



CONSULTORÍA ESTADÍSTICA PARA CIENCIAS SOCIALES

CONSULTORÍA ESTADÍSTICA PARA CIENCIAS SOCIALES

AUTORES:

UTRERA MÁRQUEZ, CLAUDIA ITZEL
CRUZ LÓPEZ, CECILIA
MÉNDEZ VIVEROS, CLAUDIA ELILÚ
HERNÁNDEZ ORTIZ, MARÍA ELENA
GÓMEZ MEDINA, LUIS ENRIQUE
ARCOS ROCHA, SONIA
MORALES GARCÍA, EMMANUEL
CÓRDOVA AGUILAR, ALMA ROSA
GABRIEL ARGÜELLES, ELISEO
CAMPOS CARMONA, MARÍA FERNANDA
HERNÁNDEZ LANDA, USBALDO
OLIVERA GÓMEZ, DANIEL ARMANDO
MARTÍNEZ DE LOS SANTOS, MONSERRAT
SÁNCHEZ JIMÉNEZ, DEMETRIO
REYES FLORES, ITZEL ALESSANDRA
SÁNCHEZ GUEVARA, MIRIAM
WANG, MAOZHU

©RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C. 2025



EDITA: RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C.
DUBLÍN 34, FRACCIONAMIENTO MONTE MAGNO
C.P. 91190. XALAPA, VERACRUZ, MÉXICO.
CEL 2282386072
www.redibai.org
redibai@hotmail.com



Sello editorial: Red Iberoamericana de Academias de Investigación, A.C.
(978-607-5893)
Primera Edición, Xalapa, Veracruz, México.
No. de ejemplares: 2
Presentación en medio electrónico digital
Formato PDF 6 MB
Fecha de aparición 26/12/2025
ISBN 978-607-5893-75-4

DICTAMEN EDITORIAL

La presente obra fue arbitrada y dictaminada en dos procesos; el primero, fue realizado por el **COMITÉ EDITORIAL RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C.** con sede en México; que sometió a los capítulos incluidos en la obra a un proceso de dictaminación a doble ciego para constatar de forma exhaustiva la temática, pertinencia y calidad de los textos en relación a los fines y criterios académicos de la misma, cumpliendo así con la primera etapa del proceso editorial. El segundo proceso de dictaminación estuvo a cargo del **COMITÉ CIENTÍFICO RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C.**; donde se seleccionaron expertos en el tema para la evaluación de los capítulos de la obra y se procedió con el sistema de dictaminación a doble ciego. Cabe señalar que previo al envío a los dictaminadores, todo trabajo fue sometido a una prueba de detección de plagio. Una vez concluido el arbitraje de forma ética y responsable y por acuerdo del Comité Editorial y Científico de la Red Iberoamericana de Academias de Investigación A.C. (REDIBAI), se dictamina que la obra **“CONSULTORÍA ESTADÍSTICA PARA CIENCIAS SOCIALES”** cumple con la relevancia y originalidad temática, la contribución teórica y aportación científica, rigurosidad y calidad metodológica, actualidad de las fuentes que emplea, redacción, ortografía y calidad expositiva.

Dr. Daniel Armando Olivera Gómez

Director Editorial

Sello Editorial: Red Iberoamericana de Academias de Investigación, A.C.

(978-607-5893)

Dublín 34, Residencial Monte Magno

C.P. 91190. Xalapa, Veracruz, México.

Cel 2282386072

Xalapa, Veracruz. México a 26 de diciembre de 2025

CERTIFICACIÓN EDITORIAL

RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C. (REDIBAI) con sello editorial **No. 978-607-5893** otorgado por la Agencia Mexicana de ISBN, hace constar que el libro **“CONSULTORÍA ESTADÍSTICA PARA CIENCIAS SOCIALES”** registrado con el **ISBN 978-607-5893-75-4** fue publicado por nuestro sello editorial con fecha de aparición del 26 de diciembre de 2025 cumpliendo con todos los requisitos de calidad científica y normalización que exige nuestra política editorial.

Fue evaluado por pares académicos externos y aprobado por nuestro Comité Editorial y Científico.

Todos los soportes concernientes a los procesos editoriales y de evaluación se encuentran bajo el poder Editorial de **RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C. (REDIBAI)**, los cuales están a disposición de la comunidad académica interna y externa en el momento que se requieran.

La normativa editorial y repositorio se encuentran disponibles en la página **<http://www.redibai-myd.org>**

Doy fe.

Dr. Daniel Armando Olivera Gómez

Director Editorial

Sello Editorial: Red Iberoamericana de Academias de Investigación, A.C.

(978-607-5893)

Dublín 34, Residencial Monte Magno

C.P. 91190. Xalapa, Veracruz, México.

Cel 2282386072

ÍNDICE

EMPRENDIMIENTO POR COVID-19 DE MUJERES CON VENTA DE PRODUCTOS ADQUIRIDOS EN PLATAFORMAS CHINAS DEL ECOMMERCE	1
INFLUENCIA DEL REZAGO SOCIAL EN LOS PATRONES ESPACIALES DEL DENGUE EN VERACRUZ (2020-2023)	12
ANÁLISIS ESTADÍSTICOS PARA LA GESTIÓN SANITARIA: INCIDENCIA DE ENFERMEDADES EN MENORES DE 1 A 14 AÑOS EN MÉXICO (2017)	21
INTRODUCCIÓN Y VENTA DE CAFÉ VERACRUZANO EN CHONGQING, CHINA	33
ANÁLISIS ESTADÍSTICO PARA LA GESTIÓN DEL RIESGO EDUCATIVO Y DEL BIENESTAR ESTUDIANTIL	44
ESTADÍSTICA ESPACIAL Y ANÁLISIS SOCIAL: DISCAPACIDAD Y REZAGO EN LOS MUNICIPIOS DE VERACRUZ (2020)	57
ANÁLISIS MULTIVARIADO APLICADO A LA SALUD PÚBLICA: PERFIL DE LA POBLACIÓN CON CÁNCER CERVICOUTERINO EN VERACRUZ	70
ANÁLISIS DE VENTAS INTERNACIONALES UTILIZANDO TÉCNICAS DE MACHINE LEARNING CON POWER BI	80

EMPRENDIMIENTO POR COVID-19 DE MUJERES CON VENTA DE PRODUCTOS ADQUIRIDOS EN PLATAFORMAS CHINAS DEL ECOMMERCE

CLAUDIA ITZEL UTRERA MÁRQUEZ¹

CECILIA CRUZ LÓPEZ²

CLAUDIA ELILÚ MÉNDEZ VIVEROS³

Centro de Estudios China – Veracruz (CECHIVER)

Universidad Veracruzana

RESUMEN

La pandemia derivada del Covid-19, propició que el Gobierno de México basado en las recomendaciones internacionales con el fin de minimizar el contagio, decretara medidas de mitigación del virus, mismas que impactaron significativamente el panorama económico y social del país. La falta de sustento en los hogares, aunado al incremento en las labores del hogar, cuidado de familiares, falta de empleo, o aburrimiento, instó el espíritu emprendedor en las mujeres, obteniendo ingresos para sus hogares además de sacar tiempo para las labores que históricamente se les han impuesto con los roles de género. El presente trabajo aborda la problemática que enfrentaron las mujeres derivado de la pérdida del empleo, acepto a como el emprendimiento con productos adquiridos en plataformas chinas de e-commerce, llevó sustento a sus hogares, generando un fenómeno social denominado “NENIS” en la ciudad de Xalapa, Veracruz.

Palabras claves: Emprendimiento, E-commerce, Mujeres emprendedoras.

¹ cutrera@uv.mx

² ceccruz@uv.mx

³ clmendez@uv.mx

INTRODUCCIÓN

El COVID-19 surgió a finales de 2019 como un brote viral que rápidamente se convirtió en una amenaza sanitaria mundial. Su alto grado de contagio y expansión obligó a los gobiernos de todo el planeta a adoptar medidas extraordinarias de contención y mitigación, que incluyeron el confinamiento de la población, la suspensión de actividades presenciales y el cierre de negocios. La gravedad de su propagación llevó a la Organización Mundial de la Salud (OMS) a declararla oficialmente pandemia el 11 de marzo de 2020, marcando el inicio de una crisis sanitaria y económica sin precedentes en el siglo XXI.

La pandemia puso en evidencia la fragilidad de los sistemas de salud, especialmente en los países en desarrollo, donde las limitaciones de infraestructura y financiamiento amplificaron los impactos del virus. Sin embargo, las consecuencias trascendieron el ámbito sanitario: las restricciones impuestas para contener la enfermedad provocaron una drástica desaceleración económica que afectó a millones de empresas y empleos, modificando las formas de producir, consumir y relacionarse.

En México, los efectos fueron inmediatos. De acuerdo con el Estudio sobre Demografía de los Negocios (2020), de los 4.9 millones de establecimientos micro, pequeños y medianos (MiPyMES) existentes antes de la pandemia, solo 3.9 millones (79.19%) sobrevivieron, mientras que más de un millón (20.81%) cerraron definitivamente. El sector terciario o de servicios, que concentra la mayor parte de la fuerza laboral femenina, fue uno de los más golpeados. Miles de mujeres perdieron sus empleos o vieron reducidos sus ingresos, evidenciando la vulnerabilidad económica que aún caracteriza su participación en el mercado laboral.

A este panorama se sumó una sobrecarga en las tareas domésticas y de cuidado —labores históricamente feminizadas—, que se intensificaron con el confinamiento. La falta de empleo, el aumento de responsabilidades en el hogar y la necesidad de mantener el sustento familiar despertaron en muchas mujeres un renovado espíritu emprendedor, impulsándolas a generar ingresos desde casa mediante actividades productivas y comerciales. Este fenómeno no solo representó una respuesta ante la emergencia, sino también una reconfiguración del papel de las mujeres en la economía digital.

Durante este periodo, las redes sociales y el comercio electrónico (e-commerce) se convirtieron en aliados fundamentales. Plataformas como Facebook, Instagram, Mercado Libre y TikTok experimentaron un notable incremento en sus actividades comerciales. Como señala González (2020), la pandemia transformó estos espacios en verdaderos mercados digitales, donde miles de mujeres comenzaron a vender productos, ofrecer servicios y crear comunidades de consumo. El e-commerce permitió que sus negocios trascendieran las limitaciones físicas y llegaran a un público amplio, diverso y en expansión, consolidando una nueva modalidad de emprendimiento basada en la conectividad.

Sin embargo, el contexto laboral general reflejó una contracción considerable. Según la Encuesta Telefónica de Ocupación y Empleo (ETOE) realizada por el INEGI (2020), la población económicamente activa disminuyó de manera significativa, al tiempo que aumentaron la subocupación y la informalidad. Ante esta situación, el emprendimiento se consolidó como una estrategia de supervivencia económica y emocional, ya que permitió a las mujeres compatibilizar su rol familiar con la necesidad de generar ingresos.

El caso de la ciudad de Xalapa, capital del estado de Veracruz, resulta particularmente ilustrativo. Según DataMéxico (2020), contaba en ese año con 488,531 habitantes, de los cuales 53.6% eran mujeres. El Censo de Población y Vivienda (2020) registró que el 43.2% de los hogares xalapeños estaban encabezados por mujeres, lo que refleja su peso decisivo en la dinámica económica y social local. Esta realidad se intensificó durante la contingencia sanitaria, cuando muchas de ellas optaron por emprender a través de redes sociales, vendiendo productos de diversa índole —ropa, accesorios, alimentos, artículos de belleza, entre otros—, en su mayoría importados desde China mediante plataformas de comercio electrónico como Shein, Romwe o AliExpress.

El aumento de las importaciones de manufacturas chinas coincidió con una expansión global del e-commerce. La competitividad de estos productos, basada en sus bajos costos y alta disponibilidad, desplazó parcialmente la producción nacional, pero también democratizó el acceso al consumo y facilitó la creación de microemprendimientos gestionados por mujeres. En este sentido, la pandemia no solo trajo consigo una crisis, sino también una oportunidad de reconversión económica, donde la resiliencia y la creatividad femenina se convirtieron en motores de transformación.

A raíz de este proceso emergió un fenómeno social que se hizo viral en México: el de las “NENIS” (acrónimo popular de “nuevas emprendedoras de negocios por internet”). Lo que comenzó como una expresión despectiva en redes sociales hacia mujeres que vendían productos en línea, terminó resignificándose como un símbolo de orgullo, independencia y empoderamiento. Las NENIS no solo desafiaron los estigmas de género asociados al trabajo femenino, sino que también construyeron una comunidad de apoyo mutuo, innovación y sororidad digital.

Estas mujeres se organizaron principalmente mediante redes sociales, donde acordaban puntos de entrega, precios, métodos de pago y canales de promoción, utilizando los recursos tecnológicos a su alcance para crear una economía paralela, ágil y dinámica. El uso intensivo del internet durante la contingencia —ya fuera para trabajo remoto, educación o comercio— facilitó este tipo de emprendimientos, evidenciando cómo la digitalización del consumo puede convertirse en un vehículo para la inclusión económica y social.

El presente capítulo aborda este fenómeno desde una perspectiva socioeconómica y de género, analizando el surgimiento del emprendimiento femenino xalapeño en el contexto del COVID-19, con especial atención a aquellas mujeres que comercializan productos de importación a través del e-commerce. Este análisis busca comprender las motivaciones, estrategias y desafíos de las emprendedoras digitales, así como el impacto de su actividad en la economía local.

Asimismo, el estudio invita a reflexionar sobre el papel que las crisis pueden tener como catalizadores de innovación y cambio social. Las mujeres emprendedoras no solo se adaptaron a un entorno adverso, sino que transformaron las limitaciones en oportunidades, redefiniendo su identidad económica y su participación en el espacio público. Este fenómeno, lejos de ser marginal, revela una tendencia global hacia la autonomía económica femenina, la digitalización de los negocios y la reconfiguración de los mercados locales en un mundo interconectado.

En suma, el caso de las NENIS xalapeñas representa un microcosmos de resiliencia y creatividad en tiempos de crisis. Desde una mirada amplia, el capítulo propone entender este movimiento como parte de una transformación más profunda: el tránsito hacia un modelo económico donde la tecnología, la equidad de género y el emprendimiento convergen para construir nuevas formas de desarrollo, sostenibilidad y empoderamiento.

MATERIALES Y MÉTODOS

La presente investigación se enmarca en un enfoque cuantitativo, descriptivo y correlacional-causal, orientado a analizar las características del emprendimiento femenino en la ciudad de Xalapa durante el periodo posterior a la pandemia de COVID-19, específicamente en mujeres que comercializan productos de importación provenientes de China mediante plataformas digitales de comercio electrónico. El propósito central fue identificar las dinámicas de venta, los canales digitales utilizados, así como los factores sociodemográficos que inciden en el uso de redes sociales como medio de emprendimiento.

Dado el contexto sanitario vigente en el momento de la recolección de datos, el estudio se desarrolló mediante encuestas autoadministradas en línea, diseñadas en Google Forms, lo que permitió garantizar el acceso seguro y la participación remota de las emprendedoras. Este procedimiento favoreció la recopilación de información en condiciones de confinamiento y mantuvo la integridad ética de la investigación.

Población y muestra

La población objeto de estudio estuvo conformada por mujeres emprendedoras residentes en la ciudad de Xalapa, Veracruz, que durante el periodo 2020–2022 emprendieron actividades comerciales mediante redes sociales y plataformas de e-commerce, particularmente Shein, AliExpress y Romwe, enfocadas en la venta de productos de importación procedentes de China.

El rango de edad considerado fue de 18 a 55 años, con base en los datos proporcionados por Statista (2021), los cuales señalan que en enero de 2021 más del 30 % de los usuarios de redes sociales en México eran mujeres y hombres de entre 25 y 35 años, mientras que la población mayor de 65 años representaba menos del 4 % de los usuarios activos en el país. Este rango etario permitió concentrar el estudio en el grupo de mujeres más representativo dentro del entorno digital y con mayor actividad comercial en línea.

La muestra se obtuvo mediante un muestreo no probabilístico por el método de “bola de nieve”, el cual consiste en identificar a las primeras participantes que cumplen con los criterios de inclusión, quienes a su vez refieren a otras personas con características similares, hasta alcanzar el tamaño muestral adecuado. Este procedimiento es especialmente útil cuando se trabaja con poblaciones de difícil acceso o en sectores emergentes, como el emprendimiento digital informal.

Criterios de inclusión y exclusión

Se establecieron los siguientes criterios de inclusión:

- Ser mujer residente en la ciudad de Xalapa, Veracruz.

- ☐ Empezar actividades de venta o reventa de productos importados desde China, adquiridos a través de plataformas como Shein, AliExpress o Romwe.
- ☐ Utilizar redes sociales (Facebook, Instagram, TikTok, WhatsApp) como principal medio para contactar clientes, promocionar o vender sus productos.

Los criterios de exclusión fueron:

- ☐ Mujeres que no residan en Xalapa.
- ☐ Aquellas que adquieran o comercialicen productos únicamente en establecimientos físicos, sin uso de redes sociales.
- ☐ Personas fuera del rango de edad establecido (18–55 años).

Instrumento de recolección de datos

Para la recolección de la información se aplicó una encuesta estructurada elaborada en formato digital mediante Google Forms, diseñada para medir variables relacionadas con el perfil sociodemográfico, la naturaleza del emprendimiento, las plataformas utilizadas, el tipo de producto comercializado, los motivos del emprendimiento y la percepción de los beneficios y dificultades experimentadas durante su actividad comercial.

El instrumento incluyó preguntas de tipo cerrado y mixto, organizadas en secciones temáticas, con escalas de medición nominal, ordinal y de intervalo. Este diseño permitió obtener datos cuantificables y comparables, garantizando la validez del análisis estadístico posterior.

Debido a las restricciones sanitarias derivadas de la pandemia de COVID-19, la aplicación del instrumento se realizó de manera virtual, enviando el enlace de la encuesta a las posibles participantes mediante redes sociales, grupos comunitarios y contactos previos identificados en la primera etapa del muestreo. Las respuestas fueron recolectadas y almacenadas automáticamente en una base de datos de Google Sheets, asegurando la confidencialidad de la información.

Una vez obtenidos los datos, se procedió a su depuración y codificación en hojas de cálculo de Excel, para posteriormente realizar el análisis descriptivo e inferencial en el software SPSS (Statistical Package for the Social Sciences) y en Power BI, a fin de generar visualizaciones gráficas que facilitaran la interpretación de los resultados.

Consideraciones éticas

Se garantizó la confidencialidad, anonimato y consentimiento informado de todas las participantes. Cada emprendedora fue informada sobre los objetivos del estudio, la naturaleza voluntaria de su participación y el uso exclusivamente académico de la información. No se solicitó ningún dato sensible ni se establecieron condiciones que comprometieran la privacidad o integridad de las participantes.

Diseño estadístico

En suma, el estudio combina un enfoque descriptivo con un diseño correlacional-causal, orientado a explorar cómo el uso de redes sociales y plataformas digitales ha impulsado el emprendimiento femenino en Xalapa tras la pandemia. El empleo de herramientas digitales como Google Forms y Excel facilitó la

obtención y visualización de los datos, mientras que la selección del método de bola de nieve permitió acceder a una población dinámica, diversa y representativa del fenómeno social conocido como las “NE-NIS”, símbolo de resiliencia y empoderamiento económico en tiempos de crisis.

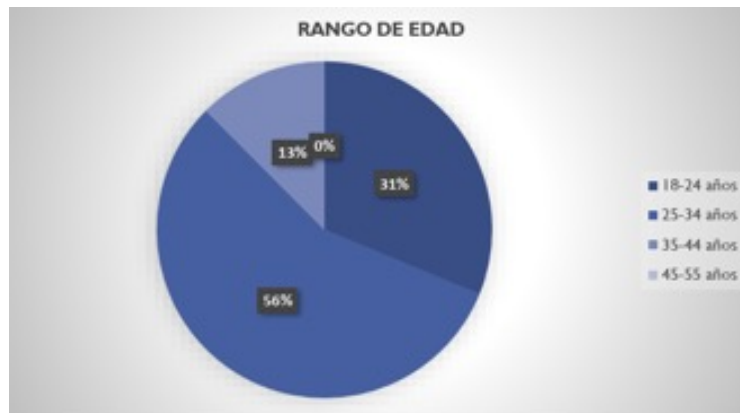
RESULTADOS

Perfil sociodemográfico de las emprendedoras

En total, participaron 32 mujeres emprendedoras xalapeñas con edades comprendidas entre 18 y 55 años, quienes adquirirían productos importados de China mediante plataformas como Shein, AliExpress y Romwe, y los comercializaban a través de redes sociales.

Más de la mitad de las participantes (56%) tiene entre 25 y 34 años, lo que refleja que el emprendimiento digital mediante plataformas chinas es predominantemente impulsado por mujeres jóvenes adultas. Un 31% se ubica entre los 18 y 24 años, mientras que el 13% pertenece al rango de 35 a 44 años.

Figura 1. Rango de edad de las emprendedoras



Fuente: Elaboración propia.

En cuanto al nivel educativo, el 53% cuenta con preparatoria concluida, y el 47% con estudios universitarios. Ninguna participante reportó tener posgrados, lo cual sugiere que este tipo de emprendimiento es más común en mujeres con formación media o superior, pero sin necesariamente contar con especializaciones avanzadas.

Figura 2. Nivel de estudios concluido.



Fuente: Elaboración propia.

Respecto a su situación familiar, el 56% indicó tener hijos, y el 44% no. Esto confirma la importancia del emprendimiento como estrategia para equilibrar responsabilidades domésticas y laborales, al permitirles gestionar su tiempo y generar ingresos propios.

Situación laboral y económica

Antes y durante la pandemia por COVID-19, las condiciones laborales de las emprendedoras se vieron modificadas. El 31% trabaja por cuenta propia, el 28% se encuentra desempleada, pero en búsqueda activa de empleo, y un 19% es estudiante o tiene empleo de tiempo completo. Solo el 22% perdió su trabajo directamente a causa de la pandemia, lo que muestra que muchas optaron por emprender como alternativa preventiva o complementaria a sus ingresos.

Figura 3. Situación laboral actual.



Fuente: Elaboración propia.

En cuanto a los ingresos mensuales, el 41% percibe entre \$2,000 y \$6,000 pesos, un 38% obtiene \$2,000 o menos, el 16% gana entre \$6,000 y \$10,000, y únicamente un 6% supera los \$10,000 pesos. Estos datos revelan que el emprendimiento digital representa una fuente de ingresos complementaria, aunque en muchos casos insuficiente para garantizar independencia económica total.

Figura 4. Rango de ingresos mensuales.



Fuente: Elaboración propia.

Acceso digital y plataformas utilizadas

Todas las participantes cuentan con teléfono celular con conexión a internet, el dispositivo principal para gestionar sus ventas. Además, el 28% dispone de computadora o laptop, el 9% de todos los dispositivos

(celular, laptop y tablet), y un 4% complementa con tablet. Esto demuestra la portabilidad y centralidad del teléfono móvil como herramienta de trabajo y comunicación para las emprendedoras digitales.

Figura 5. Acceso a dispositivos digitales.



Fuente: Elaboración propia.

El 94% ha adquirido productos en plataformas chinas de e-commerce, siendo Shein la más utilizada (71%), seguida de AliExpress (21%), mientras que Romwe y Alibaba apenas alcanzan un 2% cada una. La frecuencia de compra más común es una vez al mes (66%), lo que indica un flujo constante de adquisición y reventa.

Figura 6. Frecuencia de compra en plataformas chinas



Fuente: Elaboración propia.

Figura 7. Plataforma de e-commerce más utilizada.

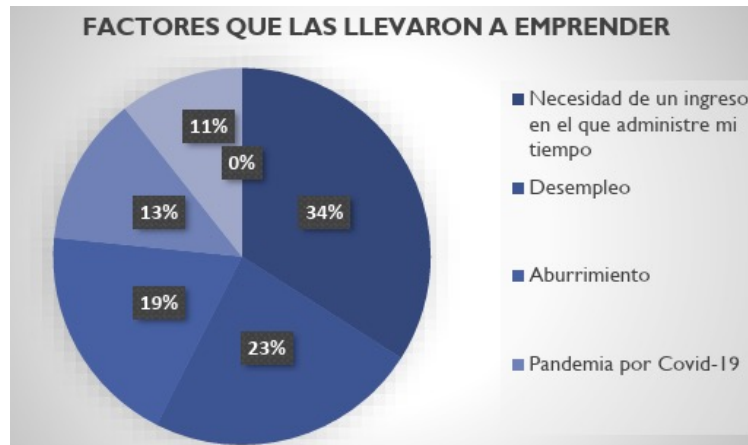


Fuente: Elaboración propia.

Motivaciones y percepción del emprendimiento

Las principales razones para emprender fueron la búsqueda de ingresos con flexibilidad horaria (34%), la pérdida de empleo (23%), y el aburrimiento durante el confinamiento (19%). En menor medida, un 13% lo hizo motivada por la propia pandemia y un 11% por seguir una tendencia o moda.

Figura 8. Factores que motivaron el emprendimiento.



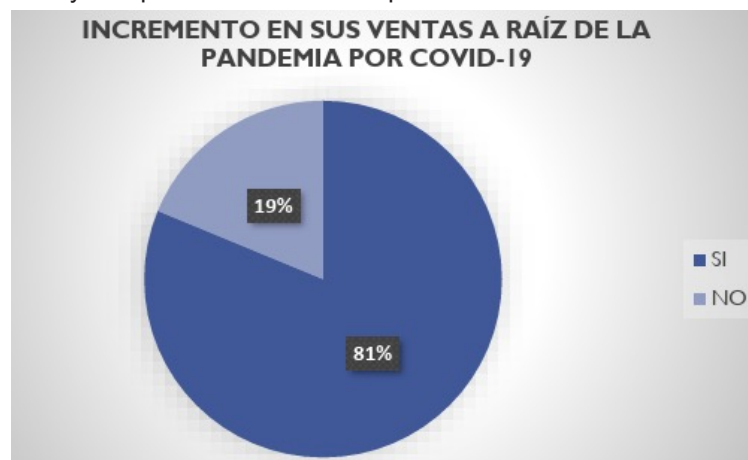
Fuente: Elaboración propia.

En cuanto a la rentabilidad del negocio, el 69% considera que su emprendimiento es rentable, principalmente gracias al uso de cupones, promociones y descuentos en las plataformas chinas. El 25% logra ganancias al revender con márgenes mayores, y solo un 6% reporta pérdidas ocasionales o falta de rentabilidad, lo que evidencia que este modelo, aunque informal, puede resultar económicamente viable.

Impacto de la pandemia y el fenómeno “NENIS”

El 81% de las participantes observó un incremento en sus ventas durante la pandemia, atribuyéndolo al aumento del uso de redes sociales, al cierre de negocios físicos y al auge de plataformas como Shein. Sin embargo, el 91% también reportó un aumento en la competencia, estimando que esta creció entre un 10% y 20%, lo que confirma la expansión acelerada del e-commerce informal durante ese periodo.

Figura 9. Incremento en las ventas y competencia a raíz de la pandemia.

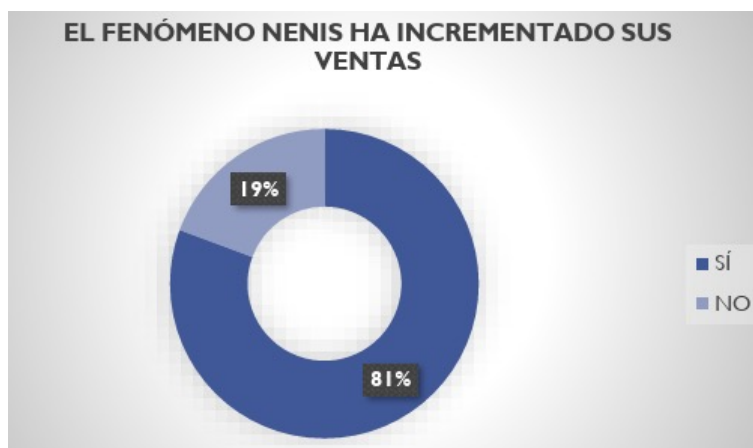


Fuente: Elaboración propia.

El fenómeno social denominado “NENIS” tuvo gran relevancia entre las encuestadas. El 81% reconoció un incremento en sus ventas tras su viralización, ya que este movimiento visibilizó el emprendimiento femenino digital. No obstante, las opiniones sobre el término fueron diversas: algunas mujeres lo adoptan

con orgullo, viéndolo como una identidad colectiva positiva, mientras que otras lo perciben como un apelativo despectivo que trivializa su esfuerzo y trabajo.

Figura 10. Impacto del fenómeno “NENIS”.



Fuente: Elaboración propia.

En conjunto, los resultados evidencian que el emprendimiento digital femenino en Xalapa durante la pandemia de COVID-19 fue un fenómeno emergente, flexible y adaptativo, liderado por mujeres jóvenes con formación media o universitaria, muchas de ellas madres, que encontraron en el e-commerce y las redes sociales una vía de sustento, independencia y empoderamiento económico.

El caso de las emprendedoras xalapeñas —particularmente aquellas vinculadas al comercio de productos chinos— muestra cómo la crisis sanitaria global impulsó nuevas dinámicas laborales y reforzó el papel de las mujeres como agentes activos de recuperación económica en contextos locales.

CONCLUSIONES

El emprendimiento de mujeres con productos adquiridos mediante plataformas chinas de e-commerce presentó un incremento durante la pandemia por Covid-19 en Xalapa. Un dato importante que generó la presente investigación es que las emprendedoras con productos adquiridos en plataformas chinas de e-commerce cuentan con una escolaridad de preparatoria a universitaria, todas las encuestadas cuentan con internet en sus hogares. Su principal medio de comunicación es el celular, dado que todas las encuestadas cuentan con un celular como dispositivo, contando con diversas aplicaciones que utilizan tanto para adquirir como para vender sus productos, en este mismo dispositivo pueden tomar las fotos que utilizarán en sus ventas, además de permitirles comunicarse con sus clientes. La mayoría de las emprendedoras utiliza SHEIN con una frecuencia de compra de por lo menos 1 vez al mes, apoyando a esta plataforma del gigante asiático misma que no cerró operaciones durante la pandemia. Los factores que fomentaron el emprendimiento fueron diversos mismos que van desde la pérdida del empleo hasta el aburrimiento, la moda o la necesidad de un ingreso en el que puedan administrar su tiempo. Para las emprendedoras resulta atractivo y redituable adquirir productos en plataformas chinas de e-commerce gracias a las estrategias de venta implementadas. El fenómeno social de las “NENIS” si bien inicia como una burla hacia las emprendedoras mexicanas, es un fenómeno que nos muestra la relevancia de las mujeres en la venta de productos mediante internet, algunas de ellas lo consideran un insulto, no obstante, la

mayoría de las mujeres refiere que es la forma de llamar de cariño a sus clientes y que es una estrategia para hacerlas sentir en confianza, incrementando gracias a ello sus ventas.

REFERENCIAS

- Campines Barría, F., de Tyler, C., y González Jiménez, T. (2021). Comercio electrónico como estrategia de venta de la administración empresarial en tiempos de pandemia del Covid-19. *Revista Científica Guacamaya*, 5, pp. 58-71. <https://www.revistas.up.ac.pa/index.php/guacamaya/article/view/2058>
- I.N.A.F.E.D. (2021). Veracruz - Xalapa. Instituto Federal para el Federalismo y el Desarrollo Municipal. <http://www.inafed.gob.mx/work/enciclopedia/EMM30veracruz/municipios/30087a.html>
- INEGI. (2020). Resultados de la encuesta telefónica de ocupación y empleo (ETOE) cifras oportunas de junio de 2020. Comunicado de prensa núm. 387/20. https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2020/enoe_ie/ETOE2020_08.pdf
- INEGI. (2020). Resultados de la segunda edición del ECOVID-IE y del estudio sobre la demografía de los negocios 2020. Comunicado de prensa núm. 617/20. https://inegi.org.mx/contenidos/saladeprensa/boletines/2020/OtrTemEcon/ECOVIDIE_DEMOGNEG.pdf
- INEGI. (2020).
- González, Y. M. (2020). El lugar de la creencia en tiempos de pandemia: Jóvenes y redes sociales virtuales. *Revista Argentina de Estudios de Juventud*, (14), e043. Doi <https://doi.org/10.24215/18524907e043>
- Lázaro Berenguer, M. (2020). Estudio de la percepción del comercio online para negocios tradicionales durante la pandemia por COVID-19. Universidad Politècnica de València. <http://hdl.handle.net/10251/158722>
- Mendoza J. (2020). COVID-19 y el empleo en México: impacto inicial y pronósticos de corto plazo. *Contaduría y Administración*, 65, pp. 214. ISSN 2448-8410. <http://dx.doi.org/10.22201/fca.24488410e.2020.3028>
- Oficina del PNUD en México (2020). Desafíos de desarrollo ante la COVID-19 en México. Panorama Socioeconómico. Economía UNAM. PNUD LAC C19 PDS No. 13. <http://revistaeconomia.unam.mx/index.php/ecu/article/download/543/576/>

INFLUENCIA DEL REZAGO SOCIAL EN LOS PATRONES ESPACIALES DEL DENGUE EN VERACRUZ (2020-2023)

MARÍA ELENA HERNÁNDEZ ORTIZ⁴

LUIS ENRIQUE GÓMEZ MEDINA⁵

Facultad de Estadística e Informática

Instituto de Investigaciones y Estudios Superiores de las Ciencias Administrativas

Universidad Veracruzana

RESUMEN

La organización Mundial de la Salud (OMS) en 2019 incluyó al dengue como una amenaza potencial entre diez enfermedades, por su dispersión en diferentes países del mundo. Veracruz es considerado susceptible por sus climas predominantes cálido subhúmedo y húmedo, así como la circulación de los cuatro serotipos. Pérez y Calles establecen que la incidencia del dengue involucra cuestiones sociales, para conocer la relación de la distribución del dengue con las variables sociales, se realizó un Análisis de Componentes Principales Geográficamente Ponderado, obtenido de los casos notificados en 2020 a 2023 (en porcentaje) y el índice de rezago social (CONEVAL), se agruparon los municipios con el método DBSCAN y se analizaron con un PERMANOVA. Los resultados mostraron que la incidencia del DNG y DCSA/DG se distribuía al norte y sur del estado de 2020-2022 y al centro en 2023. Los municipios presentaron comportamientos similares los cuales se agruparon en cinco clústeres y se sustentaron con un valor p menor al alfa establecido, obtenido de PERMANOVA. Concluyendo que el índice de rezago social tiene influencia en la distribución y gravedad de los casos de dengue.

Palabras clave: Dengue, DBSCAN, PERMANOVA, Incidencia, Estado de Veracruz.

⁴ GS20015088@egresados.uv.mx

⁵ luisgomez04@uv.mx

INTRODUCCIÓN

Las acciones del hombre son un influyente en la epidemia del dengue, como son la contaminación y sobrepoblación, así como el rezago social, provocando que países de la región de África y del Mediterráneo Oriental, sean más vulnerables por los sistemas frágiles de salud, los movimientos masivos de población, la deficiente infraestructura de agua y saneamiento. En la región de las Américas, cuarenta y seis países notifican a la Plataforma de Información en Salud de las Américas (PLISA) el número de casos de dengue grave, número de muestras y letalidad, semanalmente (OMS, 2023). En donde Brasil es el país con mayor número de casos sospechosos, con 1359 casos por 100,000 habitantes, le sigue Perú con 813 casos por 100,000 habitantes y México ocupa el tercer lugar con 179 casos por 100,000 habitantes.

De los casos reportados en México en 1978 a 1994, Veracruz fue uno de los municipios que consolidaba el 64% de los casos en el país (Narro-Robles & Gómez-Dantés, 1995), evento que ocurrió nuevamente en los años 2017 y 2018, el 84% de los casos lo conforman los estados de Chiapas, Veracruz, Jalisco, Nuevo León y San Luis Potosí, considerando al estado como altamente vulnerable a la epidemia del dengue.

El presente trabajo tiene como objetivo comprender la relación que existe en la distribución del dengue en el estado de Veracruz, con variables sociales establecidas, mediante las técnicas estadísticas espaciales, Componentes Principales Geográficamente Ponderada y Agrupación Espacial Basada en Densidad de Aplicaciones con ruido, para analizar este último análisis se emplea un PERMANOVA.

ANTECEDENTES

Desde años pasados, las epidemias han estado presente en la humanidad y su estudio ha servido a la comprensión de sus características, distribución y prevención de estas. El dengue no es la excepción, se encuentran múltiples de estudios sobre sus síntomas, causas, riesgos y sobre todo como prevenir a la población. Entre las investigaciones más actualizadas y relacionadas con el enfoque desarrollado en el presente trabajo, Pérez et al. (2003), menciona que las variables que intervienen en el aumento del dengue son consideradas en su mayoría del tipo social y ambiental, si bien el autor citado considera un tercer factor, el alojamiento de turistas en hoteles u otras residencias alternativas.

En el estudio de Cuartas et al. (2017) tiene como objetivo describir la relación espacial entre criadores positivos y potenciales en el interior de las viviendas y en el exterior del domicilio a partir del análisis exploratorio de datos espaciales y de correlación de la ciudad de Cali, Colombia. El área de estudio fue dividida en: área periurbana sobre los farillones de los ríos Cuca y Cali y el área residencial planificada. Mediante el sistema de Información Geográfica (SIG) se estudiaron las variables de interés: criaderos positivos y potenciales en el intradormitorio y extradormitorio, realizó un análisis de la tendencia espacial para describir el comportamiento de las variables en la unidad geográfica y un análisis de autocorrelación espacial local para determinar la presencia de patrones espaciales en las variables, empleado los programas GeoDa y ArcGis.

Pérez et al. (2003) registros del Sistema de Información de Problemas de Salud Notificables (SINAN) utilizó un Análisis de Varianza (ANOVA) para comparar las variables sociodemográficas y epidemiológicas entre los grupos. Los resultados fueron en cuanto a las variables sociodemográficas que los casos

predominan en el sexo femenino en personas entre 20 y 39 años. Para el nivel de estudios el porcentaje más alto en escuelas secundaria completa y conforme a la raza el porcentaje más alto fue blanca. La cantidad de casos de dengue grave se presentó la cifra más elevada en la región Sudeste con tendencia decreciente, seguida por la Nordeste con la misma tendencia y finalmente por el centro-Oeste con tendencia estacionaria.

La comparación entre el comportamiento de epidemias es un punto clave para conocer la existencia de factores que contribuyan a la susceptibilidad de un lugar en específico, Triana-Vidal et al. (2019) con base en datos obtenidos de SIVIGILA (Sistema Nacional de Vigilancia en Salud Pública) realizó un Análisis Exploratorio de Datos Espacial (AEDE) y temporal del virus del Dengue (2014-2017), Sika (2015-2017) y Chikungunya (2014-2017), con el fin de identificar localizaciones atípicas, esquemas de asociación y estructuras espaciales de la distribución.

Obteniendo así para el virus del dengue, dos consolidaciones de 2 clúster, uno comprendido del 2010-2015 con 335 municipios, ubicado en la región de la Orinoquia y un segundo del 2013-2016 con 137 municipios, ubicado al suroccidental de la región Andina. Para el Chikungunya analizó 76,602 casos en el periodo 2014-2017, los cuales fueron agrupados en

tres clústeres, en el 2014 se agruparon 49 municipios de la región Atlántica colombiana, en 2015 se dividió en dos zonas, similares a la distribución del dengue y con un aumento de riesgo en la región Andina. Finalmente, para el Sika la distribución se agrupó en dos conglomerados, semejantes a los del Dengue y Chikungunya. Concluyendo que en el estudio hay dos zonas geográficas donde circulan los tres arbovirus, en donde la zona suroccidental de la región Andina es la de mayor riesgo.

La población que consideran autores como Calles (2014); Cabezas et al. (2015), es la más susceptible en el estado de Veracruz, son los municipios localizados al sur para ello Tlalolini (2011) hace una revisión sobre la situación del dengue en los años 2008 - 2010, de la Delegación Veracruz Sur de estado. Donde tuvo como resultados que el año con mayor frecuencia de fiebre hemorrágica por dengue (FHD) y fiebre (FD) fue en 2009, con 747 casos. Al evaluar la tasa de dengue por grupo de edad de FHD y FD, se tiene la tasa más alta de FHD en 2009 en el grupo de edad de 10-19 años, con una tasa de 87/100,000 derechohabientes), le sigue el grupo de 20-59 años con una tasa de 61/100,000 derechohabientes. El grupo de edad de 10-19 años, también fue el primero desde 2008. Los municipios que presentaron mayor número de casos de FHD y FD, son Coatzacoalcos, Minatitlán, Cosoleacaque y Jáltipan y los meses de mayor frecuencia fueron agosto, septiembre y octubre.

MATERIALES Y MÉTODOS

Para el análisis estadístico, primero se obtuvieron los datos, para los casos de dengue en el estado de Veracruz, de 2020 a 2023, se solicitaron a la Plataforma Nacional de Transparencia, se obtuvo respuesta por parte de la Dirección de Salud Pública, la información del número de casos de dengue por clasificación: Dengue No grave (DNG), Dengue con Signos de Alarma y Dengue Grave (DCSA/DG), municipio, sexo y rango de edades. En donde los casos de dengue solo fueron desglosados por municipio con base a su clasificación, y únicamente se nos proporcionó un acumulado del sexo y rangos de edad por año.

Los casos de dengue por municipio fueron convertidos a porcentaje, con base a sus poblaciones correspondientes, mismas que fueron obtenidas del Censo de Población y vivienda 2020, del Instituto Nacional de Estadística y Geografía (INEGI,2020).

$$Porcentaje_{Mun} = \frac{\text{Número de casos confirmados por municipio}}{\text{Población total por municipio}} * 100$$

La segunda base de datos son las variables del Índice de Rezago Social (IRS) del estado de Veracruz, obtenidas del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, 2020). Se unió en una sola base de datos los casos de dengue por clasificación para cada año y el IRS, obteniendo así una base para cada año (4 bases de datos), con un contenido de 13 variables cada base.

Tabla 1. Información de las variables en estudio.

Nomenclatura	Nombre de la variable	Tipo de variable
CVE	Clave del municipio	Cualitativa nominal
Mun	Nombre del municipio	Cualitativa nominal
Año	Año de estudio	Cuantitativa discreta
X1	Proporción de dengue no grave	Cuantitativa continua
X2	Proporción de dengue con signos de alarma y dengue grave	Cuantitativa continua
X3	Población de 15 años o más analfabeta	Cuantitativa continua
X4	Población de 6 a 14 años que no asiste a la escuela	Cuantitativa continua
X5	Población de 15 años o más con educación básica incompleta	Cuantitativa continua
X6	Población sin derechohabencia a servicios de salud	Cuantitativa continua
X7	Viviendas con piso de tierra	Cuantitativa continua
X8	Viviendas que no disponen de excusado o sanitario	Cuantitativa continua
X9	Viviendas que no disponen de agua entubada de la red pública	Cuantitativa continua
X10	Viviendas que no disponen de drenaje	Cuantitativa continua

Con la herramienta de Google Colab se hicieron los siguientes análisis con código en Python, desde el navegador web. Primero se realizó la lectura de los datos y así mismo los archivos Shapefiles (que almacenan datos geoespaciales vectoriales), con estos datos se obtuvo el merge con cada una de las bases, haciendo uso de las librerías pandas y geopandas. Al merge se le realizó una intervención estadística exploratoria para las variables X1 y X2 en los cuatro años, con el fin de conocer como fue la distribución del dengue en cada municipio del estado, durante los cuatro años.

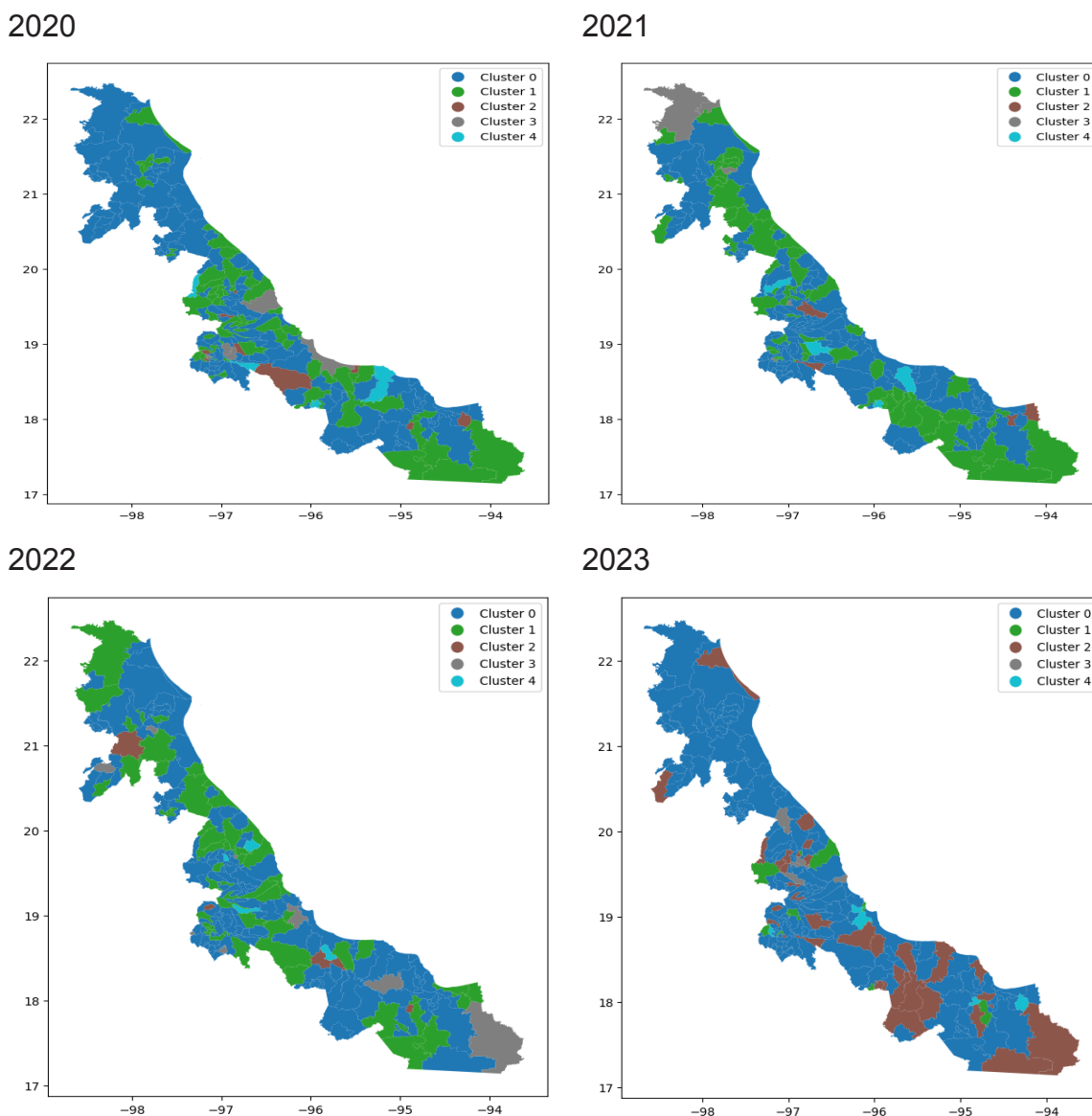
Posteriormente se realizó el GWPCA, se utilizó las variables X1 a la X10, se normalizaron para obtener el Análisis de Componente Principales, para los tres componentes más significativos se obtuvieron los valores propios y los pesos ajustados para cada PCA. Se creo la matriz de pesos espaciales de acuerdo con cada PCA, obtenida de multiplicar $PCA_1 * W$, donde W corresponde a los pesos espaciales, es así como se obtiene el Análisis de Componente Principal Geográficamente ponderado, GWPCA. Se seleccio-

naron los primeros dos componentes y se visualizaron los puntos GWPCA en un mapa para cada componente, haciendo uso de las librerías PySAL y Geoda. Finalmente, se utilizó Agrupación Espacial Basada en Densidad de Aplicaciones con Ruido (DBSCAN), para agrupar a los municipios de ambos GWPCA, visualizado en un mapa. Para conocer si los clústeres obtenidos del DBSCAN eran iguales o diferente se analizaron con un PERMANOVA, corrido desde el software RStudio 4.2.0. Se utilizó las variables X1 a X10 de la base originales y se agregó la variable clúster, que contiene el número de clúster al que pertenece cada municipio. Estos pasos se realizaron para los cuatro años de estudio y un PERMANOVA para comparar los cuatro años.

RESULTADOS

Los resultados se presentan en términos del análisis multivariado espacial desglosado por año. Mediante los resultados obtenidos de los GWPCA, se obtuvo el Análisis de clúster espacial (DBSCAN), el cual está desglosado por año para identificar el comportamiento en los diferentes años.

Figura 1. Análisis DBSCAN para los años 2020 al 2023.



Para conocer si en el periodo de estudio la incidencia del dengue en relación con las variables sociales ha sido igual, ha disminuido o aumentado, se puede ver que en año 2020, 2021 y 2022 el clúster 0, agrupa a los municipios con un grado de rezago social muy alto, alto, medio, bajo y muy bajo, con una incidencia de dengue no grave, que va desde baja a alta, pero una incidencia de dengue grave baja. En el año 2020, en el Clúster 0, se agruparon 130 municipios, pertenecientes en su mayor parte norte y sur del estado. Para el año 2021 agrupa a ciento treinta y seis municipios, de los cuales en su mayoría son del centro y sur, finalmente en el año 2022 agrupa a 133 municipios, igualmente distribuidos más al centro y sur.

2020, el Clúster 1 a sesenta y tres, el Clúster 2 a nueve municipios, el Clúster 3 a seis municipios y el Clúster 4 a cuatro municipios.

2021 el Clúster 1 a sesenta y tres, para el Clúster 2 cuatro municipios, el Clúster 3 cinco municipios y finalmente el Clúster 4 agrupo cuatro municipios.

2022, para el Clúster 1 sesenta y dos, el Clúster 2 a cinco municipios, al Clúster 3 ocho municipios y finalmente el Clúster 4 agrupa a cuatro municipios.

Resultados de PERMANOVA por Año (2020–2023)

El análisis PERMANOVA (Permutational Multivariate Analysis of Variance) es una prueba estadística no paramétrica que permite evaluar si existen diferencias significativas en la composición multivariada entre grupos, como los clústeres generados por el algoritmo DBSCAN. Esta prueba se basa en permutaciones y calcula un estadístico F, un valor de R^2 que indica la proporción de la variabilidad explicada por los grupos, y un valor p que determina la significancia estadística. En este caso, se aplicó PERMANOVA para los años 2020 a 2023 con un nivel de significancia $\alpha = 0.05$.

Juego de hipótesis:

Pregunta: ¿Existen diferencias significativas en la composición multivariada de las variables entre los distintos clústeres formados?

Hipótesis propuesta:

$$H_0: \mu_{\text{Clúster}_{1j}} = \mu_{\text{Clúster}_{2j}} = \dots = \mu_{\text{Clúster}_{ij}}$$

$$H_1: \mu_{\text{Clúster}_{ij}} \neq \mu_{\text{Clúster}_{ij}}$$

Tabla 2. Resultados del PERMANOVA

Año	R^2	F	Valor p
2020	0.03177	6.891	0.001
2021	0.02415	5.1968	0.009
2022	0.01539	3.2829	0.029
2023	0.02707	5.8427	0.002

Año 2020

En el año 2020, el análisis PERMANOVA mostró un valor de , lo que indica que aproximadamente el 3.18 % de la variabilidad total de los datos puede atribuirse a las diferencias entre los clústeres identificados por el algoritmo DBSCAN. El estadístico refleja una marcada diferencia entre los grupos en comparación

con la variabilidad interna de cada uno. Además, el valor de η^2 es inferior al nivel de significancia ($\alpha = 0.05$), por lo que se rechaza la hipótesis nula. En consecuencia, se concluye que durante 2020 existieron diferencias significativas en la composición multivariada de las variables entre los clústeres formados, lo que sugiere una clara diferenciación espacial o estructural en los patrones observados.

Año 2021

Para el año 2021, los resultados indicaron un $\eta^2 = 0.024$, evidenciando que el 2.4 % de la variabilidad total estuvo explicada por los clústeres. El estadístico sugiere que, aunque las diferencias entre grupos fueron menores que las observadas en 2020, siguen siendo relevantes desde el punto de vista estadístico. Con un valor de $p = 0.000$, se rechaza nuevamente la hipótesis nula, lo que confirma la existencia de diferencias significativas entre los grupos. Esto podría reflejar una ligera disminución en la heterogeneidad espacial o una transición hacia una distribución más uniforme de las variables analizadas en ese periodo.

Año 2022

Durante el año 2022, el valor de $\eta^2 = 0.0154$ indica que los clústeres explicaron cerca del 1.54 % de la variabilidad total, siendo este el valor más bajo del periodo analizado. El estadístico señala una menor separación entre los grupos, aunque el valor de η^2 continúa siendo inferior al nivel de significancia. Por lo tanto, se rechaza la hipótesis nula, pero con un efecto más débil en comparación con años anteriores. Este resultado sugiere que las diferencias entre los clústeres persistieron, aunque con menor intensidad, posiblemente debido a una mayor homogeneidad espacial o a una distribución más equilibrada de los factores socioambientales y epidemiológicos.

Año 2023

Finalmente, en 2023 se observó un incremento en la fuerza de las diferencias entre clústeres, con un $\eta^2 = 0.035$ y un estadístico $F = 3.5$, lo que refleja una separación nuevamente marcada entre los grupos. El valor confirma una alta significancia estadística, permitiendo rechazar la hipótesis nula. En este año, las diferencias multivariadas entre los clústeres resultaron evidentes, lo cual podría estar relacionado con la reconfiguración de los patrones espaciales de las variables analizadas, posiblemente derivada de cambios socioeconómicos, ambientales o en la dinámica de transmisión del dengue.

CONCLUSIONES

De acuerdo con el objetivo que buscaba evaluar la posible relación entre las variables de tipo social con la incidencia en municipios susceptibles, a través del método estadístico espacial GWPCA. Los resultados del análisis muestran que existe una relación entre las variables sociales y la incidencia del dengue, aunque, con un comportamiento diferente en cada año, en algunos más apegadas a la infraestructura de las viviendas, otras a la educación que tienen y en algunos casos a la escasez de servicios de salud.

Cada municipio presenta un contexto diferente, tanto para los valores de las variables del IRS y los casos de DNG, DCSA y DG, esto mediante el análisis DBSCAN, se obtuvieron así grupos con valores sociales y casos de dengue similares y otro que no pertenecen dentro un grupo. Así mismo se encontraron diferen-

cias entre cada grupo, contrastado con PERMANOVA. Lo que significa que los municipios se encuentran en mayor, mediana y menor incidencia, al igual algunos que no pertenecen a ninguno de estos grupos.

Finalmente, la investigación evidencia que existe una relación espacial de la distribución geográfica de los casos de dengue grave y/o con signos de alarma y dengue no grave con el índice de Rezago Social en el estado de Veracruz. Mediante el desarrollo de las técnicas estadísticas queda demostrando que estas variables influyen en la incidencia del dengue, como son el nivel de educación, acceso a servicios de salud, condiciones e infraestructura de las viviendas en los municipios del estado, durante los años 2020 a 2023. Mostrando que mientras las carencias sociales de los estados de Veracruz sigan aumentando la incidencia del dengue será mayor y expandida en los municipios que presentan mayores escases de estos factores. Provocando que el aumento de estos factores, así como del calentamiento global proporcionan condiciones favorables para la transmisión del dengue y no tomar medidas generara incidencia a mayor escala.

Cabe mencionar que durante los años 2020 a 2022, la población se encontraba en aislamiento, por lo que la mayor parte del tiempo los habitantes se encontraban en el entorno de sus viviendas, indicando que la relación de las variables sociales se mostrara más notoria, por lo que, el contagio del dengue dependía de las condiciones de la vivienda y el grado de estudio. En cambio, para 2023, estos factores son inversos, ya que, la población se encontraba retorno a sus trabajos y por ende a un entorno diferentes con condiciones más favorable en términos de las variables sociales (menores carencias).

Es indispensable la mención de las técnicas estadísticas espaciales utilizadas, que nos brindan una comprensión más amplia de los datos analizados mediante una componente espacial, como es el caso. En donde GWPCA considera los valores más cercanos con base a sus pesos espaciales y definiendo los valores atípicos con los valores más alejados, los cuales pueden ser mapeados y nos dan una visualización mejor de los resultados que cualquiera puede entender con solo verla. El uso de estas técnicas, aunque no es utilizado ampliamente en México es fundamental para fortalecer las investigaciones en donde dependan de la posición geográfica, como son temas de epidemiología, rezago social y economía.

REFERENCIAS

- Cabezas, C., Fiestas, V., García-Mendoza, M., Palomino, M., Mamani, E., & Donaires, F. (2015). *Dengue en el Perú: A un cuarto de siglo de su reemergencia.* Revista Peruana de Medicina Experimental y Salud Pública, 32, 146–156. <https://www.scielosp.org/pdf/rpmesp/2015.v32n1/146-156/es>
- Calles, A. (2014). *Asociación de factores socio-demográficos y clínicos con dengue grave y no grave en la ciudad de Veracruz, México* (Tesis de posgrado). Universidad Veracruzana, Región Veracruz. <http://cdigital.uv.mx/handle/123456789/46995>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). (2020). *Índice de rezago social 2020.* CONEVAL. https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice_Rezago_Social_2020.aspx
- Cuartas, D. E., Martínez, G., Caicedo, D. M., Garcés, J., Ariza-Araújo, Y., Peña, M., & Méndez, F. (2017).

*Distribución espacial de criaderos positivos y potenciales de *Aedes aegypti*.* *Biomédica*, 37(2), 59–66. <https://doi.org/10.7705/biomedica.v34i2.3471>

Da, T. R. D., Costa, A. K. A. N., Alves, K. A. N., Santos, A. N., & Cota, M. D. F. (2022). *Tendencia temporal y distribución espacial del dengue en Brasil.* *Cogitare Enfermagem*, 27, e84000.

Narro-Robles, J., & Gómez-Dantés, H. (1995). *Dengue en México: Un problema prioritario de salud pública.* *Salud Pública de México*. Recuperado el 24 de junio de 2024, de [https://www.saludpublica.mx/index.php/spm/article/view/4560/5012#:~:text=La%20distribuci%C3%B3n%20por%20grupo%20de,a%2014%20a%C3%B1os%20\(9\)](https://www.saludpublica.mx/index.php/spm/article/view/4560/5012#:~:text=La%20distribuci%C3%B3n%20por%20grupo%20de,a%2014%20a%C3%B1os%20(9))

Organización Mundial de la Salud. (2023). *Dengue – Global situation.* Organización Mundial de la Salud. <https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498>

Pavlis, M., Singleton, A., Longley, P., & Cheshire, J. (2018). *Retail areas and their catchments.* En *Consumer Data Research* (pp. 40–51). UCL Press. <https://doi.org/10.2307/j.ctvqhsn6.6>

Pérez, T. T., Íñiguez, L., Sánchez, L., & Remond, R. (2003). *Vulnerabilidad espacial al dengue: Una aplicación de los sistemas de información geográfica en el municipio Playa de Ciudad de La Habana.* *Revista Cubana de Salud Pública*, 29(4). http://scielo.sld.cu/scielo.php?pid=S0864-34662003000400009&script=sci_arttext

Sistema Nacional de Vigilancia Epidemiológica (SINAVE). (2019). *Panorama epidemiológico de dengue, 2019.* Secretaría de Salud. https://www.gob.mx/cms/uploads/attachment/file/524262/Pano_dengue_52_2019.pdf

Tlalolini, P. (2011). *Caracterización del dengue de la Delegación Veracruz Sur IMSS* (Tesis de posgrado). Universidad Veracruzana, Orizaba, Veracruz.

ANÁLISIS ESTADÍSTICOS PARA LA GESTIÓN SANITARIA: INCIDENCIA DE ENFERMEDADES EN MENORES DE 1 A 14 AÑOS EN MÉXICO (2017)

SONIA ARCOS ROCHA⁶

EMMANUEL MORALES GARCÍA⁷

ALMA ROSA CÓRDOVA AGUILAR⁸

Facultad de Estadística e Informática

Universidad Veracruzana

RESUMEN

El estudio evidencia que la salud infantil en México sigue siendo un reto importante, especialmente en los menores de 1 a 14 años, donde las enfermedades infecciosas y no transmisibles continúan representando una proporción significativa de la morbilidad y mortalidad. A pesar de los avances en la reducción de la mortalidad infantil, padecimientos como las infecciones respiratorias agudas, anomalías congénitas, intoxicaciones y tumores malignos afectan principalmente los sistemas respiratorio, digestivo y cardiovascular. Los resultados del análisis con la prueba de Kruskal-Wallis muestran que los grupos de 1 a 4 y de 5 a 9 años presentan mayor prevalencia de enfermedades. Estados como Chihuahua, Baja California, Sinaloa, Coahuila, Chiapas, México y Campeche registran las tasas más altas, asociadas con pobreza y acceso limitado a servicios de salud. Se recomienda fortalecer las estrategias de prevención, atención médica y estudios más detallados que profundicen en las causas y distribución de las enfermedades infantiles en México.

Palabras clave: Análisis multivariado, salud pública, estadística no paramétrica

⁶ gS20015073@egresados.uv.mx

⁷ emmorales@uv.mx

⁸ almcordova@uv.mx

INTRODUCCIÓN

La salud infantil en México es importante para que los niños tengan un desarrollo saludable y puedan alcanzar su potencial máximo en su vida adulta. Este bienestar evoluciona a lo largo del tiempo conforme van creciendo sin embargo este puede ser afectado por diferentes factores ya sean biológicos, sociales, ambientales o económicos, la salud infantil se trata de un derecho fundamental que en la población en menores de edad debe tener garantizado adjudicando que esto sea una responsabilidad del Estado. (Vargas y López, 2024)

En México la tasa de mortalidad infantil en 2021 fue de 13 por cada 1,000 nacidos vivos, lo que significa que mueren 12 niños por cada 1,000 nacidos vivos antes de cumplir los 5 años, de acuerdo con la Encuesta Nacional de Salud y Nutrición 2023, las Infecciones respiratorias agudas afectan alrededor de una tercera parte de los menores en la edad en México mientras que tenemos un porcentaje de 9.4% en la enfermedad diarreica igual en menores de edad. (Vargas y López, 2024)

La tasa de mortalidad infantil (TMI) es un indicador clave en México ya que esta principalmente relacionado por factores como la pobreza en donde se refleja muertes evitables, prematuras e innecesarias que pueden prevenirse mediante acciones de atención a la salud. La pobreza y salud infantil están relacionadas ya que la pobreza se asocia con los peores resultados en la salud infantil. (Bravo-García, 2024).

Este estudio tiene como objetivo Identificar agrupaciones en las enfermedades presentadas en la población de 1 a 14 años en México, empleando la información proveniente de INEGI. Estadísticas de mortalidad de 2017, en donde se tienen las distintas enfermedades en los diferentes grupos de población clasificados según el INEGI de los cuales son los siguientes 1-4 años, 5-9 años y 10-14 años.

ANTECEDENTES

Valbuena (2017) realizó un análisis del comportamiento estadístico sobre la incidencia de la enfermedad respiratoria aguda en los niños menores de 5 años en el periodo de 2012 a 2017, como resultado obtuvo que la enfermedad Respiratoria Aguda (ERA), constituye la causa más frecuente de morbilidad y mortalidad, en niños menores de 5 años, en todo el mundo, Colombia, en el años 2017, presentó un 13% de casos de VSR (Virus Sincitial Respiratorio), de los cuales, el 5.1% corresponde a casos de influenza y 29.7% a otros virus respiratorios, es decir que la incidencia del VSR es la más alta, comparada con la influenza, en ese mismo año. En el periodo de 2012 a 2017, el virus de la influenza en Colombia presentó mayor incidencia en el año 2013 con un 22% de casos reportados. cual representa un problema importante de salud.

Córdova et al., (2020) hizo un estudio descriptivo de prevalencia de infecciones respiratorias agudas en niños menores de 5 años en un centro materno-infantil de Lima, durante los meses de enero y septiembre del 2018, en una población de 4050 historias clínicas. En donde en una ficha de registro se consignaron los datos de las historias clínicas: presencia o no de infecciones respiratorias agudas, sexo, edad (categorizada en menores de 1 año, 1 año, 2 años. 3 años, 4 años y 5 años.) y el mes en el que los niños acudieron al centro materno infantil. Arrojando como resultado el 30.5% (1234) presentaron infecciones

respiratorias agudas. La prevalencia según la edad fue de 29.9% (426) en menores de un año, en el grupo de 2 años fue 29.2% (188), aumentó a 35.35 en niños de 3 años y disminuyó a 30.7% en el grupo de 4 años finalmente, alcanzó el 28.0% en niños de 5 años.

Fernández y Hernández (2015) realizaron un breve análisis sobre mortalidad en menores de 15 años por enfermedades no transmisibles en México de 2000- 2014 en donde utilizaron la información generada por la estadística de defunciones INEGI/SSA4, se consideró la lista del *Global Burden of Disease (GBD)*, misma que a partir de la Clasificación Internacional de Enfermedades en su décima revisión (CIE-10), presenta las enfermedades no transmisibles en diez subdivisiones importantes. Como resultados obtuvieron que el mayor volumen de fallecimientos se debe a las anomalías congénitas, que comprenden el 56.7% de las defunciones en menores de 15 años, después le siguen en importancia los tumores malignos con el 13% de las muertes, con el 8.3% se encuentran los trastornos mentales junto con las enfermedades del sistema nervioso central, las enfermedades cardiovasculares con el 6.2% y finalmente con un 5.9% las enfermedades digestivas.

García et al., (1998) quería comunicar las diversas tasas de mortalidad pediátricas en España durante el año 1991 así como también poder describir los grupos de causa de mortalidad para las distintas edades en donde utilizó las publicaciones del Instituto Nacional de Estadística “Movimiento Natural de la Población Española 1991” y “Defunciones” según Causa de Muerte. En donde las distintas tasas de mortalidad fueron las siguientes para la tasa de mortalidad preescolar (1 a 4 años) fue de 1.75 y la tasa de mortalidad escolar (5-14 años) fue de 3.06%, en donde los accidentes y tumores fueron las causas de muerte más frecuentes sin embargo también las anomalías congénitas fueron también responsables de un alto porcentaje de mortalidad.

MATERIALES Y MÉTODOS

A continuación, se presenta el diagrama del proceso metodológico (Figura 1), el cual resume las etapas desarrolladas durante la investigación. Este esquema permite visualizar de manera clara y ordenada las fases seguidas, desde la recolección y análisis de datos hasta la interpretación de resultados, facilitando la comprensión general del enfoque aplicado en el estudio.

Figura 1: Propuesta metodológica.



Datos

La base de datos utilizada fue sacada del sitio INEGI. Estadísticas de mortalidad 2017, contiene datos sobre las diversas enfermedades en los diferentes estados de la República Mexicana. Los datos están organizados por entidad federativa, diagnostico específico (enfermedad) y los 3 diferentes grupos de edades durante el año 2017, el primer grupo de edad es de 1 a 4 años, el segundo grupo es de 5 a 9 años y el último grupo es de 10 a 14 años. En la Tabla 1 se muestran las variables utilizadas.

Tabla 1. Descripción de variables			
Variable	Descripción	Clasificación	Escala
	Conteo de menores de edad afectados por enfermedades relacionadas con el sistema corporal en ese rango de edad.	Cuantitativa	Continua
	Conteo de menores de edad afectados por enfermedades relacionadas con el sistema corporal en ese rango de edad.	Cuantitativa	Continua
	Conteo de menores de edad afectados por enfermedades relacionadas con el sistema corporal en ese rango de edad.	Cuantitativa	Continua

Estadística descriptiva

Se realizó un análisis exploratorio de los datos textuales mediante la generación de una nube de palabras que nos permitió visualizar la frecuencia relativa de los diagnósticos más predominantes, posterior a eso se elaboró un gráfico de barras con las 10 palabras más frecuentes lo que permitió cuantificar y comparar los términos más relevantes. Los pasos para realizar este análisis fueron los siguientes:

- Se instaló las librerías y su importación
- Lectura del archivo de texto
- Se generó la nube de palabras
- Se visualizó la nube de palabras mediante una figura de tamaño personalizado

Para complementar la nube de palabras se realizó un gráfico de barras horizontales de las 10 palabras más frecuentes en donde los pasos fueron los siguientes:

- Con la ayuda del método: (most_common) se extrajeron las 10 palabras con mayor frecuencia en el texto.
- Se separaron etiquetas y valores
- Se creó una figura de tamaño 12 x 6 pulgadas y se usó un gráfico de barras horizontales para poder visualizar las palabras.
- Y finalmente se visualizó el grafico de barras horizontales.

Análisis exploratorio espacial

El análisis exploratorio espacial es una técnica estadística fundamental para estudiar la distribución geográfica de los fenómenos analizados, en este caso, la incidencia de enfermedades según los diferentes sistemas del cuerpo humano en los diferentes estados de México.

Para llevar a cabo este análisis, se elaboraron tres mapas utilizando Python los pasos fueron los siguientes:

- Se importó la base de datos a Python
- Se importó el archivo geoespacial de los diferentes estados (Shapefile)
- Se combinó el DataFrame con el GeoDataFrame espacial mediante una columna común.
- Se verificó que el GeoDataFrame tenga el sistema de referencia espacial (CRS) correcto para México en este caso es EPSG:4326
- Se realizó la creación de los tres mapas diferentes en donde se coloreó los estados según las variables de interés que en este caso fueron las variables de “1 a 4 años”, “5 a 9 años” y “10 a 14 años”.

Análisis multivariado

Para el análisis multivariado los pasos fueron los siguientes:

- Se importaron las librerías necesarias
- Se importaron los datos de población por estado y grupo de edad
- Se aplicó un algoritmo de clustering, como K-means, para identificar grupos de estados con características similares en los tres grupos de edad.
- Y finalmente se graficó el cluster en 3D, donde cada eje representa la población en uno de los grupos de edad.

Estadística no paramétrica

Debido a que los datos no cumplieron con los supuestos de normalidad requeridos para pruebas paramétricas, se optó por utilizar la prueba no paramétrica de Kruskal-Wallis. Esta prueba permite comparar las medianas de más de dos grupos independientes para determinar si existen diferencias estadísticamente significativas entre ellos. Para poder realizarla se llevaron a cabo los siguientes pasos en el software R:

- Los datos se organizaron en un DataFrame, asegurando que cada grupo a comparar estuviera correctamente identificado.
- Se aplicó la prueba de Kruskal-Wallis para comparar los grupos.
- Los resultados obtenidos se organizaron en una tabla para facilitar su interpretación.
- Si el valor p de Kruskal-Wallis es menor que 0.05, se realiza un análisis post hoc con corrección de Bonferroni para identificar entre que grupos existen diferencias.
- Se presentaron los resultados en tablas,

A partir de la nube de palabras se generó un gráfico de barras con las 10 palabras más frecuentes de las cuales fueron las que se muestran en el gráfico en donde las enfermedades más recurrentes están relacionadas con intoxicación, infecciones y metabólicas (Figura 3).

Análisis exploratorio espacial: Distribución de las enfermedades en menores de 1 a 14 años en México.

Figura 4: Mapa por enfermedades que afectan a los diferentes sistemas en la edad de 1 a 4 años



En los estados que más se presentan enfermedades son en estados como Baja California, Baja California Sur, Chihuahua, Coahuila, Campeche, Chiapas, San Luis Potosí, Colima, Michoacán y Aguascalientes. En donde las principales enfermedades afectan a los sistemas como el sistema respiratorio (Infecciones respiratorias agudas, Neumonías y bronconeumonías, Tos ferina, influenza,) sistema digestivo (Infecciones intestinales por otros organismos y las mal definidas, Intoxicación alimentaria bacteriana, Otras infecciones intestinales debidas a protozoarios, Amebiasis intestinal, Otras helmintiasis) sistema cardiovascular (Obesidad) el sistema auditivo (Otitis media aguda) sistema ocular (Conjuntivitis, Conjuntivitis epidémica aguda hemorrágica) sistema tegumentario (Quemaduras, Escabiosis, Varicela).

Estados como Sinaloa, Tamaulipas, Guanajuato y Puebla son los estados que tienen menos registros de enfermedades que afectan a los diferentes sistemas del cuerpo.

Figura 5: Mapa por enfermedades que afectan a los diferentes sistemas en la edad de 5 a 9 años



Los Estados Baja California, Baja California Sur, Aguascalientes, Chihuahua, Coahuila, Colima, Hidalgo, Chiapas y Campeche, tienen presencia de enfermedades que afectan al sistema digestivo (Amebiasis intestinal, Infecciones intestinales por otros organismos y las mal definidas, Otras helmintiasis) sistema respiratorio (Infecciones respiratorias agudas, Faringitis y amigdalitis estreptocócicas, Asma, Neumonías y bronconeumonías), sistema auditivo (Otitis media aguda) sistema tegumentario (Quemaduras, Mordeduras por perro, Varicela, Escabiosis) sistema inmunológico (Desnutrición leve,) sistema cardiovascular (Obesidad), en el rango de edad de la población de 5 a 9 años.

Guanajuato, Tamaulipas y Puebla estos estados en México no presentan un registro mayor de enfermedades en el grupo de edad de 5 a 9 años.

Figura 6: Mapa por enfermedades que afectan a los diferentes sistemas en la edad de 10 a 14 años

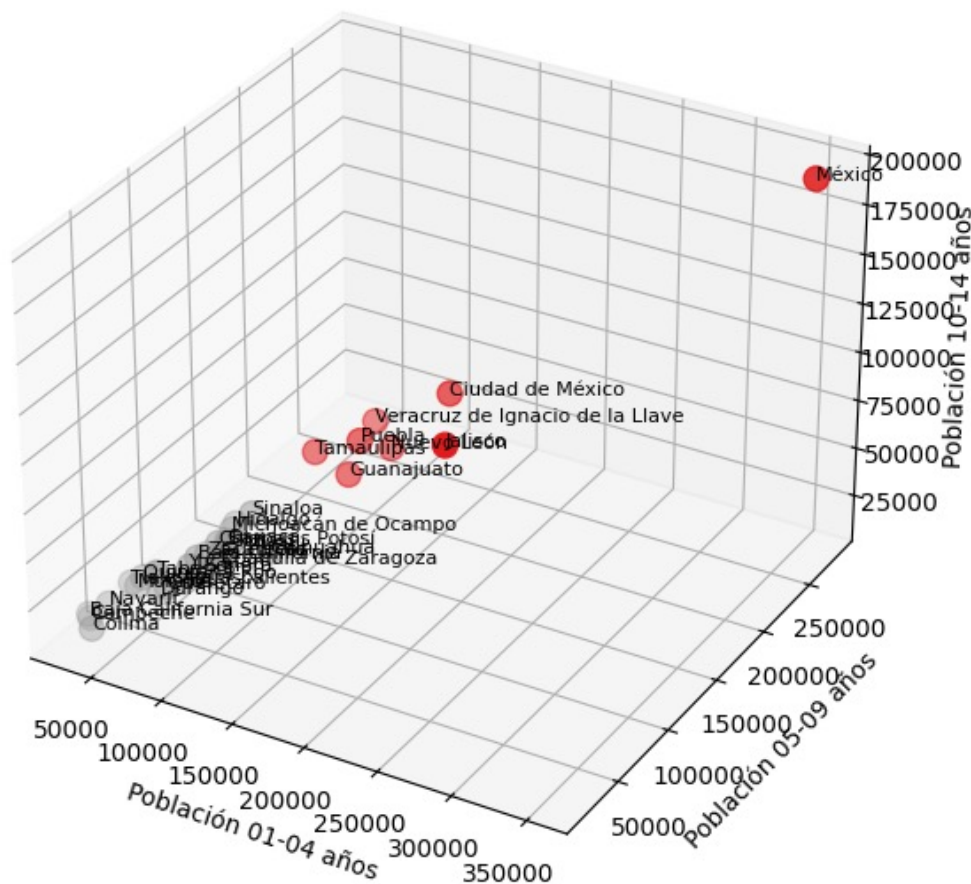


En el grupo de edad de 10 a 14 años los estados de Chiapas, Jalisco, Aguascalientes, Campeche, Colima, Coahuila, Chihuahua, Baja California, Baja California Sur, Veracruz, son los estados en donde hay más presencia de enfermedades las cuales afectan a diferentes sistemas como sistema digestivo (Amebiasis intestinal, Infecciones intestinales por otros organismos y las mal definidas, Otras helmintiasis, Otras salmonelosis, Intoxicación alimentaria bacteriana) sistema respiratorio (Infecciones respiratorias agudas, Asma, Faringitis y amigdalitis estreptocócicas, Neumonías y bronconeumonías) sistema auditivo (Otitis media aguda) sistema tegumentario (Quemaduras, Varicela, Mordeduras por perro, Herida por arma de fuego y punzocortantes, Escabiosis) sistema cardiovascular (Obesidad) sistema urinario (Infección de vías urinarias).

Durango, Nayarit y el estado de México son los estados que en este grupo de edad de 10 a 14 años no presentan muchos registros de enfermedades que afectan a los diferentes sistemas del cuerpo humano.

Análisis multivariado

A continuación, se muestra el gráfico de clúster k medias (Figura 7), para visualizar los estados más parecidos de acuerdo con las enfermedades de los menores de 1 a 14 años en México durante el 2017.

Figura 7: Clusters de estados según la edad.

Este gráfico tridimensional utiliza tres ejes para representar los grupos de edad:

- Eje X: “Población 01-04 años”
- Eje Y: “Población 05-09 años”
- Eje Z: “Población 10-14 años”

Cada punto en el gráfico representa un estado de México.

Puntos rojos (grandes): Representan los estados con mayores valores de enfermedades presentadas en los tres grupos de edad, dispersos hacia la parte superior derecha del gráfico.

Puntos negros (pequeños): Representan estados con menores cantidades de enfermedades en la población infantil, concentrados en la esquina inferior izquierda.

Estados con alta incidencia de enfermedades en menores de 1 a 14 años (en rojo):

- México: Claramente aislado en la esquina superior derecha lo que nos indica que mayores enfermedades presentadas en población infantil en los tres rangos.
- Ciudad de México, Veracruz, Puebla, Jalisco, Nuevo León, Guanajuato, Tamaulipas, forman un subgrupo donde también tiene enfermedades presentadas altas.

Estados con baja incidencia de enfermedades en menores de 1 a 14 años (en negro):

- Colima, Baja California Sur, Campeche, Nayarit, Tlaxcala, Aguascalientes, Zacatecas, etc.

Pruebas estadísticas no paramétricas: Verificar si existe diferencias entre grupos (Clúster) de acuerdo con las enfermedades presentadas en los estados.

Debido a que los datos no cumplieron con los supuestos de normalidad requeridos para pruebas paramétricas, se optó por utilizar la prueba no paramétrica de Kruskal-Wallis. Esta prueba permite comparar las medianas de más de dos grupos independientes para determinar si existen diferencias estadísticamente significativas entre ellos.

Hipótesis nula (H_0): No hay diferencias significativas en la distribución de las poblaciones entre los grupos de edad. Es decir, los estados tienen una distribución similar en las tres categorías de edad.

Hipótesis alternativa (H_1): Al menos un grupo de edad tiene una distribución significativamente diferente de los demás.

Tabla 2. Resultados de la prueba de Kruskal-Wallis

Estadístico de Kruskal-Wallis	P-value
19.3363	0.0001

El valor que nos arrojó el estadístico de Kruskal-Wallis es el resultado del cálculo que compara las medianas de los diferentes grupos, cuanto más grande sea este estadístico, mayor es la evidencia de que al menos un grupo difiere significativamente de los otros.

Como el p-value (0.0001) es mucho menor que 0.05, se rechaza la hipótesis nula. Esto significa que hay diferencias estadísticamente significativas entre al menos dos de los grupos comparados.

Resultados de la prueba post hoc de Dunn con corrección de Bonferroni:

Tabla 3. Resultados de la prueba post hoc de Dunn

	1	2	3
1	1.000000	0.127160	0.000034
2	0.127610	1.000000	0.054123
3	0.000034	0.054123	1.000000

Luego de aplicar la prueba de Kruskal-Wallis, se realizó una prueba post hoc de Dunn con corrección de Bonferroni. Los resultados mostraron comparaciones estadísticamente significativas únicamente entre los Grupos 1 y 3 ($p = 0.000034$), mientras que las comparaciones entre los Grupos 1 y 2 ($p = 0.1276$), y Grupos 2 y 3 ($p = 0.0541$), no fueron significativas.

CONCLUSIONES

Los hallazgos logrados en este estudio es que la salud en menores de edad en México continúa siendo un reto importante, especialmente en los menores de 1 a 14 años, donde las enfermedades infecciosas y no transmisibles siguen representando una parte significativa de la mortalidad en México. A pesar de los avances en la reducción de la mortalidad infantil en las últimas décadas, los datos muestran que las enfermedades como: infecciones respiratorias agudas, intoxicaciones, anomalías congénitas y tumores malignos siguen siendo las principales causas de muerte en este rango de edad, de las cuales afectan a los diferentes sistemas del cuerpo como el sistema respiratorio, sistema digestivo y sistema cardiovascular. Mostrándonos la necesidad de fortalecer las estrategias de prevención y atención oportuna en la infancia.

Estados como Chihuahua, Baja California, Baja California Sur, Sinaloa, Coahuila, Chiapas, México y Campeche presentan altas tasas de mortalidad infantil en los 3 grupos diferentes de edad lo que sugiere que factores como la pobreza, el acceso limitado a servicios de salud y las condiciones ambientales influyen significativamente en la salud de la población infantil.

El estado de México registra el mayor número de casos de enfermedades en los tres diferentes grupos de menores de edad en donde esta situación evidencia la necesidad de fortalecer acciones de prevención, diagnóstico temprano y atención médica para poder mejorar la salud de los menores de edad y puedan tener una infancia plena y sana.

Se pudo observar gracias a la prueba de Kruskal -Wallis que los grupos etarios donde prevalecen más las enfermedades son en los siguientes rangos de edad: de 1 a 4 años y de 5 a 9 años. En donde los principales sistemas afectados son el sistema respiratorio y el sistema digestivo sin embargo los sistemas menos afectados son los siguientes: sistema glandular, sistema hepático, sistema circulatorio, sistema musculoesquelético, sistema cardiovascular, sistema nervioso y sus dos derivados sistema nervioso central y sistema nervioso periférico.

La falta de estudios detallados sobre la distribución de enfermedades según los sistemas afectados y la ausencia de datos específicos por estado limitan la capacidad. Por ello, se recomienda promover investigaciones que profundicen en estos aspectos y fortalezcan la base de datos nacional.

REFERENCIAS

- Bravo-García, E. (2024). *Desigualdad social, pobreza y mortalidad infantil en México 2010–2020*. *Revista de la Facultad de Medicina (México)*, 67(5), 52–63. <https://doi.org/10.22201/fm.24484865e.2024.67.5.07>
- Córdova Sotomayor, D. A., Chávez Bacilio, C. G., Bermejo Vargas, E. W., Jara Ccorahua, X. N., & Santa María Carlos, F. B. (2020). *Prevalencia de infecciones respiratorias agudas en niños menores de 5 años en un centro materno-infantil de Lima*. *Horizonte Médico (Lima)*, 20(1), 54–60. <https://doi.org/10.24265/horizmed.2020.v20n1.08>

- Fernández Cantón, S., & Hernández Martínez, A. M. (2015). *Mortalidad en menores de 15 años por enfermedades no transmisibles: México, 2000–2014*. *Boletín Médico del Hospital Infantil de México*, 72(1), 80–83. <https://doi.org/10.1016/j.bmhmx.2015.03.001>
- García-Marcos, L., Guillen, J. J., Martínez, A., Martín, M., Barbero, P., & Borrajo, E. (1998). *Tasas de mortalidad en la infancia y sus causas en España (1991)*. *Anales Españoles de Pediatría*, 48(1), 39–43.
- Instituto Nacional de Salud Pública (INSP). (2023). *Encuesta Nacional de Salud y Nutrición 2022 (ENSANUT 2022)*. Secretaría de Salud. <https://ensanut.insp.mx/encuestas/ensanut2022>
- Valbuena, L. C. (2017). *Análisis del comportamiento estadístico sobre la incidencia de la enfermedad respiratoria aguda en los niños menores de 5 años en el periodo de 2012 a 2017*. *Revista Aire Libre*, 5.
- Vargas, M., & Reyes López, A. (2024). *Salud infantil en México: problema de política pública*. *Revista Iberoamericana de Ciencias*. <https://www.reibci.org>

INTRODUCCIÓN Y VENTA DE CAFÉ VERACRUZANO EN CHONGQING, CHINA

MAOZHU WANG⁹

ELISEO GABRIEL ARGÜELLES¹⁰

MIRIAM SÁNCHEZ GUEVARA¹¹

Centro de Estudios China – Veracruz (CECHIVER)

Universidad Veracruzana

RESUMEN

El café es uno de los productos agrícolas más relevantes a nivel mundial, con una demanda creciente y una amplia diversidad de mercados consumidores. México ocupa el undécimo lugar global en producción, destacando Veracruz como uno de los principales estados productores por la calidad de su café Arábica, reconocido en América del Norte y Europa. Sin embargo, su presencia en el mercado asiático, particularmente en China, sigue siendo limitada. Chongqing, una de las ciudades con mayor desarrollo económico y cultural del país, representa un mercado emergente con un consumo creciente de café y una industria local en expansión. El presente estudio analiza la viabilidad de introducir café veracruzano en Chongqing, considerando las características del mercado, las preferencias de consumidores y cafeterías, así como los requisitos logísticos y normativos de importación. Los resultados evidencian que el factor decisivo para ingresar exitosamente al mercado chino es la calidad del producto, acompañada de estrategias de promoción que destaquen su origen. Se concluye que la introducción del café veracruzano en Chongqing es viable, siempre que se implementen estrategias de diferenciación, se cumplan los requisitos de importación y se fortalezcan las alianzas comerciales.

Palabras claves: Mercado chino, viabilidad comercial, exportación, comercio internacional.

⁹ maowang@uv.mx

¹⁰ egabriel@uv.mx

¹¹ miriasanchez@uv.mx

INTRODUCCIÓN

El café es actualmente una de las bebidas de mayor consumo en el mundo, con un promedio de 225 millones de tazas al día, y una amplia gama de beneficios que van desde proporcionar energía hasta reducir el riesgo de desarrollar enfermedades como Parkinson (Gotteland y De Pablo, 2007), diabetes tipo 2 (Reis et al., 2019) y enfermedades cardiovasculares (Rodríguez et al., 2017). En la temporada 2020/21, el volumen de café consumido a nivel mundial fue de 133.8 millones de sacos de 60 kilogramos; sin embargo, esta cifra continúa en aumento, alcanzando 173 millones de sacos en 2022/23 (Orús, 2024; Orús, 2024b). Los principales países consumidores son Estados Unidos, Alemania, Japón, Francia, Italia, Rusia, Reino Unido, España, Polonia y Países Bajos (World Population Review, 2024).

En cuanto a la producción, la mayoría del café se cultiva en regiones tropicales cercanas al ecuador, donde las condiciones climáticas favorecen su crecimiento. Brasil, Vietnam, Colombia, Indonesia, Etiopía, Honduras, Uganda, India, Guatemala, Perú y México son los mayores productores mundiales (International Coffee Organization, 2022). En el caso de México, el café llegó en el siglo XVIII a través de Córdoba, Veracruz, introducido por inmigrantes franceses. La topografía, altitud, clima y suelos del país contribuyeron al cultivo de variedades de alta calidad, destacando la especie Arábica como la más producida. La producción nacional representa el 2.4 % del total mundial, posicionando a México como el undécimo productor global. Esta actividad se desarrolla en 15 estados, siendo Chiapas el principal con el 41 % de la producción, seguido de Veracruz con el 24 % y Puebla con el 15.3 % (Secretaría de Agricultura y Desarrollo Rural, 2018). Además, más de tres millones de mexicanos participan en actividades relacionadas con la siembra y cosecha del café.

Existen dos tipos principales de café: Robusta y Arábica. El Robusta es amargo, con mayor cantidad de cafeína y más fácil de cultivar; mientras que el Arábica es de mejor calidad, menos amargo y con menor concentración de cafeína, aunque más costoso y difícil de producir. En México, el 95 % del café corresponde a la variedad Arábica y sólo el 5 % a Robusta. El café mexicano es reconocido por su alta calidad en Norteamérica y Europa, exportándose a más de 50 países. Estados Unidos lidera las importaciones, seguido de Japón, Canadá y Alemania (Exoap, 2021).

Veracruz es el segundo mayor productor de café en México, con 139 mil hectáreas dedicadas al cultivo y alrededor de 86 mil productores, que generan un millón de sacos de café verde al año. Este estado alberga diez regiones cafetaleras destacadas, entre las que se encuentran Huayacocotla y Papantla en la zona norte; Atzacan, Misantla, Coatepec, Huatusco, Córdoba y Zongolica en la zona centro; y Tezonapa y Los Tuxtlas en la zona sur (La Jornada Veracruz, 2022).

Por su parte, Chongqing se ha consolidado como uno de los principales centros de consumo y distribución de café en China. De acuerdo con el Gobierno de China (2018), esta ciudad alberga el mayor centro comercial cafetero del país, conocido como “el repartidor de café de China”, y en 2021 contaba con 2,819 empresas dedicadas a esta industria (Asociación de la Industria del Café de Chongqing). Eventos como el “Quinto Festival Internacional de Café 2023 de Chongqing y la Exposición de Tostado de Café y Bebidas de Moda”, organizado por el Gobierno del Distrito de Yuzhong con apoyo del Comité de Comercio y la Oficina de Asuntos Exteriores, reflejan el dinamismo y la consolidación de esta cultura cafetera.

A pesar de ello, el consumo de café mexicano en China sigue siendo reducido. México exporta la mayor parte de sus granos comerciales a Estados Unidos y los de mayor calidad a países europeos como Alemania y Bélgica. En 2020, las exportaciones a la Unión Europea alcanzaron \$121,694,700 millones (Exoap, 2021), lo que muestra un patrón de comercialización concentrado en Occidente. Aunque existen esfuerzos por parte de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA) para promover el café mexicano en ferias internacionales como la de Hainan, aún queda mucho por hacer para posicionar el café veracruzano en el mercado chino.

En este contexto, diversos estudios han analizado la evolución del mercado mundial del café entre 1980 y 2009, identificando tendencias en producción, comercio y consumo a partir de estadísticas de la FAO y la OIC, sus resultados muestran un crecimiento sostenido en la producción global, con Brasil, Vietnam, Indonesia, Colombia e India como principales productores, y Estados Unidos, Alemania, Italia, Japón y Bélgica como los mayores importadores. Europa se consolidó como el mayor consumidor per cápita, destacando los países nórdicos, mientras que los precios del café mostraron una notable volatilidad durante el periodo analizado. Estos hallazgos permiten comprender la dinámica internacional del café y las desigualdades estructurales en su cadena de valor (Quintero y Rosales, 2014).

Por otra parte, investigaciones recientes se han enfocado en los cafés de especialidad y en las prácticas sostenibles que fortalecen la competitividad de los pequeños productores. Ospina (2017) y Marulanda (2017) destacan el papel de las certificaciones de calidad y origen impulsadas por la Federación Nacional de Cafeteros (FNC), la Asociación de Cafés Especiales de América (SCAA) y otros organismos internacionales. Dichas certificaciones, junto con iniciativas de comercio justo y producción orgánica, han promovido mejores condiciones socioeconómicas, estabilidad en los precios y capacitación técnica entre los caficultores. Aunque persisten barreras de acceso por falta de información y financiamiento, estas estrategias contribuyen a mejorar la sostenibilidad y el posicionamiento del café latinoamericano en los mercados internacionales.

Por todo lo anterior, este capítulo tiene como propósito analizar la viabilidad de introducir y comercializar café veracruzano en Chongqing, China, identificando las preferencias de consumidores y cafeterías, así como las condiciones comerciales y regulatorias del país asiático. Se busca, con ello, contribuir al diseño de estrategias efectivas que permitan posicionar el café veracruzano como un producto competitivo y de alta calidad en un mercado en expansión.

MATERIALES Y MÉTODOS

Esta investigación se desarrolló bajo un enfoque cualitativo, dado que este método permite obtener información profunda y contextual que contribuye a comprender mejor el fenómeno estudiado. El objetivo central fue identificar la viabilidad y las estrategias adecuadas para introducir café del centro de Veracruz en Chongqing, China. Este enfoque permitió recopilar datos significativos mediante instrumentos diseñados para captar percepciones, experiencias y expectativas de los actores involucrados en el mercado cafetero de Chongqing. Las respuestas abiertas obtenidas proporcionaron información de naturaleza cualitativa, base fundamental para el análisis interpretativo y la formulación de conclusiones. Además, el estudio es

de alcance descriptivo, ya que la información recolectada fue analizada e interpretada para responder a las preguntas de investigación planteadas.

En cuanto a la población de estudio y la muestra la investigación consideró tres poblaciones objetivo, ubicadas en Chongqing: consumidores de café, agencias de importación y cafeterías locales. Los informantes fueron seleccionados mediante un muestreo por conveniencia, que incluyó a aquellas personas y representantes de empresas dispuestos a participar y aportar información relevante. Las entrevistas se realizaron a profundidad, buscando obtener una comprensión integral de las preferencias, prácticas comerciales y percepciones sobre el café mexicano.

Para recolectar información se diseñaron tres guiones de entrevistas estructuradas adaptados a cada grupo de interés. El instrumento aplicado a las cafeterías constó de 24 preguntas sobre aspectos relacionados con el café y dos preguntas sociodemográficas. Las variables y dimensiones analizadas se organizaron en torno a categorías como experiencia con el café, requisitos del producto, percepción del café veracruzano y expectativas del mercado.

Tabla 1. Operacionalización de variables para agencias de importación

Tabla 1. Operacionalización de las variables para las agencias de importación.			
Dimensión	Variable	Indicador	Tipo de Variable
Datos Generales	Identificación de la cafetería	Nombre de la cafetería y breve introducción	Nominal
	Identificación del entrevistado	Nombre y cargo de la persona entrevistada	Nominal
Experiencia con café	Origen del café	Países o regiones de origen del café	Nominal
	Canal de obtención del café	Método de compra del café	Nominal
	Factores de selección del café	Criterios de compra de café	Ordinal
	Perfil de clientes	Tipo de clientes principales	Nominal
	Popularidad del café	Tipo de café más popular en la cafetería	Nominal
	Producto más vendido	Artículo de café con mayor demanda	Nominal
	Método de compra de café	Estrategia de adquisición de café	Nominal
	Cantidad de compra anual	Volumen de compra y planificación	Discreta
	Percepción del café veracruzano	Opinión sobre café de Veracruz	Ordinal
	Problemas en el mercado del café	Principales desafíos en el sector del café	Nominal
Requisitos del producto y de la entrega	Tipo de grano de café	Preferencia por tipo de grano	Nominal
	Criterios de calidad del café	Aspectos más valorados en la calidad del café	Ordinal
	Requisitos de origen del café	Preferencias sobre la procedencia del café	Nominal
	Método de procesamiento	Método de procesamiento preferido	Nominal
	Preferencias de sabor	Perfil sensorial deseado	Ordinal
	Método de recolección	Preferencia en la recolección del café	Nominal
	Método de tostado	Preferencia de tostado o café verde	Nominal
	Requisitos de embalaje	Tipo de empaque, tamaño y diseño del embalaje	Nominal
Expectativa del café en el futuro	Tendencias del mercado	Opinión sobre las futuras tendencias del café	Ordinal
	Presupuesto de compra	Presupuesto anual destinado a la compra de café	Continua
	Información deseada sobre origen	Datos esperados de proveedores de café	Nominal
	Potencial de regiones productoras	Regiones con mayor potencial según percepción	Nominal
	Servicios adicionales deseados	Servicios esperados de proveedores de Veracruz	Nominal
	Preferencias de café a futuro	Tipo de café que desearían comprar	Nominal

En cuanto al diseño, tratamiento y análisis de los datos recolectados, se aplicó un análisis cuantitativo de datos cualitativos, que permitió identificar patrones y tendencias en las respuestas. Se elaboraron distribuciones de frecuencia de palabras y nubes de palabras con el propósito de visualizar los términos y conceptos más representativos en las entrevistas. Todo el procesamiento y análisis se realizó utilizando el software R y la plataforma WordArt, herramientas que facilitaron la exploración sistemática del contenido y la identificación de las categorías clave que sustentan la interpretación de resultados.

RESULTADOS

Los resultados obtenidos en este estudio permiten comprender de manera integral las preferencias, percepciones y requisitos del mercado cafetero en Chongqing, China, desde tres perspectivas clave: las agencias importadoras, las cafeterías locales y los consumidores finales. Los hallazgos se interpretan en función de los objetivos de investigación, con el propósito de identificar la viabilidad de introducir el café veracruzano en este mercado y las estrategias más adecuadas para su posicionamiento.

Agencias importadoras

La información recabada con una agencia importadora —Shanghai Carmo Trading Co., Ltd., empresa con 21 años de experiencia en la importación y comercialización de café— ofreció una visión detallada sobre las dinámicas del comercio internacional y las oportunidades que pueden aprovecharse desde México. La empresa basa su modelo de negocio en la adquisición de café verde de alta calidad mediante contacto directo con productores y fincas de diversas regiones, lo que subraya la relevancia de establecer relaciones sólidas y duraderas entre exportadores e importadores.

Los factores determinantes para la importación de café se agrupan en tres categorías principales:

- **Canales de importación:** El contacto directo con los productores es el medio más valorado, pues permite garantizar la trazabilidad y la calidad del café. Además, la empresa utiliza competencias de excelencia de taza para seleccionar cafés de especialidad.
- **Criterios de selección:** La calidad del café verde es el factor decisivo en la compra, por encima del precio, el origen o la sostenibilidad, aunque estos últimos también influyen.
- **Clientes principales:** Las cafeterías y tostadores constituyen el mercado objetivo más relevante, lo que sugiere que la introducción del café veracruzano debe dirigirse inicialmente a estos sectores.

En cuanto a las preferencias de producto, la empresa no se limita a un solo tipo de grano; trabaja tanto con Arábica como con Robusta, priorizando el sabor y la satisfacción del cliente. No establece restricciones sobre el origen geográfico del café, pero sí exige información completa sobre la finca, el clima, el tipo de suelo y el proceso de cultivo, lo cual refuerza la necesidad de una narrativa sólida en torno al producto veracruzano.

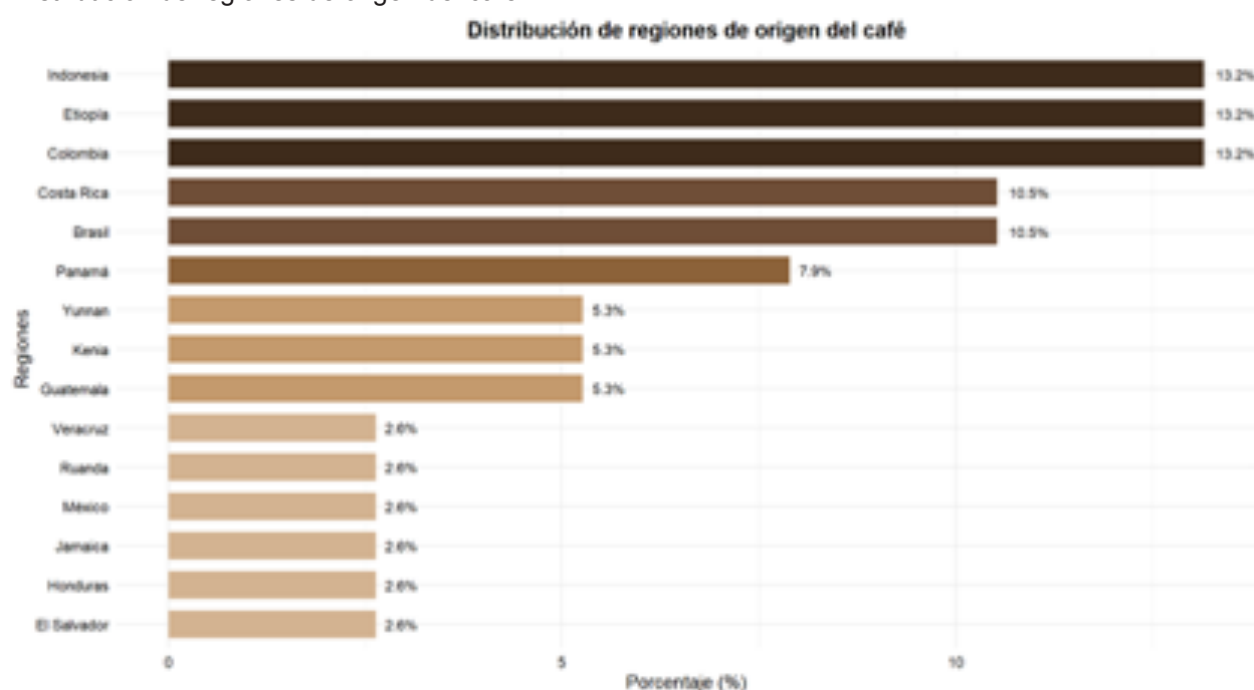
Respecto a las tendencias del mercado, el representante destacó que el consumo de café en China crece entre dos y tres veces por año, con Chengdu y Chongqing como polos emergentes. Sin embargo, la falta de promoción es un obstáculo común: muchas fincas de países productores no logran posicionarse por

no comunicar su historia ni sus atributos diferenciales. En contraste, los cafés panameños han ganado prestigio gracias a estrategias de marketing narrativo y certificaciones de calidad.

Entre los requisitos logísticos, se resaltan los tiempos de entrega y el tipo de transporte: se prefiere el marítimo (50 días) para café comercial y el aéreo (7–14 días) para café de competencia. También se recomienda el uso de sacos de yute con bolsa interior para mantener la calidad y evitar daños durante el traslado. Asimismo, los importadores valoran servicios complementarios, como capacitaciones en cata de café (Q Graders) y tostado personalizado, lo cual representa una oportunidad para el café veracruzano si se vincula con la formación sensorial y la profesionalización del producto.

Finalmente, la agencia entrevistada considera que el café mexicano tiene potencial en el mercado chino, pero requiere estrategias de promoción sólidas, comunicación efectiva con los compradores y adaptación a las condiciones arancelarias y logísticas de China.

Figura 1. Distribución de regiones de origen del café.



Fuente. Elaboración propia.

Cafeterías

El análisis de las entrevistas aplicadas a las cafeterías de Chongqing muestra una visión precisa sobre las tendencias de consumo, las preferencias sensoriales y los requerimientos técnicos del producto.

Regiones de origen del café

Las cafeterías indicaron que las principales regiones de origen del café que comercializan son Indonesia, Etiopía y Colombia, con un 13.2 % cada una, seguidas de Costa Rica y Brasil (10.5 % cada una). En menor proporción aparecen Panamá (7.9 %) y otras regiones como Yunnan, Kenia y Guatemala (5.3 %). En este panorama, México y Veracruz aparecen con una representación baja (2.6 %), lo que confirma el poco posicionamiento del café mexicano en este mercado, pero también una oportunidad para diversificar la oferta mediante cafés de alta calidad.

Preferencias de consumo

En cuanto a los productos más vendidos, las cafeterías manifestaron una clara preferencia por el Latte (33.3 %) y el Café Americano (26.7 %), que en conjunto representan el 60 % de las ventas. Esto sugiere que el consumidor de Chongqing aprecia las bebidas equilibradas y suaves, que combinan el sabor del espresso con leche o agua caliente. Otras opciones, como los métodos manuales de preparación (13.3 %), también tienen relevancia entre los clientes especializados.

El análisis de la nube de palabras sobre preferencias de sabor revela que los términos más mencionados fueron “suave”, “dulce”, “acidez”, “chocolate” y “nuez”, lo que indica una preferencia por cafés balanceados, de cuerpo medio y con notas aromáticas agradables.

Tipo de grano y procesamiento

El 62.5 % de las cafeterías prefieren el café Arábica, mientras que un 25 % opta por mezclas y solo un 12.5 % utiliza Robusta. En cuanto al método de procesamiento, el lavado es el más valorado, seguido del natural, que se asocia con sabores más intensos. Estos resultados confirman que el café veracruzano lavado de tipo Arábica tiene alta posibilidad de aceptación.

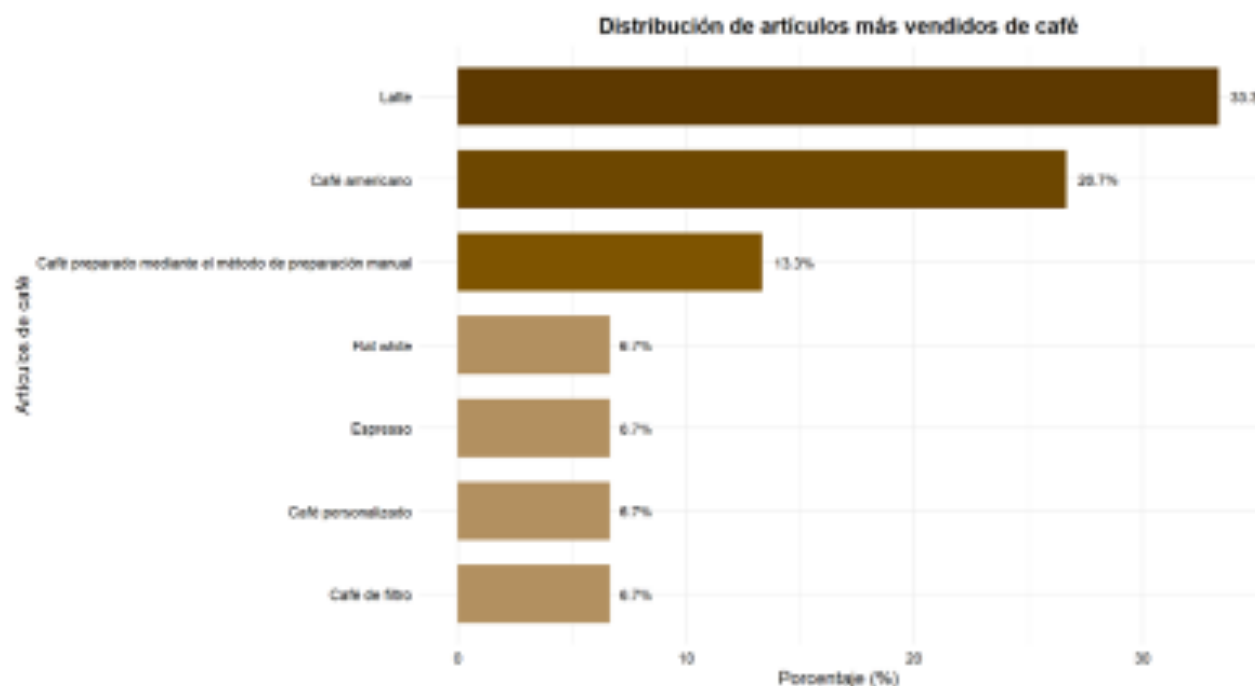
En términos de tueste, los niveles medio oscuro y oscuro son los preferidos, ya que permiten resaltar las notas de chocolate y caramelo, características que coinciden con las propiedades sensoriales del café veracruzano. Asimismo, las cafeterías valoran en la calidad del café aspectos como fragancia, aroma y sabor, identificados como los indicadores más importantes de excelencia.

Empaque e información del producto

En relación con los requisitos de embalaje, el 85.8 % de las cafeterías indicó requerir empaques ecológicos, herméticos y de alta protección. Este aspecto representa un elemento diferenciador en el mercado, donde el diseño sostenible y la presentación estética influyen significativamente en la decisión de compra. Además, la mayoría de los entrevistados consideró esencial recibir información sobre el origen, el cultivo y las características del grano, lo cual refuerza la importancia de contar con etiquetas y materiales informativos detallados.

Servicios y expectativas

Las cafeterías expresaron interés en recibir apoyo en la cultura del café, capacitaciones en catación y asesoramiento técnico de los proveedores. Estos servicios suman el 57.2 % de las preferencias, lo que muestra la apertura a colaboraciones de largo plazo con productores y exportadores. Asimismo, destacaron regiones como África (27.3 %), América (18.2 %) y Yunnan (18.2 %) como zonas de alto potencial, lo que sugiere que Veracruz podría posicionarse si comunica su calidad y sostenibilidad con estrategias adecuadas.

Figura 2. Distribución de artículos más vendidos de café.

Fuente. Elaboración propia.

Consumidores

El análisis de los datos obtenidos de los consumidores de Chongqing refuerza la visión de que el café Arábica es el más apreciado y que la calidad sensorial —particularmente el aroma, la fragancia y el sabor— determina la elección.

Origen y factores de compra

Las regiones preferidas por los consumidores son Etiopía, Colombia y Brasil, que representan el 45.1 % del total. Los principales factores de compra son la calidad (37 %), el precio (33.3 %) y el sabor (18.5 %), seguidos en menor medida por el origen y el contenido de cafeína. Estos resultados confirman que la calidad es el principal determinante del consumo, pero que la competitividad en precios es igualmente relevante para lograr penetración en el mercado.

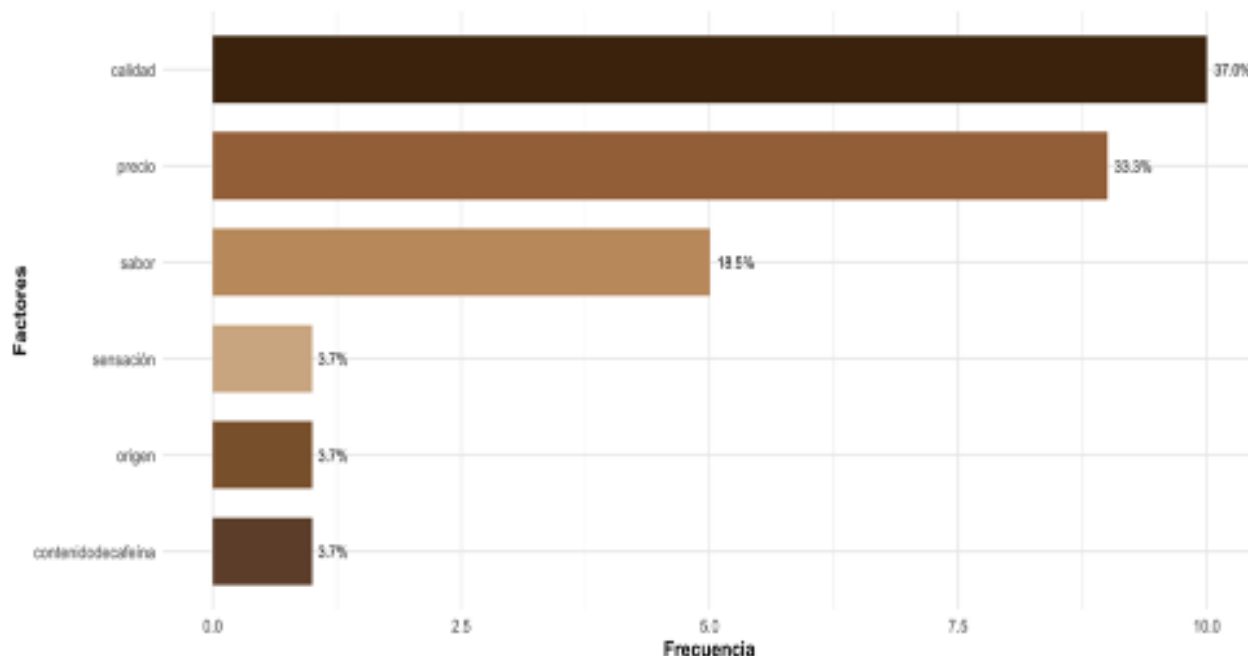
Atributos sensoriales y preferencias

Los consumidores asocian la calidad del café con tres atributos esenciales: aroma (32.3 %), sabor (32.3 %) y fragancia (19.4 %). Además, valoran la complejidad, el nivel de tostado y la sensación en boca como características que definen un buen café. En cuanto a las bebidas preferidas, el café americano (31.6 %) y el método filtrado o de goteo manual (21.1 %) son los más populares, seguidos por el latte (15.8 %) y el espresso (10.5 %).

El grano Arábica es claramente dominante con un 81.8 % de preferencia, mientras que en el método de procesamiento predominan los cafés naturales (40 %) y lavados (26.7 %), por encima de los procesos honey o mixtos. En cuanto al tostado, los consumidores se inclinan hacia niveles medio oscuro (28.57 %) y oscuro (21.43 %), que ofrecen sabores más intensos y persistentes.

Por último, en lo que respecta al empaque, los consumidores prefieren opciones sostenibles, herméticas y visualmente atractivas, que preserven la frescura del café y sean fáciles de transportar. Este factor refuerza la necesidad de que el café veracruzano adopte una presentación acorde con las tendencias globales de consumo responsable.

Figura 3. Factores importantes para las consumidores de café.



Fuente: Elaboración propia.

Los resultados muestran que la introducción del café veracruzano en Chongqing es viable y competitiva, siempre que se base en tres pilares esenciales: la calidad certificada del producto, con énfasis en la trazabilidad, el tueste adecuado y la narrativa del origen. Las estrategias de promoción diferenciadas, que comuniquen la historia, las características sensoriales y los valores culturales del café veracruzano y la adaptación a las tendencias locales, incluyendo empaques ecológicos, sabores balanceados y colaboración directa con importadores y cafeterías.

Estas evidencias respaldan la conclusión de que el café veracruzano puede consolidarse en el mercado chino, especialmente si se posiciona como un producto de alta calidad con identidad regional, respaldado por una estrategia integral de comercialización.

CONCLUSIONES

El análisis realizado demuestra que la introducción del café veracruzano en Chongqing, China, es viable, siempre que se sustente en estrategias de calidad, diferenciación y adaptación al mercado local. Los hallazgos confirman que el consumidor y las cafeterías chinas valoran profundamente la excelencia sensorial del producto, la historia del origen y la autenticidad cultural del café, factores en los que el café veracruzano tiene una ventaja competitiva clara. La calidad se posiciona como el criterio de compra más importante tanto para las agencias importadoras como para los consumidores, lo que refuerza la necesidad de mantener altos estándares de producción, procesos controlados de lavado y tostado, y certificaciones de origen que avalen su prestigio internacional.

Asimismo, la investigación revela que el mercado cafetero de Chongqing se encuentra en expansión acelerada, con un crecimiento anual estimado entre dos y tres veces, impulsado por el auge de cafeterías especializadas y una creciente cultura de consumo. En este escenario, el café veracruzano puede consolidarse como un producto diferenciado si logra comunicar eficazmente sus atributos —acidez balanceada, sabor dulce, aroma floral y notas de chocolate— y si se acompaña de empaques sostenibles, herméticos y visualmente atractivos.

Otro hallazgo relevante es la importancia de establecer relaciones directas y duraderas con los importadores y tostadores locales, mediante un intercambio constante de información, envío de muestras y colaboración técnica. La experiencia de las agencias entrevistadas resalta que los proveedores exitosos no solo venden café, sino que ofrecen conocimiento, capacitación y acompañamiento. En este sentido, las alianzas estratégicas entre productores veracruzanos y actores chinos se perfilan como un elemento esencial para el posicionamiento sostenible en el mercado.

Finalmente, se concluye que el potencial de ingreso del café veracruzano en Chongqing depende de su capacidad de adaptación a las tendencias globales, especialmente en sostenibilidad, trazabilidad y comunicación del valor social del producto. Con una adecuada planificación comercial, una narrativa sólida sobre su origen y una estrategia de vinculación que combine calidad, identidad y cooperación internacional, el café veracruzano puede consolidarse como una marca de prestigio y competitividad en el mercado asiático, contribuyendo al desarrollo económico regional y a la proyección internacional del café mexicano.

REFERENCIAS

- Exoap. (2021). *Comercio Internacional del café de México*. <https://exoap.com/exportacion-de-cafe/>
- Gobierno de China. (2018). *Industria cafetera en Chongqing*.
- Gotteland, M., y De Pablo, V. (2007). *Efectos del consumo de café en la salud*.
- International Coffee Organization. (2022). *Trade Statistics Tables*. <https://www.ico.org/>
- La Jornada Veracruz. (2022, 26 enero). *Ataque a beneficio de café afectará a los productores, advierte Consejo Regional*. La Jornada Veracruz. <https://jornadaveracruz.com.mx/especiales/ataque-a-beneficio-de-cafe-afectara-a-los-productores-advierte-consejo-regional/>
- Marulanda, A. O. (2017). *La industria de los cafés de especialidad. Relación entre certificaciones de calidad o especialidad, y los ingresos de los caficultores en Colombia (1989 - 2015)*. Repositorio Institucional EdocUR. <https://repository.urosario.edu.co/handle/10336/13770>
- Orús, A. (2022). *Volumen total de café exportado e importado a nivel mundial entre 2005/06 y 2023/24*. Statista. <https://es.statista.com/estadisticas/1287679/comercio-internacional-de-cafe-a-nivel-mundial/>
- Orús, A. (2022b). *Volumen de café consumido a nivel mundial entre 2012 y 2024*. Statista. <https://es.statista.com/estadisticas/600964/consumo-global-de-cafe-2009/>

- Ospina, A. (2017). *La industria de los cafés de especialidad. Relación entre certificaciones de calidad o especialidad, y los ingresos de los caficultores en Colombia. 1989 - 2015* [Trabajo de grado, Universidad del Rosario]. Repositorio Institucional Universidad del Rosario. <https://repository.urosario.edu.co/handle/10336/13770>
- Quintero Rizzuto, M. L., & Rosales, M. (2014). El mercado mundial del café: tendencias recientes, estructura y estrategias de competitividad. *Visión Gerencial*, (2), 291-307.
- Reis, C. E. G., Dórea, J. G., y Da Costa, T. H. M. (2019). *Effects of coffee consumption on glucose metabolism: A systematic review of clinical trials*. Journal of Traditional and Complementary Medicine, 9(3), 184-191. <https://doi.org/10.1016/j.jtcme.2018.01.001>
- Rodríguez, N. Y. M., Montero, J. M., Pérez, M. L. M., De Peralta, M. G., & Hechavarria, Y. C. (2017). *Efecto anticariogénico del café*. DOAJ. <https://doaj.org/article/cd2ba4f7ec1a4e18b5b253047a0115b2>
- Secretaría de Agricultura y Desarrollo Rural. (2018). *Producción de café en México y principales estados productores*.
- World Population Review. (2024). Coffee Consumption by Country 2024. <https://worldpopulationreview.com/country-rankings/coffee-consumption-by-country>

ANÁLISIS ESTADÍSTICO PARA LA GESTIÓN DEL RIESGO EDUCATIVO Y DEL BIENESTAR ESTUDIANTEL

MARÍA FERNANDA CAMPOS CARMONA¹²

CECILIA CRUZ LÓPEZ¹³

Facultad de Estadística e Informática

Universidad Veracruzana

RESUMEN

El presente estudio analiza los niveles de depresión, ansiedad y estrés en estudiantes de licenciatura del área de Ciencias de la Salud de una Universidad Pública. A través de un enfoque cuantitativo, se aplicaron instrumentos estandarizados para medir la presencia y severidad de estos factores emocionales, identificando su relación con variables sociodemográficas y académicas. Los resultados evidencian una alta prevalencia de síntomas asociados a la carga académica, la presión por el rendimiento y los procesos de adaptación universitaria. Se observa que la ansiedad es el trastorno más frecuente, seguida del estrés y la depresión. Estos hallazgos destacan la necesidad de fortalecer programas de apoyo psicológico y estrategias institucionales de bienestar estudiantil, orientadas a promover la salud mental, mejorar el rendimiento académico y prevenir el abandono escolar en la comunidad universitaria del área de Ciencias de la Salud.

Palabras clave: Depresión, ansiedad, estrés, regresión logística ordinal, Random Forest, Fiabilidad, Validación

¹² gS19013294@egresados.uv.mx

¹³ ceccruz@uv.mx

INTRODUCCIÓN

Se ha verificado que el ambiente universitario conlleva que el estudiante obtenga muchas más responsabilidades y estas a su vez se ven cada vez más complejas. Para Moterrosa et al. (2020) los estudiantes del área de ciencias de la salud son quienes se ven mayormente sometidos a estas sobrecargas académicas y esto se debe a las condicionantes que pueden llegar a existir en algunas materias educativas de su mapa curricular, influyendo aquí el volumen de conocimiento que deben obtener y las habilidades que deben desarrollar en las prácticas clínicas o de laboratorio, por lo tanto, son propensos a generar una mayor alteración psicológica.

La prevalencia de estrés, ansiedad y depresión en los estudiantes del área de ciencias de la salud, donde tiene una carga académica mucho más alta, es mayor, comparado con otras áreas académicas, en estos estudiantes se observa una mayor prevalencia que la sociedad común. Esto tiende a afectar el rendimiento académico y también, el aumento de la probabilidad de abandonar la carrera universitaria (Trunce, et al. 2020.)

La comprensión de la relación que pueda o no existir entre los factores y los trastornos mentales, hará posible desarrollar estrategias que mejoren el bienestar emocional y prevenir que exista un bajo rendimiento académico o incertidumbre alta de abandonar la carrera. Por esto mismo, se propuso realizar una investigación con base en un cuestionario que buscó ver la presencia de las características de los síntomas de depresión, ansiedad y estrés en los estudiantes del área de ciencias de la salud a nivel licenciatura.

Antecedentes

Gutiérrez et al. (2010) dejan claro que los estudiantes del nivel superior son quienes pueden alterar con mayor frecuencia sus condiciones de salud mental debido a las exigencias académicas, el desarrollo para el aprendizaje, etc. Entre los resultados que se obtuvieron, resaltan que se identificó una relación entre el estrés de origen académico y la presencia de depresión de los estudiantes, los estresores más señalados como generadores de alto nivel de estrés, severo y moderado, también presentaron alta tendencia en la presencia de depresión. En este caso, el 73.9% de los estudiantes con nivel severo de estrés fue originado por la competitividad con sus compañeros, cuentan con un nivel de depresión. Agregando que también influyó la carga académica, los planes a futuro y las presentaciones orales.

Un estudio realizado por Trunce et al. (2020) mostró que 29% de los estudiantes presentaron cuadros depresivos, el 53.20% ansiedad y el 47.80% estrés. También arrojó que la carrera donde se reportó una mayor presencia de niveles de depresión es nutrición y dietética contando con un 36.6% y ansiedad un 61%, mientras que para Kinesiología se reportó un porcentaje más alto de estrés (53.8%). No se observó relación significativa entre el promedio de los estudiantes y la presencia de un trastorno depresivo y estrés, pero para el caso de la ansiedad, si existe una asociación, pues se obtuvo un valor de $p = 0.016$, el cual es menor al 0.05.

Cabezas et al. (2021) destacaron en su artículo que la depresión y ansiedad se consideraban como los desórdenes que más se han presentado en los centros universitarios, asimismo, contemplaron al estrés como un asunto muy relevante e inquietante a nivel mundial. Por esto, es que no se tiene que pasar por

alto, que los estudiantes son quienes se han visto en diversas ocasiones, expuestos a la presencia de altos niveles de los trastornos ya mencionados.

Pulido et al. (2011) desarrollaron un estudio donde uno de los principales objetivos fue identificar el perfil de los estudiantes con mayores niveles de estrés académico. La población bajo estudio fueron los estudiantes de la Universidad Intercontinental, tomando una muestra de 504 individuos por medio de la selección aleatoria (participando todas las carreras a excepción de Pedagogía y Teología). El cuestionario aplicado es el SEEU (sobre estrés en estudiantes universitarios), el cual fue construido considerando la presencia o ausencia de los síntomas de estrés, contando con una consistencia interna de un Alpha de Cronbach 0.950 y una validación del constructo por medio de extracción de factores identificando 6 principales, explicando el 45% de la varianza total.

Por otro lado, Manríquez et al. (2021) realizaron un estudio donde principalmente buscaban determinar la correlación del estrés, ansiedad y depresión con el síndrome de burnout estudiantil en estudiantes de la licenciatura en enfermería de una Universidad Pública de México. La población implicada fueron los estudiantes de enfermería de una Universidad Pública y el muestreo aplicado es de tipo aleatorio estratificado de 256 estudiantes pertenecientes a 8 estratos. Para la colecta de datos sobre la presencia de ansiedad, depresión y estrés fue aplicado el instrumento DASS-21, contando con 21 ítems y 4 posibles respuestas en tipo escala de Likert. Para el caso de la presencia de síndrome de burnout estudiantil se aplicó la escala unidimensional del síndrome de burnout del Dr. Arturo Barraza Macias, el cual cuenta con 15 ítems en tipo escala de Likert. Dentro de los resultados de este estudio y respecto a la relación entre el estrés, ansiedad y depresión, se determinó que si existe la relación con el síndrome de burnout estudiantil, contando con un valor de 0.513 para ansiedad, 0.609 para depresión y 0.17 para el estrés, por lo cual se rechazó la hipótesis nula, la cual indica que no existe correlación significativa, es decir, que estos trastornos mentales si cuentan con una correlación significativa con el síndrome en los alumnos de enfermería de una universidad pública.

Vallejo et al. (2019) abordaron una investigación, donde para evaluar la sintomatología depresiva y de ansiedad utilizaron la escala de Goldberg, siendo aplicado en 4,543 estudiantes de una universidad pública en Poza Rica, Veracruz. Resaltando que, para el área de ciencias de la salud, fue la que reporto una sintomatología depresiva presente más alta que las demás áreas, aunque también se puede destacar que son los que más consumieron sustancias como alcohol (23% con sintomatología depresiva), tabaco (22% con sintomatología depresiva) o marihuana (27% con sintomatología depresiva).

Por otra parte, Ochoa et al. (2021), aplicaron la escala de Halminton (HARS) para evaluar la presencia de ansiedad y depresión, participaron 300 estudiantes de medicina, odontología, nutrición, psicología, enfermería y bioanálisis de la región Veracruz. Quienes no presentaron ansiedad son las carreras de medicina con el 50%, nutrición con el 48% y bioanálisis con el 46%; en ansiedad leve odontología, psicología y nutrición obtuvieron el 52% y enfermería el 50%; en ansiedad moderada están situados los alumnos de psicología y enfermería con el 16% y finalmente para ansiedad grave bioanálisis con el 4%. Hablando ahora de depresión, quienes no la padecen son nutrición (74%), enfermería (68%) y medicina (66%); en el rango de depresión leve odontología presento el 40%; Para el caso de psicología, presento el 22% en depresión moderada y 14% en depresión grave.

MATERIALES Y MÉTODOS

La investigación fue de tipo cuantitativa, con un enfoque descriptivo, correlacional y predictivo. El fin de la investigación fue describir, relacionar y analizar factores asociados con el bienestar psicológico y el desempeño académico de los estudiantes a nivel licenciatura del área de ciencias de la salud de la región Xalapa de la Universidad Veracruzana.

Muestra

La población objetivo fueron los estudiantes de nivel licenciatura del área de ciencias de la salud de la región Xalapa de la Universidad Veracruzana. Para esto se consideraron seis facultades participantes las cuales son: Facultades de Medicina, Odontología, Bioanálisis, Enfermería, Psicología y Nutrición. Se aplicó un muestreo no probabilístico por conveniencia.

Instrumento y recolección de datos

Se diseñó un cuestionario sobre la depresión, ansiedad y estrés y la composición de este es de 21 preguntas, donde 14 fueron en escala de Likert, 3 dicotómicas y 4 abiertas. Se recabaron datos sociodemográficos y académicos como la edad, género, carrera, semestre, materias inscritas, promedio general obtenido. Se construyeron tres constructos: estrés (preguntas 4 a la 7), ansiedad (preguntas 8 a la 12) y depresión (preguntas 13 a la 17), los cuales se conformaron de ítems negativos. Se estableció puntajes de estos se muestran en la **Tabla 1**.

Tabla 1. Puntaje para ítems negativos en escala de Likert					
	Siempre	Casi siempre	Algunas Veces	Muy pocas veces	Nunca
Ítems negativos	1	2	3	4	5

El cuestionario fue aplicado a los estudiantes inscritos en el periodo agosto 2024 – enero 2025. El tiempo de aplicación del cuestionario fue de un mes, del 11 de noviembre de 2024 al 11 de diciembre de 2024. Se dividió en dos fases, la primera con los datos recabados hasta el 25 de noviembre para la prueba piloto. La segunda fase finalizó el 11 de diciembre, donde se lograron obtener 306 observaciones en total. Dicho cuestionario fue aplicado en modalidad en línea, por medio de la aplicación forms, perteneciente a Microsoft 365.

El software Excel se utilizó para la limpieza, filtración y codificación de los datos y Rstudio versión 4.4.2 para el análisis y procesamiento de los datos.

Fiabilidad y validación

Para evaluar la fiabilidad y validez del instrumento, se utilizó el coeficiente Alfa de Cronbach, con el propósito de medir la consistencia interna de los ítems. Asimismo, se realizaron tanto un análisis factorial exploratorio (AFE) como un análisis factorial confirmatorio (AFC). El AFE permitió identificar la estructura subyacente de las variables y agrupar los ítems según sus dimensiones teóricas, mientras que el AFC se empleó para comprobar el ajuste del modelo propuesto y confirmar la validez de constructo del instrumento aplicado.

Regresión logística ordinal

Para el análisis de los factores asociados a los niveles de depresión, ansiedad y estrés, se aplicó una regresión logística ordinal, adecuada para variables dependientes con categorías ordenadas. Este método permitió identificar la probabilidad de que los estudiantes presentaran niveles más altos de malestar psicológico en función de diversas variables predictoras, como el género, la edad, el semestre cursado y la carga académica percibida. A través de los coeficientes obtenidos, se evaluó la fuerza y dirección de las asociaciones, determinando qué factores incrementan o reducen la probabilidad de presentar síntomas más severos. El modelo proporcionó una visión integral sobre la influencia de las variables personales y académicas en la salud mental del estudiantado.

RESULTADOS

Fiabilidad y validación

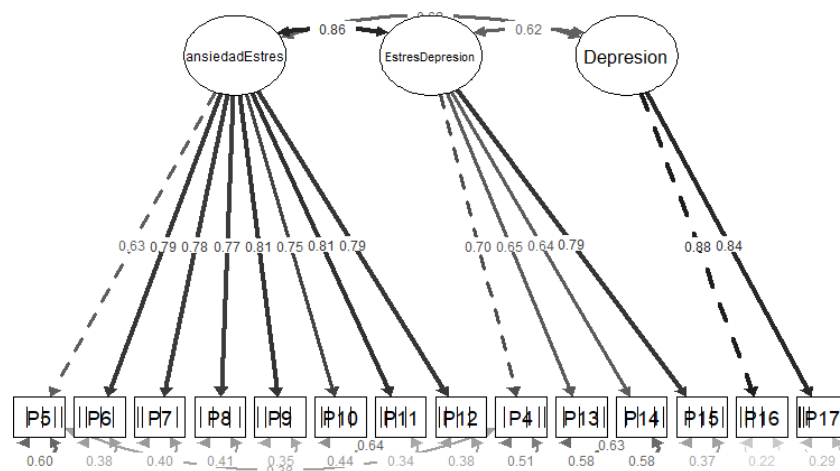
Para la fiabilidad, se optó evaluarla mediante el Alpha de Cronbach, con una ponderación alta para el cuestionario general de 0.91, contando con una muy buena consistencia interna. Para cada uno de los constructos arrojó valores aceptables como se muestra en la Tabla 2, lo cual garantiza una fiabilidad adecuada para este cuestionario.

Tabla 2. Coeficiente de Alpha de Cronbach por constructo

Constructo	Coeficiente	Interpretación
Estrés	0.8	Buena, con una fiabilidad alta.
Ansiedad	0.88	Buena, con una fiabilidad alta
Depresión	0.77	Aceptable, una fiabilidad adecuada.

En el caso de la validación, por medio de la re-especificación basada en los índices de modificación se hipotetizó un modelo revisado donde se agregó el supuesto de las correlaciones entre los ítems p3 y p14, p10 y p11, p y p5 (Figura 1), corroborando por medio de los índices de bondad de ajuste que se tiene un buen y excelente ajuste del modelo, por medio de WRMR: 0.762, CFI: 0.926, TLI: 0.906, SRMR: 0.057, AGFI: .985, GFI: 0.993, RFI: 0.954, aceptando dicho modelo.

Figura 1 Modelo de ecuaciones estructurales aplicando la sugerencia de los MI en el nuevo modelo



Estadística descriptiva

Un total de 306 participantes en este estudio, de los cuales 233 estudiantes pertenecen al género femenino y 71 al género masculino, mientras que 2 especificaron que se identifican con otro género (Figura 2), además de la edad, donde los estudiantes de 17 a 20 años (155) fueron quienes más participaron en la aplicación del instrumento, por debajo de ellos 137 estudiantes de 21 a 24 años, para el intervalo de 25 a 28 años participaron 13 estudiantes y por último para el intervalo de 29 a 32 años solo colaboró 1 estudiante (Figura 3).

Figura 2 frecuencias sobre el género de los estudiantes que participaron en la aplicación del cuestionario.

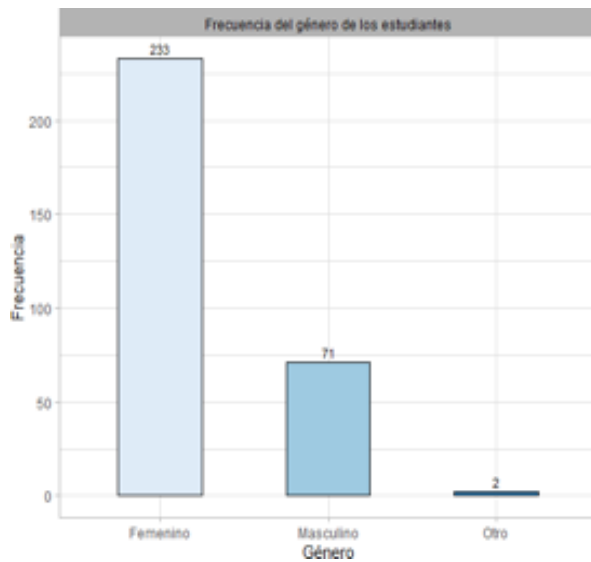
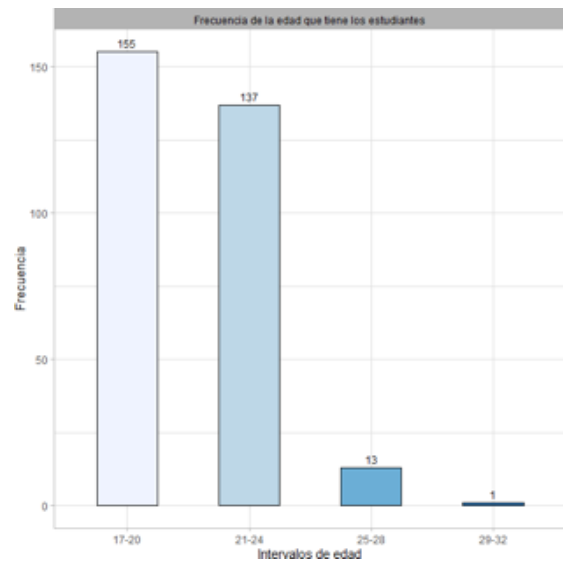


Figura 3 frecuencias sobre los intervalos de edad de los estudiantes que participaron en la aplicación del cuestionario.



Las licenciaturas participantes fueron del área de ciencias de la salud de la región Xalapa, 86 estudiantes cursan la licenciatura en Química clínica, 75 cirujano dentista, 49 enfermería, 41 psicología, 32 médico cirujano y por último 23 la licenciatura en nutrición (Figura 4).

Figura 4 frecuencias para las licenciaturas cursadas por los estudiantes participantes.

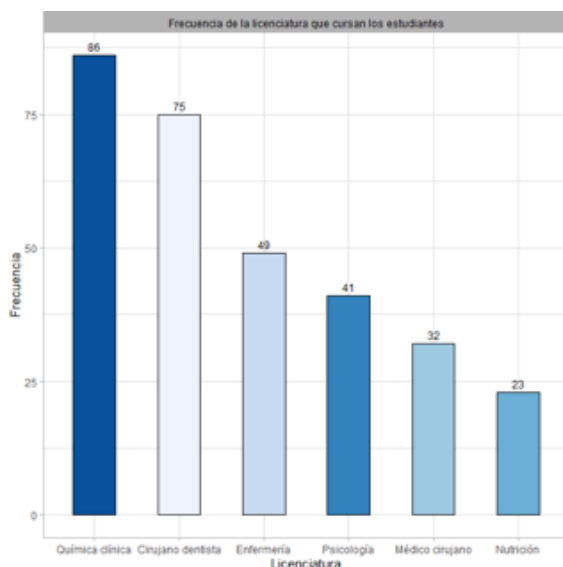
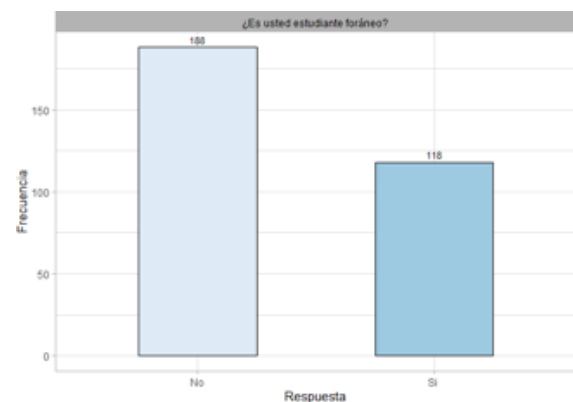


Figura 5 frecuencias respecto a si los estudiantes son foráneos o no.



188 no son estudiantes foráneos, mientras que 108 si lo son (Figura 5). Se puede resaltar que la mayoría de los estudiantes pertenecen a municipios que se encuentran en el estado de Veracruz y otros más pertenecen a estados como por ejemplo Tabasco, Puebla, Oaxaca, Hidalgo, por mencionar algunos.

Dentro del constructo “estrés” la escala de valoración fue: de 8 a 9 puntos “Estrés leve”, 10 a 12 “Estrés moderado”, 13 a 16 “estrés severo” y de 17 o más “estrés extremadamente severo”. En la Figura 6 se presenta la frecuencia de las escalas de valoración que arrojaron los puntajes para el constructo “estrés”. De acuerdo con el gráfico 102 estudiantes fueron quienes presentan estrés moderado, 97 estrés severo, 45 presentan estrés leve, 39 se ubicaron dentro de la ponderación de sin presencia de estrés y 23 en estrés extremadamente severo.

Figura 6 frecuencias de las escalas de valoración para el estrés.

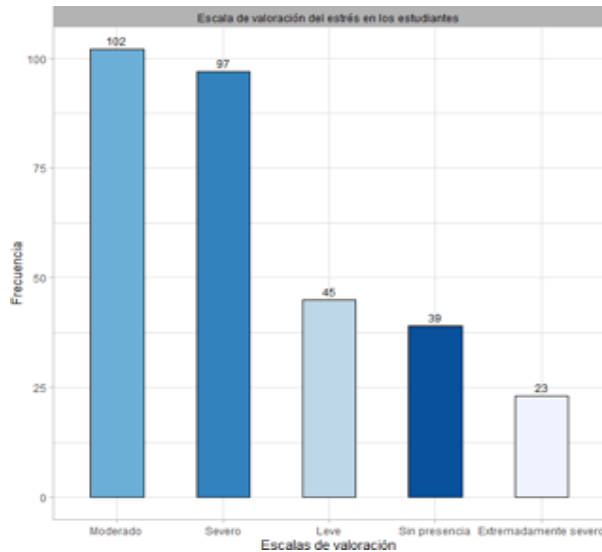
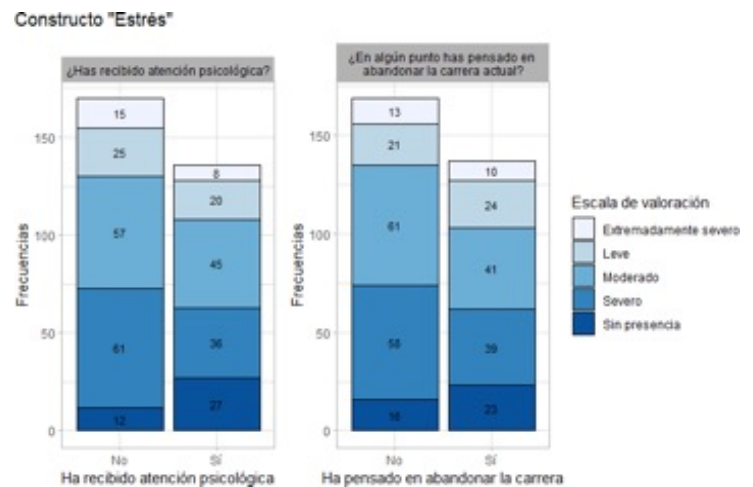


Figura 7 frecuencias de las escalas de valoración para el estrés.



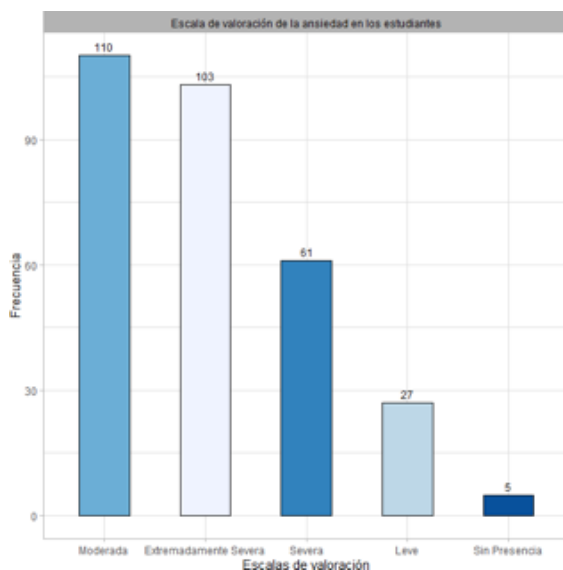
En la Figura 7 respecto a la atención psicológica que, si han recibido, 61 estudiantes están dentro de la valoración de estrés severo, 57 dentro del estrés moderado, 25 en estrés leve, 15 en extremadamente severo y 12 sin presencia de estrés. Para el caso de los estudiantes que no recibieron atención psicológica, 45 son los que se encuentran dentro de estrés moderado, 36 en estrés severo, 27 están sin presencia de estrés y 8 en extremadamente severo.

También se observa el grafico que hace referencia a si pensaron en abandonar la carrera actual, para quienes no han pensado en abandonarla, 61 están dentro de estrés moderado, 58 dentro de estrés severo, 21 en estrés leve, 18 sin presencia de estrés y 13 en estrés extremadamente severo. En el caso de quienes, si consideraron en abandonar la carrera, 41 se encuentran en estrés moderado, 39 en severo, 24 en estrés leve, 23 sin presencia de estrés y por último 10 en estrés extremadamente severo.

Para el constructo “ansiedad” la escala de valoración fue dada de la siguiente manera: de 7 a 9 puntos ansiedad leve, de 10 a 16 puntos ansiedad moderada, de 17 a 19 ansiedad severa y de 20 o más puntos se considera ansiedad extremadamente severa. La frecuencia de estas escalas de valoración está representada en la Figura 8, donde se puede mencionar que 110 son los estudiantes que se encontraron dentro de la escala de valoración de ansiedad moderada, 103 son quienes están en ansiedad extremadamente severa, 61 en ansiedad severa, 27 en ansiedad leve y por último son 5 los estudiantes que están sin presencia de ansiedad.

Para el caso de que no han recibido atención psicológica (Figura 9) son 69 estudiantes los que se encontraron en ansiedad extremadamente severa, 61 en ansiedad moderada, 28 en ansiedad severa, 9 en ansiedad leve y 3 sin presencia de ansiedad.

Figura 8 frecuencias de las escalas de valoración para la ansiedad.



Para los que, si han recibido la atención psicológica, 49 estudiantes se ubicaron en ansiedad moderada, 34 en ansiedad extremadamente severa, 33 en ansiedad severa, 18 en ansiedad leve y 2 sin presencia de ansiedad. En el segundo gráfico de la Figura 9, se observa la distribución de frecuencia de si han considerado o no el abandonar su carrera actual, donde quienes no han considerado el abandonarla son 60 estudiantes en ansiedad extremadamente severa, 59 en ansiedad moderada, 36 en ansiedad severa, 12 en ansiedad leve y 2 sin presencia de ansiedad. Mientras que para los que si han considerado abandonar la carrera son 51 estudiantes en ansiedad moderada, 43 en ansiedad extremadamente severa, 25 en ansiedad severa, 15 en ansiedad leve y 3 sin presencia de ansiedad.

Para las escalas de valoración de la depresión, al igual que la ansiedad, contaron con la misma ponderación: de 7 a 9 depresión leve, de 10 a 16 depresión moderada, de 17 a 19 depresión severa y 20 o más puntos está en depresión extremadamente severa. En la Figura 10 se muestra la distribución de frecuencias de las escalas de valoración para depresión, donde, 118 estudiantes son quienes se ubicaron dentro de la depresión moderada, 97 en depresión severa, 71 en depresión extremadamente severa, 12 en depresión leve y 8 estudiantes sin presencia de depresión.

Figura 10 frecuencias de las escalas de valoración para la depresión.

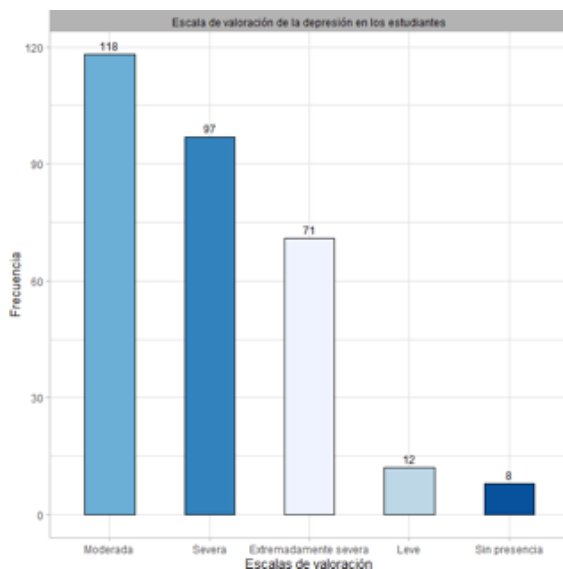


Figura 9 frecuencias de las escalas de valoración para la ansiedad respecto a las preguntas ¿Has recibido atención psicológica? y ¿En algún punto has pensado en abandonar la carrera actual?

Construto "Ansiedad"

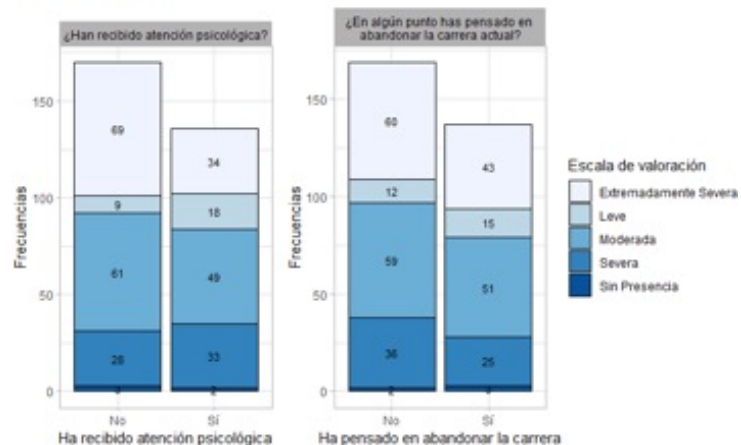
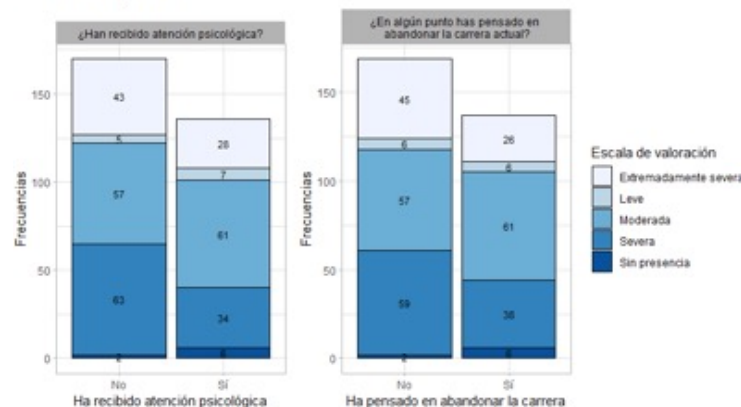


Figura 11 frecuencias de las escalas de valoración para la depresión respecto a las preguntas ¿Has recibido atención psicológica? y ¿En algún punto has pensado en abandonar la carrera actual?

Construto "Depresión"



En la Figura 11 se muestran dos gráficos que representan el cruce la escala de medición de depresión con dos variables dicotómicas. Para el primer gráfico, respecto a la pregunta “¿Has recibido atención psicológica?” quienes respondieron que no fueron 63 estudiantes en la escala de depresión severa, 57 en depresión moderada, 43 en depresión extremadamente severa, 5 en depresión leve y por último 2 estudiantes sin presencia de depresión. Para los estudiantes que si han recibido atención psicológica 61 fueron identificados en depresión moderada, 34 en depresión severa, 28 en depresión extremadamente severa, 7 en depresión leve y 6 sin presencia de depresión.

Para el segundo gráfico, sobre si en algún punto han pensado en abandonar la carrera actual, quienes respondieron que no, 59 son los estudiantes que se encontraron en la escala de depresión severa, 57 se ubicaron en depresión moderada, 45 en depresión extremadamente severa, 6 en depresión leve y 2 sin presencia de depresión. Mientras que para quienes contestaron que, si han pensado en abandonar la carrera, son 61 los que se encontraron en depresión moderada, 38 en depresión severa, 26 en extremadamente severa, 6 para depresión leve y lo mismo para 6 estudiantes que están sin presencia de depresión.

Tabla 3. Modelo de regresión logística ordinal ajustado para el estrés				
	Value	Std. Error	t. value	P valor
Atención psicológica	-0.6213	0.2103	-2.954	3.1322e-03

Regresión logística ordinal

Estrés

El objetivo de este modelo aplicado fue identificar qué factores se asocian al aumento del nivel de estrés en los estudiantes de del área de ciencias de salud a nivel licenciatura de la región Xalapa de la universidad veracruzana. El modelo de regresión logística ordinal se ajustó siempre y cuando se cumpla el supuesto de líneas paralelas, para esto únicamente consideró la variable predictora “ha tomado atención psicológica” (0 = No, 1 = Si) y quedando como variable respuesta la “escala de valoración del estrés”.

Se apreció (Tabla 3) un valor de AIC para este modelo de 907.3491 el cual, a comparación de un modelo donde se consideraron todas las variables presentadas en los estadísticos descriptivos, resultó ser más bajo, por lo que, al ser menor, se tiene un mejor ajuste del modelo a los datos. Para el caso de la atención psicológica (coeficiente de -0.79115), se puede decir que, si los estudiantes si han tomado atención psicológica, esto reduce la probabilidad de niveles altos de estrés.

Los coeficientes de regresión resultaron estadísticamente significativos para la variable atención psicológica. Para la interpretación de los coeficientes de regresión que se obtuvieron en unidades logarítmicas y se interpretó como:

- Los estudiantes que, si han tomado atención psicológica comparado con los estudiantes que no la han tomado, están asociados con un incremento de 0.537227 veces con las escalas de mayor nivel de estrés, con un nivel de confianza al 95% y manteniendo constante el resto de las condiciones.

Ansiedad

La variable predictora fue “ha tomado atención psicológica”, quedando como variable respuesta la “escala de valoración de la ansiedad”. El AIC para este modelo fue de 821.8967 el cual, a comparación de un modelo donde se consideraron todas las variables presentadas en el análisis exploratorio (AIC = 826.3578), resultó ser menor, mostrando un mejor ajuste. Dentro de los coeficientes que este modelo mostró, en la Tabla 4 para la variable “atención psicológica” (p valor $8.2585e-03 < 0.05$) resultó estadísticamente significativo.

Respecto a los coeficientes estimados, para la atención psicológica, con un coeficiente de -0.58460, se entiende que, si los estudiantes si han tomado la atención psicológica, esto reduce la probabilidad de que padezcan niveles altos de estrés.

Tabla 4. Modelo de regresión logística ordinal ajustado para la ansiedad				
	Value	Std. Error	t. value	P valor
Atención psicológica	-0.534	0.2106	-2.536	1.1208e-02

Por lo que, para su interpretación de estos coeficientes de regresión, obtenidos en unidades logarítmicas las cuales indican:

- Los estudiantes que, si han tomado atención psicológica comparado con los estudiantes que no la han tomado, están asociados con un incremento de 0.5862316 veces con las escalas de mayor nivel de estrés manteniendo constante el resto de las condiciones. Este efecto es estadísticamente significativo con un nivel de confianza del 95%.

Depresión

El modelo de regresión logística ordinal se ajustó para las variables que resultaron significativas en un primer modelo y que si cumplían con el supuesto de líneas paralelas. Únicamente se considerando las variables predictoras “Promedio”, “ha tomado atención psicológica” (0 = No, 1 = Si) y “si ha considerado en abandonar la carrera actual” (0 = No, 1 = Si) quedando como variable respuesta la “escala de valoración de la depresión”. Se apreció un valor de AIC para este modelo de 786.7423 el cual, a comparación de un modelo donde se consideraron todas las variables presentadas en los estadísticos descriptivos (AIC = 794.8669), resultó ser más bajo, por lo que, al ser menor, se tiene un mejor ajuste del modelo a los datos.

Respecto al modelo que se ajustó, arrojó coeficientes mostrados en la Tabla 5, donde para el caso del promedio el cual resultó significativo, el coeficiente es positivo (0.3513), el cual quiere decir que el promedio aumenta la probabilidad de estar en un nivel más alto de depresión. Mientras que, para la atención psicológica, con un coeficiente negativo de -0.5757, se dice que, si los estudiantes si han tomado atención psicológica, esto reduce la probabilidad de estar en niveles altos de depresión, igual que abandono de la carrera con un coeficiente negativo (-0.5164) donde, los estudiantes que si han considerado en abandonar su carrera actual reducen la probabilidad de estar en niveles más altos de depresión.

Tabla 5. Modelo de regresión logística ordinal ajustado para el estrés

	Value	Std. Error	t. value	P valor
Promedio	0.3513	0.1549	2.269	0.0233
Atención psicológica	-0.5757	0.2164	-2.660	0.0078
Abandono de la carrera actual	-0.5164	0.2145	-2.407	0.0161

De acuerdo con el p valor, se puede deducir que los coeficientes de regresión resultaron estadísticamente significativos para el promedio (p valor $0.0233 < 0.05$), atención psicológica (p valor $0.0078 < 0.05$) y abandono de la carrera actual (p valor $0.0161 < 0.05$).

Las interpretaciones de los coeficientes de regresión que fueron obtenido en unidades logarítmicas quedando interpretadas de la siguiente manera:

- Los estudiantes con promedio más alto, en comparación con los que cuentan con un promedio más bajo, están asociados con un incremento de 1.4209562 veces la probabilidad de estar en un nivel alto de las escalas de valoración de la depresión, manteniendo constante el resto de las condiciones. Este efecto es estadísticamente significativo con un nivel de confianza del 95%.
- Los estudiantes que, si han recibido atención psicológica, en comparación con los que no la han recibido, están asociados con un incremento de 0.5623205 veces la probabilidad de estar en un nivel alto de las escalas de valoración de la depresión, manteniendo constante el resto de las condiciones. Este efecto es estadísticamente significativo con un nivel de confianza del 95%.
- Los estudiantes que, si han considerado en abandonar su carrera actual, en comparación con los no han considerado en abandonar su carrera actual, están asociados con un incremento de 0.5966858 veces la probabilidad de estar en un nivel alto de las escalas de valoración de la depresión, manteniendo constante el resto de las condiciones. Este efecto es estadísticamente significativo con un nivel de confianza del 95%.

Por lo que estas interpretaciones, muestran una comprensión, más clara de cómo cada variable predictora afecta el estar en un nivel más alto de las escalas de valoración de la depresión, en este caso el promedio, la atención psicológica y el abandono de la carrera actual, contado con efectos significativos.

CONCLUSIONES

Los resultados indicaron que si existen factores que se asocian con los niveles de alguno de estos trastornos mentales, como lo pueden ser si el estudiante tomó o no atención psicológica, el promedio general obtenido, si ha considerado o no abandonar la carrera actual, aunque la asociación depende del trastorno. Por lo que estos resultados, difieren completamente con los hallazgos que obtuvo Trunce, et. al. (2020), ya que en su estudio menciona que no existe relación entre el promedio de los estudiantes con la presencia de un trastorno depresivo, mientras que para la presente investigación si se encontró relación

entre estas dos variables, pues el efecto del promedio general obtenido por el estudiante aumenta la probabilidad de estar en un nivel más alto de la escala de depresión.

Dentro de la investigación que realizó Gutiérrez, et. al. (2010) donde identificó situaciones generadoras de estrés académico, encontró algunas situaciones que si influyen como lo son la competitividad con sus compañeros, carga académica, planes a futuro; mientras que en conjunto con los resultados de esta investigación descubrieron que, si el estudiante ha tomado atención psicológica, tiende a tener menor probabilidad de estar en un nivel alto de la escala de estrés. Un factor más que influye en los niveles de estrés en estudiantes universitarios, en este caso, del área de ciencias de la salud.

Existieron limitaciones y posibles fuentes de error, debido a la muestra, pues considerando el difícil acceso a la población de interés y el tiempo requerido, se optó por un muestreo no probabilístico por conveniencia, reduciendo de esta manera la generalización de los resultados. Pero, de acuerdo con algunas literaturas, al utilizar un muestreo por conveniencia, aunque no tenga la representatividad requerida, este tipo de muestreo puede ser una base inicial de la obtención de datos para investigaciones preliminares.

Una línea de investigación futura sería conocer el verdadero comportamiento de los datos si realmente se utilizará una muestra representativa de la población de interés. El incluir más factores que puedan afectar los niveles de los trastornos de depresión, ansiedad y estrés para realizar evaluaciones e identificar que otros aspectos pueden provocar niveles altos de estos trastornos. Asimismo, ¿sería recomendable mejorar el instrumento de medición para obtener mejores índices dentro de la validación sin necesidad de aplicar IM en la re- especificación?

REFERENCIAS

- Cabezas, H. E., Herrera, C. R., Ricaurte, O. P., & Novillo, Y. C. (2021). *Depresión, ansiedad, estrés en estudiantes y docentes: Análisis a partir del COVID-19*. [PDF]. Dialnet. <https://dialnet.unirioja.es/download/articulo/8890448.pdf>
- Gutiérrez, R. J. A., Montoya, V. L. P., Toro, I. B. E., et al. (2010). *Depresión en estudiantes universitarios y su asociación con el estrés académico*. *CES Medicina*, 24(1). <http://www.scielo.org.co/pdf/cesm/v24n1/v24n1a02.pdf>
- Manríquez, M. M. I., Veloz, M. K. G., et al. (2021). *Estrés, ansiedad y depresión relacionado con burnout estudiantil en alumnos de enfermería en una universidad pública*. *Enfermería Global*, 91, e212407. <https://ciberindex.com/index.php/g91/article/view/e212407/e212407>
- Moterrosa, C. A., Ordosgoitia, P. E., & Beltrán, B. (2020). *Ansiedad y depresión identificadas con la Escala de Goldberg en estudiantes universitarios del área de la salud*. *MedUNAB*, 23(2). <https://revistas.unab.edu.co/index.php/medunab/article/view/3881>
- Ochoa, M. R. E., Castro, T. R., & Vázquez, O. N., et al. (2021). *Ansiedad y depresión en estudiantes del área de la salud de la Universidad Veracruzana*. *Ciencia Frontera*, 13(2). <https://revistas.uacj.mx/ojs/index.php/cienciafrontera/article/view/3611/3275>

- Pulido, R. M. A., Serrano, S. M. L., Valdés, C. E., et al. (2011). *Estrés académico en estudiantes universitarios*. *Psicología y Salud*, 21(1). <https://psicologiaysalud.uv.mx/index.php/psicysalud/article/view/584/1008>
- Trunce, M. S. T., Villarroel, Q. G. P., & Arntz, V. J. A., et al. (2020). *Niveles de depresión, ansiedad, estrés y su relación con el rendimiento académico en estudiantes universitarios*. *Investigación en Educación Médica*, 9(36), 8–18. <https://www.scielo.org.mx/pdf/iem/v9n36/2007-5057-iem-9-36-8.pdf>
- Vallejo, C. A. G., Díaz, M. K., & Méndez, F. M. M., et al. (2019). *Sintomatología depresiva y consumo de sustancias en universitarios de Veracruz, México*. *Revista Iberoamericana de Psicología*, 12(2). https://www.researchgate.net/profile/karina-diaz-morales/publication/337844278_sintomatologia_depresiva_y_consumo_de_sustancias_en_universitarios_de_veracruz_mexico/links/5deef94299bf10bc34eac6f/sintomatologia-depresiva-y-consumo-de-sustancias-en-universitarios-de-veracruz-mexico.pdf

ESTADÍSTICA ESPACIAL Y ANÁLISIS SOCIAL: DISCAPACIDAD Y REZAGO EN LOS MUNICIPIOS DE VERACRUZ (2020)

USBALDO HERNÁNDEZ LANDA¹⁴

LUIS ENRIQUE GÓMEZ MEDINA¹⁵

DANIEL ARMANDO OLIVERA GÓMEZ¹⁶

Facultad de Estadística e Informática

Instituto de Investigaciones y Estudios Superiores de las Ciencias Administrativas

Universidad Veracruzana

RESUMEN

La discapacidad resulta de la interacción entre las deficiencias físicas, sensoriales, intelectuales o mentales de una persona y diversas barreras del entorno que limitan su participación plena en la sociedad. Actualmente, aproximadamente 1,300 millones de personas equivalente al 16% de la población mundial, viven con algún tipo de discapacidad. En Veracruz, se reporta un total de 1,542,272 personas con discapacidad, de las cuales el 53.94% corresponde a mujeres y el 46.06% hombres. Estos grupos enfrentan importantes desafíos en salud, inclusión y bienestar. No obstante, existe poca investigación enfocada en la dimensión territorial de este fenómeno y su relación con el rezago social a nivel municipal. El objetivo de esta investigación es analizar la distribución espacial de las discapacidades en los municipios de Veracruz durante el año 2020, identificando agrupamientos de municipios con características similares. Mediante DBSCAN se identificó el clúster -1 que agrupaba a más de la mitad de los municipios con un alto rezago social ($IRS = 0.54$) que lo separaban de las concentraciones densas. Los clústeres definidos por K-means mostraron tres tendencias: media presencia de discapacidad y bajo rezago; baja presencia de discapacidad y alto rezago; y un grupo reducido con baja presencia de discapacidad y alto rezago.

Palabras claves: Clasificación no supervisada, análisis multivariado, discapacidad, rezago social

¹⁴ usbaldohernandezlanda@gmail.com

¹⁵ luisgomez04@uv.mx

¹⁶ doliverag@gmail.com

INTRODUCCIÓN

La discapacidad es la restricción o impedimento de poder realizar actividades que se consideran normales para el ser humano (OMS, 2020). Las personas con discapacidad son aquellas que tienen deficiencias físicas, mentales, intelectuales o sensoriales que a largo plazo obstaculizan su plenitud y su desarrollo en una sociedad (OPS, 2025). Según el informe mundial sobre discapacidad alrededor del 15% de la población vive con algún tipo de discapacidad.

A nivel mundial más de 1300 millones de personas (16%) vive con alguna discapacidad. Sin embargo, la discapacidad no sólo refleja una condición médica sino también un fenómeno social vinculada a la desigualdad social. En México específicamente en el estado de Veracruz estas desigualdades se ejemplifican por los contrastes territoriales de educación y servicios de salud, vivienda y educación los cuales radican directamente en el grado de rezago social y en las condiciones de las vidas de las personas con discapacidad.

El análisis espacial se ha convertido en una herramienta para explicar la distribución territorial de la discapacidad y su relación con los índices de rezago. Mediante técnicas de agrupamiento no supervisado como DBSCAN y K-means, Es posible identificar patrones de concentración, dispersión y atipicidad entre los municipios sin imponer sus puestos previos sobre la forma o cantidad de grupos. En este estudio se aplicaron ambos algoritmos a los 212 municipios de Veracruz para determinar las relaciones que existe entre los 7 tipos de discapacidad y el índice de rezago social durante el año 2020.

Los resultados revelaron una alta heterogeneidad. El análisis de DBSCAN identificó el clúster -1 que agrupaba a más de la mitad de los municipios caracterizados por un alto rezago social ($IRS = 0.54$) que lo separaban de las concentraciones densas. Por otra parte, los clústeres definidos mostraron 3 tendencias principales: municipios con baja prevalencia y discapacidad y bajo rezago, municipios con niveles medios de discapacidad y rezago intermedio y un grupo de reducido con una alta prevalencia de discapacidad y alto rezago. Las variables con mayor variabilidad fueron la discapacidad visual y la motriz junto con el índice de rezago social que influyeron más en la diferenciación territorial.

Estos resultados confirmaron que la discapacidad y el rezago social mantiene una relación espacial desigual en Veracruz las zonas rurales y montañosas concentran los mayores niveles de marginación mientras que los municipios costeros y urbanos presentan condiciones más favorables. En la aplicación el análisis multivariado y espacial permiten no sólo reconocer la distribución de la discapacidad sino también identificar regiones prioritarias para la atención específica de grupos vulnerables.

El presente estudio tiene como objetivo analizar la segmentación espacial de los municipios de Veracruz según el tipo de discapacidad y su relación con el rezago social empleando técnicas de agrupamiento no supervisado que permitirá visualizar la desigualdad territorial desde una perspectiva estadística social y geoespacial.

Antecedentes

El algoritmo DBSCAN, al igual que otros métodos de agrupación no supervisada, es ampliamente utilizado para descubrir grupos de datos de formas arbitrarias y detectar puntos atípicos o ruido. (Nasereddine, Benazi & Sayad, 2019) presentan una versión adaptativa del parámetro *Eps* que calcula su valor de manera automática mediante el algoritmo de K vecinos más cercanos (KNN), permitiendo detectar clústeres con densidades heterogéneas y mejorar la precisión frente al DBSCAN original. Este enfoque resolvió la principal limitación del algoritmo clásico y extendió su aplicabilidad a conjuntos de datos más complejos.

Por su parte, (Xie & Shekhar, 2019) proponen el sig-DBSCAN, una versión estadísticamente robusta del algoritmo que incorpora pruebas de hipótesis y simulaciones Montecarlo para identificar clústeres significativos y eliminar patrones espurios. Este método reduce falsos positivos y mejora la interpretación en datos espaciales, aportando mayor fiabilidad en la detección de agrupamientos genuinos, lo cual es relevante para aplicaciones sociales y territoriales.

En cuanto a su aplicación geoespacial, (Tu et al., 2022) emplearon DBSCAN en el análisis urbano de Wuhan, China, utilizando datos de puntos de interés (POI) para optimizar la planificación territorial. Los resultados revelaron una estructura espacial jerárquica y la existencia de zonas con alta densidad de servicios urbanos, validando el algoritmo como una herramienta eficaz para el análisis territorial.

En otro estudio, (Jibrin et al., 2025) combinaron Kernel PCA y DBSCAN para evaluar la calidad del agua subterránea, demostrando la utilidad del algoritmo en contextos de alta dimensionalidad y con presencia de valores atípicos. Esta integración metodológica evidencia la capacidad de DBSCAN para identificar patrones relevantes en datos complejos, sirviendo de referencia para su aplicación en temas sociales y ambientales.

En México, la disponibilidad de datos geoespaciales ha impulsado el uso de algoritmos de clustering para analizar problemáticas sociales y territoriales. (Carrasco-Islas & Moyotl-Hernández, 2022) aplicaron un enfoque basado en densidad mediante OPTICS derivado de DBSCAN para detectar zonas con alta incidencia delictiva en la Ciudad de México, logrando una detección más precisa de *hot spots* y evidenciando su potencial para estudios espaciales.

Por último, (Caudillo-Cos et al., 2024) propusieron la aplicación de DBSCAN y la entropía de Shannon para definir límites urbanos dentro del Sistema Urbano Nacional Mexicano, mostrando que el método permite delimitar áreas urbanas de manera más económica y automática que los métodos tradicionales, fortaleciendo la toma de decisiones en políticas públicas y planeación territorial.

MATERIALES Y MÉTODOS

Diseño del estudio

La investigación de tipo cuantitativo, con un diseño no experimental, transversal y descriptivo. Su enfoque va dirigido al análisis espacial de datos mediante técnicas de agrupamiento no supervisado (Clustering), con el propósito de identificar patrones territoriales de discapacidad en los municipios del estado de Veracruz durante el año 2020.

Unidad de análisis

El análisis corresponde a los municipios del estado de Veracruz. Se eligen únicamente aquellos municipios que cuentan con información completa sobre las variables seleccionadas.

Fuente de datos

Se utilizó la información de diversos centros de información pública entre ellos una base del INEGI (2020) sobre la diversidad de discapacidad por tipo y municipio, así como las coordenadas geográficas además del indicador de rezago social obtenido a través de un repositorio llamado CONEVAL (2020).

Etapas del análisis:

1. Análisis exploratorio de Datos Espaciales (AEDES)

En esta primera fase se realizará una exploración visual y estadística de los datos para comprender la distribución general de las discapacidades en los municipios de Veracruz. Las actividades corresponden a:

- estadísticas descriptivas por cada variable
- Mapas coropleticos para representar las discapacidades y Rezago Social por municipio permitiendo identificar visualmente concentraciones geográficas.

2. Agrupamiento sin estandarización

En esta segunda etapa se aplicarán dos métodos de agrupamiento no supervisado directamente sobre los datos originales. El objetivo es identificar grupos de municipios con características similares con respecto a la discapacidad y rezago social. Las cuales son:

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Identifica clusters basados en la densidad, es robusto a outliers y no requiere definir el número de grupos.
- K-means: genera clústeres de forma esférica y homogénea. Requiere definir el número de clústeres k , el cual será determinado con criterios como el método del codo o el índice de silueta

Variables del estudio

Nombre de la variable	Descripción	Tipo de variable	Escala de medición
Municipio	Nombre del municipio en Veracruz	Cualitativa nominal	Nominal
Latitud	Coordenada geográfica norte-sur	Cuantitativa continua	Razón
Longitud	Coordenada geográfica este-oeste	Cuantitativa continua	Razón
Porcentaje de población con discapacidad visual	Porcentaje de personas con discapacidad visual	Cuantitativa continua	Razón
Porcentaje de población con discapacidad auditiva	Porcentaje de personas con discapacidad auditiva	Cuantitativa continua	Razón
Porcentaje de población con discapacidad motriz	Porcentaje de personas con discapacidad motriz	Cuantitativa continua	Razón
Porcentaje de población con discapacidad cognitiva	Porcentaje de personas con discapacidad para concentrarse o recordar	Cuantitativa continua	Razón
Porcentaje de población con discapacidad Autonomía	Porcentaje con discapacidad en autonomía personal	Cuantitativa continua	Razón

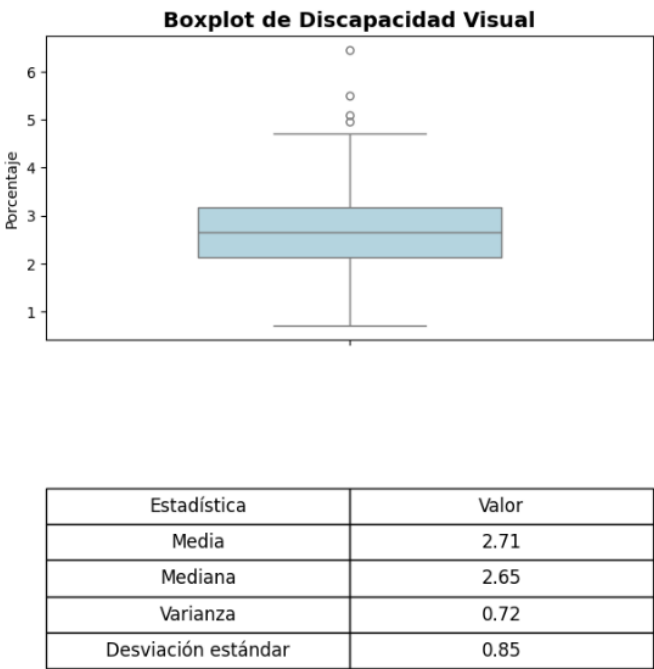
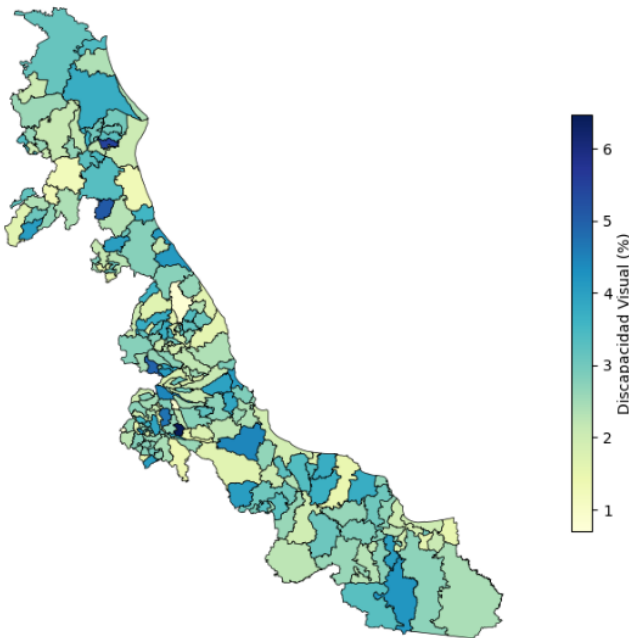
Porcentaje de población con discapacidad Comunicación	Porcentaje con dificultad para comunicarse	Cuantitativa continua	Razón
Porcentaje de población con discapacidad Mental	Porcentaje con condición mental o emocional	Cuantitativa continua	Razón
IRS	Índice de Rezago Social	Cuantitativa continua	Intervalo

RESULTADOS

Análisis Espacial Exploratorio Descriptivo

Figura 1. Mapa de porcentaje de discapacidad visual en Veracruz.

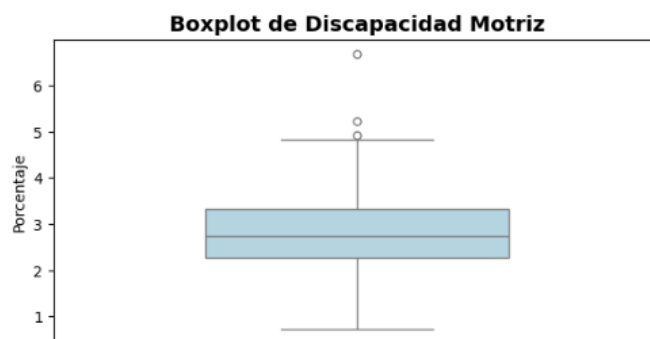
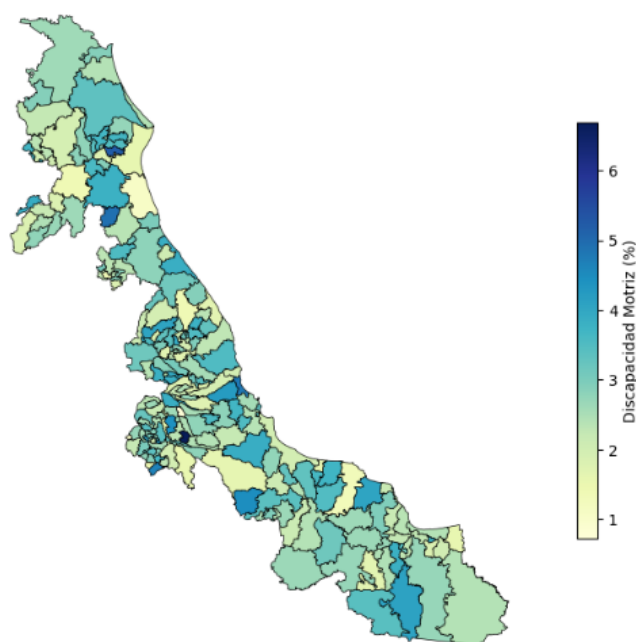
Porcentaje de Discapacidad Visual en Veracruz



La Figura 1. Muestra el porcentaje de discapacidad visual por municipio en el estado de Veracruz. Donde mediante una escala de color indica la intensidad: los claros representan los porcentajes más bajos de discapacidad visual (Cercanos al 1%), los azules corresponden a los porcentajes intermedios (entre de 3% a 4%) y los tonos más oscuros son los porcentajes más altos (Cercanos o superiores al 6%).

Se observa que la distribución espacial de la discapacidad visual no es uniforme en el estado. Los municipios con una alta contracción parecen estar más dispersos, sin embargo, existen áreas como la zona central-sur o algunos municipios al norte que presentan concentración de valores altos. En general se muestra una variabilidad que se encuentran rangos medios o bajos en gran parte del territorio.

El 50% de los municipios presentan un porcentaje de discapacidad visual igual o inferior a 2.65%, mientras que el promedio estatal corresponde a 2.70%. La desviación estándar (0.85) indica que los porcentajes se desvían en un de la media indicando una dispersión considerable. Además, el boxplot revela la presencia de datos atípicos siendo cercanos a 5.5% y 6.4%, los cuales corresponden al a municipios con una mayor presencia de esta discapacidad, siendo más alta en comparación a los demás municipios de Veracruz, indicando que hay muy pocos municipios donde esta discapacidad es más elevada.

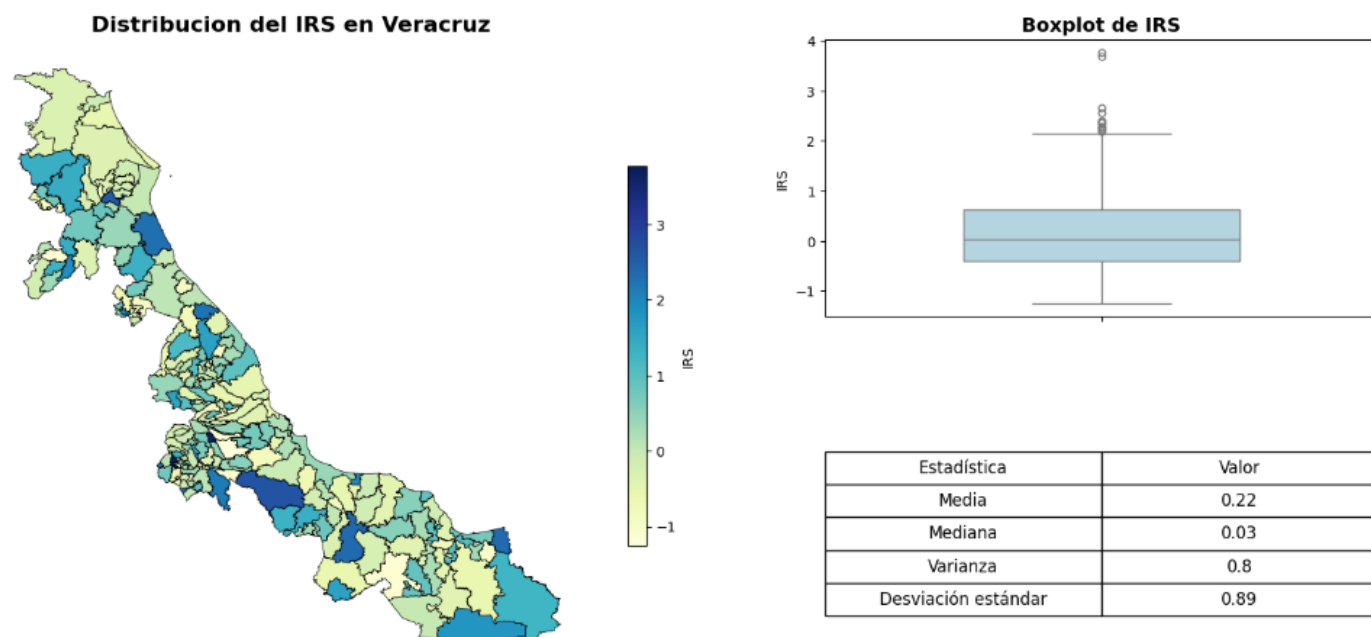
Figura 2. Mapa de porcentaje de discapacidad motriz en Veracruz.**Porcentaje de Discapacidad Motriz en Veracruz**

Estadística	Valor
Media	2.79
Mediana	2.73
Varianza	0.69
Desviación estándar	0.83

La figura 2. Muestra el porcentaje de discapacidad motriz por municipio en el estado de Veracruz. Donde mediante una escala de color indica la intensidad: los claros representan los porcentajes más bajos de discapacidad motriz (Inferiores o cercanos al 1%), los azules corresponden a los porcentajes intermedios (entre de 2.5% a 4%) y los tonos más oscuros son los porcentajes más altos (Cercanos o superiores al 6%).

Se observa que la distribución espacial de la discapacidad motriz no es uniforme en el estado. Los municipios en su mayoría presentan porcentajes moderados, lo que indica que los municipios tienen porcentajes de discapacidad motriz alrededor del promedio, sin embargo, existen zonas con porcentajes altos que se observan en diferentes zonas del estado, indicando que existen zonas con tasas elevadas que no están limitadas en una única región. En general se muestra una variabilidad que se encuentran rangos medios en gran parte del territorio lo que sugiere que la mayoría de la población tiene sus porcentajes de discapacidad oscilan entre 2.5% y 3%.

El 50% de los municipios presentan un porcentaje de discapacidad motriz igual o inferior a 2.73%, mientras que el promedio estatal corresponde a 2.79%, indicando una distribución bastante simétrica. La desviación estándar (0.83) indica que los porcentajes se desvían de la media con una dispersión moderada. Además, el boxplot revela la presencia de datos atípicos siendo cercanos a 5.2% y 6.5%, los cuales corresponden al a municipios con una mayor presencia de esta discapacidad, siendo más alta en comparación a los demás municipios de Veracruz, indicando que hay muy pocos municipios donde esta discapacidad es más elevada

Figura 3. Mapa de índice de rezago social en Veracruz.

La Figura 3. Muestra la distribución de índice de rezago social (IRS) por municipio en el estado de Veracruz. Donde mediante una escala de color indica la intensidad: Los colores claros representan los grados bajos de rezago (Cercanos a -1%), los colores azul claro corresponden a grados de rezago intermedio (entre 1%) y los tonos más oscuros corresponden a los grados más altos de rezago (Cercanos o superiores a 3%).

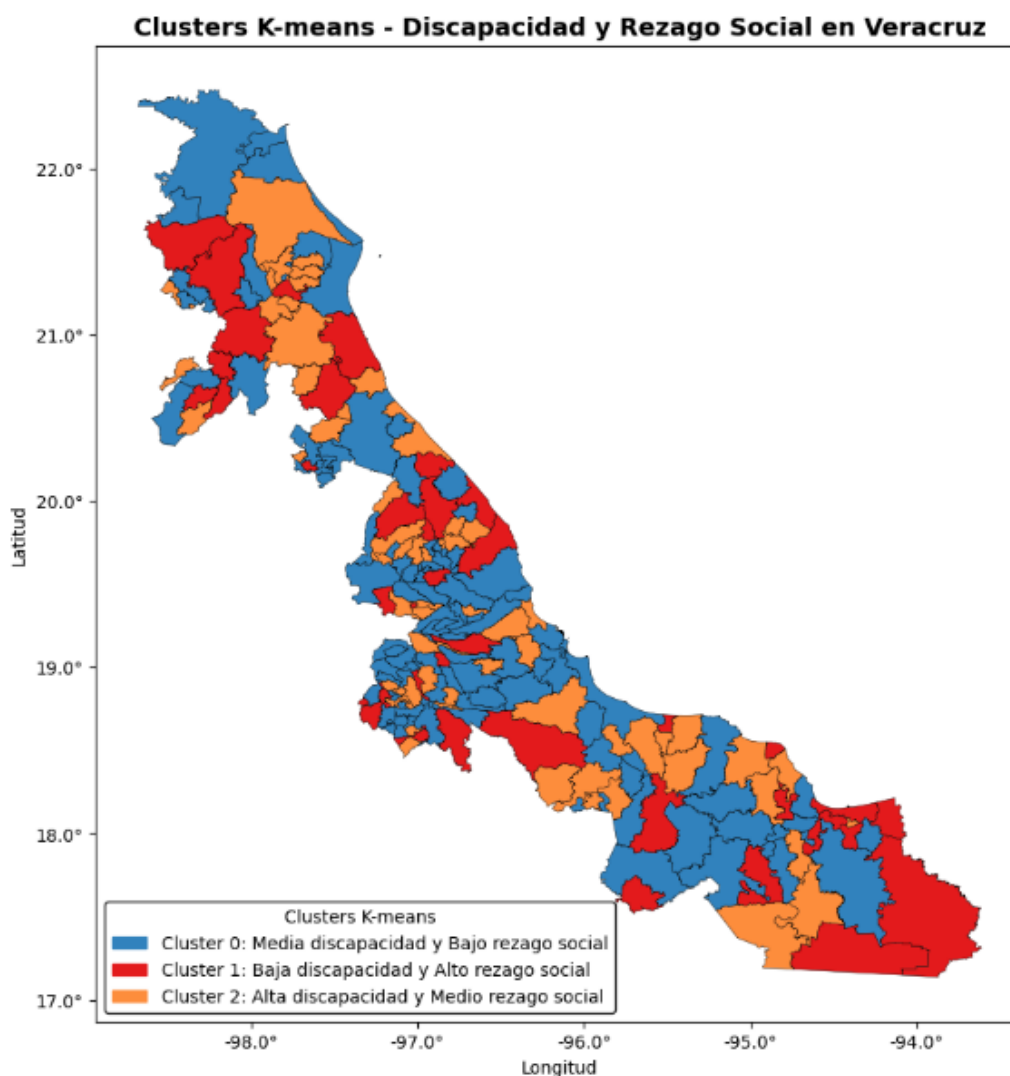
Se observa que la distribución espacial del IRS en su mayoría se encuentra en un rezago bajo a moderado, con valores cercanos a 0. Esto indica que una gran parte de los municipios tienen porcentajes de rezago relativamente bajos. Sin embargo, existen zonas con porcentajes de rezago altos que se observan en diferentes áreas del estado, específicamente en la zona central, cercana a la capital, y algunos municipios dispersos al norte del estado. Por otra parte, existen áreas con un porcentaje negativo o muy bajo de rezago, distribuidas especialmente en el sur y algunas áreas costeras.

El 50% de los municipios presentan un grado de rezago inferior o igual a 0.03, mientras que el promedio estatal corresponde a 0.22 siendo superior a la mediana, indicando que la distribución con un sesgo hacia la derecha, significando que hay más municipios con valores de rezago alto que con valores muy bajos o negativos de rezago. La desviación estándar (0.89) es alta, indicando que estos valores se desvían considerablemente de la media, con una dispersión muy alta y gran variabilidad en el grado de rezago social. Además, el boxplot revela la presencia de datos atípicos que van desde 2.1 hasta 3.4. Estos corresponden a municipios con un grado de rezago social muy alto, en comparación a los demás de los municipios. Estos atípicos que son pocos, pero muy elevados, indican la existencia de pobreza extrema y rezago que requieren de atención urgente y focalizada.

Tabla 1. Resultado del agrupamiento con K-Means

Cluster	Visual (%)	Auditiva (%)	Mental (%)	Motriz (%)	Autonomía (%)	Recordar (%)	Comunicativa (%)	IRS
0	2.43	1.35	1.12	2.54	1.06	1.07	0.92	-0.21
1	1.89	1.33	1.15	1.94	0.95	0.89	1.92	1.58
2	3.66	1.98	1.37	3.67	1.40	1.58	1.18	0.13

En la Tabla 1. Se puede observar un resumen con los resultados de los clústeres obtenidos mediante K-Means (con el número óptimo de clúster encontrados mediante el método Elbow Method $K=3$) se observa que el primer grupo (Clúster 0) presenta una prevalencia media en diferentes tipos de discapacidad (visual, auditiva, motriz, autonomía y recordar) sin embargo también cuenta con características bajas de discapacidad como prevalencia baja de discapacidad (mental y comunicativa) así como el grado de rezago más bajo (-0.21), se puede deducir que es el grupo con una prevalencia media de discapacidad y bajo rezago. El Segundo grupo (Clúster 1) está conformado por la presencia de baja prevalencia en discapacidad (visual, auditiva, motriz, autonomía, recordar) una prevalencia media en discapacidad (Mental) pero con una prevalencia alta de discapacidad (comunicativa) y un grado alto de rezago social (1.58), por último el tercer grupo (Clúster 2) se observa la alta prevalencia de discapacidades (visual, auditiva, motriz, autonomía y recordar) pero con prevalencia media de discapacidad comunicativa y un rezago moderado.

Figura 4. Clúster K- Means acerca de discapacidad y rezago social en Veracruz 2020

En la Figura 4. Muestra el mapa de estado de Veracruz clasificado en tres clústeres mediante K-Means, considerando variables de discapacidad y grado de rezago social. El Clúster 0 agrupa 108 municipios caracterizados por prevalencia media en discapacidad y bajo o nulo grado de rezago social, el clúster 1 considera a 39 municipios con una incidencia alta de discapacidad y alto grado rezago social y el clúster 2 incluye a 65 municipios con alta incidencia de discapacidad y un grado medio de rezago social.

Se observa que Clúster 0 es el que se distribuye con mayor extensión indicando que los veracruzanos presentan condiciones favorables dentro del grado de rezago social y con una prevalencia media de discapacidad. Su presencia se distribuye a lo largo del estado específicamente en el centro y sur del estado. Este grupo incluye municipios costeros y urbano como Veracruz boca del río Córdoba Orizaba Coatzacoalcos y Minatitlán, así como otros municipios con desarrollo económico o industrial. Asimismo, también se observan algunos municipios dispersos en regiones montañosas centrales indicando focos de desarrollo y menor marginación dentro de ciertas áreas rurales.

El clúster 1 se distribuye principalmente en zonas rurales y montañosas del norte, centro y sur, incluyendo a las regiones Huasteca, Totonaca, Sierra de Zongolica, Los Tuxtlas y la región Olmeca. Estos municipios se caracterizan por la baja prevalencia de discapacidad, pero con altos niveles de rezago la presencia de este clúster en áreas de difícil acceso muestra la mayor vulnerabilidad social donde el bajo número de discapacidad puede estar relacionado a la falta de atención médica o seguimiento de ciertos casos.

Así mismo el clúster 2 se localiza de manera dispersa en las diferentes regiones del estado, pero con una mayor presencia en áreas costeras del centro y sur, así como en municipios cercanos a centros urbanos como Xalapa, Córdoba, Veracruz y Coatzacoalcos. También se observan municipios en el norte próximos a Tantoyuca y Tuxpan este grupo con estas características indican que son zonas con un grado medio de rezago, pero con altos niveles de discapacidad, Lo que indica que son zonas con accesibilidad moderada a sus necesidades, tratamientos y seguimiento en salud.

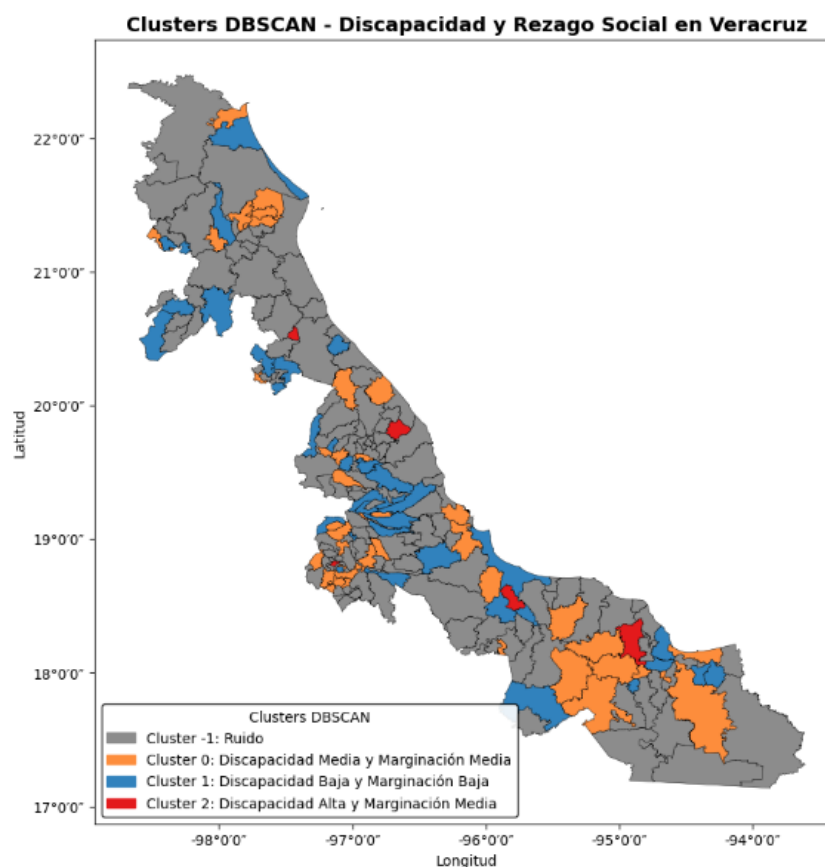
De manera general, se observa que Veracruz presenta mayores concentraciones de municipios con clústeres en condiciones favorables, con niveles de discapacidad media y bajo grado de rezago. Sin embargo, la distribución territorial del clúster 1 revela que persisten ciertas desigualdades en municipios rurales con una alta marginación. Finalmente, el clúster 2 representa a grupos de municipios que se les atribuye a pertenecer a un riesgo intermedio ya que presenta una alta prevalencia en discapacidad, pero con un rezago intermedio. Este tanto como el clúster 1 presentan un foco de atención, debido a que son de riesgo intermedio a alto. Por esto es importante la atención y seguimiento a personas con discapacidad en municipios rurales y de regiones montañosas para reducir la marginación, mejorar el acceso a servicios médicos y sociales.

DBSCAN

Tabla 2. Resultado del agrupamiento con DBSCAN

Cluster	Visual (%)	Auditiva (%)	Mental (%)	Motriz (%)	Autonomía (%)	Recordar (%)	Comunicativa (%)	IRS
-1	2.85	1.66	1.16	2.90	1.19	1.26	1.16	0.54
0	2.77	1.52	1.29	2.87	1.17	1.22	1.00	-0.02
1	2.05	1.10	1.22	2.16	0.91	0.91	0.80	-0.45
2	3.02	1.72	1.30	3.14	1.39	1.28	1.19	0.25

En la Tabla 2. Se puede observar un resumen de los resultados de los clústeres obtenidos mediante DBSCAN con parámetros (eps:0.05 y min_samples:5) Se puede observar que el grupo (clúster -1) presenta una prevalencia media de discapacidad sin embargo es el clúster que mayor rezago social captura. El algoritmo de DBSCAN identifica a este grupo como atípicos sin embargo la característica que separa a estos municipios de la distribución principal es el alto índice de rezago social (0.54) indicando que los municipios con carencias sociales tan extremas no encajan en los patrones de densidad definidos por los clústeres más grandes. Por otra parte el clúster el primer grupo (Clúster 0) presenta una prevalencia media en diferentes tipos de discapacidad (Visual, Auditiva, Mental Motriz, Autonomía y Recordar) sin embargo cuenta el grado de rezago más bajo (-0.02), se puede deducir que es el grupo con una prevalencia media de discapacidad y bajo rezago, El Segundo grupo (Clúster 1) está conformado por la presencia de baja prevalencia en discapacidad (Visual, Auditiva, Mental Motriz, Autonomía y Recordar) y un grado bajo de rezago social (-0.45), por último el tercer grupo (Clúster 2) se observa la alta prevalencia de discapacidades (Visual, Auditiva, Mental Motriz, Autonomía y Recordar) pero con el grado mas alto de rezago social en comparación a los demás grupos (0.25).

Figura 5. Clúster DBSCAN acerca de discapacidad y rezago social en Veracruz 2020

En la Figura 5. Muestra el mapa de estado de Veracruz clasificado en tres clústeres mediante DBSCAN, considerando variables de discapacidad y grado de rezago social. El cluster -1 agrupa a 129 municipios categorizados como ruido, es decir, aquellos que no presentan una densidad suficiente para formar parte de un grupo definido. El Clúster 0 agrupa 43 municipios caracterizados por prevalencia media en discapacidad y un grado medio de rezago social, el clúster 1 considera a 35 municipios con una prevalencia baja de discapacidad y bajo grado rezago social y el clúster 2 incluye a 5 municipios con alta incidencia de discapacidad y un grado alto de rezago social.

Se observa que el clúster del ruido (-1) predomina en la extensión geográfica de Veracruz, su distribución se localiza en la zona norte y sur del estado. Esto indica que la gran mayoría de los municipios tienen características que los hacen atípicos respecto a los demás clústeres densos, principalmente por los valores extremos de rezago social.

El clúster 0 se encuentra disperso, con una concentración en el sur del estado (región Olmeca), el centro costero y algunos municipios en el norte. Estos municipios presentan niveles medios de rezago y prevalencia de discapacidad ubicados en zonas con mejor desarrollo económico o áreas rurales de menor marginación.

Por su parte, El clúster 1 se ubica en focos dispersos, con presencia en el norte, centro y el sureste. Este clúster se asocia con centros urbanos o periferias donde el bienestar social es relativamente mejor Y la prevalencia de discapacidad se mantiene baja. Estos municipios muestran entornos más favorables para las personas con discapacidad teniendo mayor desarrollo social y acceso a servicios.

En contraparte, el clúster 2 es el menos extendido geográficamente y aparece en pequeños centros puntuales a lo largo del estado. Su baja representación geográfica coincide con el patrón de los datos atípicos de los boxplots, indicando que tienen la peor combinación de una alta prevalencia de discapacidad y alto rezago, los cuales requieren atención inmediata.

En general DBSCAN revela Una alta dispersión en los municipios y la presencia de valores atipicos, ya que la mayoría de estos municipios caían en el clúster -1 (ruido) con un IRS promedio de 0.54 indicando que la marginación y el rezago son las fuerzas dominantes en la diferenciación municipal que no se logran agrupar en patrones denso.

CONCLUSIONES

Los resultados permitieron demostrar que la discapacidad y el rezago social en los municipios de Veracruz presentan una distribución territorial heterogénea y desigual. A través de las técnicas de clustering no supervisado, se lograron identificar patrones espaciales que revelan distintos y variados niveles de vulnerabilidad vinculados a diferentes tipos de discapacidad como al grado de rezago social.

Los resultados obtenidos mediante camino mostraron tres clústeres: uno caracterizado por la baja prevalencia de discapacidad y bajo rezago social concentrados en regiones costeras y urbanas; otro con

niveles altos de discapacidad y un grado de rezago social intermedio; y un último grupo conformado por una alta prevalencia de discapacidad y un grado alto de rezago social localizándose principalmente en áreas rurales y montañosa.

Asimismo, el algoritmo DBSCAN confirmó la existencia de una gran cantidad de municipios clasificados con características atípicas los cuales se verían afectados y agrupados por alto grado de rezago social. Esta dispersión indica que la marginación y la desigualdad son factores dominantes en la distribución espacial de la discapacidad en el estado los clusters más densos identificar los municipios medios de capacidad y rezago, mientras que en pequeños grupos tenían cantidades de mayor vulnerabilidad.

Ambas técnicas coinciden en que la discapacidad en Veracruz mantiene una relación directa con el rezago, especialmente en las regiones rurales montañosas y del norte del estado donde la carencia de servicios básicos y la alimentación es al acceso a la salud y a la educación favorecen a las condiciones de exclusión social.

En conjunto los resultados muestran la utilidad de los métodos de agrupamiento no supervisado como la herramienta para identificar regiones prioritarias, permitiendo la toma de decisiones y atención a la población con discapacidad considerando las necesidades y limitaciones para acceder a la salud, así como mejores condiciones de desarrollo.

REFERENCIAS

- Instituto Nacional de Estadística y Geografía (INEGI). (2020). *Censo de Población y Vivienda 2020*. <https://www.inegi.org.mx/programas/ccpv/2020/>
- Organización Panamericana de la Salud. (s. f.). *Discapacidad*. OPS. <https://www.paho.org/es/temas/discapacidad>
- Organización Mundial de la Salud (OMS). (2022, 12 de mayo). *Discapacidad y salud*. <https://www.who.int/es/news-room/fact-sheets/detail/disability-and-health>
- Amroune, N., Benazi, M., & Sayad, L. (2024). An adaptive Eps parameter of DBSCAN algorithm for identifying clusters with heterogeneous density. *Computación y Sistemas*, 28(2), 465–472. <https://doi.org/10.13053/CyS-28-2-4600>
- Carrasco-Islas, J., & Moyotl-Hernandez, E. (2022). Agrupamiento basado en densidad para la detección automática de *hot spots* delictivos en la CDMX. *Research in Computing Science*, 151(8), 17–30. https://rcs.cic.ipn.mx/2022_151_8/Agrupamiento%20basado%20en%20densidad%20para%20la%20deteccion%20automatica%20de%20hot%20spots%20delictivos%20en%20la%20CDMX.pdf
- Caudillo-Cos, C. A., Montejano-Escamilla, J. A., Tapia-McClung, R., Ávila-Jiménez, F. G., & Barrera-Alarcón, I. G. (2024). Defining urban boundaries through DBSCAN and Shannon's entropy: The case of the Mexican National Urban System. *Cities*, 149, 104969. <https://doi.org/10.1016/j.cities.2024.104969>

- Díaz, O. O., Raga, G. B., Quintanar, A. I., & Mejía, J. F. (2024). Structural change points of NDVI in Mexico driven by climate oscillations. *Atmósfera*, 38, 105–127. <https://doi.org/10.20937/ATM.53201>
- Jibrin, A. M., Al-Suwaiyan, M., Yaseen, Z. M., & Abba, S. I. (2025). New perspective on density-based spatial clustering of applications with noise for groundwater assessment. *Journal of Hydrology*, 661, 133566. <https://www.sciencedirect.com/science/article/abs/pii/S0022169425009047>
- Tu, X., Fu, C., Huang, A., Chen, H., & Ding, X. (2022). DBSCAN spatial clustering analysis of urban “production–living–ecological” space based on POI data: A case study of central urban Wuhan, China. *International Journal of Environmental Research and Public Health*, 19(9), 5153. <https://doi.org/10.3390/ijerph19095153>
- Xie, Y., & Shekhar, S. (2019, August 19–21). Significant DBSCAN towards statistically robust clustering. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases (SSTD '19)* (pp. 1–10). Association for Computing Machinery. <https://doi.org/10.1145/3340964.3340968>

ANÁLISIS MULTIVARIADO APLICADO A LA SALUD PÚBLICA: PERFIL DE LA POBLACIÓN CON CÁNCER CERVICOUTERINO EN VERACRUZ

Montserrat Martínez de los Santos¹⁷

Emmanuel Morales García¹⁸

Daniel Armando Olivera Gómez¹⁹

Facultad de Estadística e Informática

Instituto de Investigaciones y Estudios Superiores de las Ciencias Administrativas

Universidad Veracruzana

RESUMEN

El cáncer cervicouterino continúa siendo una de las principales causas de mortalidad femenina en México, pese a ser prevenible mediante la detección temprana. En Veracruz, los indicadores muestran desigualdades sociales y sanitarias que limitan la prevención y el tratamiento. Este estudio aplica métodos de estadística multivariada, como el análisis de conglomerados k-medias, el análisis de componentes principales (ACP) y la georreferenciación, para caracterizar a la población femenina diagnosticada con cáncer cervicouterino según su incidencia, mortalidad, nivel educativo y acceso a servicios de salud. Los resultados revelan patrones regionales de vulnerabilidad y permiten construir un índice de evaluación social y sanitaria. Más allá de los datos, la investigación destaca la importancia de la estadística multivariada como herramienta de intervención, capaz de transformar la información compleja en conocimiento útil para la toma de decisiones y el diseño de políticas públicas que fortalezcan la equidad en salud y la atención oportuna.

Palabras clave: CaCu, VPH, análisis de conglomerados, componentes principales, desigualdad social.

¹⁷ gS19013294@egresados.uv.mx

¹⁸ emmorales@uv.mx

¹⁹ doliverag@gmail.com

INTRODUCCIÓN

La estadística es una ciencia que traduce los datos en comprensión. En el campo de la salud pública, esa comprensión puede significar salvar vidas, recursos mejor distribuidos o políticas más justas. Las técnicas multivariadas ofrecen una mirada amplia sobre los fenómenos sociales y sanitarios: permiten observar interacciones entre múltiples variables y descubrir patrones que no serían visibles a simple vista (Díaz y Morales, 2012).

En el ámbito médico, el uso de herramientas estadísticas permite identificar factores de riesgo, establecer relaciones entre variables clínicas y sociales, y evaluar la efectividad de programas de prevención. En este contexto, el cáncer cervicouterino (CaCu) aunque se trata de una enfermedad prevenible, continúa siendo una de las principales causas de muerte en mujeres de países con rezago económico (Lazcano, 2015).

En México, las neoplasias malignas constituyen una de las principales causas de muerte en mujeres; el CaCu se asocia directamente con el Virus del Papiloma Humano (VPH), cuyas cepas de alto riesgo son responsables de más del 90 % de los casos (Vargas et al., 2011). En el Estado de Veracruz, pese a la existencia de programas de detección, persisten cifras elevadas de mortalidad y brechas significativas en el acceso a servicios de salud (González, 2017).

La detección oportuna del cáncer cervicouterino, a través de pruebas como el Papanicolau, la prueba de VPH o la colposcopia, resulta determinante para reducir la mortalidad. Sin embargo, factores estructurales como la falta de acceso a servicios de salud, la marginación social, la baja educación sexual y la desinformación limitan la eficacia de estas estrategias. Por ello, resulta indispensable un abordaje estadístico que permita caracterizar la población afectada e identificar patrones que orienten acciones focalizadas de prevención.

Este trabajo busca demostrar cómo el uso de técnicas estadísticas multivariadas, particularmente el análisis de conglomerados y el análisis de componentes principales, permiten integrar múltiples variables relacionadas con la enfermedad para caracterizar la población femenina con cáncer cervicouterino en el Estado de Veracruz. Al hacerlo, se pretende resaltar el valor práctico de la estadística como instrumento de diagnóstico, comprensión y acción en contextos de salud pública.

En este sentido, la investigación no se limita a un ejercicio descriptivo, sino que propone una forma de leer los datos como una historia colectiva: la historia de cómo las condiciones sociales determinan la salud y cómo la evidencia estadística puede guiar la intervención social, con el propósito de contribuir al diseño de políticas públicas efectivas y equitativas.

Antecedentes

Diversos estudios han abordado la caracterización de poblaciones con enfermedades oncológicas mediante herramientas estadísticas y epidemiológicas. En Rodas, Cuba, Reina et al. (2018) realizaron un estudio descriptivo en pacientes con cáncer de mama, analizando variables como edad, escolaridad y ocupación, para comprender el contexto social de las afectadas. En el Hospital Gineco-Obstétrico de

Guanabacoa (Cuba), Torriente et al. (2011) desarrollaron un análisis longitudinal retrospectivo de mujeres con cáncer invasor de cuello uterino, concluyendo que los factores de riesgo más relevantes eran el inicio temprano de la vida sexual, el tabaquismo y la infección por VPH.

En México, Castillo et al. (2020) efectuaron un estudio descriptivo sobre cáncer de laringe, donde se observaron patrones de prevalencia relacionados con el tabaquismo y la edad. Por su parte, Barriga (2012) aplicó modelos de regresión de Poisson para examinar las tasas de mortalidad por cáncer cervicouterino entre 2000 y 2008, identificando una fuerte relación entre la mortalidad y las condiciones socioeconómicas, particularmente en entidades con mayor marginación como Guerrero, Chiapas y Oaxaca.

En el ámbito del análisis espacial, Reyes (2020) desarrolló un estudio cuantitativo sobre el perfil epidemiológico del cáncer cervicouterino en mujeres potosinas, aplicando técnicas de autocorrelación espacial para estimar índices de prevalencia, incidencia y mortalidad. Este trabajo evidenció la utilidad de los sistemas de información geográfica como apoyo para la visualización y análisis de datos de salud. En un contexto similar, Castañeda et al. (1998) llevaron a cabo un estudio de casos y controles en Zacatecas, encontrando que el riesgo de padecer CaCu aumentaba con el número de partos, el inicio temprano de la vida sexual y el uso prolongado de anticonceptivos hormonales.

Para el caso específico del Estado de Veracruz, Coronel et al. (2010) analizaron la incidencia de cáncer cervicouterino en pacientes del Centro Estatal de Cancerología de Xalapa, observando que la mayoría de las mujeres afectadas provenían de zonas urbanas y que un porcentaje considerable no se había realizado estudios de detección previa. González (2017) destacó las deficiencias y limitaciones en la implementación del programa estatal de detección, señalando la falta de recursos financieros y personal capacitado, lo que impide alcanzar una cobertura adecuada en las regiones de mayor marginación.

Estos antecedentes ponen de manifiesto la necesidad de abordar el cáncer cervicouterino desde una perspectiva multidimensional, combinando información médica, demográfica y social, y que permita establecer una radiografía regional del problema.

A diferencia de los análisis univariados y bivariados, las técnicas multivariadas permiten estudiar simultáneamente varias dimensiones de un fenómeno, identificando patrones, estructuras y relaciones latentes que no son visibles a simple vista (Díaz y Morales, 2012).

Entre las más utilizadas destacan el análisis de componentes principales (ACP), que resume la información de múltiples variables correlacionadas en un conjunto reducido de componentes independientes; y el análisis de conglomerados (cluster analysis), que permite clasificar observaciones (personas, regiones, instituciones) en grupos homogéneos entre sujetos dentro del mismo cluster y que a su vez sean lo más heterogéneos entre cluster, según sus características.

En el campo de la salud pública, estas herramientas se han empleado para clasificar poblaciones por riesgo, evaluar la efectividad de programas sanitarios y detectar zonas prioritarias de atención. En particular, el análisis de conglomerados ha demostrado ser útil para la planificación territorial, al permitir agrupar regiones con características similares y diseñar estrategias diferenciadas (Minguillón y Casas, 2017).

Por su parte, el ACP facilita la creación de indicadores sintéticos o scores de desarrollo, integrando variables de distinta naturaleza en una sola medida que refleje tendencias globales. En este estudio, ambas técnicas se complementan: mientras una agrupa los municipios veracruzanos según su comportamiento epidemiológico, la otra sintetiza sus principales dimensiones sanitarias y sociales.

MATERIALES Y MÉTODOS

El estudio se desarrolló bajo un enfoque cuantitativo, descriptivo y explicativo, basado en la aplicación de técnicas estadísticas multivariadas. El objetivo central consiste en determinar las características de la población femenina con diagnóstico de cáncer cervicouterino en el Estado de Veracruz, con el propósito de identificar patrones de comportamiento y establecer agrupamientos de municipios según sus condiciones sanitarias y sociales.

Fuentes de información

Los datos utilizados provienen de registros oficiales de la Secretaría de Salud y del Instituto Nacional de Estadística y Geografía (INEGI, 2019), complementados con bases de datos públicas sobre mortalidad, educación, servicios médicos y pruebas de detección. La unidad de análisis corresponde a los 212 municipios de Veracruz, los cuales se distribuyen en 10 regiones distintas. Se seleccionaron variables relacionadas con incidencia y mortalidad por CaCu, casos confirmados de CaCu, tipo de tratamiento, nivel educativo, tipo de servicio de salud, cobertura de pruebas Papanicolau y presencia de programas de prevención.

Técnicas estadísticas aplicadas

Se aplicaron dos técnicas multivariadas principales: el análisis de conglomerados k-medias y el análisis de componentes principales (ACP).

a) Análisis de conglomerados k-medias

El análisis de conglomerados k-medias es una técnica no supervisada que clasifica observaciones en grupos homogéneos con base en su similitud (Minguillón y Casas, 2017). Su objetivo es minimizar la variabilidad interna dentro de cada grupo y maximizar las diferencias entre grupos. En este caso, se aplicó para identificar patrones entre municipios con características epidemiológicas y sociales similares.

Para determinar el número óptimo de conglomerados se empleó el método del codo, que analiza la varianza explicada por el número de grupos, y el coeficiente de silueta, que evalúa la cohesión y separación de los conglomerados.

b) Análisis de componentes principales (ACP)

El análisis de componentes principales (ACP) transforma un conjunto de variables correlacionadas en un número menor de componentes lineales independientes entre sí. Esta técnica permite reducir la dimensionalidad de los datos sin perder información esencial, facilitando la interpretación y construcción de índices sintéticos (Monroy y Morales, 2012).

En el estudio, el ACP se utilizó para construir un índice de evaluación del cáncer cervicouterino, integrando variables de educación, salud y detección. Los componentes seleccionados fueron aquellos con valores propios mayores a uno, y se interpretaron según sus cargas factoriales para determinar qué variables contribuyen más a la varianza total.

Procedimiento analítico

El proceso siguió cinco etapas principales:

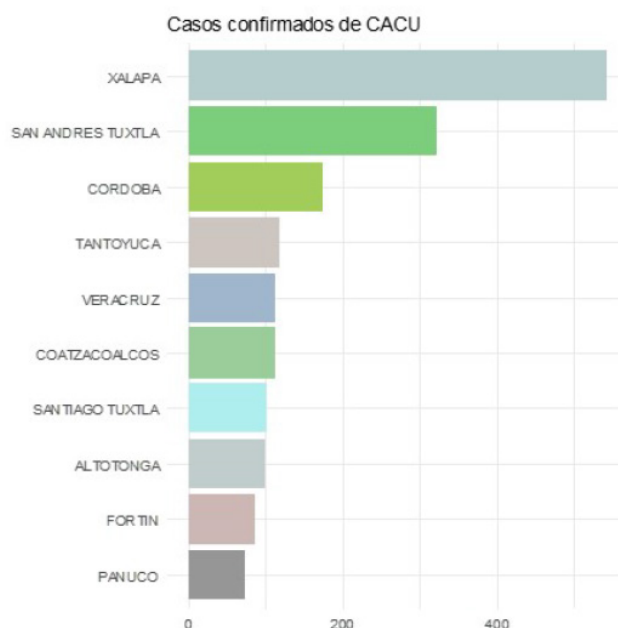
1. Depuración y normalización de la base de datos mediante escalamiento estandarizado (z-score).
2. Análisis descriptivo y exploratorio, para conocer la distribución y comportamiento de las variables.
3. Aplicación del análisis de conglomerados k-medias para identificar agrupamientos de municipios según los casos confirmados, la mortalidad y los factores sociales.
4. Aplicación del ACP, para construir el índice de evaluación sanitaria-social y mapeo de los resultados.
5. Interpretación y validación de resultados, contrastando los hallazgos con la evidencia empírica y bibliográfica existente.

El resultado de estas etapas permitió no solo clasificar los municipios, sino también proponer un marco analítico replicable para otros contextos de salud pública.

RESULTADOS

Estadística descriptiva

Figura 1. Municipios con más casos de CaCu.



En la Figura 1, se observan diez de los 212 municipios, con más casos confirmados de cáncer cervicouterino en el periodo de 2016 a 2019, siendo Xalapa, San Andrés Tuxtla y Córdoba los que encabezan dicha lista, teniendo más de 200 casos confirmados en los primeros dos lugares. Dichos municipios y agregando algunos como Veracruz, Coatzacoalcos, Papantla y Martínez de la Torre, también presentan altas cifras de pruebas de detección del CaCu, como lo son: Citología, Citologías base líquida, PCR y pruebas de detección del Virus de Papiloma Humano (VPH). Dichas pruebas según sea el caso, son para corroborar un diagnóstico o como medida de prevención. En periodo analizado, de manera general, fueron realizadas 475,869 pruebas, de las cuales un 66.5% fueron citologías, 21.56% pruebas PCR, 9.34% prueba VPH y un 2.6% citologías base líquida.

Análisis multivariado

El análisis de los datos permitió identificar patrones estadísticos y comportamientos diferenciados en las distintas regiones del Estado de Veracruz. A través de la aplicación del análisis de conglomerados k-medias, se establecieron agrupamientos de municipios según sus características de incidencia, mortalidad y condiciones sociales. Los resultados mostraron una marcada heterogeneidad territorial: mientras algunos municipios presentan tasas moderadas de casos confirmados y una adecuada cobertura de detección, otros reflejan condiciones de vulnerabilidad sanitaria y social, caracterizadas por baja escolaridad, limitado acceso a servicios médicos y elevada mortalidad.

Figura 2. Análisis clúster para cada variable en estudio.



El análisis de conglomerados permitió distinguir tres clústeres principales de acuerdo con las 4 variables en estudio (Figura 2):

- **Cluster 1:** Reunió a municipios donde predominan deficiencias estructurales en los servicios de salud y una cobertura insuficiente de pruebas Papanicolau. Este grupo se concentró principalmente en regiones de la Huasteca Alta y Papaloapan, donde la población femenina en edad reproductiva enfrenta mayores barreras, caracterizadas por baja escolaridad y limitado acceso a servicios de salud.
- **Cluster 2:** Incluyó municipios con niveles intermedios de detección y una mayor participación de los servicios públicos de salud, principalmente en zonas urbanas del centro del estado.
- **Cluster 3:** Se conformó por municipios con baja incidencia y mortalidad, ubicados en regiones con mejor infraestructura sanitaria y mayor proporción de mujeres con estudios de nivel medio y superior.

Estas diferencias reflejan no solo desigualdades en los recursos, sino también en la capacidad de respuesta del sistema de salud. Los municipios con mayor rezago educativo y menor acceso a servicios de salud son los que concentran la mayor carga de mortalidad por cáncer cervicouterino. Este hallazgo coincide con lo reportado por Barriga (2012), quien identificó que los niveles socioeconómicos bajos se asocian con una mayor probabilidad de morir por CaCu, así como con la falta de diagnósticos oportunos. Asimismo, González (2017) señala que en Veracruz los programas de detección presentan cobertura limitada en las comunidades marginadas, debido a la falta de personal capacitado y de recursos financieros.

Índice de evaluación mediante ACP

El análisis de componentes principales permitió reducir la dimensionalidad del conjunto de datos utilizado, sintetizando la información en un índice de evaluación de casos confirmados de cáncer cervicouterino. El primer componente, que explicó la mayor proporción de la varianza total un 74.7%, estuvo asociado a variables relacionadas con el tipo de prueba de detección y mortalidad.

El índice resultante permitió clasificar los municipios en función de su situación respecto al cáncer cervicouterino, a mayor score representan mayor riesgo de mortalidad a causa del CaCu y por lo contrario a menor score un riesgo menor.

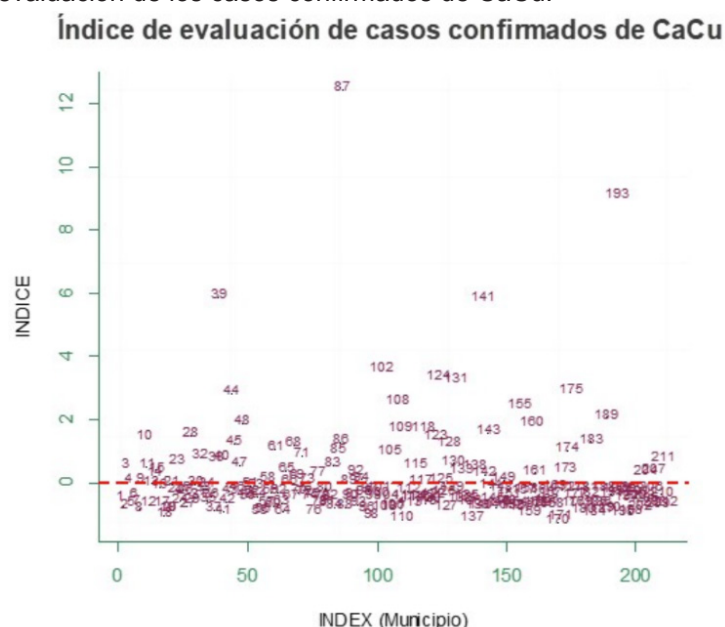
Los valores por debajo del umbral corresponden a municipios con mayor cobertura de detección, mejor nivel educativo y menor mortalidad. Reúnen pocas o nulas defunciones, pocos casos confirmados, un índice de rezago social bajo y pocas pruebas de detección.

Los municipios que oscilan en el umbral y un poco por arriba de este, son los que presentan un índice medio, estos presentan casos confirmados y muertes, pero no en cifras alarmantes, presentando pruebas de detección temprana en niveles altos y un índice de rezago social bajo.

Los valores altos (por arriba del umbral) reflejan mayor rezago social y deficiencia sanitaria. Tienen un número alto de casos confirmados, muertes e índices de rezago social muy altos y a su vez con un número de pruebas de detección oportuna muy bajo,

Este índice constituye una herramienta sintética y objetiva que facilita la interpretación de los datos y su uso en la planificación de intervenciones sanitarias.

Figura 3. Índice de evaluación de los casos confirmados de CaCu.



Aplicación del índice en políticas públicas

El índice propuesto tiene un valor práctico para la formulación de políticas públicas y la gestión de programas de salud. Desde un enfoque de intervención social, este indicador permite identificar zonas prioritarias donde las estrategias de prevención y detección temprana deben fortalecerse. Su construcción mediante técnicas multivariadas proporciona un marco metodológico robusto que integra múltiples dimensiones sociales, sanitarias y demográficas en una sola medida cuantitativa (Monroy y Morales, 2012).

Al utilizarse como herramienta de diagnóstico situacional, el índice podría incorporarse en los sistemas estatales de vigilancia epidemiológica y en los tableros de planeación de la Secretaría de Salud. Asimismo, su aplicación periódica permitiría evaluar la efectividad de las campañas de detección y realizar comparaciones temporales y regionales. De esta forma, la estadística multivariada se convierte en un puente entre la evidencia empírica y la acción gubernamental, contribuyendo a una asignación más eficiente de los recursos y al diseño de estrategias focalizadas en la reducción de la mortalidad por CaCu.

Estos hallazgos respaldan la afirmación de que el cáncer cervicouterino no debe abordarse únicamente desde la biología médica, sino también como un fenómeno social vinculado con las desigualdades estructurales. Por tanto, el conocimiento estadístico se posiciona como un instrumento esencial para la justicia sanitaria, al permitir visibilizar los territorios más afectados y orientar acciones concretas de prevención y tratamiento.

CONCLUSIONES

El análisis estadístico multivariado aplicado en esta investigación permitió caracterizar la situación del cáncer cervicouterino en el Estado de Veracruz desde una perspectiva integral, evidenciando la relación entre factores sociales, educativos y sanitarios. Las técnicas utilizadas demostraron su eficacia para re-

sumir la información y revelar patrones subyacentes en los datos, lo que permite una comprensión más amplia del problema.

Entre las conclusiones principales destacan las siguientes:

1. **La estadística multivariada como instrumento de intervención.** Las técnicas empleadas (análisis de conglomerados k-medias y análisis de componentes principales) mostraron ser herramientas efectivas para analizar fenómenos sociales complejos como el cáncer cervicouterino, donde intervienen múltiples variables correlacionadas.
2. **Relación entre rezago social y mortalidad.** Los municipios con menores niveles educativos y menor cobertura sanitaria registran las tasas más altas de mortalidad, esto no quiere decir que carecer de educación sea un diferenciador entre morir o no, si no que confirma la influencia de los determinantes sociales de la salud en la evolución del CaCu (Lazcano, 2015; Vargas et al., 2011).
3. **Importancia del diagnóstico oportuno.** La detección temprana mediante pruebas regulares reduce significativamente el riesgo de muerte, lo que subraya la necesidad de fortalecer las campañas de educación sexual y los servicios de salud preventiva (Coronel et al., 2010).
4. **Propuesta del índice de evaluación.** El índice construido a partir del ACP representa una herramienta cuantitativa útil para medir la situación regional del cáncer cervicouterino. Su aplicación puede contribuir a la toma de decisiones basadas en evidencia, la priorización territorial y la evaluación del impacto de las políticas públicas.
5. **Contribución al campo de la estadística aplicada.** Este trabajo demuestra que la estadística, más allá de su carácter analítico, tiene un potencial transformador al ser utilizada como herramienta de intervención social y sanitaria. A partir del análisis empírico, es posible orientar recursos, diseñar estrategias focalizadas y evaluar el desempeño institucional con base en información verificable.

La investigación reafirma el papel de la estadística como ciencia al servicio de la sociedad, capaz de traducir los datos en conocimiento accionable. La comprensión de las desigualdades en salud, a través de modelos multivariados, constituye un paso esencial hacia la reducción de brechas y la mejora de la calidad de vida de las mujeres veracruzanas. Como trabajo a futuro, se propone ampliar la base de datos incorporando variables de tipo temporal y geoespacial, así como extender la metodología a otros tipos de cáncer, enfermedades o tema de interés público, con el fin de fortalecer la planeación sanitaria y la equidad en salud.

REFERENCIAS

Barriga, J. J. (2012). *Tendencias de mortalidad por cáncer cervicouterino en las siete regiones socioeconómicas y en las 32 entidades federativas de México en los años 2000–2008*. *Gaceta Médica de México*, 148(1), 42–51.

- Castañeda Iñiguez, M. S., Toledo Cisneros, R., & Aguilera Delgadillo, M. (1998). *Factores de riesgo para cáncer cervicouterino en mujeres de Zacatecas*. *Salud Pública de México*, 40(4), 330–338. <https://www.scielo.org/article/spm/1998.v40n4/330-338/es/>
- Castillo López, I. Y., Govea Camacho, L. H., Sánchez Robles, E. E., & Palacios López, J. C. (2020). *Caracterización clínico-patológica del cáncer de laringe en la población mexicana*. *Revista Médica del Instituto Mexicano del Seguro Social*, 59(1), 27–33. <https://doi.org/10.24875/RMIMSS.M21000049>
- Coronel Brizio, P., Palafox Sánchez, F., Olivares Nowak, J., Díaz Domínguez, E. Z., & Ortega Luna, N. (2010). *Cáncer del cuello uterino según zona urbana, suburbana y rural en Veracruz (SESVIR 2006–2008)*. *Revista Médica UV*, 10(1), 6–10. https://www.uv.mx/rm/num_anteriores/revmedica_vol10_num1/articulos/cancer.pdf
- Díaz Monroy, L. G., & Morales Rivera, M. A. (2012). *Análisis estadístico de datos multivariados*. Bogotá, Colombia: Universidad Nacional de Colombia.
- Díaz Monroy, L. G., & Morales Rivera, M. A. (2020). *Análisis de conglomerados*. En L. G. Díaz Monroy & M. A. Morales Rivera (Eds.), *Análisis estadístico de datos multivariados* (pp. 407–439). Bogotá: Editorial Universidad Nacional de Colombia.
- González, Z. M. (2017). *Programa de detección del cáncer cervicouterino: políticas públicas y experiencias de los actores que implementan el programa en el estado de Veracruz, México*. *Salud Colectiva*, 13(3), 521–535.
- Lazcano, E. P. (2015). *Cáncer cervicouterino*. En *Observatorio de Salud Pública en México* (pp. 137–149).
- Minguillón, J., & Casas, J. (2017). *Minería de datos: modelos y algoritmos*. Barcelona: Universitat Oberta de Catalunya (UOC).
- Monroy, L. G., & Morales Rivera, M. A. (2012). *Estadística multivariada: inferencia y métodos*. Bogotá: Universidad Nacional de Colombia.
- Reyes, L. M. (2020). *Perfil epidemiológico-espacial del cáncer cervicouterino de mujeres potosinas afiliadas al Seguro Popular del periodo 2008–2018* [Tesis de maestría, Universidad Autónoma de San Luis Potosí]. <https://repositorioinstitucional.uaslp.mx/xmlui/bitstream/handle/i/7152/TesisM.FEN.2020.Perfil.Badillo%28versi%C3%B3n%20p%C3%BAblica%29.pdf>
- Torriente Hernández, S., Valdés Álvarez, O., Villarreal Martínez, A., & Lugo Sánchez, A. (2011). *Caracterización de un grupo de pacientes con neoplasia intraepitelial cervical diagnosticadas por biopsia con asa diatérmica*. *Revista Cubana de Obstetricia y Ginecología*, 37(1), 42–53. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-600X2011000100006
- Vargas Hernández, V. M., Vera Gaspar, D., Acosta Altamirano, G., & Cueriel Valdés, J. D. (2011). *Virus del papiloma humano*. En V. M. Hernández (Ed.), *Cáncer en la mujer. Volumen 1* (pp. 615–633). Editorial Alfíl, S.A. de C.V.

ANÁLISIS DE VENTAS INTERNACIONALES UTILIZANDO TÉCNICAS DE MACHINE LEARNING CON POWER BI

Eliseo Gabriel Argúelles²⁰

Demetrio Sánchez Jiménez²¹

Itzel Alessandra Reyes Flores²²

Facultad de Estadística e Informática

Universidad Veracruzana

RESUMEN

El presente capítulo analiza las dinámicas de las ventas internacionales y las exportaciones de productos a nivel global durante el periodo 2020–2022, utilizando herramientas estadísticas y técnicas de machine learning. A través de un estudio observacional, descriptivo y longitudinal, se examinaron los factores que influyen en las ventas en dólares y su impacto en las economías nacionales. La base de datos utilizada incluyó mil observaciones con información sobre productos, costos totales, precios de venta e importes globales, procesadas mediante Power BI. Se aplicaron métodos de análisis de series de tiempo, regresión lineal y clustering, lo que permitió identificar patrones de comportamiento, tendencias estacionales y agrupamientos de productos con características similares. Los resultados muestran fluctuaciones significativas asociadas con eventos globales como la pandemia de COVID-19 y la variación del valor del dólar, así como diferencias notables entre regiones y tipos de productos. Las visualizaciones revelaron que los productos domésticos y los materiales de oficina concentraron los mayores costos de exportación, mientras que las frutas y los alimentos presentaron menor participación. El uso de técnicas de machine learning y análisis predictivo demostró su eficacia para mejorar la comprensión del comercio internacional y anticipar escenarios futuros. Los hallazgos contribuyen al diseño de estrategias comerciales y políticas económicas más informadas, orientadas a reducir los riesgos derivados de la volatilidad del mercado global. Asimismo, se propone considerar la adopción de una canasta de monedas como alternativa al uso exclusivo del dólar estadounidense, con el fin de promover la estabilidad y la equidad en el sistema monetario internacional.

Palabras claves: Comercio internacional, exportaciones, series de tiempo; análisis de clúster.

²⁰ egabriel@uv.mx

²¹ sanchezjimenezdemetrio@gmail.com

²² itreyes@uv.mx

INTRODUCCIÓN

En un contexto económico global caracterizado por la interconexión y la competitividad, el análisis de las ventas internacionales y de las exportaciones constituye un componente esencial para comprender la dinámica del comercio mundial. Las transformaciones tecnológicas, la digitalización de los procesos productivos y el auge del análisis de datos han impulsado el uso de herramientas estadísticas y algoritmos de machine learning como soporte a la toma de decisiones estratégicas en el ámbito del comercio exterior. Estas técnicas permiten identificar patrones, prever fluctuaciones en los mercados internacionales y optimizar la gestión de los recursos en las empresas exportadoras.

El uso del machine learning aplicado al comercio internacional ha cobrado relevancia por su capacidad de procesar grandes volúmenes de información y generar modelos predictivos que superan las limitaciones de los métodos tradicionales. Jošić y Žmuk (2022), por ejemplo, presentan una aproximación innovadora basada en machine learning para pronosticar el comercio bilateral de Croacia con más de 180 países entre 2001 y 2019, demostrando que los algoritmos, especialmente las redes neuronales del tipo Multi-layer Perceptron, poseen una alta capacidad para anticipar los flujos de importación y exportación. Este hallazgo pone de manifiesto el potencial de estas herramientas para mejorar la precisión en la toma de decisiones comerciales y anticipar escenarios futuros.

De manera complementaria, Gabriel (2024) realiza un análisis cualitativo conceptual de publicaciones institucionales como CEPAL, Redalyc, SEICA y BID, en el que resalta la importancia de comprender la dinámica del comercio global desde una perspectiva epistemológica. En un mundo interdependiente, las empresas se convierten en actores fundamentales de la vida económica y social, y sus estrategias de exportación e importación se ven influenciadas por factores políticos, sociales y tecnológicos. El autor subraya cómo la apertura comercial y la inserción internacional de las empresas centroamericanas han permitido redefinir su papel en el comercio de bienes y servicios, impulsando el desarrollo sostenible y el bienestar económico en la región.

En este contexto, la Organización Mundial del Comercio y la OCDE (2021) ofrecen una panorámica sobre la evolución del comercio internacional de servicios, destacando su relevancia dentro de la Organización Mundial del Comercio (OMC) y los beneficios económicos derivados de la liberalización y regularización de los servicios digitales, estimados en más de 135.000 millones de dólares. Este fenómeno confirma la creciente digitalización del comercio y la necesidad de integrar herramientas de inteligencia artificial que permitan comprender sus dinámicas y anticipar sus transformaciones.

Por su parte, Orlando y Noj-Pérez (2024) enfatizan el papel de los compiladores y la optimización algorítmica en el desarrollo de la inteligencia artificial (IA) y el aprendizaje automático (AA), subrayando que el rendimiento de los algoritmos depende en gran medida de su adaptación a hardware especializado como las GPU o TPU. Esto muestra cómo la eficiencia computacional se ha vuelto un elemento clave para potenciar el análisis de datos masivos y su aplicación en contextos económicos, financieros y comerciales.

Sin embargo, más allá del potencial tecnológico, el comercio internacional continúa enfrentando desafíos estructurales. La consolidación del dólar estadounidense como moneda de referencia en el último siglo ha simplificado las transacciones internacionales, pero también ha generado una dependencia que expone

a los países y empresas a los vaivenes de la política monetaria de Estados Unidos. Una apreciación del dólar encarece las exportaciones y afecta la competitividad global, mientras que su depreciación altera las condiciones del mercado interno. Asimismo, las políticas proteccionistas, los acuerdos multilaterales, las barreras arancelarias y las fluctuaciones de los tipos de cambio influyen de manera directa en el volumen de ventas y en la rentabilidad de las operaciones internacionales.

En este escenario, comprender las relaciones entre las ventas internacionales, las variaciones monetarias y los factores comerciales es crucial para formular estrategias que fortalezcan la competitividad y minimicen los riesgos económicos. Las técnicas de machine learning representan una herramienta poderosa para alcanzar dicho propósito, ya que permiten analizar de forma simultánea un amplio conjunto de variables —económicas, políticas y financieras— y generar modelos que predicen con mayor precisión el comportamiento de las ventas internacionales.

Este capítulo tiene como propósito analizar las dinámicas de las ventas globales en dólares y las exportaciones de productos, identificando los factores que las influyen y su impacto en las economías nacionales. A partir de la integración de métodos de machine learning, se busca ofrecer recomendaciones prácticas que contribuyan a la optimización de las estrategias comerciales y a la formulación de políticas más informadas. Así, el estudio no solo aspira a describir las tendencias actuales del comercio internacional, sino también a proponer un enfoque multidisciplinario que combine la economía, la estadística y la inteligencia artificial para comprender un fenómeno complejo, dinámico y de creciente relevancia en la economía mundial.

MATERIALES Y MÉTODOS

El presente estudio se desarrolló bajo un enfoque observacional, descriptivo y longitudinal, centrado en el análisis de las exportaciones a nivel mundial durante el periodo 2020–2022. El objetivo principal consistió en identificar patrones y tendencias en las ventas y costos de exportación, a fin de comprender su comportamiento en el contexto del comercio internacional.

Los materiales utilizados incluyeron fuentes de datos históricos sobre exportaciones globales y herramientas informáticas especializadas en análisis estadístico y visualización de información. La base de datos analizada contiene 1,000 observaciones correspondientes a productos exportados durante el periodo mencionado, con cuatro variables principales:

- Productos, que identifica el tipo de bien exportado (electrónicos, textiles, alimentos, entre otros);
- Costo total, que representa el costo total de producción y envío, incluyendo mano de obra, materiales y transporte;
- Costo de venta, que indica el precio de comercialización del producto en el mercado internacional; y
- Importe de venta total global, que refleja el valor total de ventas de cada producto a nivel mundial.

Estos datos fueron procesados en Microsoft Excel para su limpieza y transformación inicial, y posteriormente analizados en Microsoft Power BI, plataforma que ofrece amplias posibilidades para la representación visual e interactiva de datos, facilitando la identificación de patrones y la comunicación de resultados.

El análisis metodológico se estructuró en cuatro fases principales: (1) procesamiento y depuración de datos, (2) análisis estadístico de series de tiempo (3) agrupamiento (clúster), y (4) modelado predictivo mediante regresión lineal.

1. Procesamiento y depuración de datos: en la primera fase, se realizó la preparación de los datos mediante la eliminación de registros incompletos, la estandarización de formatos y la creación de variables derivadas. Este proceso permitió garantizar la calidad y consistencia de la información antes del análisis. Una vez depurada, la base fue importada a Power BI, donde se construyeron visualizaciones dinámicas de los costos y ventas totales globales. Estas gráficas sirvieron como insumo para la identificación de patrones generales y la generación de modelos predictivos.
2. Análisis de series de tiempo: las series de tiempo son secuencias de observaciones registradas en intervalos regulares (por ejemplo, años), utilizadas para analizar tendencias y variaciones a lo largo del tiempo. Según Brockwell y Davis (2016), su estudio permite descomponer los datos en componentes estacionales, cíclicos y tendenciales, aportando información valiosa para la predicción de comportamientos futuros. En este trabajo, se utilizó la variable año con tres categorías (“2020”, “2021” y “2022”), lo que permitió observar la evolución temporal de los costos y ventas globales. De acuerdo con Hamilton (1994), un aspecto clave en el análisis de series de tiempo es la estacionariedad, que implica que las propiedades estadísticas —como la media, la varianza y la covarianza— permanecen constantes a lo largo del tiempo. Por su parte, Greene (2018) señala que la heterocedasticidad, entendida como la variación no constante de la varianza de los errores, puede distorsionar los resultados del modelo si no se considera adecuadamente. Asimismo, se identificaron distintos tipos de series según su estructura: univariantes, que analizan una sola variable, y multivariantes, que estudian simultáneamente varias (Tsay, 2014; Zivot y Wang, 2006). Estas técnicas permitieron visualizar el comportamiento de las exportaciones globales y su relación con variables como los costos de producción y los precios de venta.
3. Análisis de clúster: para complementar el estudio temporal, se aplicó la técnica de agrupamiento (clustering), con el objetivo de identificar grupos de observaciones que compartieran comportamientos similares en función del tiempo. Este procedimiento permitió segmentar los datos de exportación según patrones comunes en las variaciones de costos y ventas.

El proceso de clustering se realizó mediante las siguientes etapas metodológicas:

Determinación del número de clústeres (k): Se utilizó el método del codo para encontrar el número óptimo de grupos que equilibrara la varianza interna y la homogeneidad entre grupos.

- Inicialización de centroides: Se asignaron centroides iniciales representativos de cada grupo.
- Asignación de observaciones: Cada caso fue asignado al clúster con el centroide más cercano.
- Reajuste iterativo de centroides: Los centroides se recalcularon hasta alcanzar la convergencia o un criterio de estabilidad predefinido.

Este procedimiento permitió identificar similitudes en los patrones de exportación, aportando una visión segmentada del comportamiento de los productos y mercados internacionales.

4. Modelado predictivo mediante regresión lineal: la fase final consistió en la aplicación de un modelo de regresión lineal, con el propósito de estimar la relación entre los costos totales de exportación (variable dependiente) y el tiempo (variable independiente). El modelo se formuló según la ecuación general:

$$y = mx + b$$

Donde:

- ☐ y representa el costo total,
- ☐ x corresponde al año,
- ☐ m es la pendiente, que indica la tasa de variación del costo total a lo largo del tiempo, y
- ☐ b es la intersección, que representa el valor inicial o base del costo total.

El modelo fue implementado en Power BI utilizando la función de pronóstico disponible en la pestaña “Análisis”. Posteriormente, se evaluó su precisión mediante la comparación entre los valores estimados y los datos reales, apoyándose en visualizaciones interactivas para validar la calidad del ajuste y la capacidad predictiva del modelo.

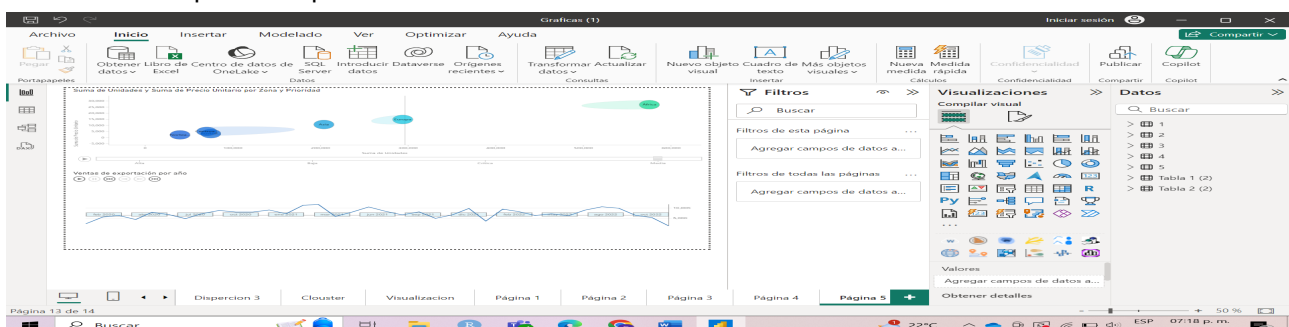
Diseño estadístico

En conjunto, el enfoque metodológico adoptado permitió integrar análisis descriptivos, temporales y predictivos en un mismo entorno visual. El uso combinado de Excel y Power BI facilitó el tratamiento de grandes volúmenes de datos y la exploración de patrones de comportamiento en las exportaciones internacionales.

Asimismo, la incorporación de técnicas estadísticas como el análisis de series de tiempo (Brockwell y Davis, 2016; Hamilton, 1994; Greene, 2018) y el agrupamiento (clustering) complementó la interpretación de las tendencias globales, mientras que el modelo de regresión lineal proporcionó una herramienta sólida para la predicción de costos y ventas futuras. Este diseño metodológico constituye una base empírica pertinente para la aplicación futura de algoritmos de machine learning en el estudio del comercio internacional.

RESULTADOS

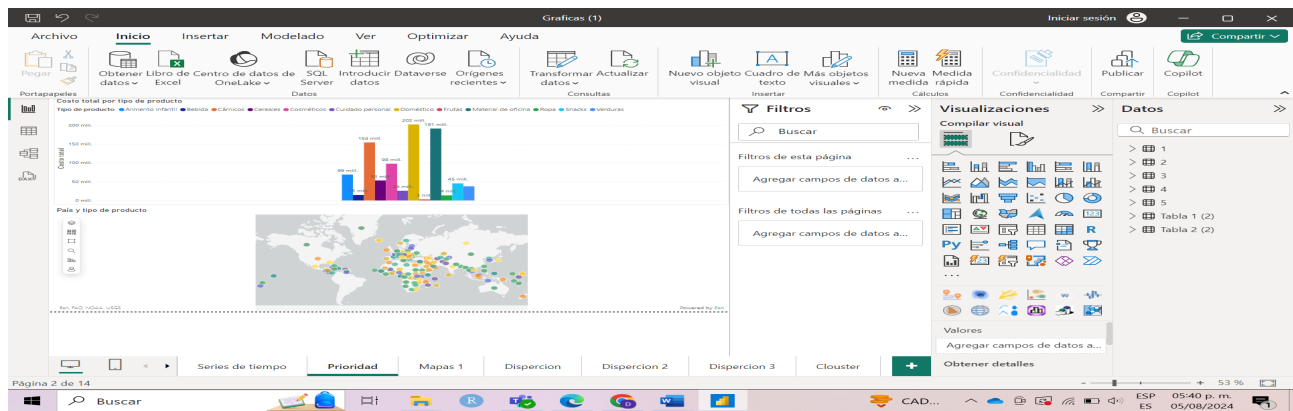
Figura 1. Ventas de exportación por año.



Fuente: Elaboración propia.

En la Figura 1 se muestran las ventas de exportación en el periodo del 2020 al 2022, presentando diferentes fluctuaciones debió a diferentes acontecimientos a lo largo de este intervalo.

Figura 2. Costo total por tipo de producto.



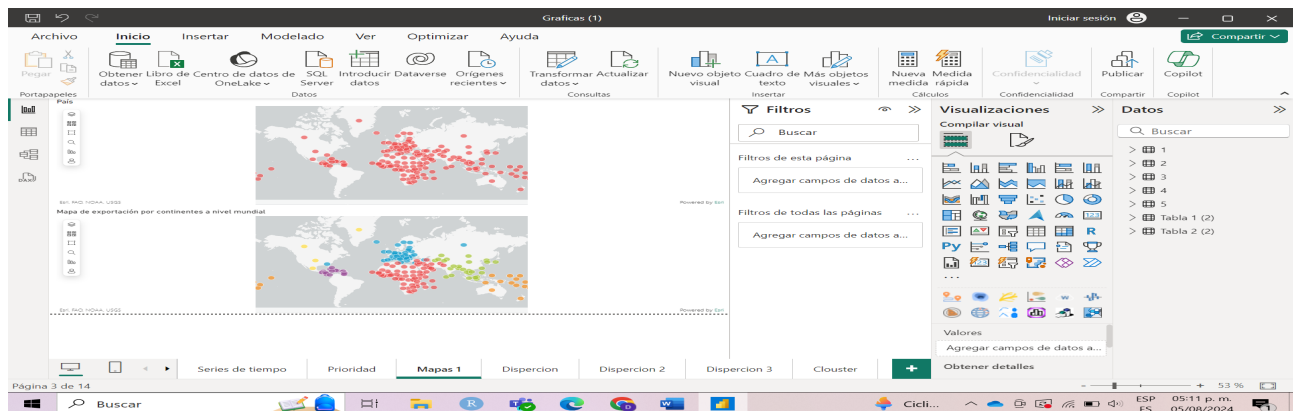
Fuente: Elaboración propia.

Los productos con mayor exportación entre el 2020 al 2022 son:

1. Domésticos ocupando el primer lugar con: \$202,134,654.04
2. Materiales de oficina en el segundo lugar con: \$191,256,052.05
3. Frutas finalmente ocupando el último lugar con: \$2,833,560.08

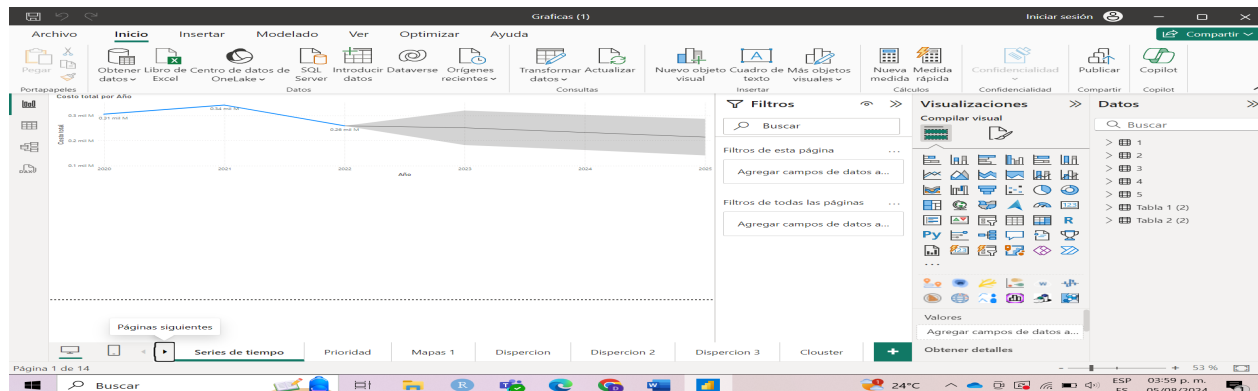
Estos datos indican que los productos domésticos tienen el mayor costo total de exportación en el período mencionado, seguidos de cerca por los materiales de oficina. En contraste, las frutas tienen un costo total de exportación significativamente menor.

Figura 3. Mapa de exportación por continente.



Fuente: Elaboración propia.

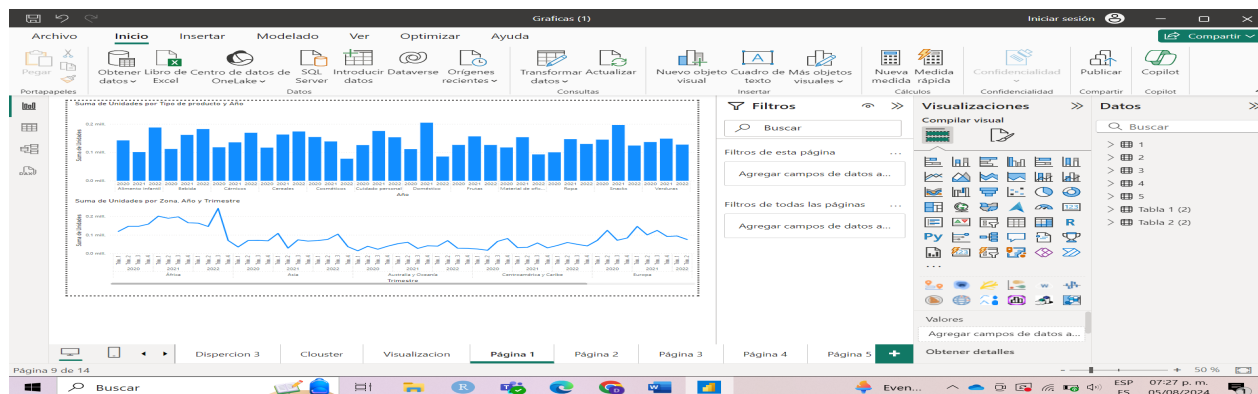
La Figura 3 muestra la exportación de productos por continente. Se puede observar que hay mayor distribución sobre la exportación en el continente africano que en los demás, A diferencia del continente europeo que muestra una mayor concentración.

Figura 4. Costo total por año.

Fuente: Elaboración propia.

La Figura 4 muestra la suma de los costos totales de las exportaciones a nivel mundial entre los años 2020 y 2023. Se observa un descenso en principios del 2020 por motivo de la pandemia, además de mostrarse un ascenso en 2021, una disminución en 2022 por la regularización del dólar, una estabilización en 2023. Es importante considerar que este es solo un indicador y otras variables pueden haber afectado las exportaciones durante este periodo.

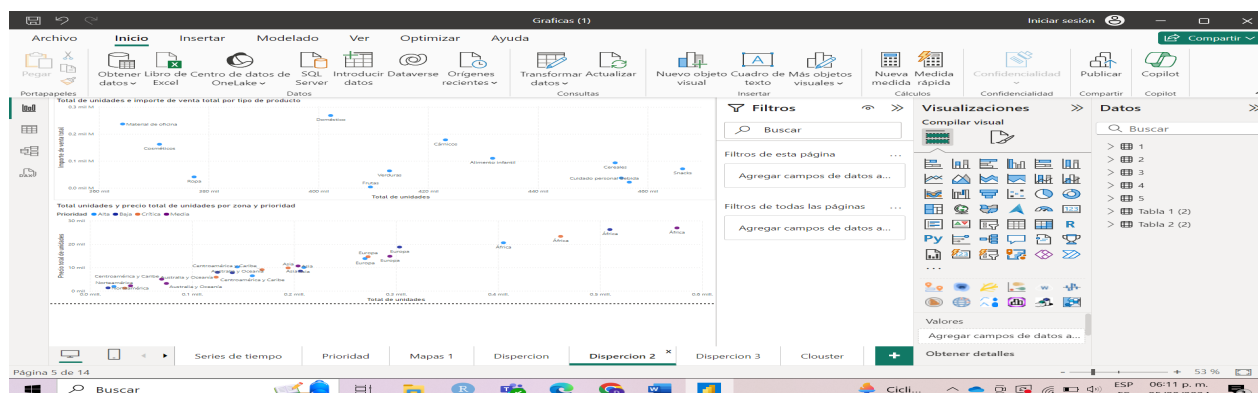
Los valores del costo total para los periodos fueron los siguientes: 2020: \$305,776,829.16, 2021: \$342,015,371.69 y 2022: \$259,284,316.53

Figura 5. Suma de unidades por zona, año y trimestre.

Fuente: Elaboración propia.

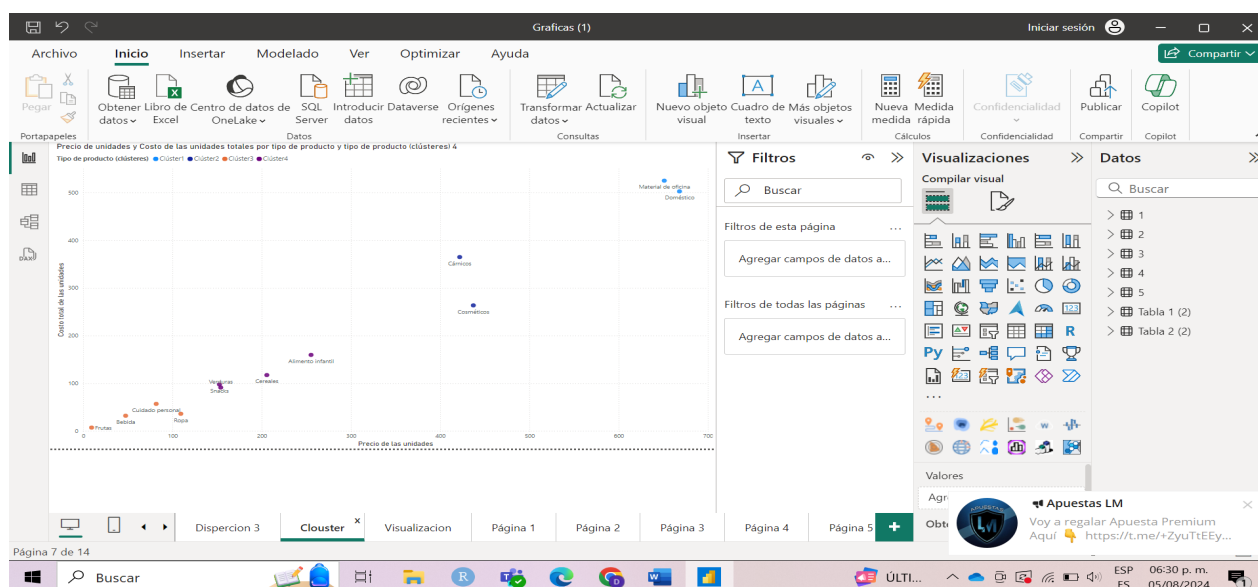
La Figura 5 muestra la suma de unidades por zona, año y trimestre. Se puede observar el año 2020, seguida de una disminución en el primer y segundo trimestre de 2021. A partir de mediados de 2021 y hasta 2022, se observa una relativa estabilidad.

Esta fluctuación en la suma de unidades podría estar relacionada con las restricciones y afectaciones que la pandemia de COVID-19 tuvo en la economía global. La recuperación económica en el último trimestre de 2020 y la primera mitad de 2021, seguida por la desaceleración de la economía en 2022, se ve reflejada en la gráfica.

Figura 6. Total de unidades e importe de venta total por tipo de producto.

Fuente: Elaboración propia.

En la Figura 6 de dispersión, se puede observar que los snacks son uno de los productos con mayor cantidad de unidades exportadas, con un total de 466,355 unidades y un importe de venta total de \$71,156,445. A diferencia del producto material de oficina presenta un total de 364,325 unidades vendidas y un importe de venta total de \$237,252,083.25.

Figura 7. Precio de unidades y costo de unidades totales por tipo de producto.

Fuente: Elaboración propia.

La Figura 7 representa un análisis clúster que agrupa productos en función de su precio unitario y su costo total. Se han identificado cuatro clústeres, cada uno con un centroide que representa el promedio de las observaciones dentro del grupo.

El clúster 1 (azul) contiene productos con un precio unitario alto y un costo total alto, etiquetados “Material de oficina” y “Doméstico”. El clúster 2 (azul oscuro) incluye productos con un precio unitario moderado y un costo total moderado, denominados “Cárnicos” y “Cosméticos”. El clúster 3 (naranja) se compone de productos con un precio unitario bajo y un costo total bajo, “Frutas”, “Cuidado personal”, “Bebida” y “Ropa”. Finalmente, el clúster 4 (morado) agrupa productos con un precio unitario moderado y un costo total bajo, como “Alimento infantil”, “Cereales” y “Snacks”.

CONCLUSIONES

El análisis realizado permitió comprender con mayor profundidad las dinámicas de las ventas internacionales y de las exportaciones de productos en dólares durante el periodo 2020–2022, un lapso marcado por fluctuaciones económicas globales derivadas de factores sanitarios, financieros y geopolíticos. Los resultados obtenidos evidencian que las variaciones en los tipos de cambio y las condiciones del mercado internacional influyen significativamente en los costos, precios y volúmenes de exportación, afectando de manera diferenciada a los sectores productivos y a las regiones del mundo.

La observación de las ventas por año muestra que el comercio internacional experimentó una caída pronunciada a inicios de 2020 como consecuencia de la pandemia, seguida de una recuperación parcial en 2021 y un proceso de estabilización en 2022. Este comportamiento fue corroborado por el análisis del costo total por año, que refleja un descenso inicial vinculado a la contracción económica y restricciones logísticas, y un repunte posterior asociado a la reactivación del mercado global.

Asimismo, el análisis por tipo de producto reveló que los bienes de consumo doméstico y los materiales de oficina concentraron la mayor proporción del costo total de exportación, mientras que las frutas y los productos alimentarios mostraron menores volúmenes, reflejando diferencias estructurales en la composición del comercio internacional. La distribución geográfica del comercio destacó una mayor diversificación de las exportaciones en África, en contraste con la concentración observada en Europa, evidenciando desigualdades regionales en los flujos comerciales.

El análisis de clúster permitió agrupar los productos en función de su precio unitario y costo total, identificando cuatro segmentos claramente diferenciados: productos de alto valor y costo, productos de valor intermedio, bienes de bajo costo y artículos de rotación masiva. Estos patrones reflejan la heterogeneidad de las cadenas globales de valor y la importancia de segmentar estrategias comerciales según la naturaleza del producto.

Metodológicamente, la combinación de análisis de series de tiempo, modelos de regresión lineal y técnicas de agrupamiento demostró ser eficaz para examinar las tendencias del comercio internacional y realizar estimaciones predictivas confiables. El uso de Power BI y el tratamiento sistemático de datos en Excel facilitaron la construcción de modelos visuales que integran componentes descriptivos y analíticos, aportando un enfoque práctico para la interpretación de información económica compleja.

Desde la perspectiva teórica y aplicada, el estudio confirma la utilidad del machine learning como herramienta complementaria para el análisis del comercio internacional. Su capacidad para identificar patrones no lineales, clasificar segmentos de productos y realizar pronósticos precisos constituye un avance relevante para el diseño de estrategias empresariales y la formulación de políticas públicas orientadas a la sostenibilidad comercial.

En conjunto, el capítulo aporta una comprensión integral de los factores que determinan las ventas y exportaciones globales, subrayando la necesidad de desarrollar estrategias basadas en evidencia empírica y modelos predictivos adaptativos. Este enfoque interdisciplinario, que combina economía, estadística y aprendizaje automático, ofrece un marco analítico robusto para comprender y anticipar los cambios en el entorno comercial global.

REFERENCIAS

- Brockwell, P. J., & Davis, R. A. (2016). *Introducción a las series temporales y a la previsión*. Salmer.
- Gabriel, R. (2024). Características de los productos y servicios que exporta e importa Centroamérica en el contexto de la globalización y el comercio internacional. *Cuadernos de Coyuntura*, 2, 60–79. <https://doi.org/10.48204/J.cc.n2.a5284>
- Greene, W. H. (2018). *Análisis econométrico*. Pearson.
- Hamilton, J. D. (1994). *Análisis de series temporales*. Princeton University Press.
- Jošić, H., & Žmuk, B. (2022). A Machine Learning Approach to Forecast International Trade: The Case of Croatia. *Business Systems Research*, 13(3), 144-160. <https://doi.org/10.2478/bsrj-2022-0030>
- OCDE & OMC. (2021). *Services domestic regulation in the WTO (Joint OECD–WTO Trade Policy Brief)*. https://www.wto.org/english/tratop_e/serv_e/oecd_wto_trade_policy_2021.pdf
- Orlando, F., & Noj Pérez, F. (2024). El impacto de los compiladores en la era de la inteligencia artificial y el aprendizaje automático. <https://doi.org/10.13140/RG.2.2.14183.66728>
- Tsay, R. S. (2014). *Análisis de series temporales financieras*. John Wiley & Sons.
- Zivot, E., & Wang, J. (2006). *Modelización de series temporales financieras con S-PLUS*. Salmer.