIBILIDAD: REVISTAS BOLIVIANAS



La Paz - Bolivia
Edificio Bloque FCPN - Campus Cota Cota
Teléfonos: 2612824 -2612844
Email: ieta@umsa.bo
Página web: <https://ieta.umsa.bo/ediciones-varianza>

COMITÉ EDITORIAL NACIONAL

**Teresa Jiménez Zamora, Lic.
(Estadístico)
Universidad Autónoma Tomás Frías, UTF
Potosí-Bolivia**

E-mail: tjimenez12000@yahoo.es

**Isaac De La Cruz Gómez, M. Sc.
(Economista)
Universidad Autónoma Juan Misael Saracho UAJMS
Tarija-Bolivia**

E-mail: delacruzgomezi335@gmail.com

**Emma Mancilla Flores, Lic.
(Estadístico)
Universidad Mayor de San Andrés
La Paz-Bolivia**

E-mail: emmartha3@gmail.com

**Wilma Peñafiel Rodríguez, M. Sc.
(Informático)
Universidad La Salle
La Paz-Bolivia**

E-mail: wilpern7@hotmail.com

**Gilberto Arando Flores, Lic.
(Estadístico)
Universidad Tomás Frías
Potosí-Bolivia**

E-mail: ggilpotosi@gmail.com

COMITÉ EDITORIAL INTERNACIONAL

Lizbeth Román Padilla, Ph.D.
(Matemático)
Facultad de Ciencias UNAM
Departamento/Facultad: Matemáticas
Ciudad: Ciudad de Mexico, Distrito Federal
E-mail: lizroman@hotmail.com

María Eugenia Castellanos Nueda, Ph.D.
(Estadístico)
Professor of Statistics
Universidad Rey Juan Carlos
Campus de Móstoles, Madrid- España
E-mail: maria.castellanos@urjc.es

Adriana D'Amelio, Ph.D.
(Estadístico)
Universidad Nacional de Cuyo
Mendoza-Argentina
E-mail: estat06@hotmail.com

José Gallardo, Ph.D.
(Informático)
Universidad Católica del Norte
Antofagasta-Chile
E-mail: jgallardo@ucn.cl

Fabio Humberto Nieto Sánchez, Ph.D.
(Estadístico)
Universidad Nacional de Colombia
Bogotá-Colombia
E-mail: fhnetos@unal.edu.co



PRESENTACIÓN

Es un orgullo presentar a la Revista Varianza N° 19, de la Carrera de Estadística dependiente de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés. La comunidad académica conformada por docentes y estudiantes, hasta la fecha ha producido una gran cantidad de documentos de investigación, y muchos de ellos han sido publicados.

A partir de la revista Varianza N°18, publicada en octubre de 2021, la periodicidad de publicación es semestral y abierta para recibir contribuciones de autores nacionales e internacionales. De esta manera, se pretende generar un espacio favorable para el ámbito académico- investigativo, de análisis, desarrollo y difusión de las técnicas y metodologías de la estadística. Por lo tanto, sus receptores son todos los profesionales e investigadores de cualquier área que involucre la temática.

La Dirección a mi cargo agradece y reconoce la contribución de los investigadores nacionales e internacionales, que enviaron sus manuscritos, para su revisión y posterior publicación. Asimismo, agradecer al comité editorial nacional e internacional por sus horas de dedicación en la revisión de los manuscritos y por supuesto, agradecer al equipo editorial y al personal administrativo de la Carrera, sin cuya participación no habría sido posible la realización de la presente publicación.

M. Sc. Juan Carlos Flores López
DIRECTOR CARRERA DE ESTADÍSTICA

ÍNDICE

Sistemas de información estadísticos e indicadores de salud <i>Autores: Roberto Bula Magallón y Elisa Aymet Mendoza González</i>	1
Una aproximación al diseño muestral óptimo <i>Autor: Ronal Edwin Condori Huanca</i>	17
Una visión de la pandemia del COVID-19 en los municipios de Bolivia y su componente espacial <i>Autora: Marisol Paredes Alarcón</i>	27
Árbol de decisión en aprendizaje automático <i>Autor: Fernando Rivero Suguiura</i>	39
Estudio de la transmisión de COVID-19 en Bolivia a través del proceso de Galton Watson <i>Autor: Dindo Valdez Blanco</i>	47
Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual <i>Autora: Carmen Vega Flores</i>	57

Sistemas de información estadísticos e Indicadores de Salud

Statistical Information Systems and Health Indicators

Roberto M. Bula-Magallón¹

Universidad de Panamá, Panamá

✉ roberto.bula@up.ac.pa

Elisa A. Mendoza-González²

Universidad de Panamá, Panamá

✉ elisa.mendoza@up.ac.pa

Artículo recibido: 2022-03-03

Artículo aceptado: 2022-03-22

Resumen

El objetivo de este artículo es el de describir el funcionamiento y la relevancia de los subsistemas estadísticos específicamente en dos importantes regiones comarcales en el país, para la producción de indicadores de salud. Se trata de hacer una reflexión sobre la importancia de estos y el papel que desempeñan los profesionales de la estadística en el desarrollo de la metodología estadística desde el planeamiento hasta la difusión de los resultados. Se emplea una metodología documental empleando como materiales dos informes monográficos de seminario como opción de trabajo de graduación de estudiantes de la Licenciatura en Registros Médicos y Estadísticas de Salud de la Universidad de Panamá, que describieron el proceso, en tres grandes fases, que se lleva a cabo para la elaboración de dos indicadores de salud relevantes en el país.

Los resultados se concretan en evidenciar de manera didáctica y sistemática los diferentes procesos para generar la información estadística destacando aquellos aspectos facilitadores, así como los que se constituyen en barreras o limitaciones en la generación de los datos estadísticos.

Palabras clave: Indicadores, Objetivos de Desarrollo Sostenible, Procesos, Producción estadística, Salud

Abstract

The objective of this article is to describe the functioning and relevance of the statistical subsystems specifically in two important county regions in the country, for the production of health indicators. It's about making a reflection on the importance of these and the role played by statistics professionals in the development of statistical methodology from planning to the dissemination of results. A documentary methodology is used using as materials two monographic seminar reports as a graduation work option for students of the Bachelor of Medical Records and Health Statistics of the University of Panama, who described the process, in three large phases, that takes carried out for the elaboration of two relevant health indicators in the country.

¹ Analista Estadístico del Instituto Nacional de Estadística y Censo y Docente-Investigador del Centro Regional Universitario de Colón, Universidad de Panamá, en la carrera de Licenciatura en Registros Médicos y Estadísticas de Salud. ORCID: 0000-0003-4934-0703.

² Docente e Investigadora, Facultad de Ciencias Naturales, Exactas y Tecnología de la Universidad de Panamá. Miembro del Instituto Interamericano de estadística (IASI). ORCID: 0000-0003-0089-6436.

The results are concretized in showing in a didactic and systematic way the different processes to generate statistical information, highlighting those facilitating aspects, as well as those that constitute barriers or limitations in the generation of statistical data.

Key words: Indicators, Sustainable Development Goals, Processes, Statistical production, Health

Introducción

En los últimos años la popularidad de la estadística se ha incrementado aún más por la situación de pandemia vivida debido al COVID-19 (Urrubarrí, 2020). En los medios de comunicación, particularmente la televisión y la prensa digital, se convirtieron en la mayor fuente de información sobre la situación del Covid-19 (Orús, 2021), destacando el rol del periodismo en la cual se reportaron estadísticas a diario, e inclusive se mostraron gráficas e indicadores de manera didáctica haciendo más comprensible los datos presentados, donde las estadísticas y su interpretación fueron las protagonistas, llamando la atención de grandes y chicos dispuestos a escuchar y monitorear la situación sanitaria y su impacto económico y social en el país tanto en los medios de comunicación como las distintas redes sociales.

Por otro lado, en el contexto de las instancias gubernamentales se realizaron importantes avances en cuanto a la instalación del Consejo Nacional de Estadística para el período 2021-2025, en el cual se estableció como principal tarea la aprobación del Plan Estadístico Nacional (PEN 2021 -2025) (González Pinilla, 2021). Estos avances son relevantes para los profesionales de la estadística y el rol de éstos en los subsistemas estadísticos nacionales, puesto que en alguna medida le imprimen sostenibilidad y estabilidad en sus puestos laborales, un hecho esperado desde la creación del Instituto Nacional de Estadística y Censo de Panamá en 2009. Tanto el consejo nacional como el plan representan un importante instrumento del compromiso de las autoridades en la dirección del quehacer estadístico, y como país, con los organismos internacionales en la producción de los datos, indicadores y estadísticas nacionales, tal como lo establecen los Objetivos de Desarrollo Sostenible (ODS), y que buscan ser fuentes válidas confiables y oportunas para la toma de decisiones que contribuyan al bienestar económico y social de la población. (Bula, 2018; Mendoza, 2018). Además, CEPAL advierte que la pandemia se ha convertido en una oportunidad para el fortalecimiento de los sistemas estadísticos y el papel de las oficinas nacionales de estadística (CEPAL, 2020).

Teniendo en cuenta que los datos son necesarios, relevantes para la elaboración de indicadores, que en conjunto proveen información útil a la toma de decisiones políticas, que además sirven de monitoreo de las estrategias, planes y programas implementados y dirigidos a resolver problemas en la sociedad, se observan importantes falencias en los sistemas de información estadísticos. Por ejemplo, en un informe de la Conferencia Estadística de las Américas de la CEPAL (CEPAL, 2021), e empieza resumiendo una serie de problemas que deben atender los países de la región Latinoamericana y del Caribe, como lo son: las políticas públicas en materia estadística y el marco metodológico, enfatizando en que:

Si bien se cuenta con diversas recomendaciones internacionales en materia de aprovechamiento estadístico de registros administrativos, no existe una metodología ni herramientas estadísticas e informáticas que permitan a los países de la región construir una infraestructura estadística a partir de registros administrativos para su potencial uso estadístico (CEPAL, 2021, p.3).

Pues en muchas ocasiones se ha observado una gran producción de datos a través de los registros administrativos que no logran ser suficientemente aprovechados para la generación de información, aún más que sean normalizados para la comparabilidad intra e inter regional, o incluso en el tiempo por lo que se requiere “un diseño y mantenimiento adecuado, su exhaustividad, y la accesibilidad a los mismos” (Gauna, 2020), y tal vez lo más recomendado es su adecuación a sistemas informáticos basados en una transformación integral de los procesos y una clara internalización del concepto de un sistema nacional, donde cada oficina estadística es parte de una gran red productora de datos estadísticos.

CEPAL (2021) describe ocho (8) fases fundamentales en el proceso institucional estadístico, entre ellos: especificación de necesidades, diseño, construcción, recolección, procesamiento, análisis, difusión y evaluación; los cuales, sirven de referentes para comprender la magnitud y complejidad de los sistemas de información estadístico, tal como se observa en la Figura 1, en donde se listan algunas sub etapas o componentes a considerar en cada fase. Es interesante denotar que todo sistema estadístico tiene un propósito, objetivos o un plan que le otorga una dirección, lineamientos y un carácter institucional para su operatividad.

Así mismo, en cada una de las fases deberá existir un responsable de su ejecución, los controles adecuados y todos aquellos elementos (normas, herramientas, insumos, o recursos) que garanticen la operatividad eficiente del sistema de información.

Figura 1.

Fases y Subprocesos de un sistema de información estadístico.

Administración de calidad y metadatos							
Especificación de necesidades	Diseño	Construcción	Recolección	Procesamiento	Análisis	Difusión	Evaluación
1.1 Identificación de necesidades	2.1 Diseño conceptual	3.1 Construcción de instrumentos de recolección	4.1 Conformación del marco muestral y selección de la muestra	5.1 Integración de datos	6.1 Preparación de borradores de resultados	7.1 Actualización de sistemas de salida	8.1 Concentración de insumos para la evaluación
1.2 Consulta y confirmación de necesidades	2.2 Diseño de las descripciones de las variables	3.2 Construcción o mejora de componentes	4.2 Preparación de la recolección	5.2 Clasificación y codificación	6.2 Validación de los resultados	7.2 Generación de productos de difusión	8.2 Evaluación
1.3 Definición de objetivos	2.3 Diseño de la recolección	3.3 Construcción o mejora de componentes de difusión	4.3 Ejecución de la recolección	5.3 Revisión y validación	6.3 Interpretación y explicación de los resultados	7.3 Gestión de la publicación de productos de difusión	8.3 Determinación de un plan de acción
1.4 Identificación de conceptos	2.4 Determinación del marco muestral y tipo de muestreo	3.4 Configuración de flujos de trabajo	4.4 Cierre de la recolección	5.4 Edición e imputación	6.4 Aplicación del control a la divulgación	7.4 Promoción de productos de difusión	
1.5 Comprobación de la disponibilidad de datos	2.5 Diseño del procesamiento y análisis	3.5 Pruebas al sistema de producción		5.5 Derivación de nuevas variables y unidades	6.5 Finalización de resultados	7.5 Gestión de soporte a usuarios	
1.6 Elaboración del Plan General	2.6 Diseño de los sistemas de producción y de los flujos de trabajo	3.6 Prueba piloto del proceso estadístico		5.6 Cálculo de ponderadores			
		3.7 Finalización del sistema de producción		5.7 Cálculo de agregaciones			
				5.8 Finalización de los archivos de datos			

Fuente: CEPAL, 2021

En concreto, se resalta que, en los procesos de producción de datos desde el sistema de información estadístico, se pueden identificar dos fuentes de datos fundamentales: los datos originados de fuentes primarias como: los registros administrativos (Gauna, 2020); encuestas y censos nacionales; y las fuentes secundarias, las cuales son datos ya procesados, resumidos, organizados en tablas o gráficas, por ejemplo, o presentados en informes, documentos o boletines, por mencionar algunos. Por otro lado, se deben señalar que los sistemas de información estadísticos deben ser dinámicos (Salcedo Cifuentes, 2015), y por lo general sistémicos, es decir, cuentan con una fase de entrada de datos, una

de almacenamiento y procesamiento de datos, y otra de salida que corresponderá a los resultados y su difusión (indicadores, estadísticas o publicaciones) (Bula, 2018; Álvarez Lauzarique, s.f.), que a su vez son ampliamente utilizados en procesos de toma de decisiones (Beynon-Davies, 2014).

En la actividad estadística, se deberán cumplir con principios y criterios estadísticos de calidad, tal como lo establece la Ley 10 de 2009 en Panamá, por la cual se Moderniza el Sistema Estadístico Nacional (SEN) y se crea el Instituto Nacional de Estadísticas y Censo (INEC). También se encuentra en Salcedo Cifuentes, (2015) que “la calidad de la información está determinada por su pertinencia, precisión, puntualidad, accesibilidad, interpretabilidad y coherencia” (p. 31). Es interesante también señalar que este conjunto de partes del sistema, no deben ser aisladas, sino por el contrario deben estar relacionadas entre sí y organizadas para lograr algo, por lo general, transformando las entradas en salidas, de acuerdo con conceptos, objetivos o propósitos previamente planteados en su diseño (Beynon-Davies, 2014). Aunado a esto, Gauna (2020) agrega que:

“... la última etapa en este proceso es la que pone a disposición de los usuarios del Sistema la información producida teniendo como criterio central que el Estado, como principal productor de información, debe contribuir a la transparencia de la gestión y a mejorar la participación ciudadana”.(p. 47)

En Panamá, el Sistema Estadístico Nacional claramente definido en la Ley 10 mencionada, es el máximo ente responsable de la organización y gestión de las estadísticas oficiales en el país, para esto en su estructura se pueden identificar subsistemas de información que podrán estar integrados en los distintos comités consultivos. Por ejemplo, el Principio 2. Coordinación del Sistema Estadístico Nacional, en sus criterios de cumplimiento se refiere a la estrecha colaboración que deberá existir entre las entidades productoras de estadísticas y los usuarios de las estadísticas oficiales (CEPAL, 2011).

En esta misma línea se podrán derivar subsistemas de información en el contexto educativo, de salud, judicial, económicos, ambientales, por mencionar algunos. La cuestión, o el reto que deriva para todo sistema estadístico nacional es lograr una coordinación eficaz, eficiente y de calidad en la producción de estadísticas oficiales. Puesto que los datos son fundamentales para realizar análisis, predicciones, modelar y anticipar en el contexto de las políticas públicas y la toma de decisiones (FAO-CEPAL, 2020).

Existen lineamientos, normativas y una ruta interna y externa para propiciar la producción de información estadística, sin embargo, la realidad es otra, puesto que aún se identifican subsistemas estadísticos que requieren ser fortalecidos, ahora en ¿qué sentido o aspectos deben ser fortalecidos?, para poder responder esta pregunta, se realiza un análisis documental con el objetivo de poder identificar las barreras, vacíos o dificultades existentes en la producción de datos en salud, lo que dará cuenta a los tomadores de decisiones de acciones que conlleven a la solución o al fortalecimiento de las estadísticas oficiales en este sector.

Materiales y Métodos

El estudio realizado es de tipo cualitativo – documental, teniendo como materiales informes monográficos de seminarios como opción de trabajo de graduación para optar por el título de licenciados en Registros Médicos y Estadísticas de Salud. Los estudiantes graduandos pertenecen a la Escuela de Estadística, de la Facultad de Ciencias Naturales, Exactas y Tecnología, de la Universidad de Panamá, cuyo campo laboral está principalmente referido a la salud. Estos estudiantes se concentraron en seguir la ruta de subsistemas de información para la producción de indicadores enunciados en las

metas y objetivos, específicamente del Objetivo 3. Salud y Bienestar de los Objetivos de Desarrollo Sostenible (ODS). El análisis realizado es descriptivo, y se trata de reflexionar sobre cada una de las tres fases de un sistema de información considerando: entrada, proceso y salida. Donde la entrada se refiere a la obtención de los datos y su fuente; el proceso al almacenamiento, fórmulas, programas informáticos y otros relacionados para el procesamiento de los datos; y la salida, a la elaboración del indicador y su difusión.

Análisis de los Resultados

El enfoque del análisis que se realiza en este artículo está dirigido a acentuar la importancia de los sistemas de información estadísticos como mecanismos para la producción de datos estadísticos e indicadores. Aún más cuando estos son necesarios en el monitoreo, seguimiento y evaluación de compromisos adquiridos en materia de las políticas públicas del país en el cumplimiento de metas, estrategias, planes y/o programas dirigidos a lograr bienestar en la población, tal como lo establecen, por ejemplo, los diecisiete objetivos de los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas.

D'Agostino (2017) presenta una reflexión sobre este tema, al señalar que precisamente son los gobiernos quienes plantean “la importancia del uso de los datos de salud para la planeación de acciones a nivel de país, subregional, regional y global” especialmente, las estadísticas vitales, y que asimismo, existe la necesidad de fortalecer los sistemas de información, más que todo por el importante incremento de datos, y aún mas interesante la acotación dirigida hacia la sociedad la cual requiere de competencias para entender su verdadero valor (p.1-2).

Sistemas de Información Estadísticos en salud y los ODS

En la Agenda Nacional de Prioridades de Investigación e Innovación para la Salud, Panamá 2016-2025 se describen las principales líneas de investigación y temas en salud, encontrándose que para la implementación de dicha agenda es necesario el fortalecimiento de tres procesos, entre ellos la monitorización y evaluación de los actores, procesos y resultados, y la rendición de cuentas interna y externa. En el primero, por ejemplo, se hace mención de que un mecanismo de monitorización pueden ser los sistemas de información, más no es el único; y en el segundo, se refiere a la cultura de reportar resultados, por ejemplo, a través de la implementación de sistemas de información e indicadores de seguimiento (Ministerio de Salud, s.f.)

Por otro lado, la Secretaría Técnica del Gabinete Social (STGS) en su Primer Informe Voluntario de los ODS de Panamá (2017) se acentúa en la necesidad de contar con mecanismos de producción de datos e indicadores periódicos, para el seguimiento, monitoreo y evaluación de las distintas acciones realizadas en el país derivada de una política pública agresiva con un cambio de paradigma de la visión como país enfocado más en el bienestar multidimensional de su población, buscando reducir las brechas entre los distintos sectores sociales y regionales. En este informe se logró establecer la metodología y criterios para identificar la producción de los indicadores requeridos en los ODS. Se realizaron distintos análisis determinando tres niveles procesales para observar cada uno de los indicadores, de la siguiente manera: Nivel I. Existe una metodología establecida y se dispone de amplios datos, Nivel II. Existe una metodología establecida, pero los datos no son fáciles de obtener y Nivel III. No existe una metodología, ni datos. Es interesante saber que, para esa fecha, en 2017, se contaban con 104 indicadores de un total de 241 priorizados (INEC, 2019) en el país.

En el Segundo Informe Voluntario de los ODS (2020), la STGS deja evidenciado la necesidad de fortalecer los sistemas de información estadísticos en el país,

Las necesidades de información manifiestas durante la crisis que se vive actualmente, a raíz de la emergencia sanitaria provocada por el Covid-19, deja en evidencia que se requiere robustecer estos sistemas integrados de información y los registros administrativos para la toma de decisiones y la reorientación de las políticas públicas; lo cual se convierte en un tema que puede ser objeto de cooperación internacional e intercambio de experiencias y conocimientos entre países. (STGS, 2020, p.18)

En el contexto de las políticas públicas dirigidas a cumplir con los compromisos adquiridos como país entorno a los ODS se determina como un desafío relevante la falta de metodologías y datos desagregados para la obtención algunos indicadores; por lo que en la informe queda establecido esto, de la siguiente manera:

Otro desafío es mejorar la disponibilidad de la data calidad/precisión, desagregar los indicadores a nivel geográfico por discapacidad, etnia, grupos minoritarios, grupos de edad, ya que en la actualidad existen limitaciones en los instrumentos de recolección de datos que utilizan las distintas instituciones generadoras de información cumpliendo con estándares metodológicos de monitoreo y seguimiento. (STGS, 2020, p.18)

En este sentido se debe tener presente que la producción de las estadísticas e indicadores de país es un proceso amplio y complejo, y que goza de un marco legal que incorporan principios y criterios, así como metodologías, mecanismos y otras herramientas para dirigir la actividad estadística integrando los subsistemas de información y las entidades a estos procesos. A su vez, es necesario comprender que los sistemas de información como tal son definidos como un conjunto de tareas o actividades que guardan un fin común, y que en el caso particular de los sistemas de información estadísticos tienen como objetivo transformar datos de entrada al sistema en información y por lo tanto de ellos generar conocimientos para la toma de decisiones, son una herramienta de apoyo a la toma de decisiones (Organización Panamericana de la Salud, 1999). Por su parte, los sistemas de información en salud integran un conjunto de subsistemas diseñados para la producción de datos que sirven de base para la generación de estadísticas e indicadores, ejemplos, de estos el sistema de vigilancia epidemiológica y el de estadísticas vitales (INSP, 2020), obviamente, en el sector salud. Ahora, también es importante aclarar que el término sistema no necesariamente implica el empleo de la tecnología. En los procesos de modernización de los sistemas de información es deseable que estos estén automatizados, sin embargo, la tecnología es una herramienta complementaria al diseño y razón de ser de los sistemas de información. Tiene como una gran ventaja la posibilidad de almacenamiento, gestión y procesamiento de grandes volúmenes de datos en poco tiempo; como una dificultad, se puede enumerar, la incapacidad de su implementación en áreas de difícil acceso debido a los requerimientos técnicos que estos requieren (Preciado Rodríguez, 2021).

El ODS 3. Salud y Bienestar y los datos

De los diecisiete objetivos de los ODS, el número tres define los principales indicadores y metas en el contexto de la salud, titulado: Garantizar una vida sana y promover el bienestar de todos a todas las edades. En el tema de salud infantil, es preocupante el número de niños que no sobrepasarán la edad de 5 años, que niños pobres tienen mayores probabilidades de morir antes de los 5 años, y en el de la salud materna, por ejemplo, la mortalidad materna en regiones en desarrollo aún es 14 veces mayor

que en las regiones desarrolladas (Naciones Unidas, s.f.).

La relevancia de contar con sistemas de información estadísticos en salud eficientes quedó evidenciada con la crisis sanitaria experimentada en el mundo; ahora más, que se presume la hipótesis de haber trastocado importantes avances logrados en materia de salud antes de la pandemia. Sin embargo, estos sistemas de salud jugaron un papel preponderante en la información que sirvió de base para la toma de decisiones en cada uno de los países para controlar la propagación y mayores efectos negativos del Covid-19.

En cuanto a las metas en este objetivo vamos a destacar tres de ellas: 1) Reforzar la capacidad de todos los países, en particular los países en desarrollo, en materia de alerta temprana, reducción de riesgos y gestión de los riesgos para la salud nacional y mundial (Meta 3.d.), 2) Para 2030, reducir la tasa mundial de mortalidad materna a menos de 70 por cada 100.000 nacidos vivos (Meta 3.1), y 3) Para 2030, poner fin a las muertes evitables de recién nacidos menores de 5 años, logrando que todos los países intenten reducir la mortalidad neonatal al menos hasta 12 por cada 1.000 nacidos vivos, y la mortalidad de niños menores de 5 años al menos hasta 25 por cada 1 000 nacidos vivos.

En el primero, destacamos la importancia de contar con datos, como una herramienta que servirá al monitoreo y evaluación de los riesgos o posibles riesgos que pueden presentarse en la población, es decir, los sistemas de información deben poder alertar posibles afectaciones a la salud de la población. Al respecto la Organización para la Cooperación y el Desarrollo Económico (OCDE) recomienda, en este punto en particular: “Emplear una metodología sólida y comprometerse con los estándares profesionales utilizados en la producción de estadísticas oficiales”, agrega que esto podrá realizarse mediante la aplicación de procedimientos y métodos estadísticos apropiados; así como “adherirse a las normas y normas internacionales”. Además, recomienda que se debe “asegurar un acceso y una difusión de datos fáciles de usar, de modo que las estadísticas se presenten en forma clara y comprensible”, sugiriendo el empleo de datos abiertos para poner a disposición de la sociedad los datos considerando las normas de confidencialidad, privacidad y seguridad de estos (OCDE, 2019).

Siendo así que los datos son el insumo principal de la estadística, y en cierto modo el motor de los sistemas de información para producir estadísticas e indicadores, se han convertido en los últimos años en la mina de oro en las organizaciones y en el mundo de la toma de decisiones evidenciadas. Tanto es así, que surgen conceptos como datos abiertos y la gobernanza de datos, exigiendo por tanto que los datos cuenten con un alto nivel de calidad, así como de sistemas que aseguren que la información sea precisa, coherente y accesible (D’Agostino, 2017).

Análisis de dos ejemplos de producción de indicadores de salud

Atendiendo el contexto de los sistemas de información en salud, visto con un enfoque sistémico, se describen dos ejemplificaciones de la descripción de los procesos para la producción de dos indicadores de los ODS de Salud y Bienestar, específicamente en áreas comarcales.

En este proceso se determinaron antecedentes de la variable de estudio, las fuentes de datos, la responsabilidad del registro, procesamiento y publicación del indicador; así como de las dificultades, barreras o posibles vacíos identificados por los estudiantes.

Razón de Mortalidad Materna en la Comarca Ngäbe Buglé (2020). Este trabajo realizado por la licenciada Hania Martínez Martínez, trata de la descripción del funcionamiento del proceso de

recolección, procesamiento y análisis de los datos específicamente relacionado con la mortalidad materna. Se plantea la situación problemática de las mujeres en esta región del país fundamentalmente indígena, la cual presenta importantes índices de mortalidad materna respecto al resto del país, probablemente relacionado con temas culturales (UNFPA, 2017). En la descripción del problema se señaló la disminución evidenciada de la razón de mortalidad materna (RMM) del 2001 a al 2010, de 385 a 278 por cien mil nacimientos vivos, respectivamente. Al 2018, este indicador a continuado disminuyendo (138.4) probablemente como resultado de la implementación de políticas sanitarias que atienden aspectos culturales de esta región, así como el fortalecimiento del sistema de salud, sin embargo, sigue siendo un indicador importante, toda vez que, a nivel del país, se sitúa en 45.5 (INEC, 2018).

Martínez-Martínez (2020), define el sistema de información estadístico como un conjunto de elementos físicos, lógicos de comunicación de datos y personal que interrelacionados permiten la recogida, el almacenamiento y transmisión respondiendo a los propósitos para el cual ha sido creado. Por otro lado, cita los principios que rigen la estadística nacional establecido en el Sistema Estadístico Nacional por Ley, tales como: Independencia, secreto estadístico, imparcialidad, transparencia, pertinencia, precisión, coherencia, comparabilidad, accesibilidad, oportunidad y puntualidad. Así como también resume cinco etapas del proceso de la actividad estadística: Planeamiento o planificación, recolección de datos, elaboración de la información, análisis de datos y publicación de los datos (Guerra Bustillo, 2003).

En lo concerniente a la contextualización, se define el indicador 3.1.1 de los ODS, Razón de Mortalidad Materna, cuya meta describe lo siguiente: De aquí al 2030, reducir la razón de mortalidad materna a menos de 70 por cada 100.000 nacidos vivos. Este indicador relaciona el número de muertes maternas por cada 100 mil nacidos vivos. A este respecto, se presentan los conceptos relativos a los nacimientos vivos (INEGI, 2015) y a las defunciones fetales (Mendoza, 2000) basados en el Manual de Clasificación Estadística Internacional de Enfermedades y Problemas relacionados con la Salud. La descripción del proceso del microsistema estadístico llevado a cabo para la producción del indicador se describe de la siguiente manera:

- Registro de datos (entrada) relativos a la defunción materna en la Comarca Ngäbe Buglé. Se emplean como fuentes de datos, los certificados de Nacimiento Vivo (del recién nacido), y el certificado de Defunción (de la madre) de la Dirección Nacional del Registro Civil de las direcciones regionales y oficinas distritales. Estos formularios están enumerados (número control), llenados por registradores auxiliares. Para la generación de este indicador, se identifica un marco legal y un manual de procedimientos, por ejemplo, la Ley 31 de 25 de junio de 2006 que regula el registro de los hechos vitales y reorganiza la Dirección Nacional del Registro Civil, el manual de procedimiento se especifican las responsabilidades, conceptos, normativas que sustentan el llenado correcto de los formularios o registros administrativos.
- Procesamiento de los datos. Los formularios tanto de nacimientos vivo como el de defunción son enviados a la Oficina de Estadísticas Vitales del Instituto Nacional de Estadísticas y Censo (INEC) y procesados en la sección de Estadísticas Vitales, en donde se elaboran las estadísticas e indicadores correspondientes atendiendo a los criterios de calidad y los principios de buenas prácticas.
- Publicación del indicador. El INEC publica los indicadores en diferentes formatos, entre ellas el informe estadístico llamado Panamá en Cifras y el Boletín de Defunciones, además de los

Anuarios estadísticos de acuerdo con los principios y recomendaciones de las Naciones Unidas, a la vez que sus datos son compartidos con sus usuarios del sector salud nacional e internacional, entre otros.

Como parte de las conclusiones se resalta que las principales causas de mortalidad materna identificadas en esta región del país están la hemorragia postparto y la sepsis puerperal, también, se refiere al tema de los subregistros como un grave problema en la gestión de los indicadores, toda vez que cerca del 50% de los nacimientos o partos no llegan a las instalaciones de salud, sino que se realizan por parteras comunitarias en sus hogares, así mismo se identifica como un grave problema sanitario la ausencia de atención hospitalaria en la región (Unicef, 2019) dificultando la accesibilidad de la población al goce del derecho a la salud.

Tasa de mortalidad neonatal de la Comarca Guna Yala (2020), trabajo realizado por las licenciadas Karen Brown y Noemí Rodríguez Sáenz, toma como marco de referencia lo relativo al indicador 3.2.2. de los ODS, para el cálculo de la tasa de mortalidad neonatal, definiendo como mortalidad neonatal aquella que acontece entre el nacimiento hasta los 28 días de vida (OMS) constituyendo un problema de salud a nivel mundial asociado a la calidad de la atención materno infantil, específicamente a problemas relacionados con la alimentación de la madre. La Comarca Guna Yala, es una importante región indígena en el país, conformada mayormente por un archipiélago de 365 islas. Entre las principales causas de muerte neonatal en la región, están las malformaciones congénitas, complicaciones por prematuridad o durante el parto, y las infecciones como la sepsis, meningitis y neumonía.

En cuanto a las bases de datos, se refiere a que:

La gestión de bases de datos debe ser planificada y bien pensada de modo que todo el esfuerzo y la inversión que se realiza dé los productos esperados, y en este caso la información o los datos esperados que sirvan a los tomadores de decisiones y a los propósitos por los cuales han sido diseñado e implementados.

Además, define la base de datos como:

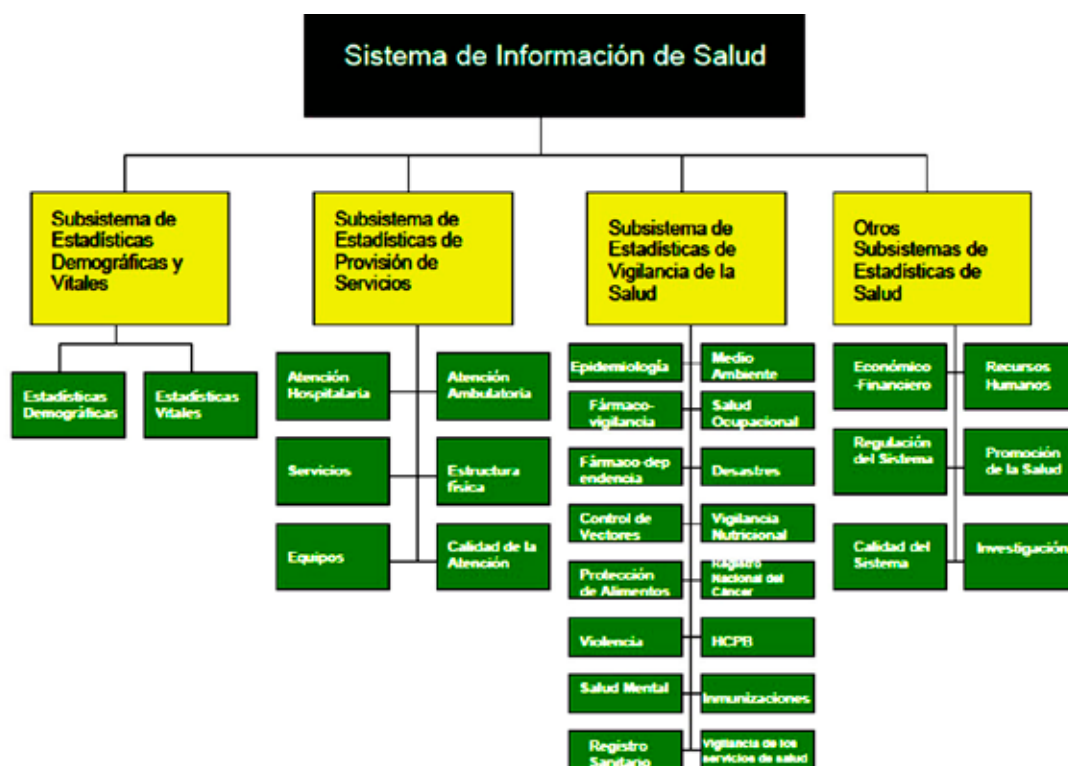
Un almacén de colección de múltiples tipos de registros de datos que nace de las necesidades del individuo y organización de poder obtener información rápida, organizada y precisa.

Otro aspecto relevante que destacar corresponde a la definición del Sistema de Información en Salud, resaltando la labor del profesional de Registros y Estadísticas de Salud (REGES), personal médico y la participación del paciente en el proceso del registro de datos, así como otro personal de salud que participa en dicho proceso. Además del personal, se refiere a la infraestructura y a la tecnología, como un componente esencial en los sistemas de información.

Es interesante también destacar que en el país existen dos subsistemas de salud, uno público segmentado por dos grandes instituciones: Ministerio de Salud y Caja de Seguro Social y, uno privado. Así como una amplia descripción de los diferentes subsistemas que integran el sistema de información de salud (Figura 2).

Figura 2.

Descripción de los subsistemas que integran el Sistema de información de salud (SIS).



Fuente: Tomado de Tasa de mortalidad neonatal de la Comarca Guna Yala (Brown, 2020)

Los procesos correspondientes del sistema de información para el cálculo de este indicador describen lo siguiente:

- Entrada de datos al SIS. Este sistema integra dos subsistemas de informáticos, uno de carácter electrónico hospitalario y otro de origen manual (registro administrativo) que posteriormente es captado en el sistema estadístico de salud, por el personal de REGES. En cualquiera de los casos, tanto nacimientos como las defunciones en las instalaciones de salud sean de la Caja de Seguro Social o del Ministerio de Salud deben ser registradas por mandato legal, en los formularios establecidos por la Dirección Nacional del Registro Civil, pues deben declararse las defunciones dentro de las cuarenta y ocho horas siguientes de ocurridas el hecho; sin embargo, se ha identificado la declaración tardía de estos hechos particularmente en regiones rurales.
- Procesamiento de datos. Los datos que son captados por las instancias sanitarias, en los registros administrativos regulados los cuales tienen un número de control el cual es reportado a la Dirección Nacional del Registro Civil y al Instituto Nacional de Estadística y Censo (INEC).
- Publicación del indicador. Los datos son procesados en la sección de Estadísticas Vitales del INEC, luego que se elabora el indicador este se publica en los diferentes mecanismos dispuestos por esta entidad, y en su página web institucional.

Entre los principales resultados obtenidos en este tema, se pudo determinar una tendencia de la tasa de mortalidad neonatal hacia la disminución, en 2018 se reportó una tasa de 6.6 por mil nacidos vivos. En esta región del país no se cuenta con Internet, lo que dificulta el registro de datos de manera

oportuna. En este sentido, se abre una ventana al subregistro de datos, introduciendo posibles sesgos a la elaboración de los indicadores.

Discusión

La generación de datos es una tarea que requiere de una detallada planificación, metodología que incluya una serie de elementos que atiendan los criterios de calidad, y los principios estadísticos (OPS, 1999); aún más cuando se trata de indicadores que sirven a la sociedad y a los tomadores de decisiones en materia de políticas públicas (STGS, 2020), puesto que ayudan a comprender mejor hechos o fenómenos que acontecen en los distintos contextos en que se analizan (Bernal, 2014). Otra connotación genérica sobre la relevancia de los datos en la actualidad es el título que se le da como la mina de oro (Caballero y otros, 2019) y que a su vez se constituye en un problema complejo, ya que se producen millones de datos por segundo, que en mucho de los casos no llegan a ser explotados (Villegas Zamora, 2019), de manera que cada vez más se hace necesario la formación de profesionales de la estadística en el que se comprenda su metodología y pueda ser aplicado en el contexto que se requiera. Además, Lahera Rol (2018) refiriéndose al personal sanitario, indica que:

En el proceso de formación se debe insistir en: la comprensión, la interpretación, la evaluación y la comunicación de los resultados sobre la base de un conocimiento estadístico básico, sin desatender, además, elementos la responsabilidad y la honradez profesionales, en tanto constituyen cualidades esenciales para su formación profesional.

En este artículo se realiza una aportación del funcionamiento de un proceso de recolección de datos, el procesamiento y la elaboración del indicador, siguiendo un modelo sistémico general en salud, que por analogía sigue el método estadístico de recolección, organización y procesamiento, y análisis de datos para la toma de decisiones (Bernal, 2014), y que a su vez, se idealiza un sistema parsimonioso, es decir, entre más sencillo y simple es mejor, sin omitir ningún aspecto relevante o que afecte la calidad de la información (Díaz Narváez, 2009) y los principios estadísticos establecidos por organismos internacionales y adaptados por ley en el país, por mencionar algunos: secreto estadístico, independencia, imparcialidad, transparencia, pertinencia, coherencia, oportunidad, puntualidad. Además, en estas normativas, se establecen metodologías que parten desde la justificación de la necesidad de datos, los conceptos que se emplean y las variables a investigar, así como los procedimientos para la recolección de datos (INEC, 2011).

Los indicadores de salud tienen características muy particulares de relevancia en las políticas sanitarias, que requieren contar con criterios de calidad y principios estadísticos de mucho valor para garantizar la veracidad y la confiabilidad de los hechos registrados, puesto que esto dependen decisiones importantes con implicaciones en los recursos destinados a las distintas regiones y sectores del país, así como las acciones que se formulen en beneficio de la sociedad (INEC, 2018). Los datos son esenciales en la construcción de indicadores y por lo tanto de información. A la vez, requieren de recursos humanos, técnicos y de toda una infraestructura para la producción de datos con criterios de calidad. Es aquí donde se requiere la conciencia política en el aporte sustancial al fortalecimiento de los sistemas de información estadísticos.

Algunas de las barreras, desafíos o vacíos que se pueden encontrar en los sistemas de información quedan evidenciados cuando el país requiere de indicadores nacionales correspondientes a los ODS, tal como se observó en el Informe Voluntario de los ODS de Panamá, en donde a raíz de la falta de metodología, datos e instrumentos en algunos contextos, para la obtención de indicadores se aprueban normativas agresivas para robustecer los sistemas de información estadísticos en el país (STGS, 2020).

Conclusiones

Los sistemas de información estadísticos se constituyen en herramientas relevantes en los procesos de recolección, procesamiento y divulgación de indicadores en el contexto que se desarrolle. Los datos deben contar con criterios de calidad que incluyan una clara definición, sustento teórico, y la metodología de las variables a registrar; así mismo, se determinan las fuentes de datos y los mecanismos que se emplearán para el registro y procesamiento de datos. En el diseño de los sistemas, se deben contemplar no solo la metodología, criterios de calidad y principios estadísticos, sino también procesos de presentación de las salidas, es decir, las publicaciones, periodicidad y otros elementos a fin de satisfacer la necesidad de datos estadísticos e indicadores, requeridos en la evaluación, monitoreo y decisiones pertinentes.

Son los profesionales de la estadística, quienes son formados con las herramientas necesarias para acompañar en el proceso del levantamiento de los sistemas de información, quienes cuentan con conocimiento, habilidades y destrezas para formular indicadores, definir variables y la presentación de resultados, incluso ir más allá de la descripción de datos, hacia el análisis estadístico y su participación en proyectos de investigación interdisciplinarios. El estadístico de salud posee cualidades profesionales en el desarrollo de la metodología estadística en el contexto de la salud, facilitando la comunicación e integración en los equipos de salud.

Finalmente, se debe acotar la necesidad de continuar fortaleciendo la cultura estadística en la sociedad y que en estos últimos años a raíz de la pandemia, la sociedad íntegra ha logrado visualizar la importancia del dato como insumo esencial para la generación de información que da cuenta del acontecer nacional y por lo tanto para la toma de decisiones sociales, económicas y de salud; y que en efecto, los datos se constituyen en una mina de oro para generar información y conocimientos, y es aquí donde el profesional de la estadística cuenta con un potencial en la aplicación de los métodos y técnicas para el análisis de datos en la era de la ciencia de los datos.

Así mismo, se reafirma que es necesario la integración de esfuerzos desde las autoridades gubernamentales, sociedad y comunidades locales, en general, para fortalecer la cultura del dato estadístico, eliminar las barreras que conllevan a la ausencia de datos o el subregistro, así como también la importancia de contar con mecanismos, metodología e instrumentos para la producción de datos de calidad, coherentes, oportunos, pertinentes, válidos y confiables para la toma de decisiones basadas en la evidencia.

Referencias Bibliográficas

- Álvarez Lauzarique, M. (s.f.) *Sistemas de información estadístico en salud. Diseño y explotación*. [Diapositiva PowerPoint]. <https://medicos.cr/web/documentos/charlasregistrosysistemas/Sistemas%20de%20Informaci%3%b3n%20Estad%3%adstico%20en%20Salud.pdf>
- Bernal Morrell, E. (2014). *Bioestadística básica para investigadores con SPSS*. ed. Madrid: Bubok Publishing S.L. España.
- Beynon-Davies, P. (2014). *Sistemas de Bases de datos*. Reverté Editorial.
- Brown, K. y Rodríguez Sáenz, N. (2020). *Tasa de Mortalidad Neonatal de la Comarca Guna Yala* [Monografía de Licenciatura, Universidad de Panamá].

- Bula M., R. (2018). Los sistemas de información estadísticos y la población con discapacidad en panamá. *Colón Ciencias, Tecnología y Negocios*, 5(1), 51-63. https://revistas.up.ac.pa/index.php/revista_colon_ctn/article/view/328
- Caballero Muñoz-Reja, I. Gómez Carretero, A., Gualo Cejudo, F. (2019). *Calidad de datos*. Ediciones de la U. <https://elibro.net/es/ereader/upanama/127087?page=1>
- CEPAL. (Noviembre, 2011). *Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe*. https://www.inec.gob.pa/sistema/sic_archivos/codigobuenaspracticass.pdf
- CEPAL. (25 de agosto de 2020). *Países destacan importancia de las estadísticas para visibilizar las vulnerabilidades y grandes desigualdades que caracterizan a la región. Comunicado de prensa*. <https://www.cepal.org/es/comunicados/paises-destacan-importancia-estadisticas-visibilizar-vulnerabilidades-grandes>
- CEPAL. (23-25 noviembre de 2021). *Documento metodológico para el aprovechamiento estadístico de registros administrativos económicos. Resumen*. https://repositorio.cepal.org/bitstream/handle/11362/47461/1/S2100660_es.pdf
- D'Agostino, M., Martí, M., Mejía, F., De Cosio, G., Faba, G. (2017). Estrategia para la gobernanza de datos abiertos de salud: un cambio de paradigma en los sistemas de información. *Revista Panamericana de Salud Pública*, 41:e27
- Díaz Narváez, V. (2006). *Metodología de la investigación científica y bioestadística: para médicos, odontólogos y estudiantes de ciencias de la salud*. RIL editores.
- FAO-CEPAL. 2020. Disponibilidad de datos estadísticos para enfrentar la pandemia de COVID-19. Santiago, FAO.
- González Pinilla, J. (5 de julio de 2021). Instalan el Consejo Nacional de Estadística para el período 2021-2025. *La Prensa*. <https://www.prensa.com/politica/instalan-el-consejo-nacional-de-estadistica-para-el-periodo-2021-2025/>
- Guerra Bustillo, C., Menéndez Acuña, E., Barrera Morera, R., y Egaña Morales, E. (2003). *Estadística*. La Habana: Félix Varela.
- Gauna, N., Roggi, C., Zuloaga, N. (2020). Los registros administrativos en la construcción y consolidación del Sistema Estadístico de la ciudad. *Población de Buenos Aires*, 17(29), 43-49. <https://www.redalyc.org/journal/740/74065319007/html/>
- INEC. (2011). *Decreto 159 de 2011. Por el cual se reglamenta la Ley 10 de 22 de enero de 2009 Que moderniza el Sistema Estadístico Nacional y crea el Instituto Nacional de Estadística y Censo*.
- INEC. (2018). *Estadísticas vitales. Volumen III. Defunciones: Año 2018*. https://www.inec.gob.pa/publicaciones/Default3.aspx?ID_PUBLICACION=975&ID_CATEGORIA=3&ID_SUBCATEGORIA=7
- INSP. (19 de noviembre de 2020). *Línea de investigación en sistemas de información en salud y carga de la enfermedad*. <https://www.insp.mx/lineas-de-investigacion/sistemas-informacion-salud.html>

- Lahera Rol, A., Pérez Olivares, I, Hunte Roberts, V., & Ruiz Batista, E. (2018). La estadística como necesidad en la investigación en Salud. *Revista Información Científica*, 97(4), 891-901. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1028-99332018000400891&lng=es&tlng=es
- Ley 31 de 25 de junio de 2006 que regula el registro de los hechos vitales y reorganiza la Dirección Nacional del Registro Civil.
- Martínez M., Hania. (2020). *Sistema de información para la producción de indicadores de la Razón de Mortalidad Materna en la Comarca Ngäbe Buglé*. [Monografía de Licenciatura, Universidad de Panamá].
- Mendoza, E. (Noviembre, 2018). *Evaluación del Sistema Estadístico Nacional de Panamá. SENACYT*. [Archivo pdf] <https://www.senacyt.gob.pa/publicaciones/wp-content/uploads/2019/06/Evaluaci%C3%B3n-del-Sistema-Estad%C3%ADstico-Nacional-de-panam%C3%A1.pdf>
- Ministerio de Salud. (s.f.) *Agenda Nacional de Prioridades de Investigación e Innovación en Salud*. Panamá.
- Naciones Unidas. (s.f.) *Objetivo 3. Garantizar una vida sana y promover el bienestar para todos en todas las edades*. <https://www.un.org/sustainabledevelopment/es/health/>
- Organización Panamericana de la Salud (OPS). (Julio de 1999). *El establecimiento de sistemas de información en servicios de atención de salud. Guía para el análisis de requisitos, especificación de las aplicaciones y adquisición*. Washington. D.C. https://www.paho.org/hq/dmdocuments/2009/HealthcareInformationSystems_Spa.pdf
- Orús, A. (19 de marzo de 2021). *COVID-19: principales medios para mantenerse informado en España 2020*. Visto en: <https://es.statista.com/estadisticas/1107008/principales-medios-para-mantenerse-informado-sobre-la-covid-19-espana/>
- Preciado Rodríguez, A., Valles Coral, M., Lévano Rodríguez, D. (2021). Importancia del uso de sistemas de información en la automatización de historiales clínicos, una revisión sistemática. *Revista Cubana de Informática Médica*. 13(1), e417 http://scielo.sld.cu/scielo.php?pid=S1684-18592021000100012&script=sci_arttext&tlng=pt
- Salcedo Cifuentes, M. y otros. (2015). *La calidad del dato en los sistemas de información*. Programa editorial Universidad del Valle. Colombia.
- Villegas Zamora, D. (2019). La importancia de la estadística aplicada para la toma de decisiones en Marketing. *Revista Investigación y Negocios*, 12(20), 31-44. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2521-27372019000200004&lng=es&tlng=es
- Torres Huertas, J. (2019). *Estadística aplicada a las Ciencias de la salud*. ed. Madrid: Dextra Editorial.
- Unicef. (Noviembre 2019). Derecho a la vida, crecimiento y desarrollo. <https://www.unicef.org/panama/informes/sitan-cap1>
- UNFPA. (19 de septiembre de 2017). Programa de Reducción de la Mortalidad Materna. Comarca Ngäbe Buglé 2001-2011. Una sistematización de experiencias. <https://panama.unfpa.org/es/publications/programa-de-reducci%C3%B3n-de-la-mortalidad-materna-comarca-ngabe-bugl%C3%A9-2001-2011-0#:~:text=El%20Programa%20de%20Reducci%C3%B3n%20de,morbi%2Dmortalidad%20materna%20e%20infantil>

Urribarri, R. (Junio 2nd, 2020). *Medios digitales y periodismo: retos para la democracia en Panamá*.
<https://cieps.org.pa/medios-digitales-y-periodismo-retos-para-la-democracia-en-panama/>

Una aproximación al diseño muestral óptimo

An approach to the optimal sampling design

Ronal Edwin Condori Huanca¹

Universidad Mayor de San Andrés, La Paz- Bolivia

✉ ronal.c.huanca@gmail.com

Artículo recibido: 2022-02-18

Artículo aceptado: 2022-03-28

Resumen

Ante la ausencia del desarrollo y cuantificación de una mayor diversidad de medidas de precisión para los diseños de muestreo en el contexto Boliviano. El presente documento aborda el desarrollo del objetivo experimental de comparabilidad y elección de diferentes diseños de muestreo potenciales a ser aplicados a conteos rápidos para fines electorales, se utilizó información de las elecciones generales de 2020 en Bolivia, seleccionando hasta dieciocho escenarios posibles de muestreo, y evaluando sus errores estándar, coeficientes de variación logarítmica, y otras medidas más. Considerando al estimador de razón, se encontró que el diseño con estratificación en base al tamaño de recinto con una selección de tipo secuencial, es el diseño más óptimo con un coeficiente logarítmico de 0.042, errores estándar cercanos al 0.018, y otras medidas de precisión.

Palabras clave: Conteo Rápido, Coeficiente de variación logarítmico, Diseño de muestreo optimo, Elecciones.

Abstract

In the absence of the development and quantification of a greater diversity of precision measures for proven designs in the Bolivian context. This document addresses the development of the experimental objective of comparability and choice of different potential sampling designs to be applied to quick counts for electoral purposes, for the same information from the 2020 general elections in Bolivia was used, selecting up to eighteen possible proven scenarios, and evaluating their standard errors, coefficients of logarithmic variation, and other measures. Considering the ratio estimator, it was found that the design with stratification based on enclosure size with a sequential type selection is the most optimal design with a logarithmic coefficient of 0.042, standard errors close to 0.018, and other more accurate measures.

Keywords: Quick Count, Logarithmic coefficient of variation, Optimal sampling design, Elections.

1. Introducción

En el contexto Boliviano, para la mayoría de las encuestas probabilísticas, la elección del diseño de muestra puede ser considerada como una decisión muy subjetiva, y se basa principalmente los siguientes aspectos:

¹ Profesional/consultor en Estadística(s), candidato a Doctor en Políticas Públicas de la UMSA: Ha ejercido laboralmente en el área de estadística en entidades privadas y en entidades públicas como el INE-Bolivia, Ministerios de Salud, Ministerio de Desarrollo Productivo, Ministerio de Educación, Ministerio de Economía, etc. <https://orcid.org/0000-0003-2557-7079>

- Información disponible: encuestas previas, marcos muestrales, cartografía, etc.
- Limitantes de planificación y ejecución: presupuesto y personal disponible, fechas comprometidas para entrega de resultados.
- Aspectos Logísticos: organización, desplazamiento en campo, limitantes urbano rural, externalidades (clima, conflictos políticos, fricciones socioculturales en los informantes, etc.).

Este análisis suele estar a cargo del personal que elabora el diseño de muestreo, en coordinación y aprobación con los financiadores, planificadores y demás miembros del equipo temático a cargo de la encuesta en la institución. Y bajo esta última etapa de diseño y consolidación metodológica, suelen ignorarse aspectos como, el tamaño de los conglomerados, estratificación, la reducción mínima de los márgenes de error de muestreo, sesgos posibles, varianzas mínimas, y otros.

1.1. Antecedentes

En los diferentes actos electorales que se celebran en cualquier país, el recuento de votos por muestreo, o conteo de votos, es una técnica ampliamente utilizada para dar un resultado preciso, confiable, y rápido, de manera de proporcionar un panorama de tranquilidad a los diferentes actores políticos, instituciones y público en general.

En la región latinoamericana el uso de esta técnica en los actos electorales ha ido desapareciendo, debido a la digitalización de los actos electorales, o de los procesos utilizados en el recuento de votos oficial que se aplican, pudiendo así entregar sus resultados incluso en menos de 48 horas posteriores al día del acto electoral (RNEC, 2019), (TEP, 2021) y otras aún con postergaciones mayores (Garzón-Sherdek, 2021), (SERVEL, 2021), (TSJE, 2021).

En los últimos años en Bolivia, estas operaciones estadísticas fueron utilizadas en las elecciones sub nacionales de 2021 (CIESMORI, 2021), (FOCALIZA, 2021), en las elecciones generales de 2020 (Página-Siete, 2020) y las de 2019 (ViaCiencia, 2019).

Según toda esta información, las operaciones estadísticas señaladas anteriormente, utilizan diseños de muestreo que en su mayoría solo mencionan el error de muestreo teórico usado en el tamaño de muestra, sin cuantificar otras medidas de precisión. Sin embargo para 2020, ya se incluyen el error estándar y los intervalos de confianza (FOCALIZA, 2021), conforme señala el último reglamento para la elaboración de estos estudios (OEP, 2020).

1.2. Problemática

En el medio Boliviano, no existe una práctica o hábito en las encuestas por muestreo y su diseño, sobre como sustentar cuantitativamente el proceso de elección del diseño de muestreo más apropiado. Incluso en las mismas encuestas multipropósito que realiza el Instituto Nacional de Estadística de Bolivia (INE), no se detallan públicamente pormenores de este tipo (INE, 2020).

Esta afección en la práctica para elaboración de estos procesos estadísticos (que incluso no es cubierta por la oficina nacional de estadística), plantea la necesidad de poder aplicar y desarrollar en el contexto Boliviano un conjunto de recomendaciones sobre la elección cuantitativa de diseños de muestreo (Deville, Sarndal, & Sautory, 1993), entre las diferentes alternativas que puedan ser utilizadas de forma posterior.

Cabe aclarar que no está en discusión la técnica de distribución del tamaño de muestra, la cual verifica por ejemplo que para un estimador estratificado de la media $\bar{y}_{st}$, establece que la afijación óptima de Neyman posee la mínima varianza respecto a una afijación proporcional y un muestreo aleatorio simple: $V_{opt}(\bar{y}_{st}) \leq V_{prop}(\bar{y}_{st}) \leq V_{ran}(\bar{y})$ (Cochran, 1977), o la disyuntiva sobre cual método de estimación es más preciso entre los métodos directos (totales, promedios o proporciones), en comparación con los métodos indirectos (razón, regresión o por diferencias), esta comparación ya está verificada teóricamente, pudiéndose expresar como: $V(\bar{y}_{reg}) \leq V(\bar{y}_R) \leq V(\bar{y})$ (Perez, 2005).

Lo que está en discusión, es la comparación de las alternativas de diseño de muestreo que se pueden aplicar a una encuesta por muestreo en específico.

1.3. Hipótesis

Es posible desarrollar una comparación de diseños de muestreo de manera cuantitativa, y así elegir una estrategia aproximadamente óptima para el diseño de muestreo aplicado en las elecciones generales de 2020.

2. Material y Métodos

2.1. El contexto de la estimación

Como se mencionó anteriormente, la aplicación de la investigación se realiza en las elecciones generales 2020, en la cual el objetivo del diseño de muestreo es el de *estimar la distribución de votos válidos por partido político, a nivel nacional, y a nivel departamental*. Este hecho plantea definir nueve (9) muestras independientes en cada uno de los departamentos, para la asignación de escaños, tanto en diputados como en senadores (Ley026, 2010).

Otro aspecto que siempre se consideran en encuestas electorales es el del contraste urbano vs. rural, o en algunos casos, capital vs. resto, o también una agrupación de las diferentes ciudades, comunidades o localidades, según su tamaño poblacional.

2.2. Los marcos de muestreo

Los marcos de muestreo utilizados están compuestos por el padrón electoral oficial para las elecciones generales de 2020 y los resultados oficiales para el mismo evento. En algunos casos se utilizaron como información auxiliar los tamaños de muestra elaborados por las instituciones citadas concretamente.

Cabe aclarar que se utilizaron varios marcos de muestreo, ya que, al modificar los estratos, conglomerados u otra etapa, estos marcos son diferentes entre sí, aunque en algunos casos pueden ser comparables a nivel de unidades primarias de muestreo (UPM).

2.3. Tipos de estratificación

Se planteó el uso de estratificación en cada uno de los departamentos del siguiente tipo:

1. Urbano y Rural
2. Capital y Resto
3. Por tamaño del recinto

El primer caso asumió la clasificación que el INE maneja sobre los centros poblados clasificados como urbanos en base al último Censo de Población y Vivienda de 2012. El segundo caso consideró al municipio capital como un estrato único dentro del departamento y el resto fue el complemento de los municipios dentro de este. Sólo en el departamento de La Paz, consideró al municipio de El Alto como un estrato separado. El último tipo de estratificación, creo estratos para los recintos de menos de 6 mesas, y otro de 6 o más mesas.

Estos tipos de estratificación son ampliamente usados, aunque existan otros métodos para fines electorales que permiten mejorar las estimaciones usando variables electorales como tal (Condori, 2021), (PEW R.C., 2021), estos no serán abordados.

2.4. Tipos de selección

Los tipos de selección utilizados en la 1ra etapa fueron:

- Aleatoria simple de UPM
- Sistemática simple de UPM
- Secuencial simple de UPM
- Aleatoria con probabilidad proporcional al tamaño de la UPM
- Sistemática con probabilidad proporcional al tamaño de la UPM
- Secuencial con probabilidad proporcional al tamaño de la UPM

Los tipos de selección variaron solo en la primera etapa, en la segunda etapa solo se seleccionaron una única mesa dentro de cada recinto, la que se realizó de forma aleatoria simple para todos los diseños. Para las selecciones proporcionales al tamaño, se utilizó como medida de tamaño de selección al número de mesas dentro del recinto.

Las selección aleatoria simple y sistemática son ampliamente conocidas, pero la selección secuencial no es muy difundida, sin embargo se la puede describir como una simbiosis de ambas, planteando un escenario intermedio, entre saltos sistemáticos con una componente aleatoria (Chromy, 1979).

2.5. Tamaños y distribución de muestra

El tamaño y distribución de muestra es tal vez la parte importante en los diseños, sin embargo, para no recaer en usar “muestras muy grandes” e inoperables, se utilizó el tamaño de muestra total y del recuento de votos de CiesMori de 2020, que es bastante modesto y más ágil en su aplicación. El tamaño de la muestra fue de 269 recintos con un total aproximado de 151 inscritos (ERBOL, 2020), y asumiendo que regularmente hay 200 inscritos por mesa, se tuvieron que haber visitado un aproximado de 755 mesas.

Por otro lado, el estudio de la iniciativa TuVotocuenta, organizado por la UMSA y la Fundación Jubileo, (LosTiempos, 2020), muestrearon 4.711 mesas en 1.200 recintos.

Para el presente estudio se aplicó el tamaño de CiesMori de 269 recintos, y en cada recinto se eligió una sola mesa. Por limitantes de información, la ficha técnica no estaba publicada oficialmente en la página oficial del Órgano Electoral Plurinacional (OEP), con lo cual se optó por una distribución potencial con un $\alpha=0.5$ entre los departamentos, y posteriormente dentro de cada departamento se aplicó una distribución proporcional.

2.6. Metodología e indicadores para la comparación de diseños de muestreo y sus estimaciones

Entre los indicadores de calidad posterior a estimación de un muestreo las métricas fueron:

- Error estándar (ee)
- Error de muestreo o coeficiente de variación (cv)
- Efecto de diseño (eff)
- Coeficiente de variación logarítmico (cvl)

Este último elemento es una medida propuesta por algunos autores, pero poco difundida (SAMHSA, 2007), debido a que en diversas ocasiones cuando los estimadores de tipo proporción $\hat{p}$, suelen tener errores de muestreo muy altos cuando las estimaciones de $\hat{p}$ presentan valores muy bajos, pero contradictoriamente para el estimador de su complemento $\hat{q} = 1 - \hat{p}$ el error de muestreo estimado es muy bajo. Esta medida de error de muestreo se expresa como:

$$cv(\hat{L}) = -cv(\hat{p})/\log(1 - \hat{p})$$

Dicha expresión se deduce de la aproximación de Taylor de primer orden para el estimador logarítmico $\hat{L} = -\log(1 - \hat{p})$ (Gutierrez, Fuentes, & Mancero, 2020). Para el presente estudio se aplicó al estimador de razón $\hat{r}$ para el porcentaje de votos recibidos a cada partido: $cv(\hat{L}) = -cv(\hat{r})/\log(1 - \hat{r})$.

El resto de estadísticas fueron definidas convencionalmente $ee(\hat{r}) = \sqrt{V(\hat{r})}$ para el error estándar, $ee(\hat{r}) = \sqrt{V(\hat{r})}$ para el coeficiente de variación, y $eff(\hat{r}) = V(\hat{r})/V_{MAS}(\hat{r})$ para el efecto de diseño, donde el último termino es la varianza del estimador bajo un muestreo aleatorio simple.

2.7. Diseño de la investigación

Dado que se planificó múltiples escenarios, y se compararon entre sí, se considera esta investigación de tipo experimental, ya que se están evaluaron diferentes diseños de muestreo a ser aplicados.

2.8. Implementación de los diseños de muestreo

Para poder implementar todos los diseños de muestreo y sus estimaciones, se simularon en base a un conjunto de 100 semillas aleatorias, a 100 muestras independientes para cada uno de los diseños propuestos anteriormente.

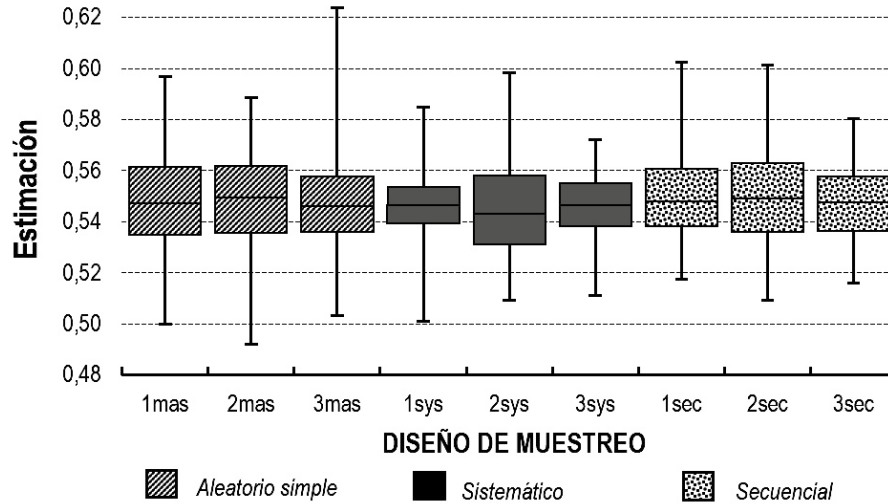
Dado que se trabajó con 3 diseños de muestreo posibles con su estratificación respectiva y 6 tipo de selección, se tuvieron 18 escenarios planteados se limitó la comparación de los estimadores y sus medidas de error solo a los resultados a nivel nacional.

3. Resultados

A continuación, se muestran las estimaciones del porcentaje de votos válidos, su error estándar y coeficiente de variación, margen de error respecto al parámetro y error cuadrático medio.

Figura 1.

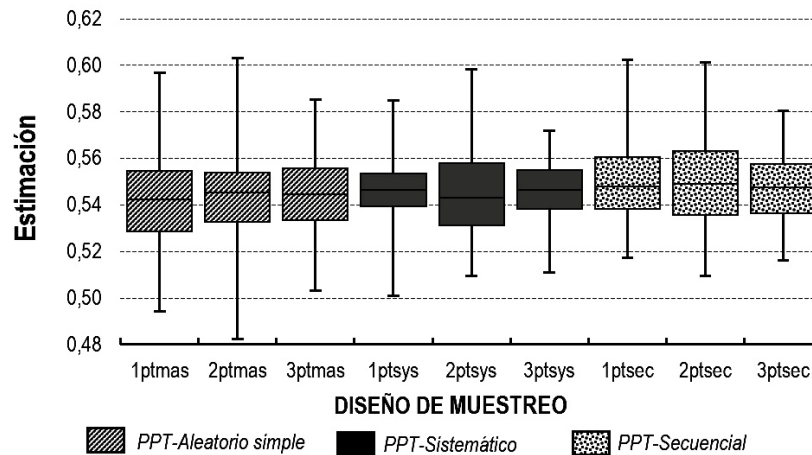
Boxplot Estimaciones de votos al partido MAS, según los diferentes tipos de diseño



Fuente: Resultados Elecciones Generales 2020, OEP.
Elaboración: Propia.

Figura 2.

Boxplot Coeficientes de variación promedio de votos al partido MAS, según los diferentes tipos de diseño



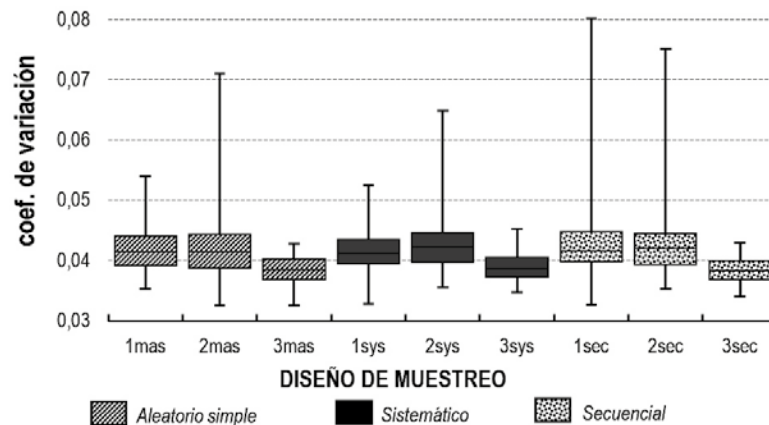
Fuente: Resultados Elecciones Generales 2020, OEP.
Elaboración: Propia.

Estas estimaciones en su mayoría bordean el resultado oficial nacional de 54.7% para el partido MAS, 14.06% a CREEMOS, y 29.16% para CC (OEP, 2020). Se puede observar en las Figuras 1 y 2, que las selecciones no son proporcionales, las que poseen menor dispersión en sus estimaciones, e incluso son más homogéneas entre sí.

Por su lado, los coeficientes de variación, son muy semejantes entre selección PPT y selección simple (Figuras 2 y 3).

Figura 3.

Boxplot Coeficientes de variación promedio de votos al partido MAS, según los diferentes tipos de diseño no proporcional

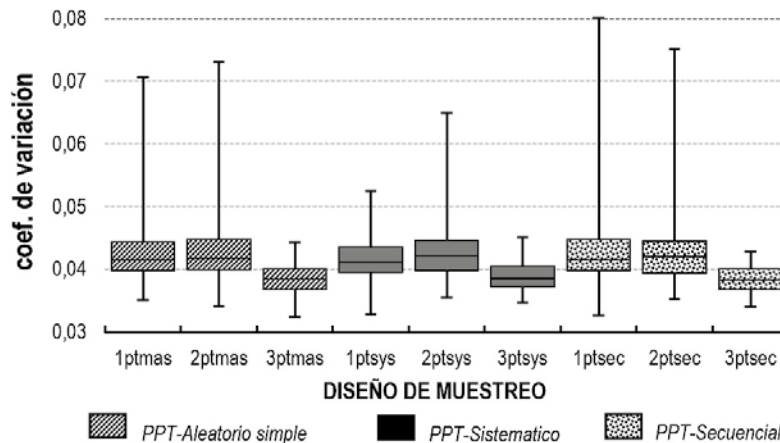


Fuente: Resultados Elecciones Generales 2020, OEP.

Elaboración: Propia.

Figura 4.

Boxplot Coeficientes de variación promedio de votos al partido MAS, según los diferentes tipos de diseño proporcional



Fuente: Resultados Elecciones Generales 2020, OEP.

Elaboración: Propia.

A continuación, se observa la dispersión de los promedios de errores estándar y coeficientes de variación, para la misma estimación. En dicho gráfico se observa que los diseños de menor error estándar y coeficiente de variación son los de la 3ra estratificación con selección secuencial (Figura 4).

Tabla 1:

Coef. de var. logarítmicos promedios, según tipo de diseño

Selección	Estratificación		
	1ra	2da	3ra
mas	0.047	0.046	0.042
sys	0.046	0.048	0.043
sec	0.048	0.047	0.042
pptmas	0.048	0.049	0.043
pptsys	0.046	0.048	0.043
pptsec	0.048	0.047	0.042

Fuente: Resultados Elecciones Generales 2020, OEP.

Elaboración: Propia

Se debe ampliar la comparación sobre las estimaciones en la composición de las cámaras de senadores y diputados para el caso Boliviano, esto de manera semejante al análisis de Stoker sobre los efectos multinivel en las medidas de precisión de muestreo (Stoker y Bowers, 2002).

5. Conclusiones

Es posible aplicar la metodología de comparación de diseños de muestreo en la temática electoral, pudiendo ostentar otras alternativas a solo observar el error teórico en función al tamaño de muestra.

Estos ejercicios permitirían elegir estrategias de muestreo con una mayor experiencia de los posibles escenarios, o incluso la posibilidad de reducir los tamaños de muestra en las encuestas oficiales que el INE sin incrementar los errores de muestreo y agilizando el procesamiento de estas.

Referencias Bibliográficas

- Chromy, J. (1979). Sequential Sample Selection Methods. *Research Triangle Institute*, 401-406, http://www.asasrms.org/Proceedings/papers/1979_081.pdf
- CIESMORI. (marzo de 2021 de 2021). *Informe Conteo Rapido Elecciones Subnacionales del 7 de marzo 2021*. Obtenido de https://www.oep.org.bo/wp-content/uploads/2021/04/Informe_Cuento_Rapido_CIESMORI_EDRM_2021.pdf
- Cochran, W. (1977). *Sampling Techniques*, 3rd Edition. New York: Jhon Wiley & Sons.
- Condori, R. (2021). Estratificacion Asimetrica en Encuestas Electorales. *Varianza Nro 18, UMSA*, 9-19, <https://drive.google.com/file/d/1IDzdNUQlg1A68UFVed96Fp3W1V9fCVTS/view>
- Deville, J.-C., Sarndal, C.-E., & Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, Vol. 88, No. 423, 1013-1020, <https://doi.org/10.2307/2290793>
- ERBOL. (19 de octubre de 2020). *ERBOL Educacion Radiofonica de Bolivia*. Obtenido de Periodico ERBOL: <https://erbol.com.bo/el-%C3%A1nfora-1/conteo-de-ciesmori-proyecta-victoria-de-luis-arce-en-primera-vuelta>
- FOCALIZA. (2021). *Informe Técnico Recuento Rápido Bolivia: Elecciones Sub Nacionales 2021*. Santa Cruz de la Sierra, https://www.oep.org.bo/wp-content/uploads/2021/04/Informe_Cuento_Rapido_Focaliza_EDRM_2021.pdf
- Garzón-Sherdek, K. (2021). Ecuador Elecciones Generales 2021 (Segunda vuelta). *Análisis de Elecciones 2021, Observatorio de Reformas Políticas en América Latina, IIJ-UNAM y Organización de los Estados Americanos (OEA)*, 1-15, <https://reformaspoliticas.org/wp-content/uploads/2021/05/Analisis-Elecciones-Ecuador-segunda-vuelta-1.pdf>
- Gutierrez, A., Fuentes, A., & Mancero, X. (2020). *Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: Una aplicacion a la migracion internacional*. Santiago: Comisión Económica para América Latina y el Caribe (CEPAL).
- INE. (2020). *Diseño de muestra Encuesta de Hogares 2020*. La Paz: Instituto Nacional de Estadística, <http://anda.ine.gob.bo/index.php/catalog/88/download/922>

- Ley026. (30 de junio de 2010). Ley del Régimen Electoral. *Estado Plurinacional de Bolivia*.
- LosTiempos. (19 de octubre de 2020). “*Tu Voto Cuenta*”: MAS obtuvo 53% y CC 30.8%, según resultados de conteo rápido. Obtenido de Periodico Los Tiempos: <https://www.lostiempos.com/actualidad/pais/20201019/tu-voto-cuenta-mas-obtuvo-53-cc-308-resultados-conteo-rapido>
- OEP. (2020). *Acta de Computo Nacional Elecciones Generales 2020*. La Paz-Bolivia: Organo Electoral Plurinacional, <https://www.oep.org.bo/wp-content/uploads/2020/11/Acta-de-computo-2020.pdf>
- OEP. (septiembre de 2020). *Reglamento de Elaboración y difusión de Estudios de Opinión en materia Electoral en Procesos Electorales*. Obtenido de Elecciones Generales 2020: https://www.oep.org.bo/wp-content/uploads/2020/10/Reg_Estudios_Opinion_EG_2020.pdf
- Página-Siete. (19 de octubre de 2020). *Página Siete*. Obtenido de Arce logra más de 50%, según el conteo rápido de dos firmas: <https://www.paginasiete.bo/nacional/2020/10/19/arce-logra-mas-de-50-segun-el-conteo-rapido-de-dos-firmas-272050.html>
- Perez, C. (2005). *Muestreo Estadístico, conceptos y problemas resueltos*. Madrid: Pearson Prentice Hall.
- PEW R.C. (30 de Junio de 2021). Behind Biden’s 2020 Victory: Methodology. Obtenido de PEW RESEARCH CENTER: <https://www.pewresearch.org/politics/2021/06/30/validated-voters-methodology/>
- RNEC. (2019). *Registraduría Nacional del Estado Civil*. Obtenido de Calendario Electoral Autoridades Locales 2019: <https://www.registraduria.gov.co/IMG/pdf/RES-14778-11-OCT-2018-CALENDARIO-ELECTORAL-AUTORIDADES-LOCALES.pdf>
- SAMHSA. (2007). *2006 National Survey on Drug Use and Health: National Findings*. Rockville: Office of Applied Studies, NSDUH Series H-32. Obtenido de <https://files.eric.ed.gov/fulltext/ED498206.pdf>
- SERVEL. (Julio de 2021). *Cronograma Electoral Elecciones Presidencial, Parlamentarias y de Consejeros regionales 2021*. Obtenido de Subdirección Registro, Inscripciones y Acto Electoral: División de Procesos Electorales: https://www.servel.cl/wp-content/uploads/2021/07/CRONOGRAMA_ELECTORAL-_ELECCIONES_DEFINITIVAS_2021.pdf
- Stoker, L., & Bowers, J. (2002). Designing multi-level studies: sampling voters and electoral contexts. *Electoral Studies* 21, 235–267, <http://www.jakebowers.org/PAPERS/stokerbowersFinalplusErratum.pdf>
- TEP. (2021). *Calendario Electoral: Elecciones Federales y elecciones estatales Mexico 2021*. Obtenido de Escuela Judicial Electoral: https://www.te.gob.mx/calendario_electoral/
- TSJE. (28 de Septiembre de 2021). *Tribunal Superior de Justicia Electoral - Republica del Paraguay*. Obtenido de Elecciones Municipales 10 de octubre 2021 - Dossier Informativo: <https://www.tsje.gov.py/elecciones-municipales-10-de-octubre-2021---dossier-informativo.html>
- ViaCiencia. (2019). *Conteo Rapido ViaCiencia Elecciones Generales 2019: Ficha Técnica*. Santa Cruz, https://www.oep.org.bo/wp-content/uploads/2019/11/Conteo_Rapido_VIACIENCIA_EG_2019.pdf

Una visión de la pandemia del COVID-19 en los municipios de Bolivia y su componente espacial

A vision of the COVID-19 pandemic in the municipalities of Bolivia and its spatial component

Marisol Paredes Alarcón¹

Universidad Mayor de San Andrés, La Paz, Bolivia

✉ mparedes15@umsa.bo

Artículo recibido: 2022-03-02

Artículo aceptado: 2022-03-24

Resumen

Este artículo examina las características, la dimensión espacial y socioeconómica de los casos de COVID-19 durante las cuatro olas de la pandemia en los 339 municipios de Bolivia, durante estos casi dos años de pandemia. Esta situación ha develado de forma clara las desigualdades sociales y económicas en las regiones geográficas del país, principalmente en los municipios.

Para obtener una mayor comprensión de este fenómeno se ha recopilado información estadística para cada una de las cuatro olas, tanto geográfica como multivariable, por medio de 16 indicadores propuestos. Los resultados de este trabajo señalan en primer lugar como el COVID-19 ha ido expandiéndose geográficamente, y a lo largo del tiempo ha manifestado una creciente intensidad desde la primera ola hasta el crecimiento exponencial de la cuarta ola. La dinámica en términos de cantidad si bien se concentró en aquellos municipios capitales, es decir urbanos, pero efectuando la agrupación por quintiles del acumulado de casos del COVID-19 se advierte la expansión hacia municipios considerados rurales y con menor población o alejados de los municipios capitales. En tanto que el análisis multivariable de indicadores revela la relación de los casos acumulados durante la primera ola del coronavirus con factores de pobreza, baja cobertura de saneamiento básico, carencias en temas relacionados a salud e incluso de poca cobertura de telefonía fija o celular por municipio

Palabras clave: COVID-19, Geografía de la Salud, análisis espacial, análisis multivariable

Abstract

This article examines the characteristics, the spatial and socioeconomic dimension of the COVID-19 cases during the four waves of the pandemic in the 339 municipalities of Bolivia, during these almost two years of the pandemic. This situation has clearly developed social and economic inequalities in the geographical regions of the country, mainly in the municipalities.

To obtain a better understanding of this phenomenon, statistical information has been compiled for each of the four waves, both geographical and multivariable, through 16 proposed indicators. The results of this work indicate in the first place how COVID-19 has been expanding geographically, and over time it has manifested a growing intensity from the first wave to the exponential growth of the fourth wave. The dynamics in terms of quantity, although it was concentrated in those capital municipalities, that is, urban, but grouping by quintiles of the accumulated cases of COVID-19, the

¹ Licenciada en Estadística. Docente en la Universidad Mayor de San Andrés. <https://orcid.org/0000-0002-8445-6202>

expansion towards municipalities considered rural and with a smaller population or far from the municipalities is noted. While the multivariate analysis of indicators reveals the relationship of the cases accumulated during the first wave of the coronavirus with factors of poverty, low coverage of basic sanitation, deficiencies in issues related to health and even low coverage of landlines or cell phones by municipality.

Key words: COVID-19, Health Geography, spatial analysis, multivariate analysis

Introducción

En Bolivia la pandemia del COVID-19 causada por el virus SARS-CoV-2, se detectó por primera vez el 10 de marzo de 2020 con dos casos en los departamentos de Oruro y Santa Cruz (Ministerio de Salud, 2020).

A partir de esa fecha se han desarrollado cuatro olas de la pandemia, cada una con sus peculiaridades. Al inicio de la primera ola no se tenía suficiente conocimiento acerca del COVID-19, y esto a escala global. Las primeras medidas adoptadas en nuestro territorio se referían al confinamiento, la suspensión de clases, el teletrabajo, cierre de fronteras, y la introducción de protocolos de cuidado y bioseguridad. En tanto que en el lado opuesto durante la cuarta ola, que a la fecha va en descenso, se caracterizó por la vacunación masiva, el inicio de clases presenciales, y en general el relajamiento de una gran parte de medidas estrictas, o la denominada nueva “normalidad”.

Una característica importante de cada una de las cuatro olas tiene que ver con el espacio geográfico de la propagación del virus, la cual no ha sido espacialmente homogénea, y en la que intervienen factores tales como elementos demográficos, características socioeconómicas o la interconexión entre municipios cercanos.

La pandemia del coronavirus ha develado de forma clara las desigualdades sociales y económicas en las regiones geográficas del país, entre ellas los municipios. Y esto guarda relación con la creciente relación entre la geografía con otras ciencias, entre ellas la salud.

Por otra parte, la estadística combinada con la geografía es un enfoque que considera una amplia gama de técnicas que proveen una perspectiva cuantitativa amplia, entre ellas los Sistemas de Información Geográfica (SIG).

Debido a lo anterior, el componente espacial es sumamente importante para estudiar como el COVID-19 se esparce por los municipios de Bolivia, en cada una de las olas, y las desigualdades que se visibilizan relacionadas a los niveles de pobreza, disponibilidad de servicios básicos, indicadores de salud, vías de transporte entre otros, que influyen en el avance de la pandemia.

El campo de estudio de la Geografía de la Salud se ha iniciado con el estudio de la distribución de las enfermedades, por medio de la elaboración de mapas temáticos que permitan mostrar la distribución espacial de las enfermedades y en otros casos su relación con otro tipo de características cercanas.

En la Geografía de la Salud, el interés radica en abordarlo desde la localización y la distribución de las enfermedades, tomando en cuenta su asociación con otro tipo de características, además de la utilización de herramientas de información geográfica y estadística que permitan generar modelos que puedan aproximar a explicaciones sobre la situación actual para luego buscar proyectar situaciones futuras (Buzai, 2020).

Diversos estudios se han realizado a nivel global sobre la pandemia del COVID-19 considerando la distribución geográfica. El trabajo de Jin et al. (2020) recopila características demográficas generales, combinadas con distribuciones temporales, geográficas y de población de la COVID-19 en China, por medio de mapas temáticos.

En países cercanos como Colombia y Ecuador también se trabajó buscando conglomerados espacio-temporales de los casos de coronavirus. El trabajo de Cuartas et al. (2020) realiza un análisis exploratorio de datos espaciales, con la disponibilidad de las coordenadas geográficas de los casos puntuales de casos de coronavirus.

En Bolivia también se desarrollaron investigaciones donde interviene el componente espacial respecto del avance de la pandemia. En la investigación efectuada por Acuña et al. (2020) se desarrolló un análisis a nivel municipal por medio de la representación geográfica de un índice de vulnerabilidad que visibiliza las condiciones estructurales y capacidad de respuesta sanitaria de los 339 municipios para afrontar la pandemia.

En resumen el COVID-19 nos presenta una gran oportunidad para estudiar el espacio geográfico, visto desde la perspectiva de los 339 municipios y de características relevantes que se interrelacionan.

El objetivo general de esta investigación es describir las características y dinámicas espaciales y socioeconómicas de los casos de COVID-19 durante las cuatro olas de la pandemia en los municipios, durante estos casi dos años de pandemia en Bolivia, y pueda servir como aporte inicial de aplicaciones que tomen en cuenta el componente espacial.

Materiales y Métodos

Se analizaron todos los casos confirmados de COVID-19 para los 339 municipios de Bolivia, desde el 10 de marzo de 2020 hasta el 13 de febrero de 2022, de los datos disponibles en el sitio oficial del Sistema Integrado de Información productiva (SIIP), del Ministerio de Desarrollo Productivo y Economía Plural, que provee datos históricos y un total acumulado nacional, departamental y municipal de los casos de coronavirus confirmados en el país.

Las coordenadas geográficas correspondientes a los límites municipales se recuperaron del sitio web de Infraestructura de datos espaciales (GeoBolivia) la cual es una iniciativa de la Vicepresidencia del Estado Plurinacional por medio del cual se provee de forma abierta y accesible una amplia información geográfica de interés.

GeoBolivia también se constituye en la punta de lanza de la Infraestructura de Datos Espaciales del Estado Plurinacional de Bolivia (IDE-EPB) que por medio de un grupo de instituciones públicas y privadas, busca compartir información ligada al territorio nacional procedente de diferentes fuentes.

Finalmente se utilizaron los indicadores de medición de los Objetivos de Desarrollo Sostenible (ODS) de Bolivia para el 2020, que son objetivos universales ratificados en 2015 por 193 países que conforman la Asamblea General de la Organización de las Naciones Unidas (ONU). Para tal fin Andersen et al., (2020) sistematizaron una variedad de información estadística que permite realizar seguimiento a los ODS en cada municipio. De esta variedad de indicadores se

seleccionaron aquellos relevantes y que guardan cierto grado de relación con los avances de la pandemia en nuestro territorio nacional.

Se utilizó el software libre QGIS para la elaboración de los mapas temáticos, y para el análisis estadístico gráfico de los indicadores el software libre R.

Resultados

El análisis de la información revisada se dio inicio con el estudio del comportamiento a lo largo del tiempo de los casos confirmados de Covid-19 desde el 10 de marzo de 2020 donde fueron detectados los primeros casos de coronavirus, hasta el 13 de febrero de 2022. Prácticamente ya son dos años de pandemia en Bolivia. En este lapso de tiempo se dieron un total de 883.642 casos confirmados de acuerdo a fuentes oficiales.

Evolución en el tiempo del COVID-19

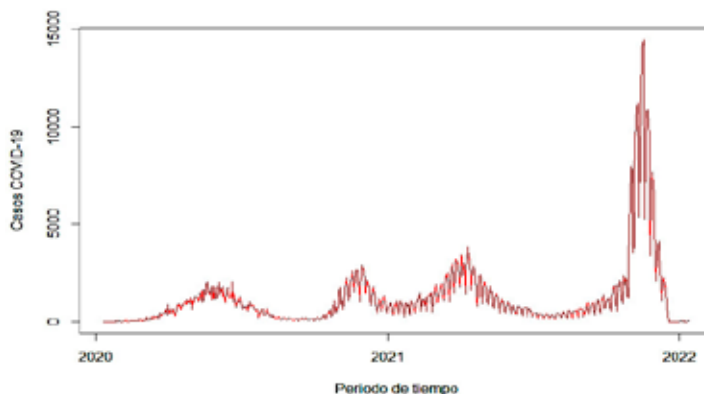
Durante estos dos años se han presentado cuatro olas en nuestro país. La Figura 1 muestra precisamente la serie temporal del comportamiento del coronavirus, en la que se advierte la evolución de casos de forma diferenciada para cada una de las olas o picos.

La primera de ellas que de forma aproximada se dio entre el mes de marzo y octubre de 2020, se ha caracterizado por el escaso conocimiento acerca del virus, con medidas de confinamiento, la suspensión de clases, el teletrabajo, cierre de fronteras, y la introducción de protocolos de cuidado y bioseguridad.

La segunda ola entre enero de 2021 y marzo del mismo año y el inicio poco diferenciado con la tercera ola entre el mes de abril de 2021 y septiembre del mismo año, se caracterizó por otro conjunto de medidas diferentes a las anteriores, además de un nuevo gobierno nacional, que dio inicio a la vacunación masiva de distintos grupos etarios de la población, y por el retorno a la denominada nueva “normalidad”.

Figura 1.

Bolivia: Casos de COVID-19 confirmados desde el 10 de marzo de 2020 al 13 de febrero de 2022.



Fuente: Elaboración propia en base a los datos disponibles en el Sistema Integrado de Información Productiva (SIIP), Ministerio de Desarrollo Productivo y Economía Plural, Bolivia.

Y finalizando con la cuarta ola, marcada por el aumento exponencial de casos en comparación a las anteriores olas, pero cuya duración en tiempo fue menor que las tres primeras, y caracterizada por la llegada al país de nuevas cepas del virus, la relajación de las medidas de bioseguridad, la ausencia de confinamiento, el inicio del retorno a clases presenciales y una serie de conflictos sociales.

Distribución geográfica de los casos de COVID-19

La situación de la pandemia ha develado una serie de desigualdades por áreas geográficas. El abordaje de ellas resulta un tema sumamente importante en el caso de los municipios de Bolivia. Para comunicar estos cambios se elaboran mapas temáticos, que puede asemejarse al uso de un histograma de frecuencias para el análisis gráfico de datos pero considerando el componente espacial.

Uno de los tipos de mapas temáticos más utilizados es el denominado de coropletas, donde se tiene una serie de áreas definidas, en nuestro caso los municipios, cada uno de los cuales tiene asociado la variable de casos confirmados de COVID-19 acumulados por cada una de las cuatro olas. Se ajustaron los periodos de tiempo de cada ola en función al descenso de cada pico, diferenciando el acumulado de casos confirmados desde el inicio de la primera ola hasta el descenso de la cuarta ola.

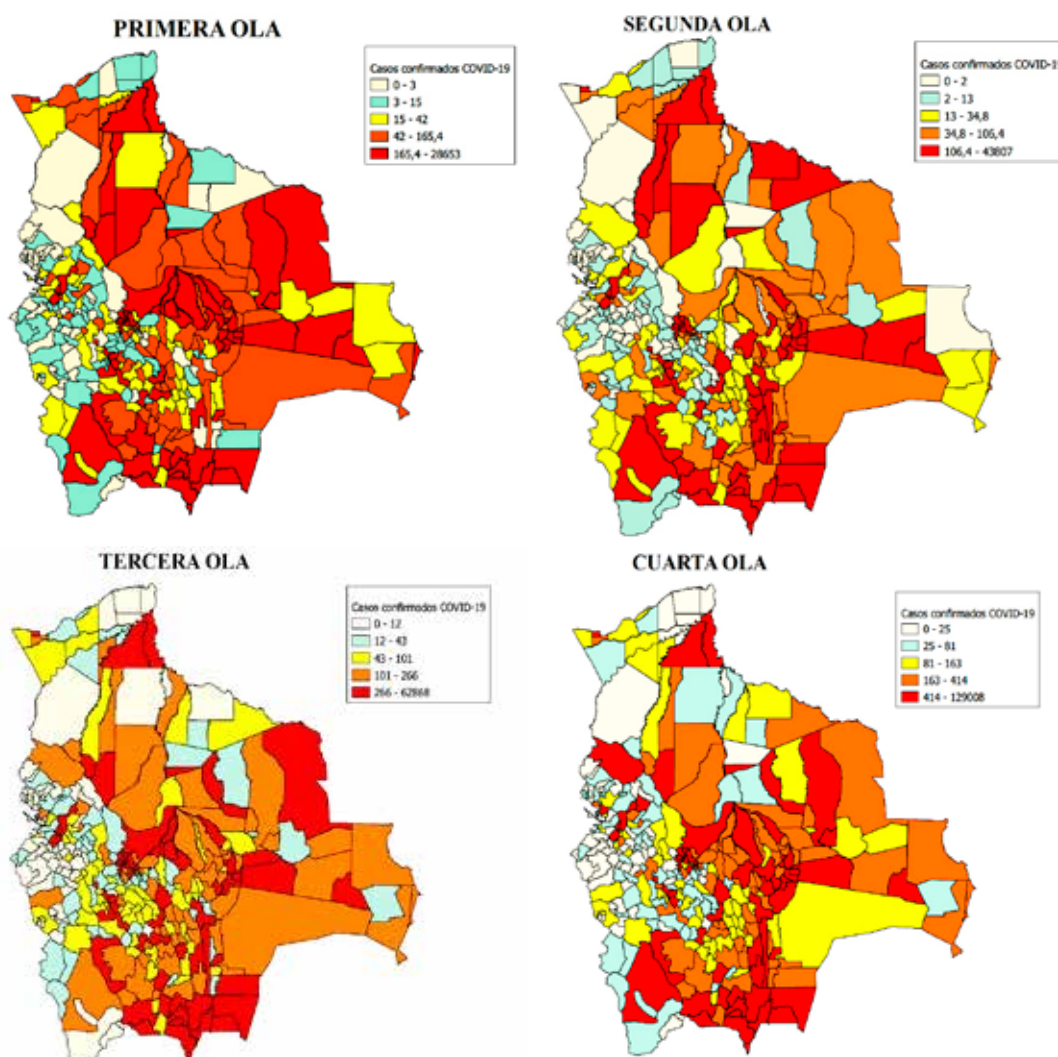
Se disponen de distintos métodos para la creación de clases o bien para resumir datos usando diferentes esquemas de clasificación geográfica. El utilizado en estos mapas es por medio de los quintiles, esto indica que en cada uno de los cinco grupos creados se tiene la misma cantidad de municipios, es decir el 20% en cada grupo.

Las diferentes tonalidades para cada una de las cuatro olas señalan la cantidad de casos acumulados. Inicia con la tonalidad roja que indica a los municipios con las mayores cantidades acumuladas de casos, le siguen en importancia las tonalidades naranja, amarilla, verde y finalmente el grupo de municipios con las menores cantidades de casos en color crema.

Los resultados de esta clasificación muestran que durante la primera ola la cantidad máxima acumulada de casos confirmados de coronavirus en un municipio fueron 28.653 casos en el municipio capital de Santa Cruz de la Sierra. A esta categoría más alta de casos corresponden en su mayoría a departamentos del oriente como Santa Cruz y Beni, aunque también están presentes en menor cantidad en los departamentos de Potosí, Oruro, Tarija, Chuquisaca y Cochabamba, básicamente en municipios capitales de departamento. Una menor cantidad en esta categoría se presentan en el departamento de La Paz, situación que va revirtiéndose paulatinamente en el resto de las olas de la pandemia, sobre todo en la cuarta ola.

El crecimiento de casos acumulados para cada una de las olas siguientes a la primera es significativo, principalmente en la cuarta ola donde el crecimiento ha sido exponencial, con el máximo de casos acumulados de 129.008 en la ciudad de Santa Cruz de la Sierra.

Figura 2.
Quintiles de casos de COVID-19 confirmados por olas



Fuente: Elaboración propia en base a los datos disponibles en el Sistema Integrado de Información Productiva (SIIP), Ministerio de Desarrollo Productivo y Economía Plural.

A pesar de las medidas de contención de la pandemia, y principalmente la aplicación masiva de vacunas, en el comportamiento de casos acumulados por municipios, se advierte que la expansión del virus ha logrado alcanzar a municipios del occidente, algunos inclusive alejados de los municipios capitales que son aquellos que concentran la mayor cantidad de casos y de población en Bolivia. También se advierte que los municipios próximos a otros departamentos y/o municipios pertenecientes a categorías de casos elevados del COVID-19 ascienden a otras categorías con una frecuencia mayor en varios departamentos.

Esto devela una influencia geográfica, pero también impulsa al estudio de características adicionales al número de casos acumulados.

Existen dinámicas particulares en el comportamiento por municipios, y éstas pueden ser apoyadas con un conjunto diferenciado de indicadores que puedan tener algún grado de correlación con los casos acumulados del COVID-19 detectados. Los indicadores que se han utilizado corresponden a la

medición de los Objetivos de Desarrollo Sostenible (ODS) de Bolivia para el 2020, de los cuales se hizo una selección.

Correlación entre los casos del COVID-19 y los indicadores seleccionados

Para facilitar el análisis, y dado que se considera que la primera ola de la pandemia puede caracterizar de forma general la identificación de los indicadores asociados a las principales desigualdades en los municipios y los casos acumulados confirmados del COVID-19, para este apartado sólo se tomaron en cuenta estos primeros casos.

Los indicadores seleccionados engloban aspectos de los 339 municipios de Bolivia relativos a pobreza, salud, acceso a servicios básicos, trabajo, vías de transporte, población entre otros, que puedan tener relación con la propagación del virus y su espacio geográfico, además de la relación con medidas restrictivas establecidas por el gobierno.

Los 16 indicadores municipales elegidos son:

P1-4: Población con necesidades básicas insatisfechas (2012)

P1-5: Índice de pobreza multidimensional

H2-1: Desnutrición crónica en niños (menores a 5 años, en porcentaje, 2016)

H2-3: Sobrepeso en mujeres (15-49 años, en porcentaje, 2016)

S3-2: Número de partos no atendidos en un establecimiento de salud (2012)

S3-3: Tasa de mortalidad infantil (menor a 1 año, por 1000 nacidos vivos, 2016)

S3-17: Tasa de fecundidad en adolescentes (15-19 años, nacimientos por 1000 mujeres, 2012)

AG6-2: Población sin cobertura de agua potable. (2017)

AG6-4: Población sin cobertura de saneamiento básico. (2017)

EN7-5: Energía limpia para cocinar (Porcentaje de hogares, 2012)

TR8-5: Tasa global de participación de mujeres (Mayor o igual a 10 años, en porcentaje, 2012)

TR8-11: Densidad de sucursales de bancos (por 100.000 habitantes, 2018)

INF9-1: Número de vías férreas y carreteras primarias que entran/salen del municipio, (2019)

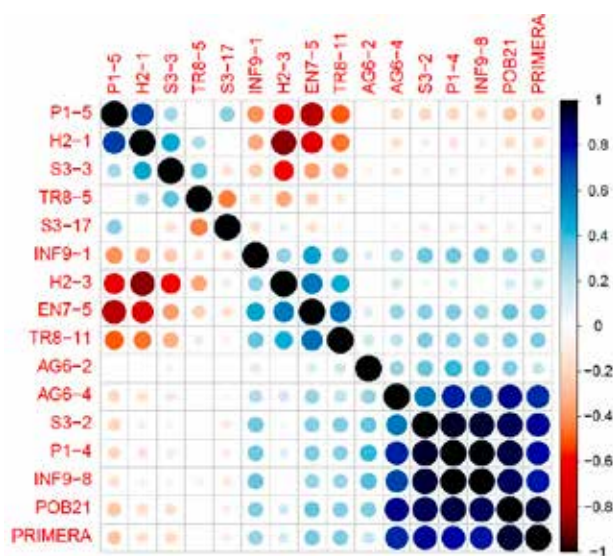
INF9-8: Número de hogares sin cobertura de telefonía fija o celular (2012)

POB21: Proyecciones de población por municipio al 2021

PRIMERA: Casos acumulados confirmados durante la primera ola 2020

Para medir el grado de relación se ha calculado la matriz de correlación de los indicadores seleccionados.

Figura 3.
Matriz de correlación de los indicadores seleccionados



Fuente: Elaboración propia

El análisis de las correlaciones mostrado en la Figura 3 señala una degradación de color en la barra lateral. Mientras más se aproxime al color azul se da una mayor correlación positiva, y si se aproxima al rojo más intenso una correlación negativa más fuerte. También el tamaño del círculo muestra este grado de correlación entre indicadores.

Resaltan las correlaciones del número de casos acumulados de la primera ola denominado “Primera”, con el tamaño proyectado del municipio, con indicadores de pobreza y sobre todo de carencias o falta de cobertura de servicios básicos. Hay una muy baja correlación con la densidad de bancos en cada municipio, o del número de vías férreas y carreteras primarias que entran/salen del municipio, que se podrían asemejar a características de movilidad que podrían haber facilitado la expansión del virus, en cuanto a conexiones con otros municipios vecinos.

Análisis de Componentes Principales entre los casos del COVID-19 y los indicadores seleccionados

El análisis multidimensional de indicadores seleccionados también puede representarse mediante un análisis de componentes principales. Desde esta perspectiva el avance de la pandemia no sólo es geográfico sino multidimensional.

El análisis de componentes principales consiste en generar nuevas variables que sean la combinación lineal de las variables originales, reduciéndolas a un número menor de componentes que explique la mayor cantidad de variación posible. De acuerdo a la Figura 4 del conjunto original de indicadores escogidos, se advierte que pueden ser explicados por los dos primeros componentes en cerca del 61%.

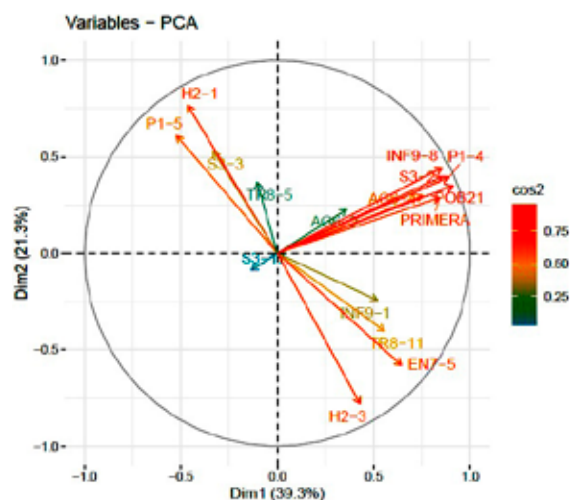
Acá se advierte de acuerdo al largo de las flechas de cada indicador, que una gran parte de ellos tiene un alto coeficiente en cada componente, aunque varios indicadores están agrupados de acuerdo a la relación presente entre ellos.

El caso que incide en esta investigación es el análisis de los indicadores cercanos al indicador “Primera” que corresponde al total de casos acumulados de coronavirus, el cual permanece cercano

a los indicadores de pobreza como son las Necesidades Básicas Insatisfechas, también al tamaño de población proyectada al 2021 de cada municipio. Se presentan asociados también indicadores de salud como el número de partos no atendidos en un centro de salud y por otra parte aspectos relativos a carencias de servicios básicos como agua potable en menor grado y en mayor grado falta de cobertura en saneamiento básico. Llama la atención la proximidad del número de casos acumulado de la primera ola con el indicador del número de hogares sin cobertura de telefonía fija o celular por municipio.

Figura 4.

Gráfica del Análisis de Componentes Principales de los indicadores seleccionados



Fuente: Elaboración propia

Discusión

Las características de propagación del virus se han ido conociendo con mayor exactitud a medida que avanzaba la pandemia. Caracterizar esta evolución de la propagación en los 339 municipios de Bolivia ayuda a entender la situación del contagio, y de manera aproximada en cómo influye en primer lugar el aspecto espacial/geográfico en esta dinámica, y en segundo lugar el grado de relación con otro tipo de variables por municipio.

Esto ayuda a conocer y posiblemente repensar las estrategias gubernamentales sobre la propagación del virus en los municipios, por la dinámica estudiada, y su relación con indicadores relevantes que muestran el efecto multidimensional de la pandemia.

La Geografía de la Salud que se dedica al análisis de la distribución espacial de enfermedades es uno de los campos de mayor dinamismo en el análisis geográfico, dadas las circunstancias actuales de la pandemia. Ante la posibilidad de incorporar metodologías estadísticas en el ámbito de los SIG y con ello avanzar al análisis espacial aplicado, es la que impulsa esta investigación.

Algunas restricciones para este análisis se manifiestan por la carencia de la identificación espacial de cada uno de los casos del COVID-19 en los municipios del país, que podría guiar hacia un análisis espacial con técnicas de análisis puntual. A pesar de lo expuesto, se puede desarrollar otros estudios respecto del análisis espacial por áreas, para la identificación de conglomerados de municipios, tomando en cuenta la correlación espacial, además de pronósticos espaciales por medio de la econometría espacial.

Conclusiones

A lo largo de prácticamente dos años de pandemia en Bolivia, la enfermedad de COVID-19 ha generado una atención constante, y esto también se da a nivel mundial.

La recopilación de toda la información para cada una de las cuatro olas, tanto geográfica como multivariable, por medio de los 16 indicadores propuestos, muestra una perspectiva particular de la situación de la pandemia en los 339 municipios.

Los resultados de este trabajo señalan en primer lugar como el COVID-19 ha ido expandiéndose geográficamente, y a lo largo del tiempo ha manifestado una creciente intensidad desde la primera ola hasta el crecimiento exponencial de la cuarta ola. La dinámica en términos de cantidad si bien se concentró en aquellos municipios capitales, esto es urbanos, pero la agrupación por quintiles del acumulado de casos del COVID-19 muestra la expansión hacia municipios considerados rurales y con menor población o alejados de los municipios capitales. La pandemia del coronavirus ha develado de forma latente las desigualdades sociales y económicas de los municipios del país.

Y precisamente por estos hallazgos es que se aborda paralelamente un conjunto de indicadores relevantes asociados a pobreza, carencia de servicios básicos, indicadores de salud, cantidad de población proyectada al 2021 entre otros.

Del análisis multidimensional de estos indicadores por medio de las correlaciones, y también del análisis de componentes principales, resalta la relación del número acumulado de casos confirmados del COVID-19 principalmente con factores de pobreza, baja cobertura de saneamiento básico, carencias en temas relacionados a salud e incluso de poca cobertura de telefonía fija o celular por municipio.

El marco de este análisis tanto de aspectos geográficos/espaciales, como del comportamiento temporal a lo largo de las cuatro olas y del estudio de su relación con las desigualdades económico-sociales por municipio que puedan explicar el avance de la pandemia, permiten tener un panorama de la situación del COVID-19 en los municipios del país que puedan aproximar el avance de la pandemia, para la toma de decisiones en la perspectiva de problemáticas de salud a nivel regional.

Referencias Bibliográficas

- Acuña, J., Andersen, L., Gonzales, L. (2020). Datos territoriales para la gestión de la pandemia: el caso del COVID-19 en Bolivia. *Revista Latinoamericana de Desarrollo Económico*, 34(1), 7-33. Recuperado de: <https://lajed.ucb.edu.bo/index.php/a/article/view/424>
- Andersen, L., Canelas, S., Gonzales, A., Peñaranda, L. (2020). *Atlas municipal de los Objetivos de Desarrollo Sostenible en Bolivia 2020*. La Paz, Bolivia: SDSN.
- Ballesteros, P., Salazar, E., Sanchez, D., Bolaños, C. (2020). Spatial and spatiotemporal clustering of the COVID-19 pandemic in Ecuador, *Revista de la Facultad de Medicina*, 69(1). Recuperado de: <https://revistas.unal.edu.co/index.php/revfacmed/article/view/86476> .
- Bogoch, I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M., Kamran Khan, P. (2020). Potencial

para la propagación global de un nuevo coronavirus de China, *Journal of Travel Medicine*, 27(2). Recuperado de: <https://doi.org/10.1093/jtm/taaa011>

Buzai, G. (2020). De Wuhan a Lujan. Evolución espacial del COVID-19. *Revista Posición*, 2020(3), 1-21. Recuperado de: www.posicionrevista.wixsite.com/inigeo

Cressie, N., Chan, N. (1989). Spatial modeling of regional variables. *Journal of the Statistical Association*, 84(406), 393-401. Recuperado de: DOI: 10.1080/01621459.1989.10478783.

Cuartas, D., Arango-Londoño, D., Guzmán-Escarria, G., Muñoz, E., Caicedo, D., Ortega, D., Fandiño-Losada, A., Mena, J., Torres, M., Barrera, L., y Méndez, F. (2020). Análisis espacio-temporal del SARS-COV-2 en Cali, Colombia. *Revista de Salud Pública*, 22(2), 1-6. Recuperado de: <https://doi.org/10.15446/rsap.V22n2.86431>

Gonzales, L., Perez, Y. (2013). La minería de datos espaciales y su aplicación en los estudios de salud y epidemiología. *Revista Cubana de Información en Ciencias de la Salud*, 24(4), 482-489. Recuperado de: <https://www.redalyc.org/pdf/3776/377645755009.pdf>

Jin, L., Zhao, Y., Zhou, J., Tao, M., Yang, Y., Wang, X., Ye, P., Shan, S., y Yuan, H. (2020). Distribución temporal, geográfica y por población de la nueva enfermedad por coronavirus (COVID-19) desde el 20 de enero hasta el 10 de febrero de 2020, en China. *Revista Clínica Española*, 220(1), 495-500. Recuperado de: <https://doi.org/10.1016/j.rce.2020.04.001>

Kulldorf, M., Athas, W., Feuer, E., Miller, B., Key, Ch. (1998). Evaluating cluster alarms: A spase-time scan statistics and brain cáncer in Los Alamos, New Mexico. *American Journal of Public Health*, 88(9), 1377-80. Recuperado de: <https://pubmed.ncbi.nlm.nih.gov/9736881/>

Ministerio de Salud. (2020). Guía para el manejo del COVID-19. Recuperado del <https://www.minsalud.gob.bo>

Tisnés, A., Margueliche, J., Haesbaert, R. (2020). Dossier – Pandemia y espacio geográfico: problemáticas, desafíos y perspectivas. *Revista del Departamento de Geografía*, 8(15), 7-15. Recuperado de: <http://revistas.unc.edu.ar/index.php/cardi/index>

Árbol de Decisión en Aprendizaje Automático

Decision Tree in Machine Learning

Fernando Oday Rivero Suguiura¹

Instituto de Estadística Teórica y Aplicada, Universidad Mayor de San Andrés, La Paz-Bolivia

✉ friverosuguiura2004@gmail.com

Artículo recibido: 2022-03-13

Artículo aceptado: 2022-04-01

Resumen

Los árboles de decisión son modelos de clasificación y regresión empleados en el aprendizaje automatizado o también denominado *Machine Learning* de uso en grandes cantidades de datos estadísticos como parte de lo que hoy es la Ciencia de Datos y el *Big Data*.

En este artículo se describe la metodología y se analiza los resultados que arroja el árbol de decisión de clasificación, con uso de variables de la encuesta de hogares, como: género, nivel educativo, ocupación e ingreso en ciudades capitales de Bolivia. Como conclusión, se puede observar que las mujeres están en menor desventaja que los hombres en cuanto a su nivel de formación educativa que influye en la ocupación y el ingreso.

Palabras clave: *Machine Learning*, minería de datos, árbol de decisión, clasificación, entropía, análisis masivo de datos.

Abstract

Decision trees are classification and regression models used in automated learning or also called Machine Learning used in large amounts of statistical data as part of what is now Data Science and Big Data.

This article describes the methodology and analyzes the results of the classification decision tree, using variables from the household survey, such as: gender, educational level, occupation and income in Bolivian capital cities. In conclusion, it can be seen that women are less disadvantaged than men in terms of their level of educational training, which influences occupation and income.

Keywords: Machine Learning, data mining, decision tree, classification, entropy, big data analytics.

1. Introducción

La Minería de Datos comenzó teniendo mucho éxito en los estudios de mercado. Junto con ella se inició la introducción de procesos inductivos basados en los árboles de decisión desarrollados en la Teoría de Decisión. El desarrollo de la informática dio pie a tecnologías especializadas como el aprendizaje automático (*machine learning*) y el reconocimiento de patrones (*pattern recognition*). Un árbol de decisión es una representación de una función multivariada y que fue posible utilizar en la vida práctica a partir del advenimiento de equipos de computación de última generación.

¹ Docente de la carrera de Estadística, Facultad de Ciencias Puras y Naturales de la UMSA. Consultor en muestreo, censos y análisis estadístico en entidades nacionales e internacionales. Magister en Ciencias de la Estadística. <https://orcid.org/0000-0001-9095-7778>

El interés por el uso práctico de los árboles de decisión, tuvo su origen en las necesidades de las ciencias sociales siendo principal el trabajo de Sonquist-Morgan (1964) el software AID (*Automatic Interaction Detection*). Este fue uno de los primeros métodos de ajuste de los datos basados en árboles de clasificación. Con ello los árboles de decisión trascendieron, de ser solo una representación ilustrativa en los cursos de toma de decisiones, para convertirse en una herramienta útil y sencilla de utilizar. Estos avances fueron mejorados por la obra de Breiman-Friedman-Olshen-Stone (1984) llamada (*Classification and Regression Trees*), (Bouza, 2012) - Universidad de La Habana – Cuba, (Santiago, 2012) - Universidad Autónoma de Guerrero – México.

Los datos contienen información oculta potencialmente útil que rara vez se hace explícita o se aprovecha. Con el uso de la minería de datos y las técnicas de aprendizaje automático, se busca determinar patrones de comportamiento en los datos, es decir, se sustenta buscar algún patrón en la información que se encuentre almacenado en las bases de datos, de tal manera que los usuarios puedan solucionar problemas al tratar analíticamente los datos y cuenten con herramientas que les permitan tomar mejores decisiones.

Los algoritmos de aprendizaje basados en árboles decisión se consideran uno de los aventajados y más utilizados métodos de aprendizaje supervisado. Las técnicas basadas en árboles de decisión, potencian los modelos de clasificación y predicción con elevada precisión, confiabilidad, robustez y facilidad de interpretación. A diferencia de los modelos lineales, éstos además consideran posibilidades de inclusión de relaciones no lineales.

Los problemas de clasificación generalmente son aquellos en los que se intentan predecir los valores de una variable dependiente categórica a partir de una o más variables predictoras continuas, discretas o categóricas, por ejemplo, será importante clasificar a las personas por la variable sexo como dependiente y a partir de ella, considerar algunas variables independientes predictivas como el nivel de educación, si trabaja o no y cuánto es su ingreso, etc., que es el ejemplo que se estudia más adelante en este artículo. (Breiman, 1984), (Friedman, 1984).

2. Metodología

2.1 Terminología del árbol de decisión

Un árbol de decisión contiene los siguientes elementos:

- **Nodo de decisión.** Aquel cuando un subnodo se divide en subnodos adicionales indicando la decisión que se tomará ante la disyuntiva.
- **Nodo de probabilidad u hoja.** Presenta múltiples resultados inciertos. Describe las probabilidades de acierto de cada una de las opciones que se plantea ante una disyuntiva.
- **Nodo terminal.** Indica el resultado definitivo como última opción en una ruta de decisión.
- **División.** Proceso de división de un nodo en dos o más subnodos.
- **Flecha.** Nexos que unen los nodos entre sí.
- **Vector.** Representa la opción por la que se opta entre todas las posibilidades que define el nodo.
- **Poda.** Esta se da cuando se reduce el tamaño de los árboles de decisión eliminando nodos opuestos a la división.
- **Rama o subárbol.** Una subsección del árbol de decisión se denomina rama o subárbol.
- **Etiqueta.** Permite la unión de nodos y flechas que denominan las acciones que se llevan a cabo.

2.2. Descripción del árbol de decisión

Un árbol de decisión en *Machine Learning* es una estructura de árbol similar a un diagrama de flujo, ver Figura 1, donde un nodo interno representa una característica (o atributo), la rama es una regla de decisión y cada nodo u hoja simboliza el resultado y el nodo superior se conoce como el nodo raíz. Este tipo de árbol se conoce como árbol de clasificación, donde cada ramificación contiene un conjunto de atributos o reglas de clasificación asociadas a una etiqueta de clase específica que se halla al final de la ramificación.

El árbol de decisión clasificatorio aprende a particionar en función del valor del atributo, éste se divide de manera recursiva buscando un mínimo local de acuerdo a medidas como la entropía en base a un conjunto de datos masivo de entrenamiento.

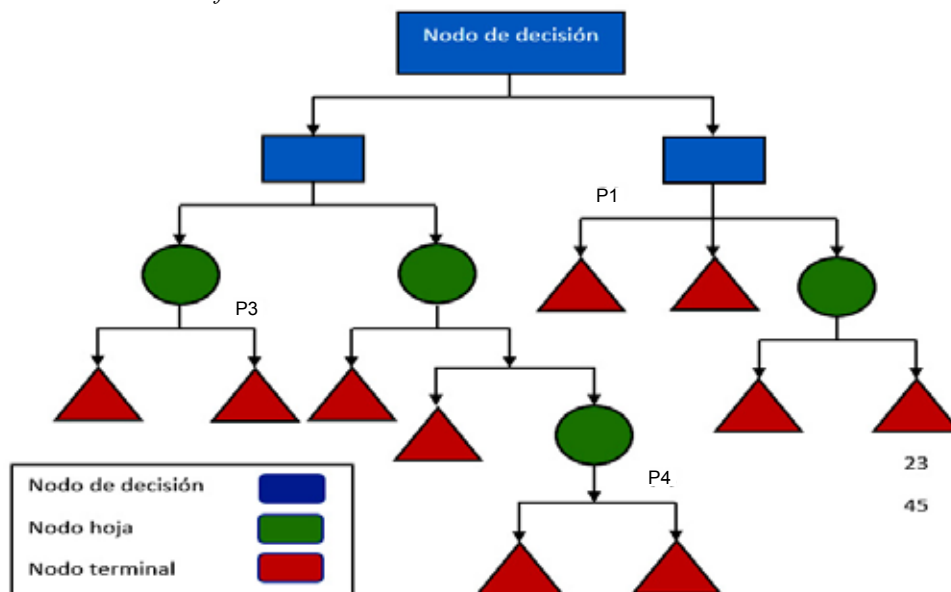
Las ventajas de utilidad de los árboles de decisión en el aprendizaje automático son:

- El costo del uso del árbol para predecir los datos disminuye con cada punto de datos adicional.
- Funciona para los datos numéricos o categóricos.
- Puede modelar problemas con múltiples resultados.
- Usa un modelo de caja blanca (lo que hace que los resultados sean fáciles de explicar).
- La fiabilidad de un árbol se puede cuantificar y poner a prueba.
- Tiende a ser preciso independientemente de si viola las suposiciones de los datos de origen.

Las desventajas que presenta son:

- La incorporación de datos categóricos con múltiples niveles provoca que los resultados se inclinen a favor de los atributos con mayoría de niveles.
- Los cálculos pueden volverse complejos al enfrentarse con la falta de certezas y numerosos resultados relacionados.

Figura 1.
Estructura de un árbol de clasificación



Fuente: Elaboración propia

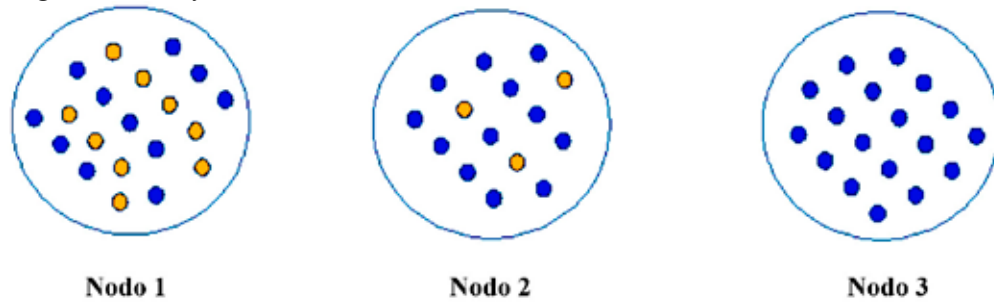
2.3. Medidas de selección

La medida de selección de atributos o características es una heurística para seleccionar el criterio que divida los datos de la mejor manera posible. También se conoce como reglas de división porque ayudan a determinar puntos de interrupción en un nodo dado. Los utilizados, son los siguientes:

Ganancia de información

La ganancia de información, es una propiedad estadística que mide homogeneidad, qué tan bien un atributo se asemeja en el entrenamiento de patrones a otro atributo según la clasificación que se busca. De acuerdo a la Figura 2, se puede observar que el nodo 3 describe muy fácilmente la información pues contiene datos con valores similares (llamado nodo puro), mientras que el nodo 2 todavía requiere regular información de entrenamiento para alcanzar la descripción del nodo 3. A su vez el nodo 1 necesitará mayor información para llegar a los anteriores, a este último se le denomina nodo impuro.

Figura 2.
Nodos con ganancia de información



Fuente: Elaboración propia.

La expresión para determinar la ganancia de información por nodo, es:

$$Ganancia(p, X) = Entropía - \sum_{i=1}^k (P_i \times Entropía)$$

Donde P_j es el conjunto de valores posibles para el atributo X , donde el valor de entropía se define luego.

Entropía

La entropía es una medida de la homogeneidad de los datos aleatorios presentes en un nodo de un árbol de clasificación. Si los datos son similares, el valor de entropía es pequeño o nulo.

De cada nodo se puede medir el valor de entropía para su respectivo control. La expresión siguiente mide entropía:

$$Entropía = - \sum_{i=1}^k P_i \log_2 P_i$$

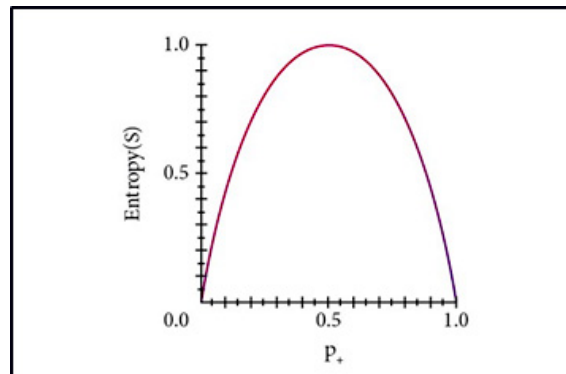
Por ejemplo, si se calcula la entropía de los nodos 1 y 3, se tiene:

$$\begin{aligned} Entropía_1 &= -(11/20)\log_2(11/20) - (9/20)\log_2(9/20) = 0,993 \\ Entropía_3 &= -(20/20)\log_2(20/20) - (0/20)\log_2(0/20) = 0 \end{aligned}$$

Se concluye que la entropía es cero si todos los elementos del grupo pertenecen a la misma clase, y la entropía es uno cuando la muestra contiene el mismo número de observaciones positivas y negativas. Si la muestra contiene un número desigual de observaciones positiva y

negativas, la entropía está entre 0 y 1. La Figura 3, presenta la forma de la función de entropía en relación con una clasificación booleana.

Figura 3.
Función entropía



Fuente: Elaboración propia

Gini

Gini es una medida de impureza de las observaciones contenidas en un determinado nodo. Si el índice de Gini vale 0, significa que el nodo es totalmente puro. Se determina de la siguiente manera:

$$Gini = \sum_{i=1}^k P_i(1 - P_i)$$

Si se determina el índice de Gini para los nodos 1 y 3 anteriores, se tiene:

$$Gini_1 = (11/20)(9/20) + (9/20)(11/20) = 0,495$$

$$Gini_3 = (20/20)(0/20) + (0/20)(20/20) = 0$$

Luego el nodo 3 es puro y el nodo 1 es impuro.

3. Resultados

3.1. Aplicación de un árbol de clasificación

Para el entrenamiento del presente árbol de clasificación, se han empleado las siguientes variables

Variable dependiente categórica:

Sexo (masculino - femenino)

Variables independientes:

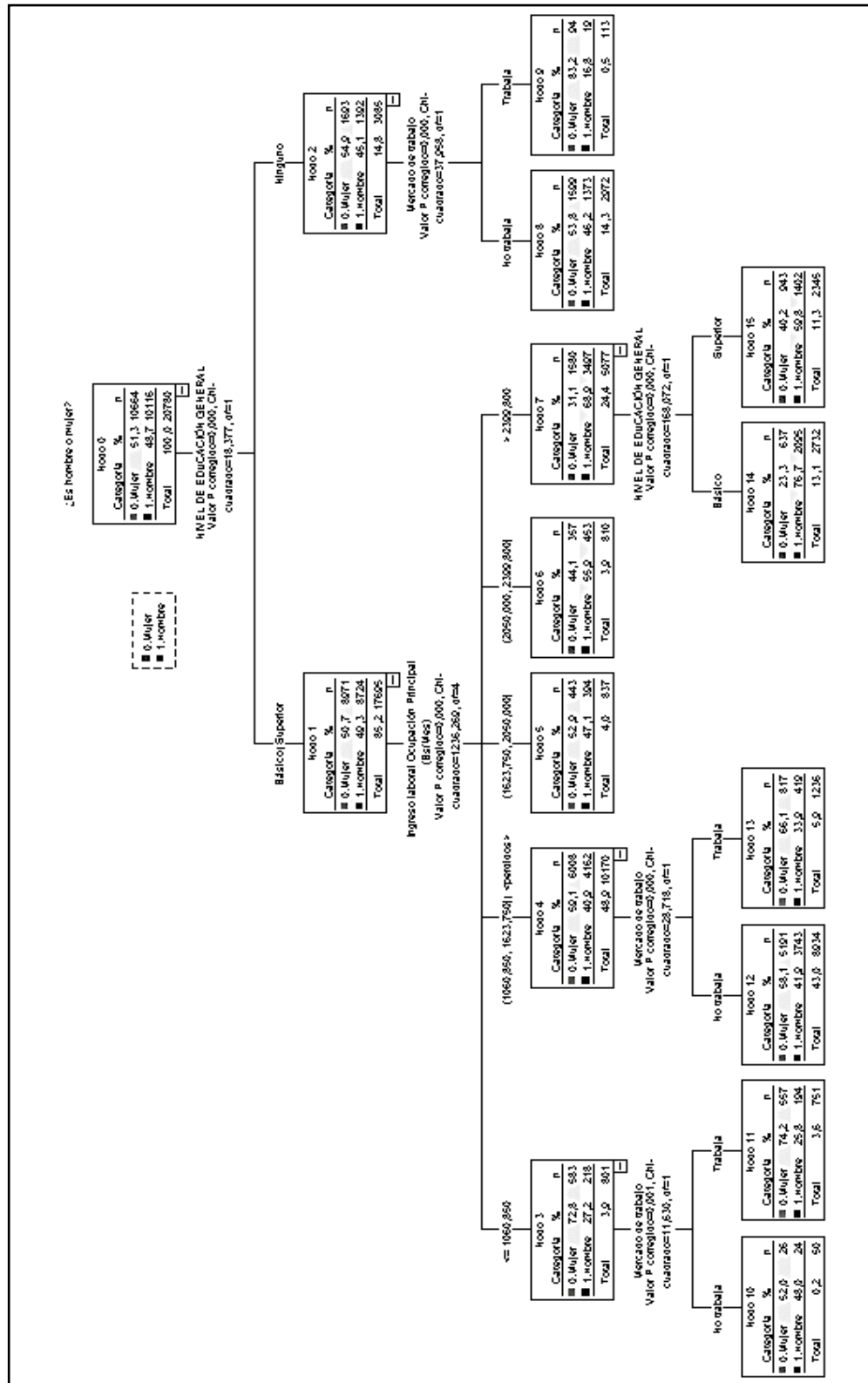
Nivel educativo(ninguno, básico y superior)

Si la persona trabaja o no

Ingreso mensual en Bs.

La información estadística corresponde a la encuesta de hogares del año 2018 de personas que habitan las ciudades capitales y El Alto. La muestra alcanza un total de 20.780 registros disponibles en la base de datos inferida a la población total. El resultado del árbol de clasificación se presenta en la Figura 4.

Figura 4.
Arbol de clasificación por sexo



Fuente: Elaboración propia - Programa SPSS

Haciendo un análisis del árbol de clasificación de la Figura 4, se tiene que: el nodo raíz, indica que el 51,3% de la población es mujer y el 48,7% es hombre con un total de 20.780 registros. La población se divide en dos grupos por el nivel educativo y se puede ver que el 50,7% de mujeres y 49,3% de hombres tienen un nivel educativo entre básico y superior. Por otro lado, se observa que el 54,9% de mujeres y el 45,1% de hombres no tienen nivel educativo, sin embargo, esta última categoría, es de solo el 14,8% en el global. Pasando a la tercera fila del árbol de clasificación, de los que tienen un nivel educativo básico y superior y cuentan con un ingreso mensual en bolivianos menor a 1.061 Bs, el 72,8% son mujeres y el 27,2% son hombres, pero apenas representan el 3,9% del total de la población. De los que tienen ingreso entre 1.061 y 1.624 Bs., el 59,1% son mujeres y el 40,9% son hombres de un total de 48,9% poblacional. Asimismo, de los que tienen ingreso mensual entre 1.624 y 2.050 Bs., el 52,9% son mujeres y el 47,1% son hombres de un total poblacional de apenas 4%.

El nodo 6 de la misma fila indica que de los que tienen un nivel educativo básico y superior y tienen ingreso mensual en bolivianos entre 2.050 y 2.400 Bs., aproximadamente, el 44,1% son mujeres y el 55,9% son hombres, representando solo el 3,9% del total de la población. De los que tienen ingresos por encima de los 2.400 Bs. y que representa el 24,4% de la población, el 31,1% son mujeres y el 68,9% son hombres.

Si se observa la parte derecha del árbol de clasificación en el nodo 8 con el 14,3% del total de la población de los que no trabajan, el 53,8% son mujeres y el resto hombres.

El nodo 12 con un 43% de la población, indica que de entre los que tienen un ingreso entre 1.061 y 1.624 Bs. mensual, no trabajan y de ellos el 58,1% son mujeres y el 41,9% son hombres. El nodo 14 describe, con una población del 13,1% son los que tienen un ingreso mensual por encima de los 2.400 Bs. y tienen sorpresivamente un nivel educativo básico, de ellos, el 23,3% son mujeres y el 76,7% son hombres. Lo mismo pasa en el nodo 15, que con una población de 11,3% son los que tienen un ingreso mensual por encima de los 2.400 Bs. y cuentan con un nivel educativo superior, de ellos, el 40,2% son mujeres y el 59,8% son hombres.

4. Discusión

Es interesante contar con un método de clasificación como el de árboles de decisión multivariante, que incluye en el análisis, muchas observaciones o datos masivos, y que va entrenando el algoritmo hasta lograr una confiable homogeneidad de observaciones en diferentes grupos excluyentes, dando lugar a una interpretación de los datos y análisis de forma sencilla. Si se compara con el método de análisis multivariante de conglomerados no jerárquico, siempre existirá en la clasificación, observaciones que pueden pertenecer a otros grupos o conglomerados de acuerdo al manejo de distancia entre objetos u observaciones. Luego, de acuerdo a lo discutido, será muy importante realizar un análisis y comparación de clasificación de observaciones mediante otras técnicas multivariantes exploratorias con el que arroja el método de árbol de decisión.

5. Conclusión

Existen otros métodos de clasificación de observaciones multivariantes que emplea la ciencia de los datos dentro del *Machine Learning* o la Minería de Datos, sin embargo, incluido este último método como es el de árbol de clasificación, no siempre son perfectos y deben ser utilizados de acuerdo a la información disponible, además, requieren siempre del cuidado de la inclusión de la información masiva que por manejar demasiada información, puede contener datos incoherentes o no válidos que

distorsionarían el análisis. Lo mismo ocurriría con la inclusión en demasía de variables y datos que pueden ser atípicos y/o faltantes.

En cuanto a los resultados obtenidos del árbol de decisión, se observa que el género mujer está en desventaja en relación a los hombres en cuanto a su nivel de formación educativa influyente en la ocupación y el ingreso.

Referencias Bibliográficas

- Bouza, C., Santiago A. *Classification and Regresssion Trees*. Universidad de La Habana, Cuba y Universidad Autónoma de Guerrero, México.
- Breiman, L., Friedman, J.H., Olsen, R. A. & Stone, C. J. (1984). *Classification and Regresssion Trees*. Wadsworth, N. York.
- Cox, D. R., Snell, E. J. (1989). *The Analysis of Binary Data*. Chapman and Hall. London.
- De la Fuente, F. S. (2011). *Análisis de conglomerados*. Universidad Autónoma de Madrid (UAM).
- Díaz, L. (2012). *Análisis Estadístico de Datos Multivariantes*. Universidad Nacional de Colombia.
- Francois, H. (2013). *Análisis de Datos con R. Colombia*. Escuela Colombiana de Ingeniería.
- Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W.C. (1999). *Análisis Multivariante*. Madrid, España. Editorial Prentice Hall.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley Interscience. New York.
- Aldás, J. & Uriel, E. (2017). *Análisis Multivariante Aplicado con R*. Alfa Centauro.
- Johnson, R. A. *Applied Multivariate Statistical Analysis*. Prentice Hall. New Jersey.
- López, A. (2012). *Análisis Multivariante para la Inteligencia de Mercados*. Tecnológico de Monterrey.
- Ruiz, M. (2013). *Árboles de Decisión y ELECTRA I*. Biblioteca Universitaria.
- Mitchell, T. M. (1997). *Machine Learning*. Texto: McGraw-Hill, Cap. III. Web: <http://www.cs.cmu.edu/~tom/mlbook.html>
- Muller, A. & Guido, S. (2016). *Introduction Machine Learning with Python*. Jupyter
- Loyola, O. (2012). *Inducción de Árboles de Decisión*. Academia Española.
- Theobald, O. (2018). *Machine Learning for Absolute Beginners*. Andriy Burkov
- Peña, D. (2002). *Análisis de Datos Multivariantes*. McGraw Hill.
- Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos*. Pearson Educación S.A., España.
- Poole, D. (1998). *AI Computational Intelligence. A Logical Approach*. Oxford University Press.
- Uriel, E. & Aldás, J. (2005). *Análisis Multivariante Aplicado*. Madrid, España. Thomson.

Estudio de la transmisión de COVID-19 en Bolivia a través del proceso de Galton Watson

Study of the transmission of COVID-19 in Bolivia through the Galton Watson process

Dindo Valdez Blanco¹

Instituto de Estadística Teórica y Aplicada, Universidad Mayor de San Andrés, La Paz-Bolivia

✉ dvaldez@fcpn.edu.bo

Artículo recibido: 2022-02-22

Artículo aceptado: 2022-03-22

Resumen

En la actualidad la pandemia del COVID-19 ha sido tema de estudio desde muchos enfoques, uno de ellos es el análisis de la evolución del virus y los casos nuevos reportados, específicamente para la determinación del promedio de la transmisión de la enfermedad y la probabilidad de extinción del mismo, la metodología del presente trabajo analiza el comportamiento del virus en Bolivia desde su aparición en el mes de abril 2020 hasta el 30 de enero del 2022, como un proceso estocástico de Galton-Watson para determinar el promedio de reproducción en cada mes y la probabilidad de extinción del COVID-19 para cada mes, de manera que se obtenga algún patrón o comportamiento sistemático que sea de utilidad para el monitoreo del virus.

Palabras clave: Procesos de ramificación, promedio de reproducción, probabilidad de extinción, procesos estocásticos

Abstract

Currently, the COVID-19 pandemic has been the subject of study from many approaches, one of them is the analysis of the evolution of the virus and the new cases reported, specifically for the determination of the average transmission of the disease and the probability extinction of the same, the methodology of this work analyzes the behavior of the virus in Bolivia from its appearance in the month of April 2020 to January 30, 2022 as a Galton-Watson stochastic process to determine the average reproduction in each month and the probability of extinction of COVID-19 for each month, so that some pattern or systematic behavior is obtained that is useful for monitoring the virus.

Keywords: Branching processes, reproduction rate, extinction probability, stochastic processes

1. Introducción

Los procesos de ramificación son una clase de procesos estocásticos que modelan el crecimiento de una población. Se utilizan ampliamente en biología y epidemiología para estudiar la propagación de enfermedades infecciosas y epidemias. En 1873, Sir Francois Galton se plantea una pregunta, respecto a los apellidos paternos ¿qué proporción de sus apellidos paternos se habrán extinguido después de r generaciones?. Su motivación original era estudiar la extinción de los apellidos familiares, un tema que preocupaba a la aristocracia victoriana en el siglo XIX. El reverendo Henry William Watson

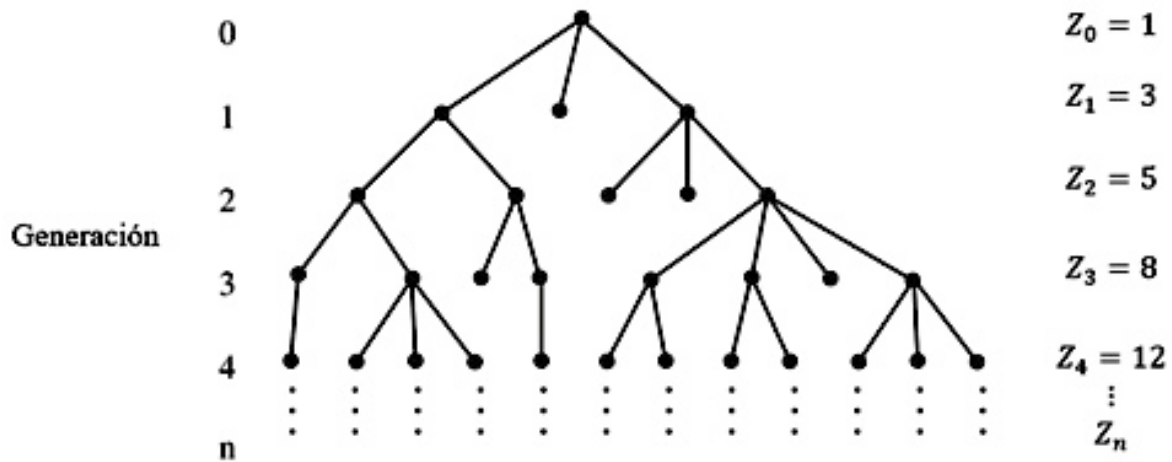
¹ Maestría en Ciencias Estadísticas, Licenciado en Estadística. Profesor de Estadística - Universidad Mayor de San Andrés.
ORCID: 0000-0003-0704-0980

respondió con una solución. El estudio de la ramificación. Los procesos surgieron de la colaboración de Watson y Galton. Sus resultados fueron descubiertos independientemente por el estadístico francés Irénée-Jules Bienaymé. Por tal razón el modelo de proceso de ramificación a veces se denomina proceso Bienaymé-Galton-Watson.

La idea general de los procesos de ramificación es la siguiente: en un instante de tiempo se identifica un conjunto de individuos que forman la generación 0, esta generación produce descendientes, cuya colección forma la generación 1. Los descendientes de la generación 1 constituyen la generación 2, y así sucesivamente. Estos procesos modelan la evolución en el tiempo de procesos con estructuras tipo árbol.

De esta manera, si se define el proceso estocástico $\{Z_n\}$ como el número de individuos en la n -ésima generación, entonces la Figura 1 muestra la evolución del proceso de ramificación de Galton-Watson con $Z_0=1$ (Athreya, 1972).

Figura 1.
Esquema de la evolución del proceso de ramificación de Galton-Watson



Fuente: Athreya, 1972

La pandemia del COVID-19 en la actualidad requiere de su estudio y modelación desde muchos aspectos, uno de ellos es el análisis de la propagación del virus, por ello el objetivo del presente trabajo es la aplicación del proceso de Galton-Watson para estimar el promedio de la transmisión de la enfermedad y la probabilidad de extinción del mismo en Bolivia.

2. Metodología

La metodología del trabajo se basa en el procedimiento de un estudio de propagación del COVID-19 en Bolivia a partir del proceso de Galton-Watson, tomando en cuenta los casos reportados desde abril de 2020 hasta enero de 2022, por tanto, es un método de estudio prospectivo. Los datos para el estudio corresponden a los reportes diarios de casos confirmados de COVID-19 en Bolivia realizados por el Ministerio de Salud y que se encuentran disponibles en el repositorio de datos COVID-19, del Centro de Ciencias e Ingeniería de Sistemas (Johns Hopkins University, 2021) situada en Baltimore, Estados Unidos.

3. El proceso de Galton-Watson

Sea $\{Z_n; n \geq 0\}$ una secuencia de variables aleatorias definidas como:

$$Z_n = \sum_{k=1}^{Z_{n-1}} X_{n,k} \quad ; \quad n \geq 1 \quad (1)$$

Donde $\{X_{n,k}; n, k \geq 1\}$ es una familia de variables aleatorias independientes e idénticamente distribuidas con una función de probabilidad $\{p_n; n \geq 0\}$ de tal forma que es independiente de Z_0 . Para cada $n \geq 0$, $X_{n,k}$ representa el número de hijos del i -ésimo individuo en la n -ésima generación, la variable aleatoria Z_n representa el tamaño de la n -ésima generación de una población. Entonces $\{Z_n; n \geq 0\}$ es conocido como el proceso de Galton-Watson.

Por la definición anterior, la distribución de Z_n queda completamente determinada por dos parámetros de entrada, la distribución p_n y Z_0 . De tal forma que $\{Z_n; n \geq 0\}$ constituye una cadena de Markov homogénea con matriz de transición $[p_{ij}]$; $i, j \geq 0$, dado por:

$$p_{ij} = P[Z_{n+1} = j \mid Z_n = i] = p_j^{*i}$$

Donde p_j^{*i} denota la i -ésima convolución de p_n .

Si se define el evento de extinción de la población como:

$$\text{Extinción} = \{\exists n \in \mathbb{N} \text{ tal que } Z_n = 0\} = \bigcup_{n \in \mathbb{N}} \{Z_n = 0\}$$

Entonces hallar la probabilidad de extinción es:

$$\eta = P(\text{Extinción}) = P(Z_n = 0) \quad ; \quad \text{para algún } n \in \mathbb{N}$$

De tal forma que:

Si $P(X = 0) = 0$ entonces $\eta = 0$.

Si $P(X=0) > 0$ entonces $\eta > 0 \rightarrow \eta \geq P(X=0) > 0$. Indica que el individuo original tiene 0 hijos.

Watson demostró en su artículo original con Galton que, si $P(X = 0) > 0$ entonces de hecho $\eta = 1$ para cualquier X . Sin embargo, su documento original tenía un error en vista que es posible tener $\eta \in (0, 1)$, dependiendo del valor de $\mu = E[X]$. Se tienen tres casos posibles:

Si $\mu < 1 \rightarrow \eta = 1$, entonces la población siempre se extingue.

Si $\mu > 1 \rightarrow \eta < 1$, entonces la población tiene la probabilidad de sobrevivir.

$$\text{Si } \mu = 1 \rightarrow \begin{cases} \eta = 1 & \text{si } P(X = 1) < 1 \\ \eta = 0 & \text{si } P(X = 1) = 1 \end{cases}$$

Según lo anterior, se puede clasificar a los procesos de Galton-Watson en 3 tipos:

Tipo sub crítico, si $\mu < 1$.

Tipo crítico, si $\mu = 1$.

Tipo supercrítico, si $\mu > 1$.

4. Tamaño medio de la n-ésima generación

La ecuación (1) se representa como una suma de variables aleatorias i.i.d (Athreya, 1972). Estos resultados pueden ser utilizados para hallar los momentos de la variable aleatoria Z_n .

Sea $\mu = \sum_{k=0}^{\infty} k p_k$ la media de la distribución de la descendencia, para encontrar el tamaño medio de la n-ésima generación $E[Z_n]$ condicionado en Z_{n-1} , por la ley de la esperanza total

$$\begin{aligned} E[Z_n] &= \sum_{k=0}^{\infty} E[Z_n | Z_{n-1} = k] P[Z_{n-1} = k] = \sum_{k=0}^{\infty} E \left[\sum_{i=1}^{Z_{n-1}} X_{n,i} | Z_{n-1} = k \right] P[Z_{n-1} = k] \\ &= \sum_{k=0}^{\infty} E \left[\sum_{i=1}^k X_{n,i} | Z_{n-1} = k \right] P[Z_{n-1} = k] = \sum_{k=0}^{\infty} E \left[\sum_{i=1}^k X_{n,i} \right] P[Z_{n-1} = k] \\ &= \sum_{k=0}^{\infty} k \mu P[Z_{n-1} = k] = \mu E[Z_{n-1}] \end{aligned}$$

Luego aplicando la fórmula de recurrencia, se obtiene:

$$E[Z_n] = \mu E[Z_{n-1}] = \mu^2 E[Z_{n-2}] = \dots = \mu^n ; n \geq 0, Z_0 = 1 \quad (2)$$

De esta manera se obtienen los tres casos mencionados anteriormente

$$\lim_{n \rightarrow \infty} E[Z_n] = \lim_{n \rightarrow \infty} \mu^n = \begin{cases} 0 & ; \text{ si } \mu < 1 \\ 1 & ; \text{ si } \mu = 1 \\ \infty & ; \text{ si } \mu > 1 \end{cases} \quad (3)$$

El proceso de ramificación es sub crítico si $\mu < 1$, critico si $\mu = 1$ y supercrítico si $\mu > 1$.

En el caso que $Z_0 \neq 1$, el número medio de reproducción de la generación n se calcula por la fórmula:

$$E[Z_n] = Z_0 \mu^n ; n \geq 0$$

5. Varianza del tamaño de la n-ésima generación

Para explorar el proceso de extinción en los casos crítico y supercrítico ($\mu \geq 1$), se considera la varianza del tamaño de la enésima generación $V[Z_n]$. Sea σ^2 la varianza de la distribución de la descendencia $X_{n,k}$. Por el teorema de la varianza total,

$$V[Z_n] = V[E(Z_n | Z_{n-1})] + E[V(Z_n | Z_{n-1})]$$

Como ya se demostró en 2,

$$E(Z_n | Z_{n-1} = k) = E\left(\sum_{i=1}^k X_{n,i}\right) = \sum_{i=1}^k E[X_{n,i}] = k$$

Entonces

$$E(Z_n | Z_{n-1}) = \mu * Z_{n-1}$$

De manera similar

$$V(Z_n | Z_{n-1} = k) = V\left(\sum_{i=1}^k X_{n,i}\right) = \sum_{i=1}^k V[X_{n,i}] = k \sigma^2$$

De esta manera se tiene

$$V(Z_n | Z_{n-1}) = \sigma^2 Z_{n-1}$$

Bajo la suposición de independencia de los $X_{n,i}$ y aplicando el teorema de la varianza total (Ospina, 2021), se tiene

$$V[Z_n] = V[\mu Z_{n-1}] + E[\sigma^2 Z_{n-1}] = \mu^2 V[Z_{n-1}] + \sigma^2 \mu^{n-1}; n \geq 1$$

Como $V[Z_0]=0$, la ecuación anterior resulta:

$$V[Z_1] = \mu^2 V[Z_0] + \sigma^2 = \sigma^2$$

$$V[Z_2] = \mu^2 V[Z_1] + \sigma^2 \mu = \sigma^2 \mu (1 + \mu)$$

$$V[Z_3] = \mu^2 V[Z_2] + \sigma^2 \mu^2 = \sigma^2 \mu^2 (1 + \mu + \mu^2)$$

De manera general el patrón anterior provee el resultado:

$$V[Z_n] = \sigma^2 * \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} n\sigma^2 & ; \text{ si } \mu = 1 \\ \frac{\sigma^2 \mu^{n-1} (\mu^n - 1)}{\mu - 1} & ; \text{ si } \mu \neq 1 \end{cases} \quad (4)$$

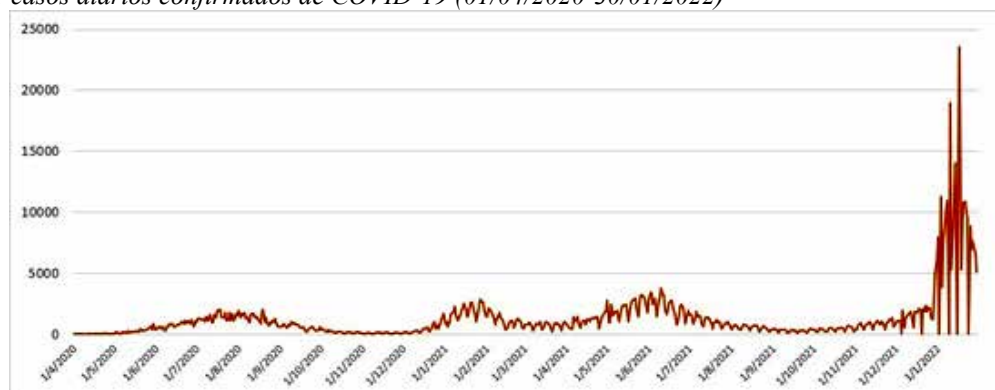
En el caso subcrítico ($\mu < 1$), tanto la media como la varianza del tamaño de la generación tienden a 0. En el caso crítico ($\mu = 1$), el tamaño medio de cada generación es uno, pero la varianza es una función de crecimiento lineal de n.

En el caso supercrítico ($\mu > 1$), la varianza crece exponencialmente. La gran diferencia entre la media μ^n y la varianza sugiere que en algunos casos tanto la extinción y el crecimiento ilimitado son resultados posibles.

6. Análisis de la reproducción y la probabilidad de extinción del COVID-19 en Bolivia

Figura 2.

Bolivia: casos diarios confirmados de COVID-19 (01/04/2020-30/01/2022)



Fuente: Elaboración propia en base a datos del Centro de Ciencias e Ingeniería de Sistemas de la Universidad Johns Hopkins

Aplicando el modelo de Galton-Watson con una distribución de Poisson se estiman el promedio de reproducción y la probabilidad de extinción para cada mes, los resultados se muestran en la Tabla 1.

Tabla 1:

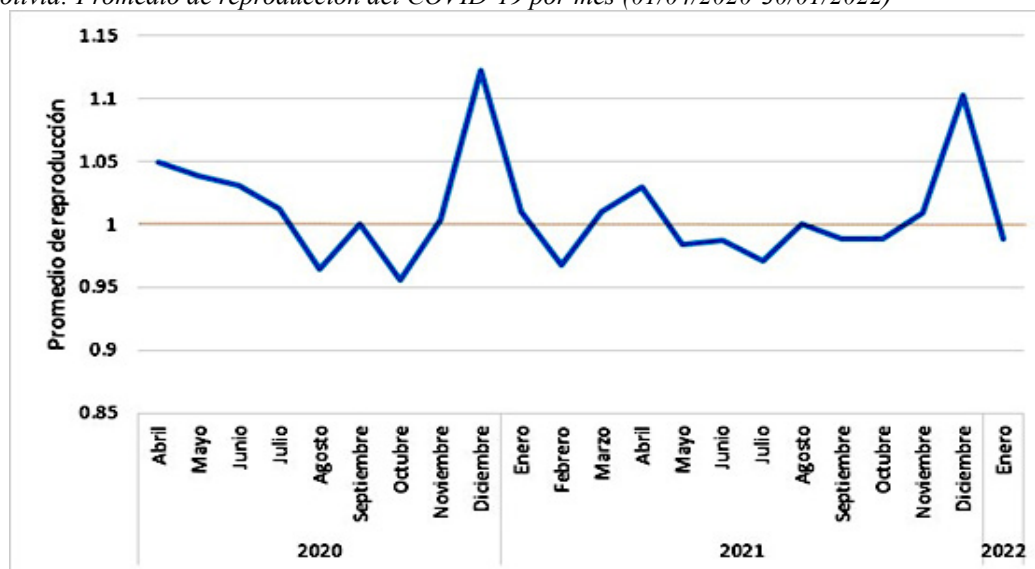
Bolivia: Promedio estimado de reproducción del COVID-19 y probabilidad estimada de extinción del virus por mes

Año	Mes	Promedio	Probabilidad de extinción
2020	Abril	1.0492	0.9077
	Mayo	1.0387	0.9264
	Junio	1.0312	0.9401
	Julio	1.0130	0.9744
	Agosto	0.9649	1.0000
	Septiembre	1.0001	0.9998
	Octubre	0.9559	1.0000
	Noviembre	1.0035	0.9930
2021	Diciembre	1.1226	0.7898
	Enero	1.0104	0.9795
	Febrero	0.9684	1.0000
	Marzo	1.0105	0.9793
	Abril	1.0297	0.9429
	Mayo	0.9845	1.0000
	Junio	0.9880	1.0000
	Julio	0.9708	1.0000
	Agosto	1.0011	0.9978
	Septiembre	0.9884	1.0000
	Octubre	0.9884	1.0000
	Noviembre	1.0098	0.9807
2022	Diciembre	1.1030	0.8192
	Enero	0.9881	1.0000

La Figura 3 muestra la variación de la tasa promedio de reproducción durante cada mes entre abril de 2020 y enero de 2022, se puede apreciar que en el 50% de los meses el promedio de reproducción se sitúa por encima de 1, y el mes con más incidencia es diciembre de 2020 y 2021, dando una pauta del comportamiento estacional de la enfermedad. De acuerdo a la ecuación 2, esto indica que la enfermedad está a punto de extinguirse en esos meses.

Figura 3.

Bolivia: Promedio de reproducción del COVID-19 por mes (01/04/2020-30/01/2022)

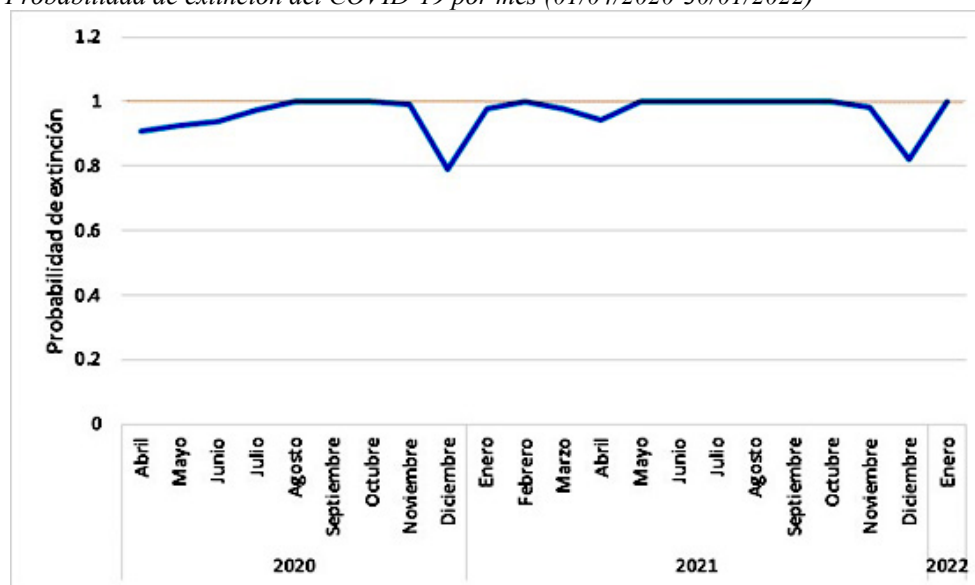


Fuente: Elaboración propia en base a datos del Centro de Ciencias e Ingeniería de Sistemas de la universidad Johns Hopkins

La Figura 3 muestra que el promedio de reproducción es mayor los meses de diciembre, indicando un patrón cíclico cada año. La Figura 4 muestra que la probabilidad de extinción del virus ha estado por encima de 1 en la mayoría de los meses, de acuerdo a la ecuación 2, esto indica que la enfermedad está a punto de extinguirse en esos meses.

Figura 4.

Bolivia: Probabilidad de extinción del COVID-19 por mes (01/04/2020-30/01/2022)



Fuente: Elaboración propia en base a datos del Centro de Ciencias e Ingeniería de Sistemas de la universidad Johns Hopkins

La Figura 4 confirma que la propagación del virus COVID-19 tiene un efecto estacional y que durante el mes de diciembre de las gestiones 2020 y 2021 se tienen la menor probabilidad de extinción del virus.

7. Discusión

El análisis de la propagación del virus en Bolivia a partir del promedio de reproducción y la probabilidad de extinción durante cada mes a partir de abril del 2020 a enero de 2022. El presente trabajo sugiere un efecto estacional de la propagación del virus durante el mes de diciembre de cada año, sin embargo, aún se necesita analizar el comportamiento en los meses posteriores.

El trabajo realizado por Crespo (2020) es un buen punto de partida de tales estudios, ante la carencia de datos se basa en simulaciones. Es importante contrastar estos resultados con otros países para tener una mayor comprensión del fenómeno de propagación del virus.

8. Conclusión

Los resultados obtenidos indican que el proceso de Galton-Watson puede ser utilizado en apoyo a la toma de decisiones y la comprensión de la evolución en la transmisión del COVID-19. La estimación del promedio de reproducción y de su varianza se realizó por máxima verosimilitud. En Bolivia, los resultados muestran que el promedio de reproducción mensual del virus COVID-19 varía entre 0.95 y 1.12, este aspecto revela que el país no está tan cerca de mitigar el virus. El comportamiento del proceso de Galton Watson se fundamenta en la estimación del promedio de reproducción μ , el mismo puede ser estimado a partir de una distribución de Poisson.

Los meses con mayor promedio de reproducción ha sido diciembre del 2020 y diciembre del 2021, aspecto que refleja la necesidad de tomar acciones para aminorar la propagación del virus durante cada fin de año.

Se recomienda el estudio del proceso de Galton-Watson con estimadores no paramétricos. La distribución de probabilidad supuesta para la aplicación es la ley de Poisson, se recomienda analizar el uso de otras distribuciones de probabilidad como la distribución geométrica.

Referencias Bibliográficas

- Athreya, P. (1972). *Branching Processes*, New York: Springer Verlag.
- Abbott, S. (2020). The transmissibility of novel Coronavirus in the early stages of the 2019-20 outbreak in Wuhan: Exploring initial point-source exposure sizes and durations using scenario analysis [version 1]. *Wellcome Open Research* 2020, 5(17), 1-12. <https://wellcomeopenresearch.org/articles/5-17>
- Blanco, L. (1996). ¿Qué es un Proceso de Ramificación? *Boletín de Matemáticas*, Nueva Serie, Santa fe de Bogotá, III, 43-50. https://www.academia.edu/33989688/Proceso_de_Galton_Watson?auto=download
- Boldog, P., Tekeli T., Vizi Z., Dénes A., Bartha F. A. y Rost G. (2020) *Risk assessment of novel*

- coronavirus COVID-19 outbreaks outside China. Journal of Clinical Medicine.* 9(2), 15-37. <https://www.mdpi.com/2077-0383/9/2/571>
- Johns Hopkins University (diciembre, 2021). *Coronavirus Resource Center: COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.* <https://coronavirus.jhu.edu/map.html>
- Crespo, M. De la Cruz, P. Quevedo, E. Valdés, J (2020). Los Procesos de Ramificación y su Aplicación al Estudio de la Transmisión de la COVID-19 en Cuba. *Revista de Ciencias Matemáticas*, 34(1), 89-97. <http://www.revinfodir.sld.cu/index.php/infodir/article/download/1042/1244>
- Guttorp, P. y Perlman, M. D. (2022). Predicting Extinction or Explosion in a Galton-Watson Branching Process with Power Series Offspring Distribution. University of Washinton. Technical Report No.626. <https://stat.uw.edu/research/tech-reports/predicting-extinction-or-explosion-galton-watson-branching-process-power-series-offspring>
- Harris, T.E. (1963). *The Theory of Branching processes*. Springer Verlag, Berlin.
- Heyde, C. C.(1970). Extension of a result of Seneta for the Super Critical. Galton-Watson Process. *The Annals of Mathematical Statistics*. Vol. 41, No.2, 739-742. <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-41/issue-2/Extension-of-a-Result-of-Seneta-for-the-Super-Critical/10.1214/aoms/1177697127.full>
- Imai, N., Cori, I. A., Dorigatti, M. Baguelin, C. A. Donnelly, S. Riley y N. M. Ferguson (2020) Report 3: Transmissibility of COVID-19. Imperial College London COVID-19 Response Team. <https://www.jstor.org/stable/2239385>
- Ospina, R.(2001). Proceso de Galton Watson. *Revista Colombiana de Estadística*, 24(1), 13-26. file:///C:/Users/xiome/Downloads/Proceso_de_Galton_Watson.pdf
- Seneta, E. (1970). An explicit-limit theorem for the critical Galton-Watson process with immigration. *Journal Royal Statistics Society. Series B* (32), 149-152. <https://www.jstor.org/stable/2984413>
- Scott, D. (1987). On posterior asymptotic normality and asymptotic normality of estimators for the Galton-Watson process. *Journal Royal Statistics Society, Series B* (49), 209-214. <https://zh.booksc.eu/book/61592903/7f6ded>
- Stigler, S. M. (1971). The estimation of the probability of extinction and other parameters associated with branching processes. *Biometrika*, 58, 499-508. <https://www.jstor.org/stable/2334385>
- Waugh, W.A. O’N. (1958). Conditioned Markov processes. *Biometrika* 45, 241-249. <https://academic.oup.com/biomet/article-abstract/45/1-2/241/264700>

Análisis factorial del grado de satisfacción de los estudiantes de la Carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

Factorial analysis of the degree of satisfaction of the students of the Systems Engineering Career of the UPEA with respect to the virtual class modality

Carmen Vega Flores¹

Carrera de Ingeniería de Sistemas, Universidad Pública de El Alto, El Alto-Bolivia

✉ vegaflorescarmen@gmail.com

Artículo recibido: 2022-03-13

Artículo aceptado: 2022-03-31

Resumen

El presente artículo tiene por objetivo identificar los factores asociados al grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA en su proceso de aprendizaje en la modalidad de clase virtual, a consecuencia de la suspensión de las clases presenciales producida por el virus Covid-19.

El estudio fue descriptivo con diseño no experimental de corte transversal con enfoque cuantitativo, los resultados obtenidos muestran a 32 variables (ítems) resumidas en 4 factores calculados a partir de la técnica multivariante del análisis factorial, se determina que la satisfacción del estudiante a la modalidad de clase virtual depende en primer lugar a las habilidades y competencias del docente con una contribución del 49.973% (1°factor), en segundo lugar a los recursos de aprendizaje virtual con una participación del 6.034% (2° factor) y en tercer lugar de la calidad del servicio de internet con un aporte del 4.686% (3°factor). En cuanto a la significancia de la dependencia del grado de satisfacción de la enseñanza virtual con relación al método de enseñanza tradicional la verificación es a través de una prueba Chi Cuadrado la significancia es del 0.000 lo que confirma la relación que existe entre ambas variables, seguidamente la prueba Spearman indica un 88.77% el mismo que es significativo con un 95% de relación positiva fuerte entre las variables la satisfacción del estudiante y la modalidad de clase virtual, finalmente al realizar la verificación de hipótesis se confirma que la técnica del análisis factorial es apropiada y admisible para la aplicación de los datos, esto es: gracias al coeficiente KMO de 0.969 (≥ 0.7) como así también el test de esfericidad de Bartlett con una significancia del test de esfericidad es 0.000 (≤ 0.05).

Palabras clave: Coeficiente KMO, esfericidad de Bartlett, factores, enseñanza tradicional.

Abstract

The objective of this article is to identify the factors associated with the degree of satisfaction of the students of the Systems Engineering career at UPEA in their learning process in the virtual class modality, as a result of the suspension of face-to-face classes produced by the Covid-19 virus.

¹ Maestría en Investigación Científica, Licenciada en Estadística. Docente de Estadística – Universidad Pública de El Alto, Carrera de Ingeniería de Sistemas. ORCID: 0000-0003-1430-2854

The study was descriptive with a non-experimental cross-sectional design with a quantitative approach, the results obtained show 32 variables (items) summarized in 4 factors calculated from the multivariate technique of factor analysis, it is determined that student satisfaction with the modality of virtual class depends firstly on the skills and competencies of the teacher with a contribution of 49.973% (1st factor), secondly on virtual learning resources with a participation of 6.034% (2nd factor) and thirdly of the quality of internet service with a contribution of 4.686% (3rd factor). Regarding the significance of the dependence of the degree of satisfaction of virtual teaching in relation to the traditional teaching method, the verification is through a Chi Square test, the significance is 0.000, which confirms the relationship that exists between both variables, then the Spearman test indicates 88.77%, which is significant with a 95% strong positive relationship between the variables student satisfaction and the virtual class modality, finally, when verifying the hypothesis, it is confirmed that the factorial analysis technique is appropriate and admissible for the application of the data, that is: thanks to the KMO coefficient of 0.969 (≥ 0.7) as well as the Bartlett sphericity test with a significance of the sphericity test is 0.000 (≤ 0.05).

Keywords: KMO coefficient, Bartlett's sphericity, factors, traditional teaching.

Introducción

Actualmente, el país se ve afectado con los cierres temporales de distintas instituciones de educación superior a causa de la pandemia del virus COVID-19, debido a esta enfermedad se han prohibido las reuniones masivas siendo uno de los principales riesgos de contagio, además de la propagación de dicha enfermedad.

A raíz de la pandemia del virus COVID-19 la mayoría de las universidades a nivel mundial dejaron de operar presencialmente. Según las estimaciones de la UNESCO y el Instituto Internacional para la Educación Superior de América Latina y el Caribe (IESALC, 2020), muestran que el cierre temporal afecta aproximadamente a unos 23,4 millones de estudiantes de educación superior y a 1,4 millones de docentes en América Latina y el Caribe; esto representa, aproximadamente, más del 98% de la población de estudiantes y profesores de educación superior de la región (IESALC, 2020).

La suspensión de las actividades presenciales ha sido extremadamente rápida, se inició el 12 de marzo en Colombia y Perú, en Argentina se recomendó la suspensión de actividades académicas a partir del 14 de marzo y en el país vecino de Chile la suspensión da inicio el 16 de marzo. En Bolivia, mediante un comunicado del Comité Ejecutivo de la Universidad Boliviana (CEUB) se da inicio a la interrupción de las actividades de forma presencial desde el 12 de marzo de 2020.

Es en esta situación que el internet es una herramienta primordial que permite realizar clases virtuales, a partir del uso de esta, existe una variedad de herramientas digitales que facilitan la enseñanza docente, no obstante, las herramientas digitales no reemplazan al docente, pero son un complemento esencial en todo este proceso de enseñanza y aprendizaje, bajo estas circunstancias es necesario su uso a fin de que, los alumnos continúen su proceso de aprendizaje y formación profesional.

La carrera de Ingeniería de Sistemas perteneciente a la Universidad Pública de El Alto (UPEA), por medio de un Honorable Consejo Universitario (HCU) decide dar la continuidad a las actividades académicas con la modalidad de clase virtual. La carrera actualmente no cuenta con su propia plataforma virtual, comparte la plataforma de la universidad a través de la plataforma Moodle, de

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

modo que, el estudiante puede acceder, revisar los recursos y actividades que los docentes transmiten y, por consiguiente, ellos podrán disponer de todo el material para su correspondiente retroalimentación.

En la presente gestión 2022, aún se continua con las clases de modalidad virtual, definitivamente habrá un antes y un después de la pandemia provocada por el virus COVID-19 en nuestro sistema educativo superior.

En vista de todo lo expuesto, es importante realizar esta investigación para identificar los factores que están asociados al grado de satisfacción de los estudiantes de la Carrera de Ingeniería de Sistemas con respecto a la modalidad de clase virtual y así llevar a cabo un análisis para una mejora continua de la nueva modalidad de educación no presencial.

De esta manera se formula la siguiente pregunta de investigación: ¿cuáles son los factores que están asociados al grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA en el desarrollo del proceso de aprendizaje en la modalidad de clase virtual?

Metodología

Análisis factorial

El análisis factorial (en adelante AF) tiene la finalidad de reducir un conjunto de variables numéricas $X_1, X_2, \dots, X_p$ interrelacionadas, en un conjunto de factores o factores latentes independientes $k < p, F_1, F_2, \dots, F_k$, que expliquen suficientemente a las variables reales perdiendo el mínimo de información, de tal manera que siempre serán, en número, inferiores a las variables iniciales. La relación que existe entre las variables observadas $X_1, X_2, \dots, X_p$, cuando los valores son grandes en la matriz de correlaciones, la matriz es considerada adecuada para realizar un análisis factorial, si existe algún tipo de relación entre las variables, los datos puedan ser reordenados a un conjunto menor de factores que pueden llegar a ser consideradas como variables que reúnan y resuman la correlación observada en el conjunto de datos. El modelo factorial se define de la siguiente manera:

$$\begin{aligned} X_1 &= w_{11}F_1 + w_{12}F_2 + \dots + w_{1k}F_k + u_1 \\ X_2 &= w_{21}F_1 + w_{22}F_2 + \dots + w_{2k}F_k + u_2 \\ &\vdots \\ X_p &= w_{p1}F_1 + w_{p2}F_2 + \dots + w_{pk}F_k + u_k \end{aligned} \quad (1)$$

Donde:

$F_1, F_2, \dots, F_k$: son los factores comunes.

$u_1, u_2, \dots, u_k$: factores únicos o factores específicos.

w_{jk} : carga factorial o saturación de la variable j en el factor h o también denominado peso del factor h en la variable j .

Por el modelo definido anteriormente en (1) se observa a cada una de las p variables iniciales son una combinación lineal de los k factores comunes ($k < p$) además de ser un factor único para cada variable. Así también, se tiene que todas las variables originales están influenciadas por factores comunes de tal manera, que existe un único factor que es específico para cada variable.

Matricialmente se expresa el modelo factorial como:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pk} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix} \quad (2)$$

$$\mathbf{X} = \mathbf{WF} + \mathbf{u}$$

Diseño de la investigación

La presente investigación tiene paradigma positivista, la cual, busca identificar los factores de las variables mediante el método estadístico multivariante del análisis factorial que corresponde en presentar factores para dar una validez a los resultados obtenidos. El enfoque de la investigación es cuantitativo, adecuado para sistemas numéricos. El estudio presenta una investigación de tipo descriptivo, ya que explica y describe las estructuras y características de enseñanza de la modalidad de clase virtual además de la satisfacción del estudiante. El diseño es no experimental transeccional o trasversal debido a que se obtuvo la información en un solo momento.

El instrumento de recolección de información es la encuesta (escala Likert), la información fue recolectada durante la pandemia mediante la herramienta Google Forms. Para medir la confiabilidad del instrumento se utilizó el estadístico de fiabilidad Alfa de Cronbach dando un resultado del 0.965 lo que significa que el instrumento tiene un nivel excelente de confiabilidad.

Variables de estudio

Las variables definidas para la presente investigación son las siguientes:

Variable 1: Satisfacción del estudiante. se tomaron en cuenta 18 ítems en escala de Likert del cuestionario con el objetivo de definir la variable 1 en 4 dimensiones siguientes: Calidad del servicio de internet, Expectativa del estudiante, Docencia y Rendimiento, de tal forma que sean medibles.

Variable 2: Modalidad de clase virtual. se tomaron en cuenta 14 ítems en escala de Likert del cuestionario con el objetivo de definir la variable 2 en las dimensiones medibles siguientes: Recursos de aprendizaje virtual, Comunicación Virtual y Competencias.

Población y muestra

La carrera de Ingeniería de Sistemas de la UPEA, cuenta con una población estudiantil de 3500 estudiantes matriculados durante la gestión II-2020.

El tipo de muestro empleado en la investigación es el muestreo probabilístico, por lo cual la muestra estuvo conformada por 381 alumnos de ambos sexos donde, el 55.9% de los encuestados son hombres y 44.1% son mujeres, así también 22.3% de los estudiantes son de cuarto semestre, 18.4% de los estudiantes están cursando el tercer semestre, 28.1% son de primer semestre que corresponde al mayor porcentaje de estudiantes entrevistados en la muestra, 0.5% de los encuestados son de

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

octavo y décimo semestre, además de edades comprendidas entre 16 a 41 años. El cálculo del tamaño de muestra para esta investigación considera a un nivel de confianza del **95%** y un error máximo de estimación del **5%** .

Resultados

Se presentan los resultados principales de la investigación; para determinar la significancia de la dependencia del grado de satisfacción de la enseñanza en la modalidad virtual en relación al método de enseñanza tradicional se considera lo siguiente:

Tabla 1

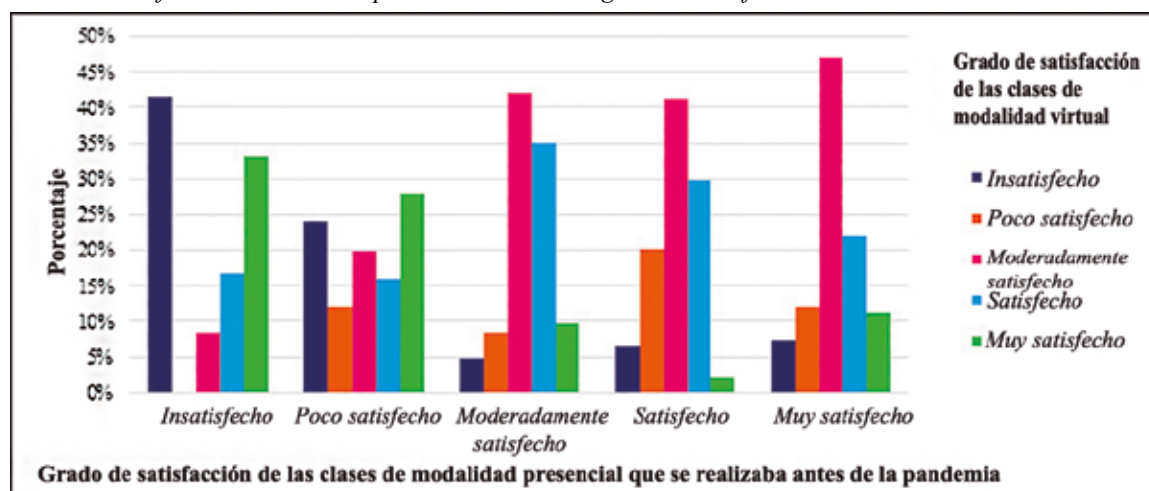
Tabla de contingencia del grado de satisfacción en el desarrollo de las clases presenciales antes de la pandemia y el grado de satisfacción de las clases en la modalidad virtual

Grado de satisfacción en el desarrollo de las clases de modalidad virtual	Grado de satisfacción del desarrollo de las clases de modalidad presencial que se realizaba antes de la pandemia.					Total
	Insatisfecho	Poco satisfecho	Moderadamente satisfecho	Satisfecho	Muy satisfecho	
Insatisfecho	5 41.7%	6 24.0%	4 4.8%	9 6.5%	9 7.3%	33 8.7%
Poco satisfecho	0 0.0%	3 12.0%	7 8.4%	28 20.3%	15 12.2%	53 13.9%
Moderadamente satisfecho	1 8.3%	5 20.0%	35 42.2%	57 41.3%	58 47.2%	156 40.9%
Satisfecho	2 16.7%	4 16.0%	29 34.9%	41 29.7%	27 22.0%	103 27.0%
Muy satisfecho	4 33.3%	7 28.0%	8 9.6%	3 2.2%	14 11.4%	36 9.4%
Total	12 100.0%	25 100.0%	83 100.0%	138 100.0%	123 100.0%	381 100.0%

Fuente: Elaboración propia

Figura 1

Grado de satisfacción de las clases presenciales versus el grado de satisfacción de las clases virtuales



Fuente: Elaboración propia

Las hipótesis de prueba son:

H_0 : El grado de satisfacción de las clases presenciales antes de la pandemia es independiente del grado de satisfacción de las clases virtuales realizadas en el semestre II-2020.

H_a : El grado de satisfacción de las clases presenciales antes de la pandemia está relacionado con el grado de satisfacción de las clases virtuales realizadas en el semestre II-2020.

Para llevar a cabo la prueba de hipótesis se utilizó la prueba de independencia Chi Cuadrado, los resultados se muestran en Tabla 2.

Tabla 2
Prueba de independencia Chi Cuadrado

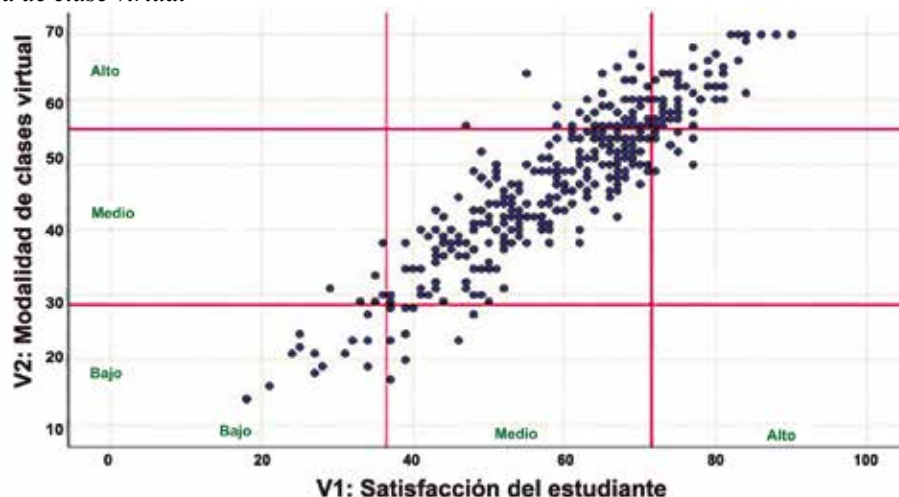
	Valor	Grados de libertad	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	68,727	16	0,000
N de casos válidos	381		

Fuente: Elaboración propia

De acuerdo con los resultados en la Tabla 2, teniendo en cuenta la significación de la prueba Chi se observa que: el resultado que se obtiene es menor al 5% (<5%) por lo tanto se acepta la hipótesis alterna (H_a) y se rechaza la hipótesis nula (H_0), de esta manera, se confirma y se verifica que existe una relación entre el grado de satisfacción de las clases presenciales antes de la pandemia con el grado de satisfacción de las clases virtuales realizadas durante en el semestre II-2020.

Con el fin de docimar la relación que existe entre la satisfacción del estudiante y la modalidad de clase virtual, se evalúa como sigue: en base a las variables del estudio, se plantea una prueba t-student para determinar si la correlación que existe entre la variable 1: Satisfacción del estudiante y la variable 2: Modalidad de clase virtual es significativa.

Figura 2.
Diagrama de dispersión de la evaluación de la variable 1: Satisfacción del estudiante y la variable 2: Modalidad de clase virtual



Fuente: Elaboración propia

La correlación de Spearman es del 88.7%, el mismo que es significativo con un 95% de confianza.

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

Tabla 3
Prueba t-student para el coeficiente de correlación de Spearman

	Valor	Grados de libertad	Significación asintótica (bilateral)
Coeficiente de correlación	0,887	379	0,000
N de casos válidos	381		

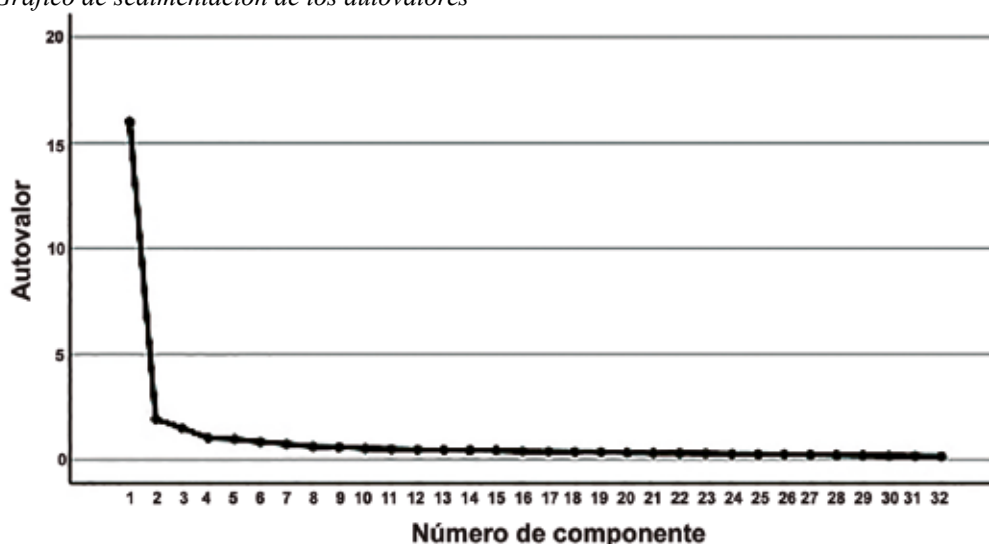
Fuente: Elaboración propia

Por los resultados observados en la Tabla 3, se verifica que existe una relación entre las variables: satisfacción del estudiante y modalidad de clase virtual además se puede evidenciar que presenta un grado de relación positiva y fuerte del 88.7%.

Con el propósito de realizar un análisis factorial a partir de los datos obtenidos con la encuesta aplicada a los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA, se tiene en cuenta lo siguiente: con respecto a realizar un análisis factorial, en primer lugar, se utilizó el método de rotación Varimax (Lacava et al., 2016); aceptando al menos un 50% de varianza explicada (Méndez y Rondón, 2012), cada variable no debe saturar a más de un factor y cada factor rotado sature por lo menos 3 variables (Lloret, Ferreres, Hernández y Tomás, 2014).

La Tabla 4 muestra los resultados del análisis factorial, según el análisis se obtiene 4 factores que sintetizan las variables utilizadas en el estudio para medir la satisfacción de las clases en la modalidad virtual, en total la varianza acumulada explicada por los cuatro factores es del 63.893%.

Figura 3.
Gráfico de sedimentación de los autovalores



Fuente: Elaboración propia

El gráfico de sedimentación de la Figura 3, muestra que las 32 variables del estudio se sintetizan en 4 factores, la distribución de las variables en los cuatro factores se detalla en la Tabla 4.

La Tabla 4 muestra los cuatro componentes y las variables asociadas a los mismos, el 1° componente que se obtuvo por medio de la técnica estadística multivariante del análisis factorial se compone de 19 variables, el 2° componente se forma de 6 variables, el 3° componente en 4 variables y el 4° componente en 3 variables.

Tabla 4
Estructura factorial de las variables del estudio

Ítems	Factor			
	1	2	3	4
p22	,739	,210	,175	,104
p21	,727	,317	,189	,151
p27	,709	,393	,088	,082
p26	,698	,349	,239	,062
p24	,696	,402	,165	,143
p23	,684	,370	,152	,120
p25	,681	,277	,135	,022
p17	,674	,424	,103	-,064
p18	,653	,495	,121	-,019
p20	,652	,067	,333	,380
p15	,588	,424	,214	,170
p12	,581	,190	,333	,433
p31	,571	,460	,202	,216
p16	,563	,553	,049	-,055
p13	,557	,322	,320	,403
p37	,556	,355	,292	,495
p19	,536	-,044	,430	,029
p28	,520	,369	,368	,137
p39	,507	,472	,235	,362
p33	,226	,788	,166	,060
p32	,190	,748	,233	,044
p35	,373	,660	,171	,050
p34	,495	,626	,204	,083
p30	,427	,609	,263	,089
p10	,324	,373	,304	,112
p7	,093	,153	,819	,064
p29	,234	,167	,777	,141
p8	,133	,350	,678	,003
p36	,349	,453	,491	,314
p14	,222	,197	,115	-,739
p40	,405	,231	,323	,594
p38	,453	,265	,234	,490
Número de ítems	19	6	4	3
Autovalores	15,991	1,931	1,499	1,024
% de varianza	49,973	6,034	4,686	3,201
% acumulado	49,973	56,007	60,692	63,893

Nota: Método de extracción: cuadrados mínimos no ponderados. Método de rotación: Varimax con normalización Kaiser.
Fuente: Elaboración propia

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

Tabla 5

Nuevas componentes identificados en el análisis factorial

Componente 1	<i>Habilidades y competencias de la docencia en el entorno virtual</i>
Componente 2	<i>Recursos de aprendizaje virtual</i>
Componente 3	<i>Calidad del servicio de internet</i>
Componente 4	<i>Satisfacción de las clases virtuales</i>

Fuente: Elaboración propia

La Tabla 5 muestra las nuevas componentes, nombradas e identificados de la siguiente manera:

El 1° componente tiene que ver con las habilidades de los docentes para la enseñanza virtual, así como las competencias que estos tienen, lo que repercute en la motivación y la satisfacción de los estudiantes. Este componente está relacionado con las dimensiones 3 y 7.

El 2° componente del análisis factorial está relacionado con los recursos de aprendizaje virtual mencionados en la dimensión 5.

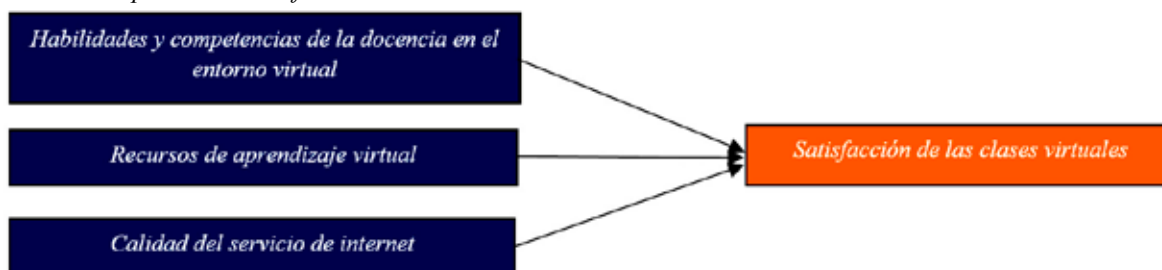
El 3° componente tiene que ver con la calidad del servicio de internet mencionado en la dimensión 1.

El 4° y último componente del análisis factorial se ha convenido en denominar como satisfacción de las clases virtuales, en vista que las tres variables que lo componen hablan de ello, cabe recalcar que en la operacionalización de variables del estudio no se había contemplado esta dimensión, sin embargo, este hecho refleja que este 4to componente es importante en el análisis.

Si bien el estudio es descriptivo, el análisis factorial muestra que el cuarto componente es la respuesta de los tres primeros explicado en la Figura 4.

Figura 4.

Nuevas componentes identificadas



Fuente: Elaboración propia

Contrastación de la hipótesis

La hipótesis enunciada para la investigación:

La técnica estadística multivariante del análisis factorial permite indicar los factores principales asociados al grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA en su proceso de aprendizaje en la modalidad de clase virtual.

Para realizar la contrastación de la hipótesis, la verificación se la hará a través de las pruebas de Kaiser Meyer Olkin (KMO), como también la prueba de esfericidad de Bartlett, por consiguiente, se definen las siguientes hipótesis:

H_0 : La técnica estadística multivariante del análisis factorial no es admisible para los datos.

H_a : La técnica estadística multivariante del análisis factorial es admisible para los datos.

La prueba de Kaiser Meyer Olkin (KMO) considera valores (≥ 0.7) como significativa, es decir, la técnica del análisis factorial es aceptable, a la vez analizar la prueba de esfericidad de Bartlett con una significancia (≤ 0.05) el mismo que valida el análisis factorial para determinar si es factible realizar dicho análisis (Pérez y Medrano, 2010). Los resultados de las pruebas se muestran en la Tabla 7.

Tabla 6.
Pruebas de validación del análisis factorial

<i>Prueba de KMO y Bartlett</i>		
<i>Medida Kaiser-Meyer-Olkin de adecuación de muestreo</i>		0,969
<i>Aproximación Chi-cuadrado</i>		8777,386
<i>Prueba de esfericidad de Bartlett</i>	<i>Grados de libertad</i>	496
	<i>Significancia</i>	0,000

Fuente: Elaboración propia

La Tabla 6 anterior se obtiene un coeficiente KMO de 0.969 (≥ 0.7) indica que la relación entre las variables del análisis es alta, muy buena. El test de esfericidad de Bartlett pone a prueba la hipótesis nula de que las variables analizadas no están correlacionadas en la muestra o, dicho de otro modo, que la matriz de correlación es la identidad (las correlaciones entre las variables son cero), en este caso la significancia del test de esfericidad es 0.000 (≤ 0.05) con lo que se rechaza la hipótesis nula, y se acepta la hipótesis de investigación corroborando así que la técnica estadística multivariante del análisis factorial es admisible para los datos, es decir permite identificar los factores principales asociados al grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA.

Discusión

Los resultados encontrados en la investigación también muestran aspectos positivos y negativos con relación a las satisfacción de las clases de modalidad virtual: aspectos positivos en la satisfacción del estudiante se concentran en las dimensiones: docencia (45% de satisfacción alta) y recursos de aprendizaje virtual (41% de satisfacción alta), aspectos negativos en la satisfacción del estudiante se concentran en las dimensiones: calidad del servicio de internet (13% de satisfacción baja) y comunicación virtual (14% de satisfacción baja). En general la satisfacción de los estudiantes respecto a las clases en la modalidad virtual es moderada (76.9%) y no se aprecian diferencias significativas entre hombres (75.6%) y mujeres (78.6%). Se concluye también que los estudiantes de 21 a 30 años son los que tienen los niveles más altos de satisfacción con las clases virtuales (24.1%), en comparación de los estudiantes menores a 20 años (13.4%) y los estudiantes mayores a 30 años (14.3%), este hecho permite concluir que es necesario acompañar y asesorar a los estudiantes más jóvenes y más adultos de la carrera en vista de los resultados obtenidos.

En el caso de los estudiantes de cuarto y quinto año de la carrera se tiene un mayor porcentaje con nivel de satisfacción alto (40%) en contraposición a los estudiantes de primero a tercer año de estudio (17.25%), este hecho permite concluir que las clases en modalidad virtual se adecúan de mejor manera a los estudiantes de últimos años de la carrera.

Así también, el estudio realiza la comparación de la satisfacción por factores socioeconómicos, el

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

grado de satisfacción es más alto en los estudiantes que pasan las clases en su domicilio propio (20.3%) en comparación con los estudiantes que pasan clases en domicilio de terceros o en algún local de internet, también es más alto en los estudiantes que tienen una conexión a internet fijo con un plan a domicilio (24.4%) y no así en los casos con conexiones de datos móviles o por modem.

De acuerdo al tipo de financiamiento de la conexión a internet, la satisfacción es más alta en los estudiantes que son financiados por un tercero (padre, tutor u otro) en relación los que pagan por cuenta propia o tienen un pago compartido. Los estudiantes que tienen un dispositivo electrónico propio para pasar las clases virtuales son los que tienen un grado de satisfacción más alto (20%) y no así los que disponen de dispositivos compartidos.

Conclusiones

Se observa la relación entre las variables: satisfacción del estudiante y modalidad de clase virtual se reporta una relación directa siendo el coeficiente de correlación de Spearman ($Rho=0,887$) representando esta una correlación alta de las variables y significativa ($p=0.00<0.05$), por lo cual cabe concluir que existe una relación funcional entre la satisfacción del estudiante y la modalidad de clase virtual de la carrera de Ingeniería de Sistemas de la UPEA, esto indica que a mayor satisfacción en la modalidad de clase virtual existe mayor satisfacción del estudiante.

Aplicando la prueba Chi cuadrado de independencia puede concluirse que existe una dependencia significativa entre el grado de satisfacción de la enseñanza en la modalidad virtual y el grado de satisfacción con el método de enseñanza tradicional ($Chi=68.727$, $p=0.00<0.05$).

Por medio del análisis factorial aplicado a las variables de estudio se encuentran cuatro factores principales ordenadas de la siguiente manera: habilidades y competencias de la docencia en el entorno virtual, recursos de aprendizaje virtual, calidad del servicio de internet y satisfacción de las clases virtuales. Se establece que la satisfacción del estudiante respecto a la clase en la modalidad virtual depende, en primer lugar, de las habilidades y competencias del docente (con una contribución de 49.973%), en segundo lugar, de los recursos de aprendizaje virtual (con una participación del 6.034%) y en tercer lugar, de la calidad del servicio de internet (con un aporte del 4.686%).

Referencias Bibliográficas

- Alaminos Chica, A., Francés García, F. J., Penalva García, C., y Santacreu Fernández, O. A. (2015). *Análisis multivariante para las ciencias sociales I*. PYDLOS ediciones. <https://fddocuments.ec/reader/full/analisis-multivariante-para-las-ciencias-sociales-i>
- Álvarez, J., Chaparro, E. y Reyes, D. (2015). Estudio de la satisfacción de los estudiantes con los servicios educativos brindados por instituciones de educación superior del Valle de Toluca. REICE. *Revista Iberoamericana sobre la Calidad, Eficiencia y Cambio en Educación*. 13(2), 5-26. <https://www.redalyc.org/pdf/551/55138743001.pdf>
- Bonilla, L. A. (2016). Deliberación entorno a la Educación Virtual. *Interconectando Saberes*, (1), 77-89. https://scholar.google.com/citations?view_op=view_citation&hl=es&user=3F5cbRoAAAAJ&citation_for_view=3F5cbRoAAAAJ:Tyk-4Ss8FVUC

- Cid, A. D., Méndez, R., y Sandoval, F. (2011). *Investigación Fundamentos y Metodologías*. Pearson Educación. <https://mitrabajodegrado.files.wordpress.com/2014/11/cid-investigacion-fundamentos-y-metodologia.pdf>
- Food and Agriculture Organization of the United Nations, & Nations, F. A. O. U. (2014). *Metodologías de E-Learning*. FAO. <https://www.fao.org/publications/card/es/c/e503d1b7-cf4f-4af3-9157-0b1cbc20b1c7/>
- Duggleby, J., (2001). *El tutor online. La enseñanza a través de Internet*, Ediciones Deusto.
- Duran, R., (2015). *La Educación Virtual Universitaria como medio para mejorar las competencias genéricas y los aprendizajes a través de buenas prácticas docentes* [Tesis]. <https://www.tesisenred.net/handle/10803/397710#page=1>
- García, L., (2001). *La educación distancia, De la teoría a la práctica*. Ariel S.A.
- Gonzales, M., Pino, M., y Penado, M. (2017). Estudio de la satisfacción percibida por los estudiantes de la UNED con su vida universitaria. *RIED. Revista Iberoamericana de Educación a Distancia*, 20(1), 243- 260. <http://dx.doi.org/10.5944/ried.20.1.16377>
- Goñi, E., y Infante, G. (2010a). Actividad físico-deportiva, autoconcepto físico y satisfacción con la vida. *European Journal of Education and Psychology*, 3(2), 199-208. <https://www.redalyc.org/articulo.oa?id=129315468004>
- IESALC (2020) *La Educación superior virtual en América Latina y El Caribe. Instituto Internacional para la Educación Superior en América Latina y el Caribe (IESALC) (2020). COVID-19 y educación superior: De los efectos inmediatos al día después*. América Latina y el Caribe. IESALC.<https://www.iesalc.unesco.org/wp-content/uploads/2020/05/COVID-19-ES-130520.pdf>
- Holgado, C. (2016). Nuevos tiempos, Universidad y TIC'S ¿Que aportan el Internet al profesor de Lenguas Modernas? (1.a ed.). Área de innovación y Desarrollo, S.L. <https://doi.org/10.17993/DideInnEdu.2016.14>.
- Lacave, C., Molina A., Fernández, M. y Redondo A. (2016). Análisis de la fiabilidad y validez de un cuestionario docente. *Revista Industrial*, 1(9), 136-146. <https://doi.org/10.15381/idata.v24i2.20348>
- Méndez, C. y Rondón, M. (2012). Introducción al análisis factorial exploratorio. *Revista Colombia Psiquiatría*, 41(1), 197-207. <https://www.redalyc.org/pdf/806/80624093014.pdf>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). (1998). *Declaración Mundial sobre la Educación Superior en el Siglo XXI: Visión y Acción y Marco de Acción Prioritaria para el Cambio y el Desarrollo de la Educación Superior aprobados por la Conferencia Mundial sobre la Educación Superior: La Educación Superior en el Siglo XXI: Visión y acción*. http://www.unesco.org/education/educprog/wche/declaration_spa.htm.
- Pérez López C. (2004). *Técnicas de análisis multivariante de datos*. Pearson Educación.
- Sampieri, R. H., Collado, C. F., Lucio, P. B., Valencia, S. M., & Torres, C. P. M. (2014). *Metodología de la investigación*. McGraw-Hill Education.

Análisis factorial del grado de satisfacción de los estudiantes de la carrera de Ingeniería de Sistemas de la UPEA con respecto a la modalidad de clase virtual

- Sánchez Quintero, J. (2018). *Satisfacción estudiantil en educación superior: validez de su medición*. Universidas Sergio Arboleda. <https://repository.usergioarboleda.edu.co/bitstream/handle/11232/1027/SATISFACCI%C3%93N%20ESTUDIANTIL.pdf?sequence=4&isAllowed=y>
- Santoveña, S. (2002). Repercusión de la metodología didáctica con apoyo de entornos virtuales de aprendizaje en el rendimiento de los estudiantes. *Revista de Currículum y Formación de Profesorado*, 16(3), 409-430. <https://www.redalyc.org/pdf/567/56725002021.pdf>
- Silvio, J. (2006). Hacia una educación virtual de calidad, pero con equidad y pertinencia. *Revista de Universidad y Sociedad del Conocimiento*, 3(1), 1-14. [file:///C:/Users/PC/Downloads/73608-Text%20de%20l'article-85552-1-10-20070925%20\(1\).pdf](file:///C:/Users/PC/Downloads/73608-Text%20de%20l'article-85552-1-10-20070925%20(1).pdf)
- Torres, C. A. B., Bernal, C. A., y Palma, O. F. (2010). *Metodología de la investigación*. Pearson Educación.
- UNESCO (1999). *Los docentes, la enseñanza y las nuevas tecnologías: Informe mundial sobre la educación*. Santillana/UNESCO.
- Valdez, B, E. (2018) La educación virtual y la satisfacción del estudiante en los cursos virtuales del Instituto Nacional Materno Perinatal 2017. [Tesis] https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/21504/Valdez_BEB.pdf?sequence=1
- Zambrano, J. (2016). Factores predictores de la satisfacción de estudiantes de cursos virtuales. *RIED. Revista Iberoamericana de Educación a Distancia*. 19(2), 217-235. http://espacio.uned.es/fez/eserv/bibliuned:revistaRied-2016-19_5055/Factores_predictores_satisfaccion.pdf.

REVISTA VARIANZA

ISSN 2789-3510 versión impresa

ISSN 2789-3529 versión en línea

MISIÓN

La misión de la revista Varianza es publicar artículos de investigación e interacción social con originalidad y alta calidad, que cubran diferentes áreas de la estadística teórica, aplicada y de estudios interdisciplinarios.

VISIÓN

La visión de la revista Varianza es convertirse en un referente de publicación de la ciencia estadística en investigación científica, con aplicación en diferentes campos del conocimiento humano en un contexto nacional e internacional.

POLÍTICA EDITORIAL

La revista Varianza es una publicación internacional, periódica de carácter electrónico, fue presentada por primera vez el año 2001 por el Instituto de Estadística Teórica y Aplicada, IETA de la Carrera de Estadística de la UMSA.

El objetivo, es contribuir al conocimiento sobre la enseñanza-aprendizaje de la estadística y establecer un foro permanente de discusión de ideas, conceptos, procedimientos y modelos concernientes al área de estadística y sus aplicaciones.

El alcance de la revista es especializado en contribuciones de artículos del ámbito estadístico, y aplicaciones estadísticas.

A partir del 2021 esta revista se edita dos veces al año durante los meses de abril y octubre. Los trabajos publicados son producto de contribuciones originales y de alto rigor académico y científico, cuyo aporte son investigaciones de profundidad estadística teóricos y aplicados. Sin embargo, publica también temas selectos provenientes de disciplinas afines en estadística aplicada.

Se publican artículos en español donde el resumen y palabras clave deben ser escritos en español e inglés.

MODELO DE FINANCIAMIENTO

La revista tiene una asignación presupuestaria para su publicación impresa con recursos provenientes de la Universidad Mayor de San Andrés, bajo el presupuesto de la carrera de Estadística y el IETA (Instituto de Estadística Teórica y Aplicada). La publicación en la revista es gratuita bajo la modalidad Open Access.

PUBLICACIÓN

Los tipos de artículos que admite la revista son:

- Originales,
- Originales cortos,
- Reportes de caso,
- Artículos de revisión,

- Artículos de reflexión.

La aceptación de un artículo depende de la calidad, originalidad, rigor científico y ética de los escritos, mismos que son verificados mediante revisión de procesos ciegos de evaluación por pares. Las publicaciones son gratuitas y en ningún caso se cobrarán honorarios a los autores para la publicación de sus artículos.

La revista Varianza cuenta con registro ISSN 2789-3510 versión impresa; 2789-3529, versión en línea con depósito legal No. 4-1-285-2021 P.O. y sus artículos son presentados en formato APA. Asimismo, se encuentra indexada en Revistas Bolivianas (REVBOL).

PROCESO DE ARBITRAJE

El proceso inicial de revisión, pasa por el Comité Editorial, que revisa los requisitos básicos, así como los aspectos éticos para ser publicados en la revista Varianza. Esa revisión se realiza en un plazo de un mes. Al concluir el proceso, se notifica al autor o autores, vía correo electrónico.

La evaluación de los manuscritos se realiza con la ayuda de un formulario donde los evaluadores registran las observaciones y recomendaciones. Los evaluadores pueden aceptar o rechazar el artículo de manera definitiva o condicional dada la aclaración y/o rectificación de parte del autor. La comunicación con el autor es responsabilidad exclusiva del editor. Algunos aspectos que se toman en cuenta en la evaluación, son:

1. La originalidad e innovación en conceptos y técnicas estadísticas.
2. Si constituye una contribución en el área.
3. Su pertinencia y rigor científico.
4. Si los referentes teóricos y empíricos son apropiados.
5. Si hace un análisis bien fundamentado y está coherentemente argumentado.
6. Su aporte a la ciencia estadística y aplicaciones.
7. Se juzgará la calidad de la presentación, verificando si el resumen sintetiza el artículo en forma clara y adecuada.
8. Ajustes al formato en las normas APA séptima edición.

La revisión de los artículos sigue un proceso bajo la modalidad de “doble ciego” con la participación de dos evaluadores con la especialidad requerida para realizar una apropiada evaluación. Esta revisión busca la originalidad e innovación de las publicaciones.

El Comité Editorial de la revista Varianza está conformado por un Consejo Editorial Nacional e Internacional cuya misión es la revisión de los artículos en la modalidad doble ciego.

ÉTICA DE PUBLICACIÓN

La revista Varianza tiene compromiso con la ética de la investigación, promueve los siguientes aspectos:

- Evitar conflictos de intereses,
- Evaluar objetivamente los manuscritos,
- Respetar los criterios de evaluación de los evaluadores,
- Conservar la confidencialidad de los autores y evaluadores, durante todo el proceso de revisión.

CONFLICTOS DE INTERESES

Para evitar conflictos de intereses en los procesos de evaluación de los manuscritos enviados, el Comité Editorial no selecciona como evaluadores a colegas que pertenezcan a la misma institución o a la misma red de investigación que los autores. Además, se solicita a los autores que declaren cualquier tipo de interés relacionado con algún miembro o miembros del Comité Editorial.

PRINCIPIOS ÉTICOS DE LA INVESTIGACIÓN

Los manuscritos recibidos son evaluados mediante el sistema de doble ciego, para evitar posible pérdida de objetividad por parte de los árbitros. Por otro lado, el Comité Editorial, se reserva el derecho de pasar por un detector de plagio, los manuscritos recibidos, a fin de identificar faltas éticas.

Cuando se detecta alguna violación a los principios éticos de la investigación, no se publica el artículo y el Comité Editorial informa las razones al autor o autores.

PROPIEDAD INTELECTUAL

Los autores al momento de tener la aceptación de la publicación, deben firmar la autorización de la publicación de la revista Varianza, en la que se estipula que son legítimos propietarios del artículo a publicar, que es una contribución original y que no existe problemas de derechos de autor con terceros y/u otros conflictos de naturaleza ética. Todo el contenido de la revista, excepto aquello que expresamente sea identificado, está bajo la licencia *Creative Commons*.

LICENCIAMIENTO

La revista Varianza se encuentra bajo licenciamiento *Creative Commons* atribución CC BY <https://creativecommons.org/licenses/by/4.0/>.



Atribución CC BY

La licencia permite que otros distribuyan, mezclen, adapten y construyan sobre su trabajo, incluso comercialmente, siempre que le reconozcan la creación original. Esta es la licencia más complaciente que se ofrece. Recomendado para la máxima difusión y uso de materiales con licencia.

INSTRUCCIONES PARA LOS AUTORES

Título

El título debe ser conciso e informativo de la investigación en mayúscula con no más de 12 palabras que contengan ya los descriptores, además debe ser redactado en sentido afirmativo, en idioma español e inglés.

Autoría

Debe ir a continuación del título, señalar el nombre y apellidos del (los) autor(es) acompañado de la(s) nota(s) a pie de página, según sea la posición en la autoría. A pie de página y con la numeración correspondiente se debe informar: profesión, filiación institucional, un breve curriculum vitae y ORCID.

Resumen

El resumen debe contener el objetivo, metodología, material, métodos, resultados de la investigación haciendo énfasis en los logros alcanzados.

Palabras claves

Sirven para identificar el artículo en bases de datos internacionales de manera que un potencial usuario pueda llegar en forma efectiva al artículo. Van debajo del resumen, mínimo cuatro (4) y máximo siete (7) palabras clave que no deben hacer parte del título del artículo. Deben estar ordenadas alfabéticamente y separadas por comas que ayuden a identificar los aspectos importantes del artículo.

Abstract

Se escribe el resumen del artículo en inglés.

Key words

Se escriben las palabras clave en inglés.

Introducción

La introducción debe contener:

- **Problema:** Debe describir claramente lo que se resolverá con la investigación. Debe enunciar claramente el qué y el porqué de la investigación. Se debe desarrollar en uno o dos párrafos iniciales.
- **Revisión de la literatura:** Expone el marco referencial que da sustento al trabajo de investigación. A través de las citas se provee reconocimiento de estudios anteriores que se relacionan específicamente con el trabajo.
- **Objetivo o hipótesis:** Debe describir el objetivo en forma clara, debe indicar en forma inequívoca qué es lo que el investigador intenta observar y medir, redactados en forma afirmativa y sujetos a una sola interpretación. La hipótesis debe expresar de manera clara, precisa y concisa una relación o diferencia entre dos o más variables, incluyendo, si corresponde, las variables del estudio y su efecto.

Materiales y métodos

Debe describir el universo de estudio, instrumentos y procedimientos con la precisión necesaria para permitir a los lectores una comprensión clara del artículo y la posibilidad de reproducir lo entendido. Debe especificarse el número de observaciones de los objetos a estudiar, los métodos y técnicas estadísticas para la generación y análisis de resultados.

Resultados

Presentarlos en secuencia lógica y que cuenten con un análisis estadístico o interpretativo en relación con el objetivo del estudio.

Discusión

La discusión debe ser sobre la base de los objetivos y los resultados para posteriormente comparar

con el marco referencial. Discutir los aspectos nuevos y limitaciones que tiene su estudio, enunciando proyecciones o nuevas hipótesis si corresponde.

Conclusiones

Estas deben responder a los objetivos del estudio, limitándose a los datos encontrados sin citar referencias.

Referencias Bibliográficas

Contiene la referencia de libros y artículos consultados para el artículo científico. Las fuentes bibliográficas deben ser citadas a lo largo del texto, en formato APA séptima edición.

Presentación

Los artículos deben ser presentados en letra *Times New Roman* tamaño 12, espacio simple, los márgenes externos son: margen izquierdo 2,5 cm., derecho 2,0 cm., superior e inferior 2,0 cm.

***Dirección: Calle 27 de Cota Cota
Bloque F.C.P.N. - Primer Piso
Email: ieta@umsa.bo
Página web: <https://ieta.umsa.bo/ediciones-varianza>
2022***

La Paz - Bolivia