

MODELOS ESTADÍSTICOS

para la
investigación
científica

Aplicaciones
en las áreas
económico-
administrativas



Duniesky Feitó Madrigal
Malena Portal Boza
Ismael Plascencia López
(coordinadores)

Modelos estadísticos para la investigación científica

*Aplicaciones en las áreas
económico-administrativas*



**COMUNICACIÓN
CIENTÍFICA**

Ediciones Comunicación Científica se especializa en la publicación de conocimiento científico de calidad en español e inglés en soporte de libro impreso y digital en las áreas de humanidades, ciencias sociales y ciencias exactas. Guía su criterio de publicación cumpliendo con las prácticas internacionales: dictaminación de pares ciegos externos, autenticación antiplagio, comités y ética editorial, acceso abierto, métricas, campaña de promoción, distribución impresa y digital, transparencia editorial e indexación internacional.

Cada libro de la Colección Ciencia e Investigación es evaluado para su publicación mediante el sistema de dictaminación de pares externos y autenticación antiplagio. Invitamos a ver el proceso de dictaminación transparentado, así como la consulta del libro en Acceso Abierto.



www.comunicacion-cientifica.com

[DOI.ORG/ 10.52501/cc.131](https://doi.org/10.52501/cc.131)



Modelos estadísticos para la investigación científica

*Aplicaciones en las áreas
económico-administrativas*

Duniesky Feitó Madrigal
Malena Portal Boza
Ismael Plascencia López
(coordinadores)



Modelos estadísticos para la investigación científica. Aplicaciones en las áreas económico-administrativas / Duniesky Feitó Madrigal, Malena Portal Boza, Ismael Plascencia López (coordinadores). — Ciudad de México: Comunicación Científica, 2023.

254 páginas: Ilustraciones, (Colección Ciencia e Investigación).

ISBN: 978-607-59749-9-6

DOI: 10.52501/cc.131

1. Economía de la ciencia — Estadísticas científicas. 2. Política Económica — Planificación económica — Estadísticas Económicas. Feitó Madrigal, Duniesky, autor. II. Portal Boza, Malena, autor. III. Plascencia López, Ismael, autor. IV. Título. V. Serie.

LC: HA32

Dewey: 330.124.5

La titularidad de los derechos patrimoniales de esta obra pertenece a los autores D.R. Duniesky Feitó Madrigal, Malena Portal Boza, Ismael Plascencia López (coordinadores), 2023. Su uso se rige por una licencia Creative Commons BY-NC-ND 4.0 Internacional, <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.es>

Primera edición en Ediciones Comunicación Científica, 2023

Diseño de portada: Francisco Zeledón • Interiores: Guillermo Huerta

Ediciones Comunicación Científica S.A. de C.V., 2023

Av. Insurgentes Sur 1602, piso 4, suite 400

Crédito Constructor, Benito Juárez, 03940, Ciudad de México,

Tel. (52) 55 5696-6541 • móvil: (52) 55 4516 2170

info@comunicacion-cientifica.com • www.comunicacion-cientifica.com

 [comunicacioncientificapublicaciones](#)  [@ComunidadCient2](#)

ISBN 978-607-59749-9-6

DOI 10.52501/cc.131



Esta obra fue dictaminada mediante el sistema de pares ciegos externos.
El proceso transparentado puede consultarse, así como el libro en acceso abierto,
en <https://doi.org/10.52501/cc.131>

Índice

<i>Prólogo</i>	9
I. Introducción a la estadística <i>Alfredo Valadez García, Josué Aarón López Leyva</i>	11
II. Estadística descriptiva y modelos de tabulación cruzada <i>Edgar Jiménez Cerra, Malena Portal Boza</i>	29
III. Modelos estadísticos de elección discreta <i>Alfredo Valadez García, Martín Arturo Ramírez Urquidy</i>	70
IV. Modelos de ecuaciones estructurales con estimación PLS <i>Duniesky Feitó Madrigal, Magdelis Moreno Ortega</i>	90
V. Modelos de cointegración y ciclo común <i>Jorge Quiroz Félix, Carlos Alberto Flores Sánchez, Martha Ofelia Lobo Rodríguez</i>	123
VI. Diseño y validación de encuestas <i>Malena Portal Boza, Edgar Jiménez Cerra</i>	145
VII. Sistemas complejos aplicados a la investigación. El enfoque de redes sociales <i>Nadxilli López Pérez, Martha Cecilia Jaramillo Cardona</i>	173

VIII. Índice de confianza ciudadana. Percepción sobre economía y seguridad en la Zona Metropolitana de Tijuana <i>Ismael Plascencia López, Oswaldo Plascencia López, Abraham Neftalí Sodi Díaz</i>	202
<i>Anexos</i>	226
<i>Sobre los coordinadores y autores</i>	241

Prólogo

Hoy en día las universidades tienen el compromiso de formar profesionistas integrales que sean creativos y desarrollen capacidades para generar conocimientos, transformar el entorno y contribuir al avance de las ciencias y la innovación. Al mismo tiempo, las exigencias de un mercado laboral cada vez más competitivo y selectivo, acompañado de la introducción de tecnologías relacionadas con la generación, procesamiento y transmisión de información, obligan a que los estudiantes cuenten con las herramientas necesarias que les permitan enfrentar los nuevos retos de la transformación digital y el manejo de los grandes datos para la generación de soluciones en un entorno de sistemas complejos.

En esta importante labor, la *Estadística* como soporte de la investigación científica se convierte en el eje principal de estructuras, proyectos y sistemas más sólidos para el estudio de las problemáticas que atienden los programas de posgrado a través de sus líneas de generación y aplicación del conocimiento. Sin embargo, cuando se realiza investigación aplicada en las áreas económico-administrativas, los estudiantes se inclinan por estudios cualitativos derivado en gran medida del temor a enfrentarse al planteamiento de modelos matemáticos y al desarrollo de habilidades para comprender, aplicar e interpretar las herramientas cuantitativas, ya que en muchas ocasiones se asocian con la formación de las carreras de ingeniería, informática, física y matemática. Aun cuando los programas educativos han incorporado en sus planes de estudio materias como Análisis de Datos, Econometría o

Métodos Cuantitativos, por mencionar algunas, se identifica cierta apatía, angustia, falta de claridad y comprensión en los temas y contenidos relacionados con la estadística al momento de aplicarlos en sus trabajos de tesis. Por ello, el objetivo de este libro es presentar el desarrollo de diferentes modelos estadísticos en un lenguaje más amigable para el lector, con los fundamentos metodológicos necesarios para desarrollar investigación cuantitativa y sobre todo con la interpretación y análisis de datos a partir de ejemplos de investigaciones realizadas en las áreas económico-administrativas, en las que se combinan modelos estadísticos de primera y segunda generación.

Al preparar esta edición hemos tomado en cuenta las necesidades y demandas identificadas en la experiencia educativa de la Facultad de Contaduría y Administración, en especial de los estudiantes de posgrado de los diversos programas educativos, así como de los investigadores, docentes y profesionistas que han mostrado interés en incorporar técnicas estadísticas de avanzada tanto en el campo académico como profesional.

El colectivo de autores participantes en esta obra espera que sea de gran utilidad para los lectores y que se logre un crecimiento en la generación de proyectos investigativos con la aplicación de modelos estadísticos que impacten de manera directa en la propuesta de soluciones, en la discusión científica y en un mayor bienestar y calidad de vida de nuestra sociedad. Agradecemos la colaboración de investigadores de diversas unidades académicas, quienes hicieron posible este libro; a nuestra *alma mater*, la Universidad Autónoma de Baja California, y en especial al doctor Ismael Plascencia López, director de la facultad, por apoyar la generación de materiales bibliográficos de calidad para su divulgación entre la comunidad científica y todos los interesados.

LOS COORDINADORES

I. Introducción a la estadística

ALFREDO VALADEZ GARCÍA*
JOSUÉ AARÓN LÓPEZ LEYVA**

DOI: <https://doi.org/10.52501/cc.131.01>

Resumen

El presente capítulo presenta las generalidades de la importancia de la Estadística para las áreas económico-administrativas. Un aspecto relevante de este apartado es que no se limita el estudio económico, sino que, aparte de ello, muestra de qué manera las técnicas y herramientas estadísticas son de utilidad en la salud, deporte, la educación, la robótica y automatización, diseño de productos y servicios, mercadotecnia, comercio, sustentabilidad, medio ambiente, astronomía, agricultura, ganadería, política, entre otros. Se explica la manera en que la teoría estadística y su aplicación están presentes y juegan un rol sumamente importante en todas las disciplinas actuales del conocimiento, y realmente sería complejo asumir que los avances científicos y el progreso tecnológico modernos no tengan relación con la estadística aplicada. También, se explica de manera general, como la estadística tanto inferencial como descriptiva tienen una importante cabida en la toma de decisiones de las ciencias económico-administrativas.

Palabras clave: *estadística, toma de decisiones, análisis cuantitativo, aplicación.*

* Doctor en Ciencias Económicas. Profesor de tiempo completo en la Escuela de Administración y Negocios del Centro de Enseñanza Técnica y Superior (Cetys Universidad), campus Tijuana, México. ORCID: <https://orcid.org/0000-0002-2841-4972>

** Doctor en Ciencias en Electrónica y Telecomunicaciones. Profesor investigador de tiempo completo en la Escuela de Ingeniería del Centro de Enseñanza Técnica y Superior (Cetys Universidad), campus Ensenada, México. ORCID: <https://orcid.org/0000-0002-3004-5686>

Antecedentes e importancia general de la estadística

De acuerdo con las evidencias históricas, el surgimiento de la estadística radica, de manera general, en el hecho de recopilar datos relacionados con un grupo de personas en una localidad dada, es decir, como una actividad del Estado, de ahí su nombre. Así, en el ámbito del Estado, el objetivo era cuantificar algunos aspectos que podrían servir para mejorar diversas áreas de la vida cotidiana de los ciudadanos, con parámetros como la edad, sexo y estado civil, entre otros, los cuales se relacionan con los censos (una de las principales evidencias históricas). De hecho, fue tanto el éxito de la estadística aplicada al Estado, que su uso se extendió a otras disciplinas, y gracias a ella, a lo largo de la historia se han realizado importantes aportaciones a la humanidad, desde aspectos vinculados con la salud y la seguridad pública hasta políticas a favor del desarrollo personal y profesional. Obviamente, durante este periodo se desarrollaron técnicas, métodos e instrumentos que ayudaron significativamente a describir de mejor manera a la población en estudio. Con el transcurso del tiempo, se desarrolló la teoría de la probabilidad, inferencia estadística, entre otras ramas de las matemáticas y estadística que permitían un mejor análisis de la información y, por lo tanto, mejores argumentos para la toma de decisiones de manera inteligente (Freedman, 1999).

De modo particular, existe evidencia de que en Egipto en el año 3050 a. C. ya existía información sobre la población y riqueza con el objetivo de conocer la factibilidad del desarrollo de infraestructura, es decir, la construcción de las pirámides. En Babilonia, en el año 3000 a. C., también tenían registros de datos generales y detalles relacionados con el comercio y la actividad agrícola. En China, en el año 2000 a. C., también se tiene evidencia de registros de datos relacionados con la riqueza. Asimismo, en Israel y Grecia, en el año 1000 y 540 a. C., respectivamente, se tiene evidencia histórica de diversos censos para fines de política pública tributaria, social, militar y de salud. Sin duda, se considera que existe más evidencia de la aplicabilidad de la estadística que aún no ha sido descubierta, pero con las diversas aportaciones en la era moderna, basta para concluir que la estadística ha estado en la vida del ser humano desde hace muchas generaciones con impactos diversos (Desrosières, 2015; Timeline of statistics, 2013).

En la actualidad, no es la excepción. Básicamente todas las actividades del ser humano están relacionadas con la estadística. Desde aspectos muy personales como las finanzas propias, hasta aspectos mundiales como el intercambio de materiales entre regiones del mundo. Además, prácticamente todas las empresas del sector primario (agricultura, ganadería, pesca, explotación de los recursos forestales y minería), secundario (transformación de los recursos naturales o las materias primas para convertirlos en productos destinados a cubrir necesidades de consumo) y terciario (generación de servicios) fundamentan, de alguna manera, su competitividad en la estadística. Por lo tanto, debido a amplitud de su aplicabilidad y los aspectos particulares de cada problemática presentada en el día a día, se han generado diversas ramas de la estadística que buscan de cierta manera especializar sus contribuciones (Gelman y Vehtari, 2021; Desrosières, 2013).

A grandes rasgos, la estadística se divide en dos ramas: la descriptiva y la inferencial. En particular, la estadística descriptiva busca facilitar la exposición, especificación, clasificación, visualización y presentación cuantitativa, ya sea numérica o gráfica de todo el cúmulo de datos obtenidos en un estudio dado. De hecho, su principal aportación se da cuando se manejan grandes cantidades de información. De aquí el hecho de las diversas opciones gráficas para representar datos, tales como tablas, histogramas, gráficas de barra, gráfica de sectores, tablas dimensionales y gráfico de cajas, entre otras (Surbhi, 2019; Stapor, 2020).

Por otra parte, la estadística inferencial busca, como su nombre lo indica, deducir, argumentar, concluir, derivar, entender, razonar y especular modelos matemáticos que permitan inferir o proyectar efectos futuros. De esta manera, podemos pensar que antes de la aplicación de la estadística inferencial, se debe apelar a la estadística descriptiva, de acuerdo con la problemática y población abordadas. Derivado de lo anterior, en la actualidad existen numerosas ramas de la estadística con aplicaciones específicas, y sin duda, conforme las problemáticas se tornen más complejas, nuevas ramas de la estadística surgirán, tanto de manera disciplinaria como multidisciplinaria. A continuación, se describirán de manera general algunas ramas de la estadística, solamente como muestra de la importancia aplicada en diversas áreas del conocimiento (Byrne, 2007; Cotton, 1977).

Una rama muy importante es la bioestadística. Como su nombre lo trasluce, su objeto de estudio es cualquier organismo biológico, ya sea del reino animal, vegetal, fúngico o incluso el ser humano. Debido a lo anterior, la bioestadística es clave para la investigación y desarrollo de diversas disciplinas, tales como la medicina, nutrición, industria farmacéutica y zoología, entre otras. Históricamente hablando, la bioestadística nació en el campo de la medicina, cuando Pierre Charles-Alexandre Loui empleó métodos matemáticos para cuantificar las variables de las enfermedades de sus pacientes. Después de ello, se estudió la epidemiología con métodos estadísticos generales, para después elaborar mapas epidemiológicos basados en análisis estadísticos. Además, la estadística ha contribuido exitosamente a la descripción e inferencias de enfermedades tan comunes como raras, como infecciones desconocidas, tabaquismo y cáncer de pulmón, entre otros (Torres y Normando, 2021; Lopes *et al.*, 2014).

La estadística también se aplica al deporte y es conocida como estadística deportiva, aunque es importante clarificar que no solamente se aplica en el contexto deportivo, sino también en la amplitud de la cultura física, con el objetivo de generar conocimiento que pueda ayudar a la persona que practica alguna actividad física, principalmente a los profesionales o que están en la preparación profesional, a mejorar el desempeño integral. De manera particular, dicha información cuantitativa permite entender mejor el desempeño del deportista según las diversas etapas de preparación, es decir, describir el grado de desarrollo con las herramientas estadísticas. Además, proporciona un marco de referencia frente a otros competidores y permite establecer pronósticos y metas de desempeño físico, entre otras ventajas competitivas. Como ocurre en otras ramas de la estadística aplicada, la estadística descriptiva e inferencial también se usan en el sector deportivo (Santos-Fernandez *et al.*, 2019; Secular, 2021).

En el ámbito económico, se le conoce como econometría a la aplicación de la estadística para estimar parámetros económicos a través de diversos métodos de regresión y analizar la relación entre variables macroeconómicas. Por ejemplo, a través de modelos econométricos se pueden pronosticar la tasa de crecimiento, el gasto público y su efecto, proyectar la demanda de recursos, entre otros aspectos. Para lograr lo anterior, la econometría dispone de la economía matemática, estadística económica y matemática, y

obviamente, de la teoría económica. De esta manera, la econometría pretende expresar las teorías económicas mediante una estructura o forma matemática, con el propósito de verificar empíricamente por medio de métodos estadísticos (Farebrother y Common, 1978; Wooldridge, 2003).

También la estadística se aplica al análisis de procesos, como los vinculados con la química, sustancias, materiales, productos, etc. La geografía también es apoyada por la estadística a través de los Sistemas de Información Geográfica (SIG), los cuales permiten procesar una gran cantidad de datos relacionados con el planeta y sus elementos, tales como el clima, relieves, entre otros. De hecho, gracias a la estadística se pueden establecer pronósticos y políticas de orden mundial. En este mismo sentido, la estadística apoya el análisis del universo a través del análisis estadístico espacial. En fin, se puede expandir muchísimo la aplicabilidad de la estadística en las ciencias y disciplinas, como la educación, la salud, la robótica y automatización, diseño de productos y servicios, mercadotecnia, comercio, sustentabilidad, medio ambiente, astronomía, agricultura, ganadería, política, entre otros. Finalmente, se puede concluir que la teoría estadística y su aplicación están presentes y juegan un rol sumamente importante en todas las disciplinas actuales del conocimiento. Realmente, sería muy complicado concebir los avances científicos y técnicos actuales sin relación con la estadística aplicada (Handbook of Computational Statistics, 2005; Astuti *et al.*, 2020).

Toma de decisiones basadas en la estadística

De manera informal y en ocasiones inconscientes, día a día tomamos decisiones personales basadas en la estadística, desde acudir a un restaurante en un día y horario determinados, hasta la mejor ruta de transporte. Lo anterior es basado en nuestra experiencia, opiniones, etc. Sin embargo, se formaliza conscientemente para problemas empresariales, industriales y sociales, ya que es parte del análisis de la competitividad sostenible de cualquier institución. De manera específica, al considerar cualquier disciplina y ciencia, la toma inteligente de decisiones es una actividad crucial para el éxito de cualquier proyecto. De hecho, la toma de decisiones implica, en principio, la existencia de diversas alternativas u opciones, las cuales deben

ser evaluadas científicamente para determinar la mejor alternativa que satisfaga una problemática particular. Así, la toma de decisiones es uno de los pasos más importantes de la planeación estratégica, ya que permite establecer las respuestas a las siguientes preguntas: ¿qué hacer?, ¿cuándo hacerlo?, ¿cómo hacerlo?, ¿dónde hacerlo?, ¿quién lo hará?, entre otras interrogantes que la estadística nos puede ayudar a responder. Además, es importante mencionar que la toma de decisiones basada en la estadística está estrechamente relacionada con la metodología de la investigación aplicada en escenarios particulares. En específico, una vez que se han establecido las hipótesis de una investigación, ya sean nulas o alternativas, es necesario tomar una decisión en relación con las hipótesis que serán aceptadas y rechazadas. Pero, ¿cómo se realiza esto? Con la ayuda de las herramientas y métodos estadísticos que permitan tomar una decisión objetiva, racional y estructurada, a lo cual se le denomina, toma de decisiones inteligentes. Solamente para clarificar, la estadística (y la probabilidad) en la toma de decisiones se debe de considerar como un indicador primordial, mas no como un factor crucial, ya que existen otros indicadores externos que también deben de considerarse para tomar la decisión correcta (Stefan, 2022; Hatton *et al.*, 2021; Bucknell *et al.*, 2011).

Aportes de los modelos estadísticos a las ciencias económico-administrativas

El objetivo de la estadística aplicada a la administración, los negocios, las finanzas, la economía, la contabilidad y disciplinas afines es proporcionar a los estudiantes de dichas áreas del conocimiento una visión introductoria de las aplicaciones de la estadística descriptiva e inferencial. Uno de los ejes rectores de este libro es que integra casos prácticos en cada una de las metodologías, para poner en la práctica lo que la teoría dicta. A su vez, servirá de guía para el desarrollo de investigaciones de la comunidad estudiantil de posgrado, a partir de que existen dentro de los proyectos de tesis, diversos estudios multidisciplinarios.

Como marco introductorio, podemos señalar que una primera ramificación que tiene la estadística es: descriptiva e inferencial. Ambas son útiles

para los estudios en las ciencias económicas y administrativas. Todos los puntos expuestos a continuación se desarrollan mediante fórmulas matemáticas y metodologías sofisticadas; en este capítulo 1 sólo vamos a explicar su concepto. La lista de elementos destacables en las disciplinas administrativas incluye:

- *La probabilidad.* Las posibilidades se definen como objetos matemáticos. En su concepción original, se aplica cuando todas las posibilidades son igualmente probables (por ejemplo, sacar una determinada carta de una baraja normal). Sin embargo, no suele suceder que todas las posibilidades sean igualmente probables. Para solventar este límite se aplica el estudio e interpretación de la frecuencia de un evento. Por ejemplo, la probabilidad de emprender o no un negocio con base en determinantes demográficos.
- *Los espacios muestrales* son la información que obtenemos de la observación y experimentación de la realidad por estudiar. Estos resultados pueden variar en dependencia del ámbito en que se realiza la observación o el experimento.
- *Distribuciones y densidades de probabilidad.* La experimentación de la realidad se produce al tomar datos de manera aleatoria que consideremos muestras representativas de la realidad por estudiar. En esta toma de datos nos interesan determinados aspectos del objeto de estudio y otros aspectos no. Es lo que llamamos “variables aleatorias”; por ejemplo, si tiramos dos dados nos puede interesar el resultado total de los dos y no como se comporta cada uno de ellos por separado, aunque el resultado total sea la suma de estos comportamientos separados y el cálculo se haga con base en los ocho lados de cada uno y cuál es la probabilidad de que salgan determinados números.
- *Funciones de variables aleatorias.* Aplica fórmulas matemáticas para deducir la probabilidad de una variable a partir de la información de un muestreo de variables aleatorias.
- *Muestreo.* El estudio sobre el que se aplican las fórmulas y análisis estadísticos se denomina muestra y se realiza de manera metódica y sistemática, para que sea representativo de una realidad a la que llamamos “población infinita” porque puede alcanzar cualquier límite.

El método para tomar estas muestras debe tener en cuenta sus características concretas para intentar que sean representativas del universo por estudiar; pero los resultados serán más fiables cuanto mayor sea el número de muestras estudiadas. Hay innumerables casos de estudio de mercado y demás análisis de encuestas que se apoyan en la teoría del muestreo.

- *Inferencia o estimación estadística.* Sirve para dar valor a un parámetro a partir de la información que nos proporciona la muestra. Esta estimación se puede hacer de manera *puntual*, al utilizar un solo criterio de muestra (por ejemplo, la estatura media medida con la estatura de los sujetos encuestados); o *por intervalos*, al analizar datos diferentes y deducir sus consecuencias para calcular el parámetro buscado.
- *Regresión y correlación.* Predecimos determinadas variables en función de otras variables. Por ejemplo, predecir el nivel de ventas de un producto según su precio, lo que va a gastar una familia en función de su nivel de ingresos o cómo influye la publicidad en las ventas de determinados productos.
- *Análisis de varianza.* Cuando se tienen muestras diferentes y se obtienen resultados diferentes o contradictorios, debemos analizar si esas diferencias se deben a alguna característica concreta de cada muestra tomada o si son totalmente aleatorias. Se estudia a la varianza como medida de dispersión clave para comprender el comportamiento de los datos, particularmente en series históricas.

En el entorno internacional actual de los negocios y de la economía, se cuenta con un inmediato y sencillo acceso a enormes cantidades de información estadística. Los directivos y los encargados de tomar decisiones que tienen éxito, valoran y comprenden la información y saben usarla de manera eficiente. En esta sección se proporcionan ejemplos ilustrativos de algunos de los usos de la estadística en los negocios y en la economía.

Algunas aplicaciones de las técnicas planteadas anteriormente se han desarrollado en distintas investigaciones. Fernández-Morales (2021) aborda la dirección y planificación del turismo, donde incluye contenidos variados que recorren temáticas que van desde las variables aleatorias y sus tipos hasta las relativas a la estimación y la verificación de hipótesis. Adicionalmente, presenta un caso de técnicas multivariantes.

Tapia Toral y Jijón (2018) abordan cómo la estadística descriptiva y la inferencial apoyan el análisis administrativo. En cuanto a lo descriptivo, la información recabada, ya sea de tipo cuantitativo o cualitativo, le permite al administrador presentar de manera ordenada y resumida las características de un fenómeno social o económico. Esto lo hace a través de gráficas y tablas para definir el comportamiento de una o más variables. Este tipo de análisis se centra en describir (de ahí el nombre de estadística descriptiva), con base en un índice, promedio, porcentajes, grados de dispersión y demás medidas, los cuales pueden describir datos como ventas, consumo, producto interno bruto, número de sucursales de una franquicia, tasas de mortalidad y tendencia de crecimiento de pobreza, entre otros.

Estos autores también exploran cómo la estadística inferencial mediante el análisis de una muestra, permite sacar conclusiones a partir de una población. Un ejemplo de lo anterior es, con una encuesta de salida, estimar quién resulta el candidato ganador de una elección, con sólo analizar un número determinado de electores (una muestra de esa población).

Lo anterior puede reforzarse cuando nos adentramos a un tema de la administración como la producción. La importancia dada actualmente a la calidad hace del control de ésta una aplicación importante de la estadística para la producción. Para vigilar el resultado de los procesos productivos se usan diversas gráficas de control estadístico de calidad. En particular, para vigilar los resultados promedio se emplea una gráfica x-barra.

Otro campo dentro del área económico-administrativa que emplea comúnmente la estadística para su análisis son las finanzas. Fernández-Morales (2016) aplica modelos de probabilidad de uso frecuente en finanzas, problemas de estimación y verificación de hipótesis y procesos estocásticos. Partiendo de estos tres grandes segmentos de bloques, usan diversas aplicaciones que sirven de complemento a tipos de interés, tipos de cambio, índices bursátiles, riesgos y demás tópicos financieros.

Cada vez que realizamos alguna compra en un supermercado, la tecnología actual permite que con los escáneres electrónicos en las cajas de los comercios minoristas se vayan recolectando datos para diversas aplicaciones en la investigación de mercado. Luego entonces, agencias de investigación en esta área compran esos datos a los negocios, los procesan y después venden los resúmenes estadísticos a los fabricantes, quienes gastan cientos de miles

de dólares por producto para obtener este tipo de datos y, de esa manera, hacer estrategias más precisas sobre las necesidades y demandas del mercado (Anderson, 2008).

Los economistas suelen hacer proyecciones acerca del futuro de la economía. Para ello se apoyan en una variedad de información estadística. Por ejemplo, para pronosticar las tasas de inflación, emplean información estadística sobre indicadores como el índice de precios al consumidor, la tasa de desempleo y la utilización de la capacidad de producción. Estos indicadores estadísticos se utilizan en modelos computarizados de pronóstico que predicen las tasas de inflación, tipos de cambio o el producto interno bruto (Anderson, 2008). Aunado a ello, hay una rama de la economía llamada econometría, que se encarga de la medición de variables y fenómenos económicos. En ella se destacan las técnicas estadísticas de correlación y regresión lineal.

Principales técnicas estadísticas aplicadas a la investigación económico-administrativa

Para llevar a cabo la investigación en el campo de las ciencias ya mencionadas, es útil apoyarse en técnicas validadas que comprueben las hipótesis planteadas en diversos estudios. Por ejemplo, ¿cómo analizar si la escolaridad influye en la tenencia de tarjetas de crédito? O, ¿cómo evaluar si los apoyos económicos a madres solteras emprendedoras impactan favorablemente en el negocio?

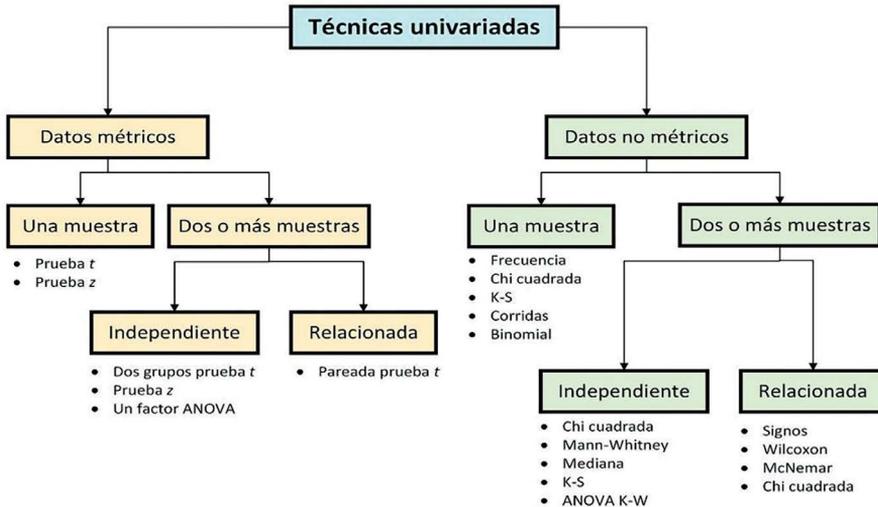
Para ordenar las principales técnicas aplicadas en la investigación de los negocios, se categorizan a continuación los dos principales grupos, que son las técnicas univariadas y las multivariadas.

Técnicas univariadas

Se conoce como técnicas univariadas aquellas que son empleadas cuando hay una sola medición para cada elemento de la muestra, o bien cuando existen varias mediciones para cada elemento; pero cada variable se analiza

por separado. El siguiente diagrama (Figura 1) muestra las diversas técnicas univariadas en el estudio estadístico.

Figura 1. Conjunto de técnicas univariadas



Fuente: Adaptado de Malhotra (2008, p. 442).

De acuerdo con el diagrama anterior, podemos destacar que la clasificación del estudio univariado se divide en datos métricos y datos no métricos. El primero de ellos se refiere a un dato cuantitativo y se denomina de escala métrica o simplemente cuantitativo métrico cuando permite realizar cualquier tipo de operación matemática. Además, se pueden cuantificar las distancias entre ellos. Algunos ejemplos de datos métricos son temperatura en grados centígrados, resultados obtenidos en un examen, distancia en kilómetros entre dos ciudades y peso de un bebé recién nacido, entre otros.

Por otro lado, los métodos paramétricos y no métricos se utilizan a menudo en diferentes tipos de datos. Las estadísticas paramétricas generalmente requieren datos de intervalo o razón. Ejemplos de este tipo de datos son la edad, ingresos, altura y peso, donde los valores son continuos, constantes, y los intervalos entre valores son significativos.

Por el contrario, las estadísticas asimétricas se utilizan normalmente en datos nominales u ordenados. Las variables se denominan nominales cuando sus valores no tienen un valor numérico. La variable nominal común

en la investigación de las ciencias sociales y también en las administrativas es el sexo, cuyos valores potenciales son categorías discretas-binarias como “masculino” y “femenino”. Otras variables nominales comunes en este tipo de investigaciones son la raza, la situación conyugal, el nivel educativo y la situación laboral.

De la Figura 1 podemos resaltar una de las técnicas más utilizadas para resumen y presentación de datos: las tablas de frecuencia. Esto nos sitúa en datos no métricos y para una muestra. Pensemos en el siguiente ejemplo: contamos con datos de las edades de los clientes de una cafetería en el último mes y la intención es tomar esa información para la adopción de decisiones y estrategias en búsqueda de un mejor manejo de esa cafetería. Los datos de este caso se muestran a continuación en una tabla de frecuencia.

Tabla 1. *Tabla de frecuencias*

Grupos de edad	Frecuencia	Frecuencia relativa
15-18 años	150	26.09%
19-22 años	253	44.00%
23-26 años	98	17.04%
27 o más años	74	12.87%
Σ	575	100%

Fuente: Elaboración propia.

Con el uso de tablas de frecuencia, se pueden organizar e interpretar de manera más eficiente una serie de datos no métricos para una muestra. Es decir, es más presentable y resumida la información descriptiva de la edad de 575 clientes en una tabla como la anterior que en un listado carente de sus frecuencias absolutas y relativas.

Otra metodología comúnmente utilizada en las técnicas univariadas para una muestra es la prueba Z. En este test se asume que hay una distribución normal estándar bajo la hipótesis nula. Ésta evalúa la media de una población normalmente distribuida con varianza conocida. Por ejemplo, el supervisor de una fábrica de dulces desea saber si el peso medio de un lote de cajas de caramelos es igual al valor objetivo de 12 onzas. Partiendo de datos históricos recolectados por la compañía, el supervisor sabe que la máquina de llenado tiene una desviación estándar de 0.6 onzas, así que utiliza este valor como la desviación estándar de la población en una prueba Z de una muestra.

Una prueba similar a la descrita en el párrafo anterior es la prueba t , la que tiene como propósito identificar si la media de una población objetivo es igual a un valor objetivo. Un ejemplo de cuándo aplicar dicha prueba puede ser para responder una cuestión como la siguiente: ¿la estatura promedio de las universitarias de UABC es mayor a 1.68 centímetros?

Una propiedad importante de la prueba t es su robustez ante los supuestos de normalidad de los datos analizados pertenecientes a una población. Dicho de otro modo, cuando contamos con muestras grandes, estas pruebas suelen ser válidas incluso cuando se viole el supuesto de normalidad. Esta particularidad la convierte en uno de los procedimientos más útiles para obtener inferencias sobre las medias de las poblaciones. No obstante, con un tamaño de muestra pequeño y distribuciones no normales y muy asimétricas (revisar pruebas como la asimetría y curtosis), la sugerencia es recurrir a pruebas no paramétricas.

Una técnica sobresaliente en el grupo de las univariadas para dos o más muestras es el análisis de la varianza (Anova) para un factor. Es una fórmula estadística utilizada para comparar las varianzas entre las medias (o el promedio) de diferentes grupos. En una variedad de contextos la utilizan para determinar si existe alguna diferencia entre las medias de los distintos grupos.

Por ejemplo, para estudiar la efectividad de diferentes medicamentos para la diabetes, los científicos diseñan y experimentan para explorar la relación entre el tipo de medicamento y el nivel de azúcar sanguínea resultante. Dividimos la población de la muestra (un conjunto de personas) en varios grupos y cada grupo recibió un medicamento en particular durante el periodo de prueba. Al final del periodo de prueba, se miden los niveles de azúcar sanguínea para cada uno de los participantes individuales. Luego, para cada grupo se calcula el nivel medio de azúcar sanguínea. Por último, Anova ayuda a comparar las medias de estos grupos para averiguar si son estadísticamente diferentes o si son similares.

El resultado de Anova es el estadístico F . Este valor nos indica la diferencia entre la varianza dentro del grupo y la varianza entre grupos, lo que finalmente arroja una cifra que permite concluir que la hipótesis nula sea aceptada o rechazada. Si existe una diferencia significativa entre los grupos, la hipótesis nula no es compatible y la razón F será mayor.

Aunado a lo anterior, la prueba de Anova en el análisis de los datos de una determinada encuesta requiere que se cumplan algunos supuestos, como la distribución normal de la información, la independencia de los casos y la igualdad de varianzas.

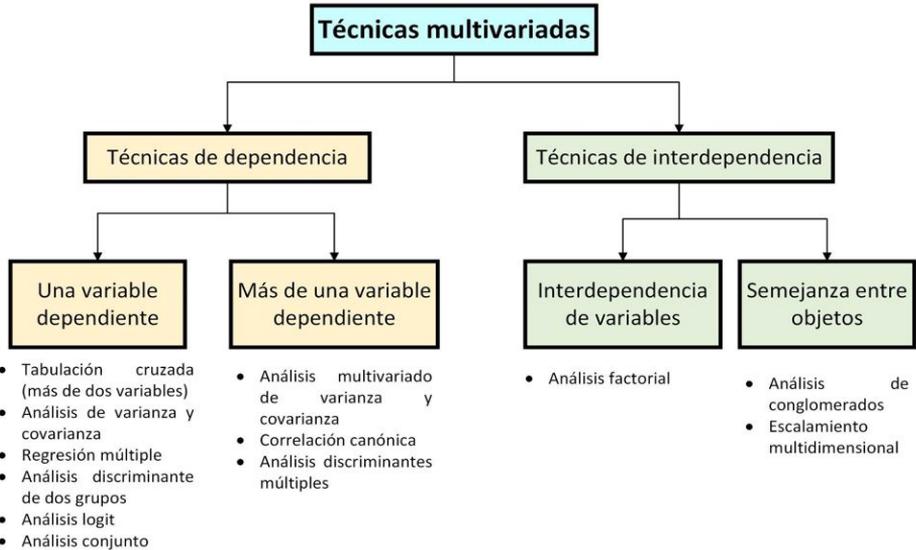
Por último, dentro del grupo de técnicas univariadas exploramos el rubro de los datos no métricos; para ello se abordará la prueba de ji al cuadrado. Es sin duda la más conocida y probablemente la más utilizada para el análisis de variables cualitativas. Su nombre lo toma de la distribución ji al cuadrado de la probabilidad, en la que se basa. La prueba de ji al cuadrado permite evaluar la independencia entre dos variables nominales u ordinales y ofrece un método para verificar si las frecuencias observadas en cada categoría son compatibles con la independencia entre ambas variables. Para evaluarla se calculan los valores que indicarían la independencia absoluta, lo que se denomina frecuencias esperadas, al compararlas con las frecuencias de la muestra. Habitualmente, H_0 indica que ambas variables son independientes, mientras que H_1 indica que las variables tienen algún grado de asociación.

La prueba de ji al cuadrado utiliza una aproximación a la distribución ji al cuadrado, para evaluar la probabilidad de una discrepancia igual o mayor a la que exista entre los datos y las frecuencias esperadas según la hipótesis nula. La exactitud de esta evaluación depende de que los valores esperados no sean muy pequeños y, en menor medida, de que el contraste entre ellos no sea muy alto.

Técnicas multivariadas

Las técnicas multivariadas se utilizan cuando se requieren dos o más mediciones para cada elemento y las variables se analizan simultáneamente. Este tipo de técnicas se clasifica en relaciones de dependencia y de interdependencia (Figura 2).

Figura 2. Conjunto de técnicas multivariadas



Fuente: Adaptado de Malhotra (2008, p. 443).

Una técnica comúnmente utilizada en las relaciones de dependencia es la tabulación cruzada. Se conoce como tabla cruzada o tabla de contingencia al tabulado, donde se analiza en conjunto a dos variables (pueden ser más, pero típicamente se emplean dos). Un ejemplo de lo anterior sería estudiar el plan de suscripción a una plataforma de *streaming* y sexo. La tabla se muestra a continuación:

Tabla 2. Tabla de contingencia

Plan de suscripción	Género		Total
	Femenino	Masculino	
Básico	80	145	225
Estándar	130	115	245
Premium	55	70	125
Total	265	330	595

Fuente: Elaboración propia.

Se puede observar en la tabla previa que un total de 595 respuestas fueron analizadas, divididas en 265 del género femenino y 330 pertenecientes al masculino. Luego entonces, estos datos se cruzan para identificar cómo se comporta la variable Plan de suscripción, analizada al interior de los

datos del género. Por ejemplo, se puede apreciar que del total de 225 suscritos al plan básico, la mayoría (145) son hombres, no así con el plan estándar, donde 130 mujeres encuestadas superan a los 115 hombres con su mismo tipo de suscripción.

Otro instrumento válido para identificar las relaciones entre variables es la regresión lineal múltiple. Esta metodología ayuda a identificar cómo una o algunas variables independientes impactan o tienen efecto explicativo en una variable dependiente. Es decir, nos revela cómo Y cambia a medida que X lo hace en una cantidad constante, todo ello mediante una recta de ajuste.

El análisis de regresión lineal múltiple es altamente utilizado para poder identificar cómo una variable responde ante cambios en otras variables. En las disciplinas económicas y administrativas, esto no es la excepción: considere el caso de querer saber qué determinantes (X) explican el comportamiento de las ventas de un determinado negocio (Y). Esta metodología se explicará más detalladamente en otro capítulo.

Dentro de las técnicas utilizadas cuando hay relaciones de interdependencia, tenemos el análisis factorial. Este método también conocido como análisis de componentes principales (ACP) tiene como finalidad explicar las posibles correlaciones entre algunas variables. Para ello, se considera el efecto de otros llamados factores, que no son observables. Por tanto, lo que hace este análisis es reducir. Así, tomamos un número considerado de variables y , por medio de esta técnica, conseguimos reducirlas a un tamaño más fácil de manejar. Con ese fin, se utiliza una serie de combinaciones lineales de las observadas con otras que no son visibles.

Un caso de la aplicación de este método en el marketing es cuando se quiere conocer la intención de compra. Para ello, analizamos diversas variables socioeconómicas, emocionales o personales. Una vez que se cuenta con ella, se reduce su número con el análisis factorial y podremos interpretarla mejor.

Dicho lo anterior, a diferencia de lo que ocurre con otras técnicas como el análisis de varianza o el de regresión, en el análisis factorial todas las variables cumplen el mismo papel: todas ellas son independientes en el sentido de que no existe *a priori* una dependencia conceptual de unas variables sobre otras. Fundamentalmente, con el análisis factorial se pretende

simplificar la información que nos da una matriz de correlaciones para hacerla más fácil de interpretar. Se pretende encontrar una respuesta a preguntarnos: ¿Por qué unas variables se relacionan más entre sí y menos con otras? Hipotéticamente es porque existen otras variables, dimensiones o factores que explican por qué unos ítems se relacionan más con unos que con otros.

Referencias

- Anderson, D., Sweeney, D., y Williams, T. (2008). *Estadística para administración y economía*. Cengage Learning.
- Astuti, D., Prabowo, A., Hidayati, N. A., y Khasanah, U. (2020). Educational statistics textbooks to develop collaborative skills and critical thinking. *Journal of Physics: Conference Series*, 1613. <https://doi.org/10.1088/1742-6596/1613/1/012045>
- Bucknell, T., Taylor-Roe, J., Killick, S., y Thompson, S. (2011). Usage statistics for decision making. *Serials*, 24(1), 17–20. <https://doi.org/10.1629/2417>
- Byrne, G. (2007). A statistical primer: Understanding descriptive and inferential statistics. *Evidence Based Library and Information Practice*, 2(1), 32-47. <https://doi.org/10.18438/b8fw2h>
- Cotton, J. (1977). Review of descriptive and inferential statistics. *Contemporary Psychology: A Journal of Reviews*, 22(9), 725–725. <https://doi.org/10.1037/016314>
- Desrosières, A. (2013). The history of statistics as a genre: Styles of writing and social uses. *BMS Bulletin of Sociological Methodology/ Bulletin de Méthodologie Sociologique*, 119(1), 8–23. <https://doi.org/10.1177/0759106313486398>
- Desrosières, A. (2015). History of statistics. En *International encyclopedia of the social & behavioral sciences* (2ª ed., pp. 423–427). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.03050-6>
- Farebrother, R. W., y Common, M. S. (1978). Basic econometrics. *Journal of the Royal Statistical Society, Series A (General)*, 141(3), 417-418. <https://doi.org/10.2307/2344828>
- Fernández-Morales, A. (2021). *Técnicas estadísticas para dirección y planificación del turismo: Cuestiones y aplicaciones*. Departamento de Economía Aplicada, Universidad de Málaga, 1-17. <https://riuma.uma.es/xmlui/bitstream/handle/10630/23431/TEDPTCA.pdf?sequence=1>
- Fernández-Morales, A., Cisneros-Martínez, J. D., y Scott McCabe. (2016). Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. *Tourism Management*, 56, 172-190. <https://doi.org/10.1016/j.tourman.2016.04.004>
- Freedman, D. (1999). From association to causation: Some remarks on the history of statistics. *Statistical Science*, 14(3), 243–258. <https://doi.org/10.1214/ss/1009212409>

- Gelman, A., y Vehtari, A. (2021). What are the most important statistical ideas of the past 50 years? *Journal of the American Statistical Association*, 116(536), 2087-2097. <https://doi.org/10.1080/01621459.2021.1938081>
- Handbook of Computational Statistics [Reseña] (2005). *Technometrics*, 47(3), 383–384. <https://doi.org/10.1198/tech.2005.s314>
- Hatton, G. E., Pedroza, C., y Kao, L. S. (2021). Bayesian statistics for surgical decision making. *Surgical Infections*, 22(6), 620-625. <https://doi.org/10.1089/sur.2020.391>
- Lopes, B., Ramos, I. C., Ribeiro, G., Correa, R., Valbon, B., Luz, A., ..., Junior, R. A. (2014). Biostatistics: Fundamental concepts and practical applications. *Revista Brasileira de Oftalmologia*, 73(1), 16–22. <https://doi.org/10.5935/0034-7280.20140004>
- Malhotra, N. K. (2008). *Investigación de mercados* (5ª ed.). Pearson.
- Santos-Fernandez, E., Wu, P., y Mengersen, K. L. (2019). Bayesian statistics meets sports: A comprehensive review. *Journal of Quantitative Analysis in Sports*, 15(4), 289-312. <https://doi.org/10.1515/jqas-2018-0106>
- Secular, S. (2021). The numbers game: The NBA v. Motorola, real-time statistics, and the rise of online fantasy sport. *International Journal of the History of Sport*, 38(1), 79–94. <https://doi.org/10.1080/09523367.2021.1876670>
- Stapor, K. (2020). Descriptive and inferential statistics. En *Intelligent systems reference library* (vol. 176, pp. 63–131). Springer. https://doi.org/10.1007/978-3-030-45799-0_2
- Stefan, A. M. (2022). Statistics for making decisions. *The American Statistician*, 76(1), 87–88. <https://doi.org/10.1080/00031305.2021.2020003>
- Surbhi, S. (2019). Difference between descriptive and inferential statistics. Regression Analysis: An Intuitive Guide, 1–31. <https://statisticsbyjim.com/regression/predictions-regression/>
- Tapia Toral, M. y Jijón Gordillo, E. (2018). *Estadística aplicada a la administración y la economía*. Centro de Investigación y Desarrollo, Universidad de Guayaquil.
- Timeline of statistics. (2013). *Significance*, 10(6), 23–26. <https://doi.org/10.1111/j.1740-9713.2013.00707.x>
- Torres, D. y Normando, D. (2021). Biostatistics: Essential concepts for the clinician. *Dental Press Journal of Orthodontics*, 26(1). <https://doi.org/10.1590/2177-6709.26.1.e21spe1>
- Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2ª ed.). <https://doi.org/10.1198/jasa.2006.s154>

II. Estadística descriptiva y modelos de tabulación cruzada

EDGAR JIMÉNEZ CERRA*
MALENA PORTAL BOZA**

DOI: <https://doi.org/10.52501/cc.131.02>

Resumen

Este capítulo ofrece una guía para los interesados en comprender la estadística descriptiva y el uso de los elementos que la componen. La estructura del texto inicia con el abordaje conceptual desde el punto de vista de diferentes autores y la explicación a profundidad de temas como las medidas de tendencia central, de dispersión, los gráficos y los modelos de tabulación cruzada. En un segundo momento se presenta un ejemplo práctico que engloba los diferentes aspectos descritos en la primera parte, con la finalidad de que el lector pueda replicarlo y, de esta forma, desarrollar este apartado tan importante en el ámbito de la investigación cuantitativa de una forma más sencilla y correcta.

Palabras clave: *estadística descriptiva, medidas, dispersión, tendencia central, tabulación cruzada.*

La humanidad se ha caracterizado siempre por la búsqueda de respuestas a los fenómenos que se suceden en su entorno, ya sean de índole social, natural, económica, entre otras, con la finalidad de entenderlos mejor y con

* Doctor en Ciencias Administrativas. Profesor de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-1285-3345>

** Doctora en Ciencias Económicas. Profesora de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0002-4237-1534>

ello mejorar su desempeño e incrementar su calidad de vida. La estadística ha sido uno de los elementos fundamentales en este proceso, al agrupar los métodos que permiten recopilar, manejar, describir y analizar la información de forma que los resultados obtenidos permitan tomar decisiones o arribar a conclusiones con rigor científico.

Cualquier investigación requiere, como parte inicial de la presentación de sus resultados, un apartado donde se describan los datos incluidos en el estudio y sus principales características. Para esto existe la rama de la estadística conocida como estadística descriptiva, enfocada en presentar al lector la información recopilada de forma tabular o gráfica, con la finalidad de resumir el comportamiento de la muestra seleccionada de una población determinada.

Este capítulo pretende ofrecer una guía que permita a los interesados comprender la estadística descriptiva y el uso de los elementos que la componen. La estructura del texto inicia con el abordaje conceptual desde el punto de vista de diferentes autores y culmina con un ejemplo práctico que engloba los distintos aspectos descritos en la primera parte.

Medidas de tendencia central de los datos

Al realizar el análisis del objeto de estudio de una investigación, resulta de mucha utilidad resumir el comportamiento de los datos mediante el empleo de cálculos matemáticos simples. La mayor parte de los valores que componen el conjunto de datos recopilado muestra una tendencia a conglomerarse en torno a un valor central. A éstos se les conoce dentro de la literatura académica como medidas de tendencia central. De ellas las más utilizadas son la media aritmética, la mediana y la moda (Viedma, 2018).

Es importante conocer qué estadísticos descriptivos aplicar de acuerdo con las características del conjunto de datos disponibles, ya que una selección errónea puede conducir a la obtención de resultados equivocados. Para los estudios de corte cualitativo es recomendable el empleo de herramientas estadísticas como la tabla de frecuencias y los diagramas de barra o circulares. En cambio, para los cuantitativos se pueden utilizar los antes mencionados, acompañados del análisis de medidas como la moda, me-

diana, media y el rango intercuartil, entre otros. Es relevante tomar en consideración la cantidad de valores diferentes en nuestro grupo de datos, pues esto pudiera incidir en la precisión de estas medidas al no resumir de forma suficiente el conjunto.

El primero de los estadísticos descriptivos en ser analizado es la media aritmética o el promedio. Ésta es considerada la medida de tendencia central más representativa en la estadística descriptiva y, a su vez, la más utilizada. Su valor se calcula con la suma de la totalidad de los datos del conjunto, dividida entre la cantidad de elementos que lo componen, como queda reflejado en la ecuación 1.

$$\sigma = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

Donde:

σ = la media aritmética

X_i = el valor para la variable en el caso i -ésimo

n = la cantidad total de datos

Esta medida de tendencia central debe utilizarse cuando el conjunto de datos presenta una distribución normal, en la que la cantidad de valores atípicos es baja. De lo contrario, se recomienda utilizar la mediana, valor central que al ordenar los elementos incluidos en la muestra de forma ascendente deja por debajo del mismo el 50% de los casos, lo que indica que el otro 50% quedará por encima. Es una medida central que se puede utilizar siempre, aunque se recomienda su uso cuando las variables incluidas en el estudio son de tipo intervalo o razón. Para el cálculo de la mediana se debe proceder según los siguientes pasos:

1. Ordenar la tabla de frecuencias de menor a mayor
2. Realizar el cálculo de las frecuencias absolutas acumuladas
3. Definir el intervalo crítico (IC) al dividir en dos la cantidad de elementos de la muestra
4. Realizar el cálculo según la ecuación 2:

$$M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} * a_i \quad (2)$$

$$a_i = L_{i+1} - L_{i-1} \quad (3)$$

Donde:

M_e = la mediana

L_{i-1} = el límite inferior del IC

L_{i+1} = el límite superior del IC

N = total de elementos de la muestra

N_{i-1} = el total de elementos por debajo del valor del IC

a_i = la amplitud del IC

n_i = la frecuencia del IC

A partir de esta fórmula se puede explicar lo abordado con anterioridad acerca de que la mediana es más utilizada cuando existen múltiples valores atípicos en la muestra, ya que para su cálculo sólo se consideran valores centrales de la distribución y las frecuencias de estos valores extremos.

Por su parte, la medida de tendencia central que se calcula de forma más fácil es la que permite definir cuál es el elemento de la muestra que presenta una frecuencia más elevada, o sea, que más se repite dentro del conjunto, la moda. Ésta es la más empleada en las investigaciones de corte cualitativo y normalmente se utiliza como complemento a las medidas anteriores. Para determinarla, en el caso de las variables categóricas sólo se deben calcular las frecuencias de cada elemento de la muestra y observar cuál es el valor mayor dentro del conjunto resultante.

Sin embargo, si la muestra está dividida en clases o subgrupos, esto influye en la forma de calcular el valor de la moda. Se determina la clase o clases con mayor frecuencia n_i y se procede a realizar la operación definida en la ecuación 4:

$$M_o = l_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) * C \quad (4)$$

Donde:

l_i = el límite inferior de la clase con mayor frecuencia absoluta

Δ_1 = la diferencia entre la mayor frecuencia absoluta y la anterior

Δ_2 = la diferencia entre la mayor frecuencia absoluta y la siguiente

C = la amplitud de la clase con mayor frecuencia absoluta

Esta medida de tendencia central es utilizada principalmente para estudios donde se busca definir las cualidades nominales que identifican a un grupo de individuos o subgrupos incluidos en una población.

Medidas de dispersión

Hasta este momento se ha mostrado un grupo de estadísticos de tendencia central que permiten al investigador valorar el conjunto de datos incluido en la muestra desde posiciones de concentración. Sin embargo, es necesario conocer el nivel de variabilidad de los valores alrededor de estas medidas de tendencia central y ampliar el contexto del investigador con una visión mayor del comportamiento de las variables, más cercana a la realidad del fenómeno analizado (Cobb y Moore, 1997). Con este fin, son utilizadas las medidas de dispersión; las más comunes son el rango y la desviación estándar.

El rango es la más sencilla de todas, al solo brindar información sobre la diferencia o distancia existente entre los valores máximo y mínimo de la muestra. Esta medida se recomienda emplearla como complemento de otras, de manera que permita demostrar la dispersión de los datos. La forma de calcularla es simple y queda reflejada en la ecuación 5.

$$\text{Rango} = \text{Valor M\u00e1ximo} - \text{Valor M\u00ednimo} \quad (5)$$

Por su parte, la desviaci\u00f3n est\u00e1ndar s\u00ed permite obtener una mayor claridad en cuanto a la variabilidad de los datos, ya que indica la distribuci\u00f3n de los valores con respecto a la media aritm\u00e9tica. El valor de este estad\u00edstico se calcula al aplicar la ra\u00edz cuadrada a la varianza, y su f\u00f3rmula depender\u00e1 de si se aplica a toda una poblaci\u00f3n (ecuaci\u00f3n 6) o a una muestra espec\u00edfica (ecuaci\u00f3n 7).

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \quad (6)$$

$$s = \sqrt{\frac{\sum (x - \mu)^2}{n-1}} \quad (7)$$

Donde:

σ = la desviación estándar para una población

s = desviación estándar para una muestra

x_i = cada dato de la variable

μ = la media aritmética

N = la cantidad de datos de la población

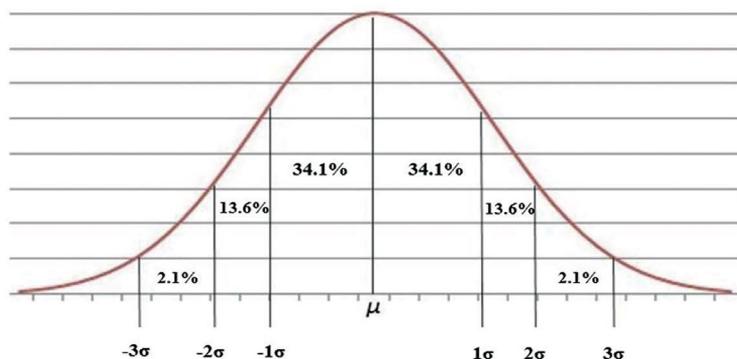
n = la cantidad de datos de la muestra

A este tipo de cálculo no recomiendan hacerlo de forma manual por su complejidad y es mejor realizarlo mediante el empleo de los diferentes *softwares* para el análisis estadístico. Valores elevados de desviación estándar representan una alta variabilidad de los datos incluidos en la muestra y, por tanto, una baja concentración en torno a su valor promedio. Por el contrario, un valor bajo mostraría una mayor cercanía a la media aritmética de los elementos del conjunto. Algunas de las aplicaciones de este estadístico son la comparación de datos de poblaciones o muestras diferentes, la medición del riesgo en el ámbito inversionista y el análisis en un conjunto de las tendencias de las variables a convergir o dispersarse en el tiempo.

De igual forma, mediante el empleo de algún *software* de análisis estadístico puede realizarse la representación gráfica de la distribución de los datos alrededor de la media aritmética μ (Figura 1).

Según Triola (2000), la imagen obtenida se asemeja a una campana o es simétrica con respecto a la media aritmética, como la observada en la Figura 1. Se puede afirmar que el 68.2% de los datos de la muestra se concentra en el intervalo entre $(\mu - \sigma)$ y $(\mu + \sigma)$, el 95.4% entre $(\mu - 2\sigma)$ y $(\mu + 2\sigma)$, y el 99.6% entre $(\mu - 3\sigma)$ y $(\mu + 3\sigma)$.

Figura 1. Representación gráfica de la desviación estándar de una población



Fuente: Elaboración propia, con base en Triola (2000).

Otra de las formas de analizar el comportamiento de los datos recopilados como parte de la investigación es mediante el empleo de las herramientas de representación gráfica. Éstas permiten arribar a conclusiones al investigador de una manera más ágil y sencilla, y contribuyen notablemente a la divulgación de los resultados.

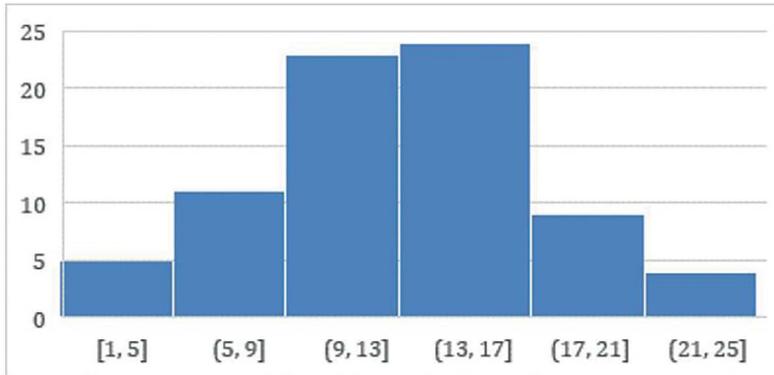
Herramientas gráficas

Los gráficos o diagramas son considerados resúmenes del comportamiento de las variables o sus posibles interacciones entre las incluidas en una investigación. Estas herramientas son altamente recomendadas para utilizar en el ámbito de la estadística descriptiva, ya que los elementos visuales que las componen facilitan en gran medida la comprensión del fenómeno estudiado (Minnaard *et al.*, 2002). Además, la representación gráfica no sólo ilustra las tendencias, sino que permite, mediante la interpretación de la ubicación de cada dato en el sistema de coordenadas, arribar a conclusiones más profundas. Entre los gráficos más utilizados para la descripción del conjunto de datos se encuentran los diagramas de barra, de caja y los histogramas.

Histograma

Los histogramas de frecuencias son gráficos empleados para representar intervalos de variables cuantitativas continuas. Ello se realiza mediante el uso de rectángulos o barras cuya altura corresponde a la frecuencia absoluta o relativa de los datos incluidos en la muestra por el eje de las ordenadas y a los valores de la variable analizada por el eje de las abscisas. De forma específica, el eje horizontal se extenderá hasta el valor máximo de la variable, y cada barra representa la segmentación de ese valor de acuerdo con las frecuencias que aparecen representadas en el eje vertical (Figura 2).

Figura 2. Ejemplo de histograma



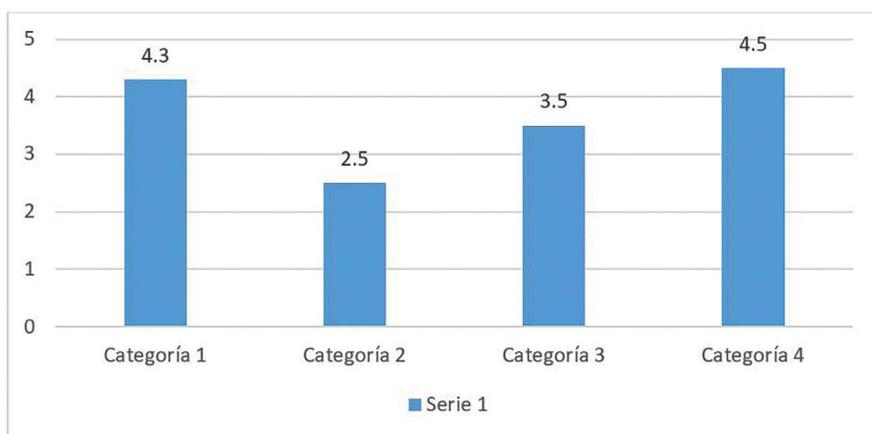
Fuente: Elaboración propia.

Esta herramienta permite desde un enfoque visual comprobar el centro, la normalidad y la amplitud de un conjunto de datos, así como la existencia de valores atípicos en el mismo. Esto influye directamente en las herramientas estadísticas que se utilizarán para realizar análisis más agudos. Es importante aclarar que debe utilizarse cuando la variable por graficar es de tipo cuantitativa y continua, como la edad, el nivel de ingresos y otros. En el caso de querer graficar variables cualitativas con valores discretos como el color de la piel, sabor de un producto, la carrera estudiada o la profesión ejercida, se deberán emplear otras herramientas como el diagrama de barras o el gráfico circular.

Diagrama de barras

Este tipo de gráfica es de los más empleados y se utiliza para representar visualmente las principales características de un conjunto de datos asociados a variables discretas, ya sean cualitativas o cuantitativas. A diferencia del histograma, el diagrama de barras se puede construir mediante barras verticales u horizontales para representar las frecuencias relativas o absolutas correspondientes a las categorías de la variable o variables en cuestión (Figura 3).

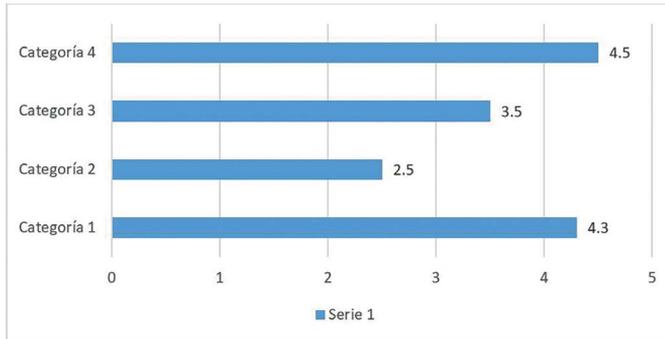
Figura 3. Diagrama de barras



Fuente: Elaboración propia.

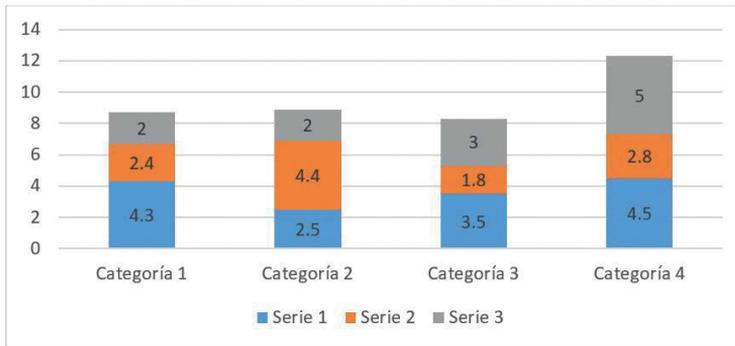
Otro elemento característico de esta herramienta es que normalmente las barras están separadas, aunque varios *softwares* permiten la edición de este tipo de gráficos y podrían presentarse casos donde no tengan ese detalle. De igual forma, pueden utilizarse líneas de tendencia para mostrar el comportamiento de las diferentes categorías en el tiempo. Dentro de los diagramas de barras los tipos más utilizados son los horizontales (*vid.* Figura 4), que se emplean cuando la extensión del nombre de cada categoría es mayor; el de barras proporcionales, que busca resaltar porcentajes que componen un total en cada categoría (*vid.* Figura 5); y el de barras comparativas (*vid.* Figura 6), que permite la comparación entre valores de una misma variable en las diferentes categorías.

Figura 4. Diagrama de barras horizontales



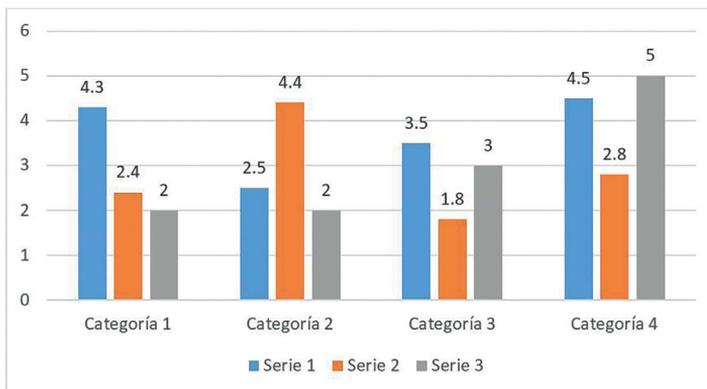
Fuente: Elaboración propia.

Figura 5. Diagrama de barras proporcionales



Fuente: Elaboración propia.

Figura 6. Diagrama de barras comparativas

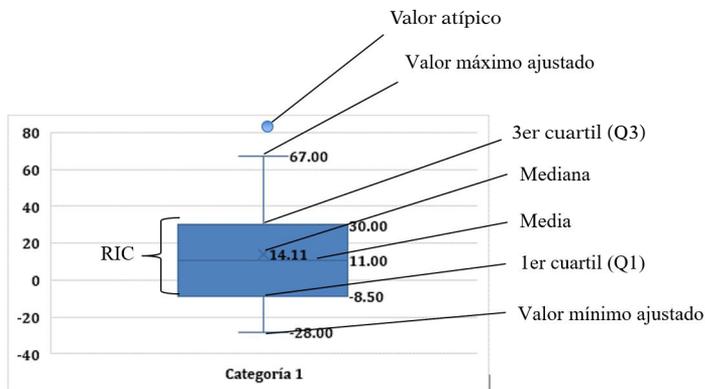


Fuente: Elaboración propia.

Diagrama de caja y bigotes

Otra de las herramientas empleadas en la estadística descriptiva son los diagramas de caja y bigotes (Tukey, 1977). Éstos permiten interpretar de una manera simple la información contenida en la muestra de una investigación. Para la conformación de este tipo de gráficos, se utilizan varias medidas que describen el conjunto de datos, la media y la mediana, ya vistas en este capítulo; los cuartiles, valores que dividen a un grupo ordenado de datos en cuatro partes iguales; y el rango intercuartílico (RIC), que en el gráfico es representado por la caja y agrupa el 50% de los valores incluidos en la muestra (Figura 7).

Figura 7. Diagrama de caja



Fuente: Elaboración propia, con base en Tukey (1977).

Al analizar el RIC, se debe tomar en cuenta que este elemento muestra la variabilidad existente en el conjunto de valores incluidos en la muestra, por lo que sus dimensiones están directamente relacionadas con la dispersión de los valores de la muestra. Así, mientras más pequeño sea, entonces existirá una mayor concentración de los valores incluidos en el estudio alrededor del valor central, lo que indica que siguen una distribución normal. Otra consideración importante sobre este componente es la ubicación del valor de la mediana. Si ésta se sitúa en el centro, se puede argumentar que la muestra es simétrica y los valores de las diferentes medidas de tendencia central coinciden. En cambio, si la mediana divide la caja en dos partes de

dimensiones diferentes, se argumenta que la muestra es asimétrica. Esta asimetría puede ser positiva/segada a la derecha, en caso de que la parte de mayor dimensión sea la superior a la mediana, lo que indica una mayor concentración de los valores en la parte inferior, o negativa/segada a la izquierda en el caso opuesto.

Por su parte, los bigotes, líneas que se prolongan desde la caja, indican la varianza esperada de los datos incluidos en la muestra y su tamaño no puede exceder en 1.5 veces el del RIC, por lo que sus extremos coinciden con los valores mínimo y máximo del conjunto si éstos no exceden ese valor. Otros elementos que deben analizarse en el momento de utilizar este tipo de diagrama son los valores atípicos. Éstos están representados por puntos ubicados por encima o por debajo de los valores extremos de los bigotes y que superan la varianza esperada para el conjunto. En caso de existir varios valores atípicos, la normalidad del conjunto puede estar comprometida. Choonpradub y McNeil (2005) recomiendan utilizar este tipo de diagrama cuando la distribución de los datos sea continua.

Hasta este momento se han abordado herramientas utilizadas para la realización de la descripción de la muestra y, por consiguiente, de una población, su composición y la interpretación de su comportamiento; pero desde un enfoque unidimensional o individual. Todo esto es muy útil para conocer el contexto en que se desarrolla la investigación. Sin embargo, tanto las personas como los fenómenos que se estudian no existen aislados, y la profundización en esos posibles vínculos que pueden existir entre ellos o las relaciones que se establecen de acuerdo con sus características permiten realizar análisis más complejos con más de una variable de forma simultánea. En este sentido, una de las técnicas más utilizadas es la creación de tablas de contingencia o tablas cruzadas.

Modelos de tabulación cruzada de datos

Las tablas de contingencia o tablas cruzadas son utilizadas en la estadística para estudiar la incidencia de una variable (independiente) sobre los valores de otra (dependiente). Este tipo de estudios puede ser bivariado, cuando sólo se analiza la posible relación entre dos variables, o multivariado, si se analiza

el efecto de dos o más variables independientes sobre la dependiente. Para que los resultados sean correctos, las variables por analizar tienen que ser categóricas, ya sean ordinales o nominales. Es importante resaltar que no se recomienda aplicar esta herramienta cuando existen muchos valores perdidos dentro de las respuestas recopiladas, o cuando el número de entrevistados no sea representativo para generalizar los resultados.

Para enfrentar este tipo de investigación, desde el planteamiento del problema y con base en su estructura debe definirse con claridad cuál será la variable dependiente, la que se desea explicar, y cuál o cuáles las que de alguna forma se probará su influencia sobre aquélla. En ese sentido, se plantean dos hipótesis estadísticas: la hipótesis alternativa, que afirma la existencia de una posible relación o asociación entre los elementos estudiados (H_1), y la hipótesis nula, que señala la total independencia de éstos (H_0). H_1 y H_0 son mutuamente excluyentes, o sea, la aceptación de una implica el rechazo de la otra. Esta definición guiará al investigador para la interpretación de la tabla mediante el análisis de las frecuencias absolutas y relativas, o porcentajes de casilla y los residuos, o sea, la diferencia entre el modelo teórico y empírico.

Este proceso de contrastación de hipótesis comienza con el planteamiento teórico de la posible relación de dependencia. Luego se analizan las variables categóricas implicadas mediante técnicas estadístico-matemáticas, para determinar la aceptación o rechazo de H_0 , que representa la negación de H_1 . Al contrastar la hipótesis nula, se asume por los autores la independencia de los fenómenos que se estudian. Finalmente, debe expresarse de forma explícita y mediante texto la interpretación de los resultados encontrados, o sea, la existencia o no de asociación.

Para evaluar la relación entre las variables mediante el análisis de los porcentajes, éstos deben generarse en la dirección de la ubicación de la variable independiente o factor causal. El análisis se realizará sobre la base de la comparación de las diferencias porcentuales para cada categoría de las variables independientes implicadas en correspondencia con la categoría de la variable dependiente. De igual forma, debe considerarse analizar estos porcentajes en relación con el promedio de los mismos o total marginal. En este sentido, se emplea la tabla de frecuencias relativas.

Sean A variable dependiente y B variable independiente, categóricas y cada una con dos categorías que las describen (A_1 y A_2 ; B_1 y B_2) respectiva-

mente, al realizar el contraste de estas variables se obtendrá una tabla como la siguiente:

Tabla 1. *Tabla de frecuencias relativas (porcentajes)*

		Variable B		TotalF
		B_1	B_2	
Variable A	A_1	P_{11}^f P_{11}^c P_{11}	P_{12}^f P_{12}^c P_{12}	p_{1+}
	A_2	P_{21}^f P_{21}^c P_{21}	P_{22}^f P_{22}^c P_{22}	p_{2+}
Total		p_{+1}	p_{+2}	p_{++}

Fuente: Elaboración propia, con base en López-Roldán y Fachelli (2015).

Donde:

A = Variable supuesta dependiente

A_i = Categorías de la variable A

B = Variable supuesta independiente

B_j = Categorías de la variable B

P_{ij}^f = Porcentaje por fila (cantidad de casos por casilla entre el total de casos por fila $\times 100$)

P_{ij}^c : Porcentaje por columna (cantidad de casos por casilla entre el total de casos por columna $\times 100$)

P_{21} : Porcentaje total (cantidad de casos por casilla entre el total de casos $\times 100$)

p_{+j} y p_{i+} : Porcentaje de categoría o distribución marginal relativa

El cálculo de las diferencias para cada casilla ij quedaría establecido de acuerdo con la direccionalidad de la investigación según la ecuación 8:

$$Dif_{ij} = p_{i+} - P_{ij}^f \quad \text{ó} \quad Dif_{ij} = p_{+j} - P_{ij}^c \quad (8)$$

Luego del análisis de los porcentajes, se deben evaluar las frecuencias absolutas obtenidas al realizar el cruce de las categorías en la tabla de contingencias.

Tabla 2. *Tabla de contingencia*

		Variable B		TotalF
		B_1	B_2	
Variable A	A_1	$f_{o_{11}}$ $f_{e_{11}}$ R_{11} Re_{11}	$f_{o_{12}}$ $f_{e_{12}}$ R_{12} Re_{12}	$TotalF_1$
	A_2	$f_{o_{21}}$ $f_{e_{21}}$ R_{21} Re_{21}	$f_{o_{22}}$ $f_{e_{22}}$ R_{22} Re_{22}	$TotalF_2$
TotalC		$TotalC_1$	$TotalC_2$	TotalT

Fuente: Elaboración propia, con base en López-Roldán y Fachelli (2015).

Donde:

A = Variable supuesta dependiente

A_i = Categorías de la variable A

B = Variable supuesta independiente

B_j = Categorías de la variable B

$f_{o_{ij}}$ = Frecuencia observada en la celda ij -ésima

$f_{e_{ij}}$ = Frecuencia esperada en la celda ij -ésima

R_{ij} = Residuo en la celda ij -ésima

Re_{ij} = Residuo estandarizado en la celda ij -ésima

TotalF = Valor total de filas

TotalC = Valor total de columnas

TotalFi = \sum Valores fila - i ésima

TotalCj = \sum Valores columna - j ésima

TotalT = \sum Valores tabla

La ubicación de las variables según su comportamiento como dependiente o independiente no es obligatoria y además no incide en el resultado del proceso; es una recomendación sobre la base del planteamiento de la hipótesis de asociación. De igual manera, las variables pueden contar con más de dos categorías que las expliquen. En la Tabla 2, los valores que se

obtienen al cruzar una categoría con otras, se explican de la siguiente manera: las frecuencias esperadas ($f_{e_{ij}}$) representan lo que se conoce como modelo teórico, pues están asociadas al planteamiento probabilístico de la existencia de la asociación. Al mismo tiempo, se obtienen las frecuencias observadas ($f_{o_{ij}}$), o modelo empírico, lo que ocurre en la realidad. Al comparar el modelo empírico obtenido con el modelo teórico de referencia mediante el cálculo de la diferencia entre uno y otro, se obtienen los residuos (R_{ij}).

$$R_{ij} = f_{e_{ij}} - f_{o_{ij}} \quad (9)$$

A este respecto debe tomarse en consideración que si al calcular los residuos R_{ij} se obtienen valores de 0 en todos los casos, entonces los valores observados son idénticos a los valores esperados, lo que lleva a aceptar la H_0 . Sin embargo, si los valores son muy diferentes de 0, ya sea en sentido positivo o negativo, eso supone la existencia de dependencia entre las variables analizadas. Al mismo tiempo, la existencia de una variabilidad elevada entre los valores obtenidos no permite definir con claridad una escala que permita aprobar o rechazar las hipótesis estadísticas planteadas, por lo que todos los valores R_{ij} obtenidos deben estandarizarse (de Rada, 2009) mediante la aplicación del estadístico ji al cuadrado de Pearson, el cual se calcula según la ecuación 9:

$$\chi^2 = \sum \frac{(f_{o_{ij}} - f_{e_{ij}})^2}{f_{e_{ij}}} \quad (10)$$

Esta medida de asociación es utilizada con la finalidad de comprobar si la relación entre dos variables es significativa. Antes de usarla es importante considerar algunos elementos, según Reynolds (1984), como que el tipo de muestreo aplicado en el momento de definir los participantes debe ser aleatorio simple, las categorías de las variables tienen que ser mutuamente excluyentes y no debe utilizarse el ji al cuadrado cuando al solicitar en el *software* que se utilice, la cantidad de celdas con frecuencia esperada menor que 5 supere el 20%.

Sobre el valor del ji al cuadrado de Pearson, deben realizarse algunas consideraciones. Un valor elevado indica la existencia de grandes diferencias

entre las frecuencias observadas y las esperadas, o sea, en términos asociativos existe relación entre las variables estudiadas. Aunado a este resultado, se debe tomar en cuenta el valor correspondiente a la significación asintótica bilateral, pues esta medida indica la probabilidad de cometer un error al afirmar que existe relación. En las investigaciones donde se emplean encuestas es recomendable no considerar valores superiores al 5% de probabilidad de equivocación, o sea, se trabaja con un 95% de certeza o nivel de confianza. Por tanto, el valor que se obtenga debe ser menor o igual a 0.05, para entonces afirmar que la relación entre las dos variables existe y es significativa, y a su vez, rechazar la hipótesis nula planteada (de Rada, 2009).

Como resultado de la estandarización, se obtienen los residuos estandarizados Re_{ij} . Al analizar los valores de esta medida, debe considerarse que un valor por encima de +1.96 y menor que -1.96 confirma que existe relación entre las variables analizadas con un nivel de confianza superior al 95%. El incremento de estos valores explica un fortalecimiento de la relación y el signo va a indicar su dirección (Haberman, 1973).

Al utilizar el estadístico ji al cuadrado para tablas de dimensiones diferentes de filas y columnas, se corre el riesgo de que los resultados puedan ser no concluyentes, por lo que deben utilizarse algunas medidas de asociación global definidas para este tipo de tablas. En este sentido, se utiliza el coeficiente V de Cramér, que permite evaluar el grado de relación y su intensidad con un único valor (Cramér, 1946), el cual se calcula según la ecuación 10.

$$V = \sqrt{\frac{\chi^2}{n*(k-1)}} \quad (11)$$

Donde:

n = total de observaciones

k = es el menor valor entre el número de filas y el número de columnas

El rango de comportamiento del V de Cramér es entre 0 y 1. El primero de estos valores es expresión de máxima independencia entre las variables estudiadas, y el segundo indicador de una asociación perfecta. Sin embargo, la literatura académica describe que valores menores a 0.2 indican una rela-

ción débil, valores entre 0.2 y 0.6 una relación moderada, y superiores a 0.6 una relación fuerte entre los campos implicados (Cramér, 1946).

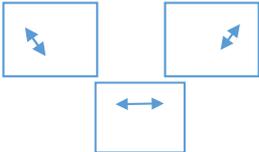
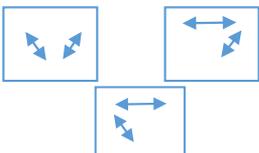
Modelos de tabulación cruzada multidimensional

Los fenómenos de la naturaleza se pueden clasificar como complejos, por lo que resulta muy difícil analizarlos al tomar solo en cuenta dos factores por separado. En este sentido, las tablas de contingencia multidimensionales brindan la oportunidad de introducir una tercera variable al análisis, después de obtener resultados significativos en cuanto a la relación entre dos indicadores. Esto permite al investigador evaluar dicha relación de acuerdo con el comportamiento de otros factores importantes, lo que incide directamente en la profundidad del estudio realizado y, por consiguiente, en la calidad de los resultados obtenidos. Al momento de incluir una tercera variable o variable de control, como se le conoce comúnmente, se debe definir con claridad el modelo que sirva como base para la selección de los indicadores por estudiar y, por tanto, al planteamiento de las hipótesis que se busca comprobar en relación con los mismos.

Para la realización de este tipo de análisis, se debe aplicar el procedimiento descrito anteriormente para las tablas bidimensionales a lo interno de cada categoría de la variable moderadora. Esto significa descomponer la relación original en subgrupos que permitirán su caracterización de una forma más detallada. La interpretación de estas tablas debe realizarse mediante la comparación entre las tablas parciales.

La comparación entre un conjunto de variables genera diferentes modelos de interrelación en función de las asociaciones que se presentan como parte del análisis. Al considerar variables A , B y C , López-Roldán y Fachelli (2015) plantean que se pueden encontrar los siguientes modelos:

Tabla 3. Modelos de interrelación en análisis de múltiples variables

Modelo	Descripción	Ilustración
Modelo de independencia mutua	No existe relación entre las variables estudiadas.	
Modelo de asociación condicional	Existe relación entre dos de las variables implicadas en el estudio, y esta relación es independiente de la tercera variable. Esto puede suceder para cualquier combinación de variables.	
Modelo de doble asociación	Resultan asociaciones parciales entre dos de las tres variables, cada variable es independiente de forma condicional. Este tipo de modelo permite la identificación de relaciones espurias.	
Modelo de asociación homogénea	Resultan dobles asociaciones entre las 3 variables implicadas, siendo cada pareja independiente de la 3ra variable.	
Modelo de interacción o asociación triple	Los 3 pares de variables están relacionados y la intensidad de cada relación varía para cada categoría de la 3ra variable.	

Fuente: Elaboración propia, con base en López-Roldán y Fachelli (2015).

Caso práctico. Factores sociodemográficos que inciden en el uso del crédito hipotecario en México

Introducción

En la actualidad, uno de los principales problemas que enfrenta la humanidad es el acceso a una vivienda propia y digna, a pesar de que este tema es considerado un derecho fundamental del ser humano. En el año 2020 la Organización de las Naciones Unidas (ONU) estimaba que alrededor de

1 800 millones de personas, cerca del 20% de la población mundial, carecía de una vivienda adecuada (ONU, 2020). En el caso de México, el artículo 4to de la Constitución Política de los Estados Mexicanos establece el derecho que corresponde a todas las familias a poseer una vivienda digna y decente, que cubra las necesidades básicas en cuanto a salubridad, habitabilidad, seguridad, así como que considere los elementos necesarios para garantizar la prevención de sus habitantes ante posibles desastres y fenómenos naturales (CPEUM, 2022).

Sin embargo, la crisis global de la vivienda es visible de igual manera en la República Mexicana. Según la Encuesta Nacional de la Vivienda (2020), sólo el 57% de los domicilios son propios, o sea, más de 14 millones de familias no tienen los medios para adquirir una vivienda o construirla. Otro elemento llamativo es que el 57.3% fueron construidas por sus propietarios y sólo el 31.6% fueron adquiridas con alguna fuente de financiamiento asociada a un crédito hipotecario. Este producto financiero es reconocido a nivel internacional como una de las fuentes principales para la adquisición de una vivienda, por lo que resulta llamativo el bajo porcentaje de la población mexicana que opta por esta herramienta para adquirirlas.

Diversas son las investigaciones que profundizan en este tema y buscan identificar las principales características sociodemográficas de los pobladores, con la finalidad de que las políticas emitidas por los gobiernos y los productos ofertados por las instituciones financieras se elaboren con base en evidencias científicas y puedan incidir en la disminución de las familias sin una vivienda adecuada. En este sentido, se ha identificado que los principales factores que inciden en el uso del crédito hipotecario son el nivel de ingresos (Carmelo *et al.*, 2019; Herrera, 2022; Sandoval, 2014; Tramontin, 2020; Vasquez y Osorio, 2022), la situación laboral (Carmelo *et al.*, 2019; Sandoval, 2014), la edad (Camelo *et al.*, 2019; Herrera, 2022), el nivel educativo (Camelo *et al.*, 2019; Vasquez y Osorio, 2022) y el tamaño de la localidad donde residen (Vasquez y Osorio, 2022). Con base en lo planteado, esta investigación tiene como objetivo evaluar si el nivel socioeconómico de los mexicanos, medido a través del nivel de escolaridad y el nivel de ingresos, se encuentra asociado a la tenencia del crédito hipotecario. En esta línea también se pretende validar si las relaciones antes descritas se encuentran influenciadas por el género de los derechohabientes.

Fuente de información

La fuente de información utilizada son los microdatos de la Encuesta Nacional sobre las Finanzas de los Hogares (Enfih), cuyos resultados fueron presentados por el Banco de México y el Instituto Nacional de Estadística y Geografía (Inegi) en el año 2019. El objetivo de esta encuesta era generar información estadística sobre las finanzas de los hogares mexicanos, en particular sobre los flujos y los acervos de sus activos y pasivos. Sus resultados son el perfecto complemento a los datos emitidos por las instituciones financieras bancarias y no bancarias que otorgan los créditos.

Población y muestra

La población estudiada estuvo conformada por los 38 467 registros incluidos en los microdatos resultantes de la aplicación de la Enfih. Sobre el total de los campos de esta base de datos se seleccionaron para este estudio los más relevantes, que se describen como parte del proceso de selección de las variables.

Selección de las variables

El cuestionario de la Enfih 2019 está conformado por 12 secciones temáticas, de las cuales las primeras cuatro son respondidas por el responsable del hogar y el resto se contestan por cada uno de los miembros que residen en el domicilio. En el caso específico de este estudio y por el objetivo que se plantea, se realizó una revisión profunda del cuestionario y se seleccionaron aquellas preguntas que generan la información correspondiente a las principales características sociodemográficas de los residentes en los hogares, que se consideran las variables independientes, mientras las variables dependientes se definen a partir de la pregunta de si la vivienda en la que residen fue adquirida mediante el uso de un crédito hipotecario (Tabla 4).

Tabla 4. Operacionalización de las variables

Variables Independientes	Pregunta Enfh 2019	Concepto
Género	2.3 (NOMBRE) es hombre (NOMBRE) es mujer	1 - Hombre 2 - Mujer
Edad	2.4 ¿Cuántos años cumplidos tiene (NOMBRE)?	00 - Menos de un año 01 - 96 años de edad 97 - 97 años y más
Nivel escolar	2.6 ¿Hasta qué año o grado aprobó (NOMBRE) en la escuela?	00 - Ninguno 01 - Básico 02 - Medio Superior 03 - Superior
Tamaño de localidad	Tamaño de localidad	1 - 100 000 y más habitantes. 2 - 15 000 a 99 999 habitantes. 3 - 2 500 a 14 999 habitantes. 4 - menor de 2 500 habitantes.
Nivel de ingresos	4.8 Considerando sueldo, comisiones, horas extras o propinas, ¿cuánto gana o recibe por trabajar? Si tiene más de un trabajo, por favor también considérela.	
Variable dependiente	Pregunta Enfh 2019	Concepto
Uso de crédito	4.15 ¿Para financiar la compra (construcción) de esta vivienda obtuvieron un crédito de vivienda con Infonavit, Fovissste o bancario?	1 - Sí 2 - No

Fuente: Elaboración propia, con base en los datos de la Enfh (2019).

Resultados

Características generales de la muestra

Para describir la muestra seleccionada en el estudio, se toman los valores de las variables asociadas a las características sociodemográficas y se analizan mediante el empleo de las medidas descritas al principio de este capítulo. En este sentido, se debe iniciar clasificando las variables del estudio para seleccionar de forma correcta los estadísticos descriptivos por aplicar.

Análisis descriptivo de variables continuas

En este caso existen dos variables continuas: la edad y el nivel de ingresos, para describirlas, como se explicó al inicio, se pueden emplear varios estadísticos (Tabla 5)

Tabla 5. Estadísticos descriptivos para las variables edad y nivel de ingresos

		EDAD	NIVEL_ING
N	Válido	38,467	25,146
	Pérdidos	0	13,321
Media		43.02	4939.76
Mediana		41.00	1800.00
Moda		18	1500
Desviación estándar		17.211	13,246.637
Rango		79	720,000
Mínimo		18	0
Máximo		97	720,000
Suma		1,655,039	124,215,139
Percentiles	25	28.00	1,000.00
	50	41.00	1,800.00
	75	55.00	5,000.00

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

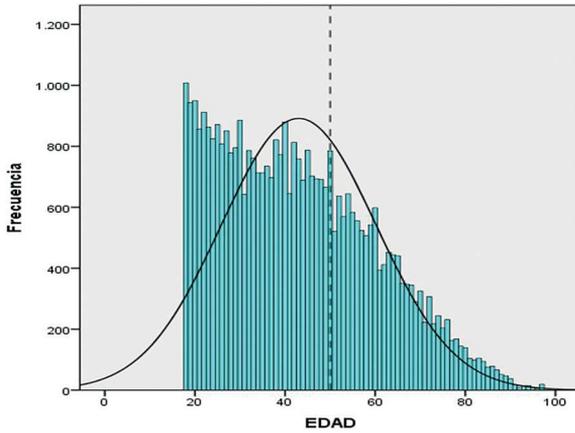
Sobre la base de la información que generan estos estadísticos, se puede apreciar que la edad promedio de la muestra es de 43.02 años, y la mediana es de 41 años. Al tomar en cuenta el valor de la desviación estándar para esta última variable (17.211), podemos afirmar que la muestra en cuestión no está normalizada, por lo que se recomienda utilizar como valor central de referencia la mediana. Por su parte, la moda es igual a 18 años, lo que nos permite interpretar que la mayor cantidad de personas encuestadas tiene esa edad. Otros datos interesantes son el rango de edad, que resulta en 79, y los cuartiles que dividen el conjunto en cuatro partes iguales.

Lo descrito anteriormente se puede reforzar con la representación gráfica de la edad mediante el empleo del histograma de la Figura 8. En ella podemos identificar que la mayor parte de los elementos de la muestra se encuentra por debajo de los 40 años, y que con línea discontinua aparece señalada la media aritmética de este conjunto igual a 43.02, o sea, la edad promedio de la muestra. Visualmente se puede confirmar que la muestra no está normalizada, ya que los valores no se distribuyen de forma uniforme a ambos lados de la media aritmética.

De igual forma, se puede graficar esta variable mediante un diagrama de caja como muestra la Figura 9. Esta imagen expone con claridad los cuartiles en los que se divide el conjunto de datos. Además, se puede observar que el 75% de los datos se encuentra por debajo de la edad de 55 años, el

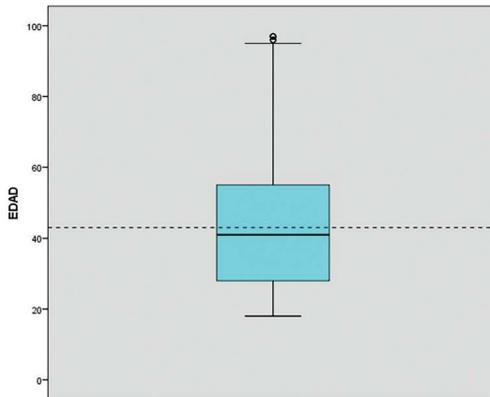
máximo ajustado está en los 95 y el mínimo en 18, así como se aprecian valores atípicos sobre el bigote superior. Con una línea discontinua se representa la media aritmética y se visualiza la diferencia con la mediana, línea continua que divide la caja y que evidencia por su ubicación que la muestra es asimétrica, pues los datos no están distribuidos de igual manera a ambos lados de la misma. De forma particular en este caso, se considera asimétrica negativa ya que el lado superior es menor que el inferior.

Figura 8. Edad (histograma)



Fuente: Elaboración propia, a partir del software SPSS versión 23.

Figura 9. Edad (diagrama de caja)



Fuente: Elaboración propia, a partir del software SPSS versión 23.

La otra variable continua en cuestión es el nivel de ingresos. En este caso se puede apreciar en la Tabla 5 que la media aritmética es de \$4 939.76 pesos; sin embargo, la mediana tiene un valor de \$1 800 pesos. Al analizar la desviación estándar, se observa que su valor es muy elevado (13 246.637), por lo que se debe considerar la mediana como medida central. En cuanto a la moda para esta variable, su valor es de \$1 500 pesos, o sea, el monto de ingresos que más se repite en la muestra seleccionada.

Análisis de variables categóricas

Dentro de las variables seleccionadas existen tres que cumplen con esta condición: el género, el nivel escolar y el tamaño de la localidad. Para la descripción de este tipo de variables se recomienda el empleo de las tablas de frecuencia, que permiten contabilizar las apariciones de las categorías que las conforman.

En este sentido, se analiza en primer lugar la variable género. Del total de individuos incluidos en la muestra el 47.64% fueron hombres y el 52.36% mujeres, como se puede apreciar en la Tabla 6. Visualmente se recomienda utilizar un diagrama de barras para ilustrar los valores de las categorías incluidas (*vid.* Figura 10).

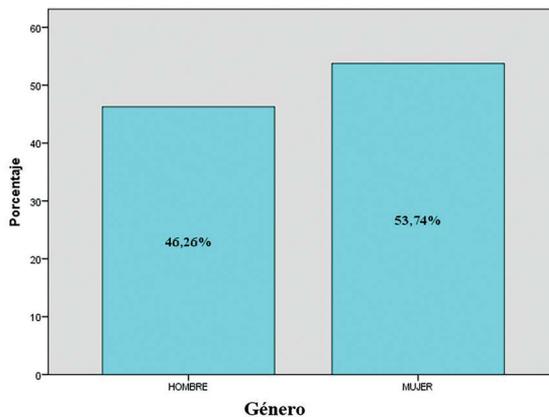
Tabla 6. Género

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Hombre	26,827	47.6	47.6	47.6
	Mujer	29,489	52.4	52.4	100.0
	Total	56,316	100.0	100.0	

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Asimismo, al analizar el nivel escolar de la muestra, se puede observar en la Tabla 7 que del total de encuestados el 48.7% tiene nivel básico, el 24% medio superior y el 22.7% nivel superior. Para ilustrar gráficamente esta variable se utiliza un diagrama de pastel, ya que su descripción va más asociada a los porcentajes que representa cada categoría del total de individuos (*vid.* Figura 11).

Figura 10. Género (diagrama de barras)



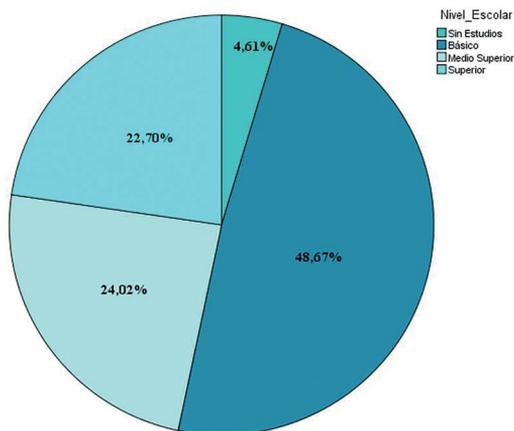
Fuente: Elaboración propia, a partir del software SPSS versión 23.

Tabla 7. Nivel escolar

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sin Estudios	1,775	4.6	4.6	4.6
	Básico	18,721	48.7	48.7	53.3
	Medio Superior	9,239	24.0	24.0	77.3
	Superior	8,732	22.7	22.7	100.0
	Total	38,467	100.0	100.0	

Fuente: Elaboración propia, a partir del software SPSS versión 23.

Figura 11. Nivel educativo (diagrama de pastel)



Fuente: Elaboración propia, a partir del software SPSS versión 23.

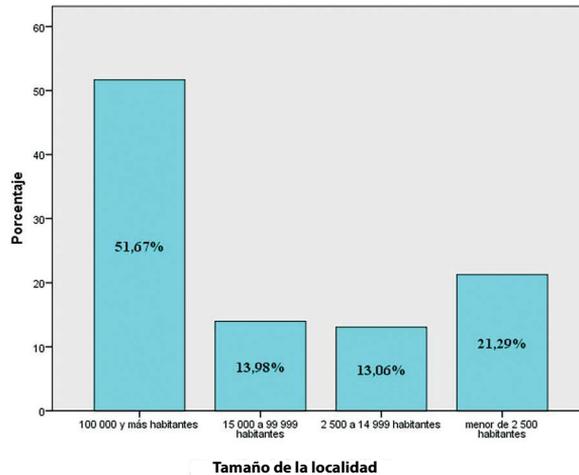
Con respecto al tamaño de la localidad donde residen, se puede observar en la Tabla 8 la distribución de los elementos de la muestra en las cuatro categorías consideradas para esta variable. Se aprecia con claridad que la mayor parte (51.7%) de los incluidos en este estudio radican en localidades urbanas de más de 100 000 habitantes, el 27.1% en localidades entre 2 500 y 100 000 habitantes y el 21.3% en localidades pequeñas de menos de 2 500 habitantes. De forma visual lo podemos exponer mediante el empleo de un diagrama de barras como el de la Figura 12, donde se describe la cantidad de individuos por categoría.

Tabla 8. *Tamaño de localidad*

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	100,000 y más habitantes	19,877	51.7	51.7	51.7
	15,000 a 99,999 habitantes	5,378	14.0	14.0	65.7
	2,500 a 14,999 habitantes	5,023	13.1	13.1	78.7
	menor de 2,500 habitantes	8,189	21,3	21.3	100.0
	Total	38,467	100.0	100.0	

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Figura 12. *Tamaño de localidad (diagrama de barras)*

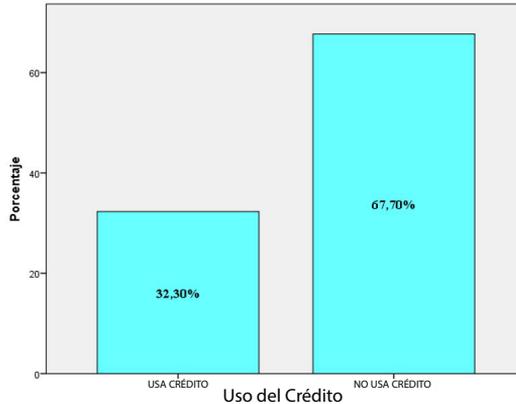


Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Con respecto al análisis de nuestra variable dependiente, se puede observar en la Figura 13 que el 67.7% de los encuestados no utilizó crédito

hipotecario para financiar la adquisición o construcción de la vivienda donde residen.

Figura 13. *Uso del crédito para el financiamiento de la adquisición o construcción de la vivienda*



Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

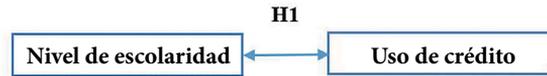
Factores sociodemográficos que determinan el uso del crédito hipotecario

Con la finalidad de identificar la asociación existente entre el nivel socioeconómico de los mexicanos y el uso o no del crédito hipotecario en México, se decide emplear la tabulación cruzada de datos. En este sentido y tomando como base la bibliografía consultada, se plantean las siguientes hipótesis estadísticas:

- H_1 : Existe relación de dependencia entre el nivel escolar de los individuos y su decisión de emplear un crédito hipotecario para financiar la compra de su vivienda en México.
- H_2 : Existe relación de dependencia entre el nivel de ingresos de los individuos y su decisión de emplear un crédito hipotecario para financiar la compra de su vivienda en México.

La hipótesis 1 plantea la existencia de asociación entre el nivel de escolaridad de los individuos y el uso del crédito hipotecario. El modelo teórico asociado a esta hipótesis queda definido según la Figura 14.

Figura 14. Modelo hipótesis 1



Fuente: Elaboración propia.

Para la validación de la hipótesis se procede a analizar en primer lugar los valores de la prueba del ji al cuadrado y el coeficiente V de Cramér. Los resultados obtenidos se muestran en la Tabla 9 y 10.

Tabla 9. *Uso del crédito según nivel de escolaridad: prueba de ji al cuadrado*

	Valor	gl	Significación asintótica (bilateral)
Ji al cuadrado de Pearson	2022.501 ^a	3	.000

^a. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 367.90.
Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Tabla 10. *Uso del crédito según nivel de escolaridad: Media de asociación V de Cramér*

		Valor	Significación aproximada
Nominal por nominal	Phi	.295	.000
	V de Cramér	.295	.000
N de casos válidos		23,263	

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Se observa que el valor de ji al cuadrado es muy elevado, y la posibilidad de cometer error al afirmar que existe relación de dependencia entre estas dos variables es nula, por lo que se puede rechazar H_0 . De igual forma, el valor del coeficiente V de Cramér es superior a 0.2, por lo que la relación es significativa de nivel moderado.

Para comentar los resultados del efecto del nivel de escolaridad en el hecho de adquirir o no un crédito hipotecario, se muestra la Tabla 11, en donde se sitúan en la fila la variable dependiente Uso del crédito y en las columnas la variable independiente Nivel de escolaridad, con sus respectivos porcentajes. Como se puede apreciar, sólo el 32.3% de la muestra tiene un crédito hipotecario, a diferencia del 67.7% que mencionó que no cuenta con

este tipo de producto financiero. Los datos demuestran que el tener nivel superior en comparación con las personas sin estudios o de nivel básico incrementa la probabilidad de adquirir un crédito hipotecario en un 42% y en un 21.9%, respectivamente. Para profundizar en este aspecto se emplea el análisis marginal, que representa la variación de los porcentajes de las categorías de la variable independiente, donde se pueden visualizar las posibles diferencias (*vid.* Tabla 12). Al realizar un análisis detallado de las distribuciones condicionales, podemos apreciar que las personas con nivel superior usan 17.8% más créditos hipotecarios que las categorías medio superior con 10.3% y en menor medida se encuentran las personas con educación básica (-11.3%) y sin escolaridad (-24.2%). Estos resultados sin lugar a dudas confirman el supuesto de que a medida que aumenta el nivel de escolaridad se incrementa la proporción de personas que adquieren un crédito en el país, lo que permite aceptar nuestra hipótesis de investigación.

Tabla 11. *Uso del crédito según Nivel de escolaridad*

		Nivel de escolaridad				Total
		<i>Sin Estudios</i>	<i>Básico</i>	<i>Medio Superior</i>	<i>Superior</i>	
USO DEL CRÉDITO	% dentro de Nivel Escolar	8.1%	21.0%	42.6%	50.1%	32.3%
	Residuo estandarizado	-14.4	-21.1	13.5	23.2	
NO USO DEL CRÉDITO	% dentro de Nivel Escolar	91.9%	79.0%	57.4%	49.9%	67.7%
	Residuo estandarizado	9.9	14.5	-9.3	-16.0	
Total	% dentro de Nivel Escolar	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente: Elaboración propia, a partir del software SPSS versión 23.

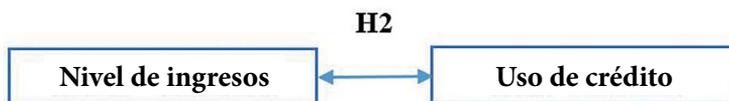
Tabla 12. *Diferencias marginales Uso del crédito - Nivel de escolaridad*

	Nivel de escolaridad				Total
	<i>Sin Estudios</i>	<i>Básico</i>	<i>Medio Superior</i>	<i>Superior</i>	
USO DEL CRÉDITO	-24.2%	-11.3%	10.3%	17.8%	0
NO USO DEL CRÉDITO	24.2%	11.3%	-10.3%	-17.8%	0
Total	0	0	0	0	0

Fuente: Elaboración propia, a partir del software SPSS versión 23.

Otro de los factores más tratados en investigaciones previas sobre la situación socioeconómica de las personas es el nivel de ingresos. En concordancia con lo anterior, nuestra segunda hipótesis plantea que el nivel de ingresos es un factor determinante para tener un crédito hipotecario en México. El modelo que soporta esta hipótesis se describe a continuación:

Figura 15. Modelo hipótesis 2



Fuente: Elaboración propia.

Para poder validar el planteamiento de nuestro modelo, se requirió realizar una modificación de la variable original nivel de ingreso, pues al ser una variable cuantitativa no permite llevar a cabo la tabulación cruzada de la información. En consecuencia y después de su recodificación en el *software* SPSS, se establecieron los siguientes cuatro cuartiles para el análisis del nivel de ingresos: Q1 (mayor o igual a 0 y menor a 1 000 pesos al mes), Q2 (mayor o igual a 1 000 y menor a 1 800 pesos al mes), Q3 (mayor o igual a 1 800 y menor a 5 000 pesos al mes) y Q4 (mayor o igual a 5 000 pesos al mes). Al realizar el análisis estadístico, los resultados son los que se pueden observar en la Tabla 13 y 14:

Tabla 13. *Uso del crédito según Nivel de ingresos (agrupado): prueba de ji al cuadrado*

	Valor	gl	Significación asintótica (bilateral)
Ji al cuadrado de Pearson	1158.306 ^a	3	.000
Razón de verosimilitud	1171.225	3	.000
Asociación lineal por lineal	1133.253	1	.000
N de casos válidos	14,654		

^a. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 1,059.46.

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Tabla 14. *Uso del crédito según Nivel de ingresos (agrupado): medida de asociación V de Cramér*

		Valor	Significación aproximada
Nominal por Nominal	Phi	.281	.000
	V de Cramér	.281	.000
N de casos válidos		14,654	

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

El valor del χ^2 al cuadrado obtenido es elevado, al mismo tiempo que la significación asintótica bilateral se encuentra por debajo de 0.05, por lo que se puede rechazar H_0 . El valor del coeficiente V de Cramér es superior a 0.2, lo que indica que la relación es significativa de nivel moderado.

Al igual que en nuestro análisis anterior, la Tabla 15 muestra los resultados de la tabulación cruzada entre la variable dependiente uso del crédito y el nivel de ingresos, con sus respectivos porcentajes en cada una de sus categorías. Como se puede apreciar y de manera similar a cuando se evaluaba el efecto de la escolaridad, el 65.8% de los encuestados no cuenta con crédito hipotecario, a diferencia del 34.2% que afirmó contar con este tipo de producto financiero. En el análisis marginal (*vid.* Tabla 16), se puede identificar una marcada diferencia entre los estratos más elevados del nivel de ingresos ($Q4 = 20\%$ y $Q3 = 3.3\%$) en comparación con los niveles inferiores $Q2$ y $Q1$, en relación con la tenencia de crédito.

En términos probabilísticos el encontrarse en el estrato de ingreso $Q4$ incrementa la probabilidad de acceso al crédito en un 35.6%, comparado con las personas con menor ingreso en el país (estrato $Q1$). Estos resultados permiten corroborar que al igual que la escolaridad, el nivel de ingreso es una variable estrechamente relacionada con el acceso al mercado de crédito hipotecario en México. Al mismo tiempo, los datos de nuestro análisis permiten afirmar que a medida que aumenta el nivel de ingreso, se incrementa la tenencia de crédito hipotecario, para aceptar de esta manera nuestra hipótesis alternativa H_2 .

Tabla 15. *Uso del crédito según Nivel de ingresos (agrupado)*

		Nivel de ingresos				Total
		Q1	Q2	Q3	Q4	
USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	18.6%	28.8%	37.5%	54.2%	34.2%
	Residuo estandarizado	-17.4	-5.2	3.4	20.5	
NO USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	81.4%	71.2%	62.5%	45.8%	65.8%
	Residuo estandarizado	12.6	3.7	-2.5	-14.8	
Total	% dentro de NIVEL ING (agrupado)	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

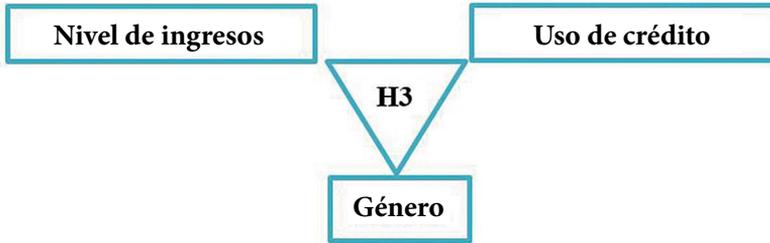
Tabla 16. *Diferencias marginales Uso del crédito - Nivel de ingresos (agrupado)*

		Nivel de ingresos				Total
		Q1	Q2	Q3	Q4	
USO DEL CRÉDITO		-15.6%	-5.4%	3.3%	20%	0
NO USO DEL CRÉDITO		15.6%	5.4%	-3.3%	20%	0
Total		0	0	0	0	0

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Siguiendo con nuestros objetivos de profundizar en el análisis de la tabulación cruzada, pasamos a abordar el efecto multidimensional que trae consigo la incorporación de una tercera variable en el modelo, conocida en la bibliografía como variable moderadora. En este sentido, se pretende robustecer los análisis anteriores con la incorporación del enfoque de género, para corroborar de esta forma si la adición de esta variable evidencia una relación espuria, inexistente, o bien de menor intensidad entre el nivel de escolaridad y la tenencia del crédito hipotecario (Figura 16). En concordancia con lo anterior, el modelo propuesto introduce una nueva hipótesis, que afirma que la decisión de utilizar un crédito hipotecario para financiar la compra de su vivienda tiene una relación de dependencia con el nivel de ingresos de los individuos según el género en México (H_3).

Figura 16. Modelo de interacción Nivel de ingresos, Uso de crédito y Género



Fuente: Elaboración propia.

La Tabla 17 presenta el test ji al cuadrado para evaluar el efecto de interacción del género en la relación de dependencia entre el uso del crédito hipotecario y el nivel de ingresos. Como se puede apreciar, tanto para hombres como para mujeres, los estadísticos de significancia confirman el efecto moderador de esta variable en el modelo teórico ($\text{sig} < 005$). De igual manera, el V de Cramér (*vid.* Tabla 18) es superior en ambos casos (0.293 y 0.292) al valor obtenido en el modelo sin el efecto de interacción (0.281).

Tabla 17. Tabla de contingencia entre Uso del crédito, Nivel de ingresos (Agrupado) y Género: Pruebas de ji al cuadrado

	Género	Valor	gl	Significación asintótica (bilateral)
Hombre	Ji al cuadrado de Pearson	720.011 ^b	3	.000
Mujer	Ji al cuadrado de Pearson	531.548 ^c	3	.000

b. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 578.66.

c. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 384.08.

Fuente: Elaboración propia, a partir del software SPSS versión 23.

Tabla 18. Tabla de contingencia entre Uso del crédito, Nivel de ingresos (Agrupado) y Género: Medida de asociación V de Cramér

	Género	Valor	Significación aproximada
Hombre	Nominal por Nominal	Phi	.292
		V de Cramér	.292
	N de casos válidos		8,443
Mujer	Nominal por Nominal	Phi	.293
		V de Cramér	.293
	N de casos válidos		6,211

Fuente: Elaboración propia, a partir del software SPSS versión 23.

La Tabla 19 cruza la información referida con el uso de crédito relacionado con el nivel de ingreso y la incorporación de nuestra variable moderadora. Para la interpretación de los resultados, debemos retomar los datos del ejercicio anterior en que el porcentaje de tenencia de crédito hipotecario era del 34.2%. Cuando analizamos entre hombres y mujeres, vemos cómo disminuye en un 1.6% en el caso del primer grupo y aumenta en un 2.2% en el segundo grupo. Si bien las discrepancias son pequeñas, los resultados obtenidos permiten apreciar que efectivamente la variable género tiene un impacto en el comportamiento de la relación ya validada. Ahora bien, cuando se analizan las distribuciones condicionales por nivel de ingreso, se tienen en cuenta los valores de los residuos estandarizados. De forma específica se realiza una comparación entre los estratos Q1 y Q4 para ambos géneros, al existir la mayor diferencia entre sus valores, lo que indica fortaleza en la relación, pero en direcciones opuestas. Se puede observar que las diferencias en hombres ($52.2\% - 14.8\% = 37.4\%$) son mayores que en mujeres ($57.7\% - 21.3\% = 36.4\%$). De esta forma, podemos afirmar que el uso del crédito hipotecario en los hombres se ve más acentuado entre los niveles superiores (Q4) e inferiores de ingreso (Q1) en comparación con las mujeres, lo que permite confirmar nuestra hipótesis 3 de investigación.

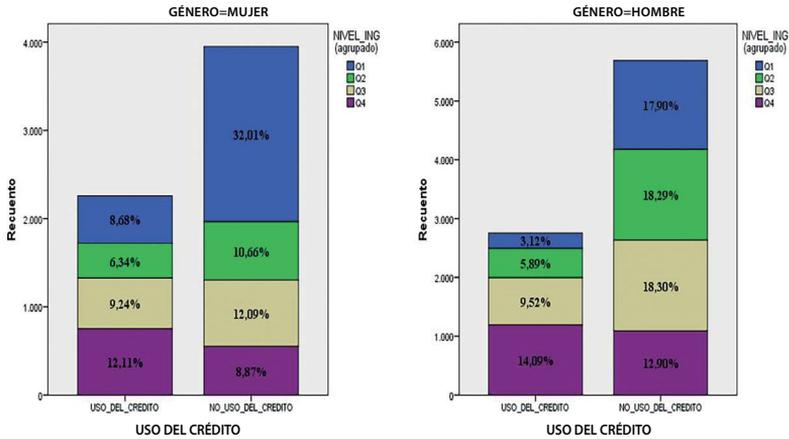
Tabla 19. *Tabla de contingencia Uso del crédito, Nivel de ingresos (Agrupado) y Género*

Género		Nivel de ingresos				Total	
		Q1	Q2	Q3	Q4		
Hombre	USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	14.8%	24.4%	34.2%	52.2%	32.6%
		Residuo estandarizado	-13.1	-6.5	1.4	16.4	
	NO USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	85.2%	75.6%	65.8%	47.8%	67.4%
		Residuo estandarizado	9.1	4.6	-.9	-11.4	
Mujer	USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	21.3%	37.3%	43.3%	57.7%	36.4%
		Residuo estandarizado	-12.5	.5	4.2	12.8	
	NO USO DEL CRÉDITO	% dentro de NIVEL ING (agrupado)	78.7%	62.7%	56.7%	42.3%	63.6%
		Residuo estandarizado	9.5	-4	-3.2	-9.7	

Fuente: Elaboración propia, a partir del software SPSS versión 23.

La Figura 17 muestra gráficamente que el uso del crédito hipotecario es mayor en mujeres que en hombres y que la brecha en cuanto al uso del crédito entre ambos géneros es mayor entre los cuartiles de ingreso Q4 y Q1.

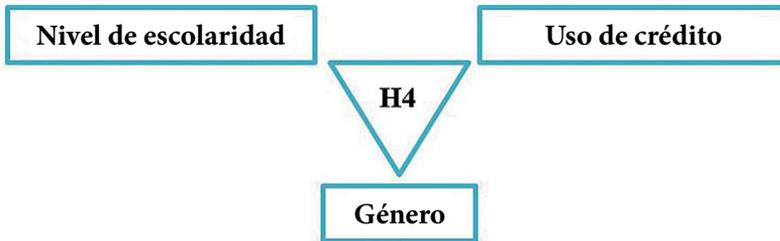
Figura 17. Gráfico de barras apiladas Uso del crédito, Nivel de ingresos (Agrupado) y Género



Fuente: Elaboración propia, a partir del software SPSS versión 23.

Por otra parte, también se busca comprobar si el género del individuo, como variable de control, incide de alguna manera en que las personas según su nivel de escolaridad utilicen un crédito hipotecario para financiar la compra de su vivienda (Hipótesis 4). Esta afirmación será contrastada mediante la misma técnica multivariante; en este caso específico se analiza si el sexo de la persona incide en el comportamiento de la relación bivariable entre el uso del crédito y el nivel de escolaridad. El modelo de interacción que soporta la hipótesis 4 se define mediante la Figura 18.

Figura 18. Modelo de interacción Nivel de escolaridad, Uso de crédito y Género



Fuente: Elaboración propia.

Las pruebas de ji al cuadrado (Tabla 20) muestran la existencia de una asociación significativa entre las variables implicadas: el valor del estadístico es muy elevado y la significación asintótica bilateral es menor a 0.05, lo que permite rechazar la hipótesis nula H_0 . Así mismo los valores del coeficiente V de Cramér (Tabla 21) está entre 0.2 y 0.6, lo que indica una asociación de nivel moderado. Si se compara con la obtenida al evaluar la relación entre el nivel de escolaridad y el uso del crédito (0.295), se puede observar como en el caso de los hombres hay un incremento del grado de asociación (0.305) y una atenuación para las mujeres (0.286), o sea, que desde esta medición se puede inferir que el género del individuo está relacionado y afecta este vínculo.

Tabla 20. *Tabla de contingencia entre Uso del crédito, Nivel de escolaridad y Sexo: pruebas de ji al cuadrado*

	Sexo	Valor	gl	Significación asintótica (bilateral)
Hombre	Ji al cuadrado de Pearson	1007.460 ^b	3	.000
Mujer	Ji al cuadrado de Pearson	1018.770 ^c	3	.000

b. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 146.02.

c. 0 casillas (0.0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 221.54.

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

Tabla 21. *Tabla de contingencia entre Uso del crédito, Nivel de escolaridad y Sexo: medida de asociación V de Cramér*

	Sexo		Valor	Significación aproximada
Hombre	Nominal por Nominal	Phi	.305	.000
		V de Cramér	.305	.000
	N de casos válidos		10845	
Mujer	Nominal por Nominal	Phi	.286	.000
		V de Cramér	.286	.000
	N de casos válidos		12418	

Fuente: Elaboración propia, a partir del *software* SPSS versión 23.

La Tabla 22 ilustra el cruzamiento de las tres variables implicadas en el análisis. Al interpretarla es importante comenzar por la comparación del porcentaje de uso del crédito obtenido, como parte del análisis de la relación bivariada ya probada (32.3%), y el correspondiente a cada categoría de la variable de control, o sea, para hombres (32.5%) y mujeres (32.1%), lo que

indica un incremento del 0.2% para los primeros y una atenuación del 0.1% para las féminas. Esto revela, a pesar de que las diferencias son pequeñas, que como en el caso de los ingresos, la variable género también tiene influencia sobre la asociación entre el nivel de escolaridad y el uso del crédito hipotecario. Para profundizar en este análisis, se examinan las distribuciones condicionales por el nivel de escolaridad y se toman en cuenta las categorías de la variable nivel de escolaridad entre las que la diferencia de los valores de los residuos estandarizados sea mayor (Básico y Superior). Esto indica mayor fortaleza en la relación, pero en direcciones opuestas, que es lo que se busca demostrar. Se observa con claridad que en el caso de los hombres las diferencias ($50.0\% - 20.1\% = 29.9\%$) son mayores que en las mujeres ($50.3\% - 21.7\% = 28.3\%$), lo que permite afirmar que el uso del crédito hipotecario es mayor en los hombres con nivel superior y básico en comparación con las mujeres, para validar la hipótesis 4 planteada.

Tabla 22. *Tabla de contingencia Uso del crédito, Nivel de escolaridad y Sexo*

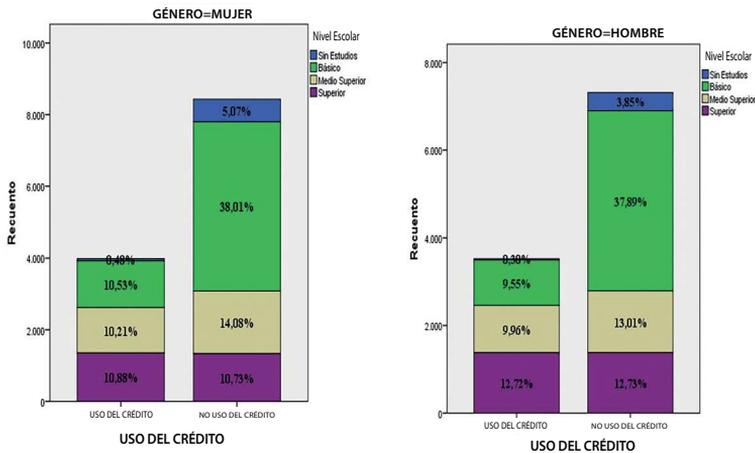
Sexo		Nivel escolar				Total	
		Sin Estudios	Básico	Medio Superior	Superior		
Hombre	USO DEL CRÉDITO	Recuento	32	1036	1080	1379	3527
		% dentro de Nivel escolar	7.1%	20.1%	43.4%	50.0%	32.5%
		Residuo estandarizado	-9.4	-15.6	9.5	16.1	
	NO USO DEL CRÉDITO	Recuento	417	4109	1411	1381	7318
		% dentro de Nivel escolar	92.9%	79.9%	56.6%	50.0%	67.5%
		Residuo estandarizado	6.6	10.8	-6.6	-11.2	
Mujer	USO DEL CRÉDITO	Recuento	60	1308	1268	1351	3987
		% dentro de Nivel escolar	8.7%	21.7%	42.0%	50.3%	32.1%
		Residuo estandarizado	-10.9	-14.3	9.6	16.7	
	NO USO DEL CRÉDITO	Recuento	630	4720	1748	1333	8431
		% dentro de Nivel escolar	91.3%	78.3%	58.0%	49.7%	67.9%
		Residuo estandarizado	7.5	9.8	-6.6	-11.5	

Fuente: Elaboración propia, a partir del software SPSS versión 23.

El análisis gráfico del resultado obtenido se aprecia en la Figura 19, donde se evidencia que efectivamente el uso del crédito hipotecario es superior en el caso del género masculino y que las principales diferencias radican en los niveles básico y superior entre hombres y mujeres.

Como conclusiones del caso práctico, podemos afirmar que, en efecto, existen aspectos sociodemográficos que tienen una relación de dependencia con el uso o no del crédito hipotecario, con la finalidad de financiar la compra o construcción de la vivienda en México. Específicamente, quedó demostrado que el nivel educativo y el nivel de ingresos tienen relación de dependencia con la variable dependiente estudiada. Se pudo confirmar que la decisión de usar el crédito hipotecario para el financiamiento de la adquisición o construcción del hogar de los mexicanos está asociada a su nivel de ingresos y se comportará según el género de las personas. En este caso se acentúa para las mujeres y se atenúa para los hombres. Por otra parte, en el análisis asociado al nivel de escolaridad, se aprecia que el impacto del género es inverso y favorece a los hombres sobre las mujeres.

Figura 19. Gráfico de barras apiladas: Uso del crédito, Nivel de escolaridad y Sexo



Fuente: Elaboración propia, a partir del software SPSS versión 23.

Conclusiones

Como se ha mostrado a lo largo de este capítulo, la estadística descriptiva es una herramienta imprescindible en cualquier investigación, ya sea de corte cualitativo o cuantitativo. La transformación de los datos brutos, mediante la aplicación de las técnicas descritas, sin importar si provienen de fuentes primarias o secundarias, permite al investigador facilitar a los lectores de su estudio la comprensión de forma significativa de la muestra o población analizada.

En este sentido, el empleo de las diferentes herramientas gráficas descritas permite, de una forma visual y más sencilla, interpretar los valores ocultos en la infinidad de datos recogidos. Por su parte, el empleo de las tablas de contingencia como componente inicial del proceso investigativo mostró su importancia, al permitir establecer mediante su uso la existencia de relaciones de dependencia entre variables implicadas. De igual forma, los diferentes conceptos abordados y la interpretación de los estadísticos aplicados son relevantes para el principal objetivo propuesto en este libro: el apoyo a estudiantes y docentes de las áreas económico administrativas para el desarrollo correcto de estudios científicos a todos los niveles.

Referencias

- Camelo, M., Amaya, J., y Parra, J. (2019). Determinants of the use of mortgage credit by households in Bogota. *Ecos de Economía*, 22(47), 38-57. <https://doi.org/10.17230/ecos.2018.47.2>
- Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press.
- Choonpradub, C., y McNeil, D. (2005). Can the box plot be improved. *Songklanakarin Journal of Science and Technology*, 27(3), 649-657. <https://www.thaiscience.info/Journals/Article/SONG/10986896.pdf>
- Cobb, G. W. y Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104, 801-823. <https://doi.org/10.2307/2975286>
- Constitución Política de los Estados Mexicanos. (2022). <https://www.diputados.gob.mx/LeyesBiblio/pdf/CPEUM.pdf>
- De Rada Igúzquiza, V. D. (2009). *Análisis de datos de encuestas: Desarrollo de una investigación completa utilizando SPSS* (Vol. 137). Editorial UOC.

- Encuesta Nacional de la Vivienda (2020). https://www.inegi.org.mx/contenidos/programas/envi/2020/doc/envi_2020_presentacion.pdf
- Haberman, S. J. (1973). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *The Annals of Statistics*, 1(4), 617-632. <http://www.jstor.org/stable/2958307>
- Herrera, A. (2022). *Determinantes que afectan la tasa de morosidad en los créditos hipotecarios otorgados por el BIESS, periodo 2010-2019* (Bachelor's thesis, PUCE-Quito).
- Instituto Nacional de Estadística y Geografía. (2019). *Encuesta Nacional sobre las Finanzas de los Hogares*. <https://www.inegi.org.mx/programas/enfih/2019/#Microdatos>
- López-Roldán, P., y Fachelli, S. (2015). *Metodología de la investigación social cuantitativa*. Dipòsit Digital de Documents, Universitat Autònoma de Barcelona.
- Minnaard, V., Rabino, C., Garcia, M., Moro, L., y Minnaard, C. L. (2002). El uso de gráficas en la escuela: otro lenguaje de las ciencias. *Revista Iberoamericana de Educación*, 29. ISSN: 1681-5653.
- Organización de Naciones Unidas (2020). *Vivienda y Covid*. <https://onuhabitat.org.mx/index.php/vivienda-y-covid19#:~:text=Se%20estima%20que%20alrededor%20de,asentamientos%20informales%20o%20barrios%20marginales>
- Reynolds, M. R. (1984). Estimating the error in model predictions. *Forest Science*, 30(2), 454-469. <https://doi.org/10.1093/forestscience/30.2.454>
- Sandoval, A. (2014). *Crédito hipotecario como opción de financiamiento para adquirir una vivienda en México*. UNAM.
- Tramontin, J. (2020). Factores determinantes de la demanda de crédito hipotecario en Argentina: un análisis comparativo 2004-2019. https://www.unsam.edu.ar/escuelas/economia/sidi/archivos_posters/TRAMONTINponencia.pdf
- Triola, M. (2000). *Estadística elemental*. Pearson Educación.
- Tukey, J. (1977) *Exploratory Data Analysis*. Addison-Wesley. https://doi.org/10.1007/978-0-387-32833-1_136
- Vasquez, A. y Osorio, A. (2022). *Determinantes para la selección de la alternativa de financiación de vivienda no Vis en Colombia*. Colegio de Estudios Superiores de Administración-CESA. https://repository.cesa.edu.co/bitstream/handle/10726/4532/MBA_1026263321_2022_1.pdf?sequence=6&isAllowed=y
- Viedma, C. (2018). *Estadística descriptiva e inferencial*. IDT.

III. Modelos estadísticos de elección discreta

ALFREDO VALADEZ GARCÍA*

MARTÍN ARTURO RAMÍREZ URQUIDY**

DOI: <https://doi.org/10.52501/cc.131.03>

Resumen

Este capítulo tiene como objetivo desarrollar una presentación, exploración y explicación puntual de la metodología de análisis estadístico que se aplica cuando la variable por explicar es binaria o discreta. El concepto de modelo de elección discreta implica que se analizará una variable no continua, la que por lo regular posee dos únicos valores entendidos como la ocurrencia o no ocurrencia de un evento y que sirven para establecer la relación en términos de probabilidades entre la variable dependiente dicotómica y la variable o variables explicativas, que pueden ser tanto cuantitativas como cualitativas, nominales u ordinales. Para clarificar este ejercicio, se expone un caso práctico mediante datos del Censo Nacional de Población y Vivienda 2020, con vistas a aplicar la regresión logística y así identificar los determinantes que explican el emprendimiento en la frontera norte de México.

Palabras clave: *elección discreta, regresión logística, probabilidad, razones de momios.*

* Doctor en Ciencias Económicas. Profesor de tiempo completo en la Escuela de Administración y Negocios del Centro de Enseñanza Técnica y Superior (Cetys Universidad), campus Tijuana, México. ORCID: <https://orcid.org/0000-0002-2841-4972>

** Doctor en Ciencias Económicas. Profesor de tiempo completo en la Facultad de Economía y Relaciones Internacionales de la Universidad Autónoma de Baja California (UABC), campus Tijuana, México. ORCID: <https://orcid.org/0000-0003-0571-0477>

Introducción a los modelos de elección discreta

El método de interés en este capítulo es la técnica de regresión logística, la cual es parte de la familia de los modelos de elección discreta. Estos modelos se usan cuando la variable explicativa no es continua, sino discreta o binaria. Un ejemplo de lo anterior puede ser un análisis en que usted esté interesado en evaluar mediante un ejercicio econométrico el votar o no votar por cierto partido político o en la ocurrencia de los eventos de deserción o no en la escuela.

Observe cómo en los ejemplos anteriores hay un claro énfasis en que la variable explicativa no puede estar sujeta a un estudio típico de regresión lineal, pues la recta de puntos donde se ajusta el modelo de mínimos cuadrados ordinarios (MCO) no tendría datos continuos que atravesar.

Por lo anterior, y en virtud del alcance, versatilidad y ajuste que permite estimar el comportamiento de una variable dependiente dicotómica en función de una o más variables independientes, además de conservar una capacidad explicativa y predictiva sobre un determinado fenómeno, los modelos de elección discreta son tan empleados en el campo de la investigación de varias disciplinas como economía, sociología, salud pública, finanzas y demografía, entre otras. Esta técnica es particularmente útil para explicar el comportamiento de las variables cualitativas o categóricas, propias de los fenómenos sociales y económicos como el que se está analizando en este capítulo, pues permite modelar decisiones de los individuos, como emprender, en este caso concreto, ya sea en territorio nacional o en Estados Unidos.

Se puede enfatizar en que este tipo de herramientas econométricas pertenece a la familia de los llamados análisis multinivel. Dicho escrutinio es una respuesta a la necesidad de analizar la relación entre los individuos y los diversos contextos en los que se desenvuelven. Las hipótesis de partida de estos modelos nos dicen que los individuos pertenecientes a un mismo contexto tenderán a ser más similares en su comportamiento entre sí, que respecto a su pertenencia a distintos contextos (Andréu, 2011). En este punto, es importante no perder de vista que esta investigación analiza justamente acciones de individuos (emprendedores) en distintos contextos (migratorios, sociodemográficos, económicos).

En concreto, los modelos logit o de regresión logística son conocidos como modelos de elección discreta cualitativa y son vastos en utilidad, ya que sirven para establecer la relación en términos de probabilidades entre la variable dependiente dicotómica y la variable o variables explicativas, que pueden ser cuantitativas o cualitativas, nominales u ordinales (Gujarati y Porter, 2010).

Aunado a lo previo, como refuerzo de la justificación de nuestro modelo, se sabe que el objetivo pilar del análisis multinivel es modelar estadísticamente la influencia que ejercen las variables contextuales sobre las actitudes o los comportamientos medidos en el plano personal.

Desarrollo del modelo matemático de elección discreta

El desarrollo del modelo matemático de elección discreta se apoya en los efectos marginales, los que constituyen una diferencia fundamental respecto a los modelos lineales, pues la influencia de las variables explicativas sobre la probabilidad de elegir la opción dada por $y_i = 1$ (la derivada parcial, $dy_i/dx_i = \beta_k$ en los modelos lineales) no es independiente del vector de características x_i . Una primera aproximación a la relación entre las variables explicativas y la probabilidad resultante es calcular los efectos marginales sobre la variable latente (y^*).

De esa manera, el efecto marginal expresa variaciones en la variable Y , cuando la variable X cambia en una unidad, algo similar a lo que se estima en los estudios de elasticidad de la demanda. Otra medida factible derivada de lo anterior es el Promedio de los efectos marginales (PEM, “Average Marginal Effect” o “AME”), que expresa el efecto promedio de la variable independiente sobre la probabilidad de que suceda la categoría de contraste de la variable dependiente ($y = 1$) (Mood, 2010).

El desarrollo formal propuesto por Hosmer *et al.* (2013) de esta metodología se especifica representando la media condicional de Y dada cierta X , cuando se utiliza una distribución logística. Esto es formalizar el modelo de regresión logística a través de:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

Posterior a ello, si Y se codifica como 0 o 1, la expresión para $\pi(x)$ dada en la ecuación 1 proporciona (para un valor arbitrario de $\beta = (\beta_0, \beta_1)$, el vector de parámetros) la probabilidad condicional de que Y sea igual a 1 dado x . Esto se denota como $\pi(x)$.

De ello se deduce que la cantidad $1 - \pi(x)$ da la probabilidad condicional de que Y sea igual a cero, dado x , $Pr(Y = 0|x)$. Así, para aquellos pares (x_i, y_i) , donde $y_i = 1$, la contribución a la función de verosimilitud es $\pi(x_i)$. Y para aquellos pares donde $y_i = 0$, la contribución a la función de verosimilitud es $1 - \pi(x_i)$, donde la cantidad $\pi(x_i)$ denota el valor de $\pi(x)$ calculado en x_i . Una forma conveniente de expresar la contribución a la función de verosimilitud del par (x_i, y_i) es a través de la expresión:

$$(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2)$$

Partiendo de que las observaciones son independientes, la función de verosimilitud se obtiene como el producto de los términos dados en la ecuación (1.2) de la siguiente manera:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3)$$

El principio de máxima verosimilitud establece como estimación de β el valor que maximiza la expresión en la ecuación 3. Sin embargo, matemáticamente es más fácil trabajar con el logaritmo de la ecuación 3. Esta expresión, la log-verosimilitud, se define como:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (4)$$

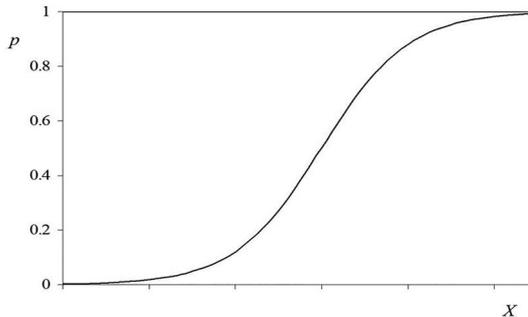
Para encontrar el valor de β que maximiza $L(\beta)$ diferenciamos $L(\beta)$ con respecto a β_0 y β_1 e igualamos a cero las expresiones resultantes. La forma general de estas ecuaciones, conocidas como ecuaciones de probabilidad, es:

$$\sum [y_i - \pi(x_i)] = 0 \quad (5)$$

Cabe destacar que este tipo de regresión lo que arroja son razones de momio (odds ratio) y no coeficientes de interpretación convencional, como los que ofrece un modelo de mínimos cuadrados ordinarios (MCO). Estos valores, más comúnmente planteados en la literatura estadística como odds ratio, se pueden definir como la probabilidad de que ocurra un evento dividido por la probabilidad de que no ocurra. Los odds fluctúan entre 0 (cero) e ∞ (infinito), es decir, únicamente asumen valores positivos, ya que no existen probabilidades negativas.

La regresión logística utiliza los odds ratio, ya que éstos son medidas estandarizadas que permiten comparar la intensidad de influencia que las variables independientes tienen sobre la dependiente. Aun, cuando éstas últimas sean variables ordinales, de escala, continuas o logaritmos, el arrojar resultados en forma de odds ratio permite comparar a las variables independientes, escenario que no siempre se logra con la regresión típica de MCO, como se observa en la siguiente figura:

Figura 1. Estimación gráfica del modelo de regresión logística



Fuente: Ortega *et al.* (2022).

Estos odds, en su definición, son sucesos resultantes de un cociente de probabilidades de ocurrencia entre sus probabilidades de no ocurrencia, bajo unas circunstancias, que alguna literatura denomina c .

$$\text{Odds}_c (\text{evento}) = \frac{P (\text{evento})}{1-P (\text{evento})} \quad (6)$$

Entonces, la medida más relevante para la interpretación del modelo logístico es el valor odds ratio, que es la exponencial del valor B del modelo de regresión de mínimos cuadrados ordinarios (MCO), y se define como el indicador del cambio en los odds resultantes del cambio de una unidad en el predictor.

Cuando la variable predictora es categórica, el odds ratio de ocurrencia de un suceso es fácil de explicar. Para el caso de esta investigación, se busca predecir la probabilidad de que una persona emprenda algún negocio en este lado de la frontera (fronterizo), o bien en territorio fronterizo de Estados Unidos (transfronterizo). Por ello, los odds de ser un emprendedor son resultado de la probabilidad de emprender dividida por la probabilidad de no hacerlo, es decir, quedaría así:

$$\text{Odds}_{\text{emprender}} = \frac{P (\text{emprender})}{P (\text{no emprender})} \quad (7)$$

Para calcular el cambio en el odds resultante del cambio de una unidad en la variable predictora, es necesario en primer lugar, calcular los odds de emprender con presencia de alguna variable independiente (x) y después los odds de emprender con ausencia de esa variable independiente.

Entonces, se puede demostrar que los coeficientes obtenidos en la regresión logística son medidas que cuantifican el riesgo de presentar cierta característica respecto a no presentarla con base en la variable de estudio, de manera que:

$$\text{Exp} (\beta) = \text{OR} \quad (8)$$

Donde β es el coeficiente resultado de la regresión logística asociado a una cierta variable participante en el modelo. Cuando la variable independiente tratada es numérica, este valor se interpreta como el cambio en el riesgo al incrementarse en uno el valor de la variable, mientras que el resto de variables permanecen constantes (Moral, 2016).

Supuestos de los modelos de elección discreta

Field *et al.* (2012) hacen hincapié en que la regresión logística comparte algunos supuestos con la regresión lineal, entre ellos enlistan los tres siguientes:

- *Linealidad*: en regresión lineal se asume que la variable respuesta tiene una relación lineal con las variables predictoras. En regresión logística, la respuesta es categórica y por ello este supuesto se viola. Por tanto, he ahí la razón de usar el modelo logit de los datos. Así, el supuesto de linealidad en regresión logística es que existe una relación lineal entre cada variable predictora continua y el logaritmo de la variable respuesta.
- *Independencia de los errores*: los distintos casos de los datos no deben estar relacionados. Por ejemplo, no se debe medir a la misma gente en diferentes puntos del tiempo. En este aspecto, nuestro estudio si bien es cierto, dado las condiciones de probabilidad de emprender, la población sujeta de estudio no es necesariamente la misma.
- *Multicolinealidad*: aunque no es un supuesto, en su concepto estricto, la multicolinealidad es un problema como en la regresión lineal. Las variables predictoras no deben estar altamente correlacionadas.

Cabe señalar que, en estos escenarios de estudio, otras posibilidades latentes de regresión serían el modelo logit y el probit, puesto que estas metodologías guardan similitudes y discrepancias. De acuerdo con Ucedo Silva (2013), la regresión logit utiliza una función de distribución acumulativa (FDA) de tipo logístico. Mientras, la regresión logística se basa en la suposición de que la variable dependiente categórica refleja una variable subyacente cualitativa (éxito y fracaso) y deduce la función de enlace a partir de la distribución binomial.

Este autor también menciona en su tesis que la regresión probit emplea una FDA de distribución normal. Por ello la literatura también la identifica con el nombre de “normit”. Este tipo de regresión descansa en el supuesto de que la variable dependiente categórica (Y) refleja una distribución subyacente cuantitativa, la cual se vuelve a clasificar para que se transforme en una variable binaria y deduzca la función de enlace de la distribución normal estándar acumulada.

La siguiente tabla muestra para cada función de enlace sus correspondientes funciones de distribución acumulativa (FDA) —las cuales son las funciones inversas de cada uno—, su media y la varianza.

Tabla 1. *Funciones de enlace de respuesta binaria*

Nombre del enlace	Función del enlace	Función de distribución acumulativa	Media	Varianza
Logística	$Logit(p)=\ln(\frac{p}{1-p})$	$\frac{e^{\eta}}{1 + e^{\eta}}$	0	$\frac{\pi}{3}$
Probit	$\Phi^{-1}(p)$	$\Phi(\eta)$	0	1
Log - Log	$Log(-\log(1-p))$	$e^{-e^{\eta}}$	-0.577	$\frac{\pi}{6}$

Fuente: Elaboración propia, con datos de Ucedo Silva (2013).

Entonces, esta técnica econométrica —que corresponde a los modelos de elección discreta— puede aplicarse al censo nacional, pues el instrumento capta las respuestas de los individuos miembros de los hogares censales y permite construir variables dicotómicas, como en este estudio. Los resultados del censo se utilizan para poder determinar las probabilidades de que un individuo residente en la frontera norte de México emprenda un negocio, controlado por determinantes individuales, institucionales y contextuales.

Una vez realizado el modelo de elección discreta, se evaluará con las distintas pruebas de significancia si es estadísticamente significativo. Entre ellas, se destaca la prueba de la curva ROC (*receiver operating characteristic*, por sus siglas en inglés), que sirve para representar la sensibilidad en función de (1 - especificidad). La curva ROC ofrece un mejor resumen de la capacidad predictiva que una tabla de clasificación, porque presenta la potencia predictiva para todos los posibles valores de referencia π_0 (González Ferrer *et al.*, 2015).

Para una especificidad dada (al fijar un valor en el eje de abscisas), la mayor potencia predictiva corresponde a la sensibilidad más elevada (mayor valor en el eje de ordenadas). De ese modo, cuanto mayor sea el área bajo la curva ROC mayor será la potencia de predicción del modelo, es decir, se busca que sea cercana a 1.

Otra prueba eficaz en la post estimación del modelo de regresión logística es la matriz de valores correctamente predichos, en la que se buscará que

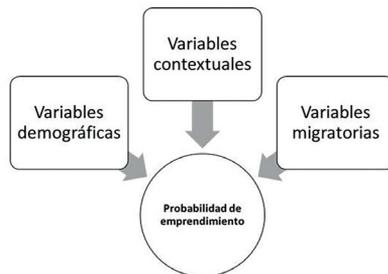
se acerque lo más posible a un 100%. Ello nos señalará que hay poco o nula diferencia entre los valores predichos y los esperados, es decir, que las estimaciones clasificadas por el modelo fueron correctas.

Caso práctico. Análisis de la probabilidad de emprendimiento fronterizo y transfronterizo: el caso de México-Estados Unidos 2020

El estudio desarrollado a continuación consiste en conocer si las condiciones individuales, no pecuniarias y contextuales explican el emprendimiento fronterizo y transfronterizo. Para ello se implementó una metodología de carácter cuantitativo. En este sentido, el objetivo general de este capítulo es mostrar claramente los pasos seguidos para construir las variables y los modelos estadísticos que se utilizarán para dar cuenta del impacto, en términos de probabilidad, que tienen los determinantes seleccionados sobre la decisión de emprender en la frontera de México con Estados Unidos.

A fin de ejemplificar mejor el modelo, nos apoyamos en el siguiente esquema (Figura 2), el cual ilustra las relaciones existentes entre nuestras variables dependientes e independientes.

Figura 2. Representación de la relación entre variables del modelo



Fuente: Elaboración propia.

La figura previa ilustra que en el eje central de nuestro modelo se localiza la probabilidad de emprendimiento, misma que se ve afectada por tres factores: el migratorio, contextual y demográfico. En la sección siguiente de

este trabajo, se detallarán las variables independientes que forman parte de estos tres factores.

Con vistas a conocer lo anterior, se utilizó el Censo de Población y Vivienda 2020, ya que esta fuente de información proporciona datos útiles para identificar, entre otras cuestiones, si el encuestado nació en otra parte del país o del mundo, si ha emprendido algún negocio en su ciudad o en Estados Unidos y demás variables de interés para implementar un modelo de elección discreta.

Profundizando en el instrumento de información, el Censo Nacional de Población y Vivienda 2020 tiene dentro de sus principales características las siguientes (Instituto Nacional de Estadística y Geografía [Inegi], 2020a):

Población objetivo

Para el Cuestionario Básico: Residentes habituales del territorio nacional, las viviendas particulares y colectivas.

Para el Cuestionario Ampliado: Residentes habituales del territorio nacional, las viviendas particulares, y migrantes internacionales.

Unidad censal: Residente habitual y la vivienda.

Temática censal: El Cuestionario Básico se conforma de 38 preguntas, que abordan los siguientes temas:

Características de la vivienda:

- Material en pisos
- Número de dormitorios
- Número de cuartos
- Disponibilidad de energía eléctrica
- Disponibilidad y abastecimiento de agua entubada
- Abastecimiento de agua no entubada
- Equipamiento para almacenar agua
- Disponibilidad de sanitario
- Admisión de agua al sanitario
- Disponibilidad de drenaje

- Bienes electrodomésticos y medios de transporte
- Tecnologías de la información y la comunicación

Características de las personas:

- Sexo
- Edad
- Parentesco
- Lugar de nacimiento
- Autoadcripción afromexicana o afrodescendiente
- Afiliación a servicios de salud
- Religión
- Tipo y grado de discapacidad
- Lengua indígena
- Habla española
- Asistencia escolar
- Escolaridad
- Alfabetismo
- Entidad o país de residencia en 2015
- Municipio de residencia en 2015
- Causa de la migración
- Situación conyugal
- Condición de actividad económica
- Hijas(os) nacidas(os) vivas(os)
- Hijas(os) fallecidas(os)

Variable dependiente

Para este estudio, la variable dependiente (Y) será el emprendimiento, la cual se divide en dos categorías: fronterizo y transfronterizo. La primera hace alusión a quienes han decidido emprender algún negocio en territorio nacional, o mejor dicho, dentro de su entidad de residencia mexicana. En tanto, se entenderá como emprendimiento transfronterizo a quien reside en México y la actividad de iniciar un negocio la lleva a cabo en Estados Unidos. Es decir, este caso práctico de análisis tiene dos variables dependientes.

Siguiendo con lo anterior, Inegi en su glosario de la Encuesta Nacional de Ocupación y Empleo (ENOE), define al trabajador independiente como “la persona que dirige su propia empresa o negocio, de manera que no tiene un jefe o superior a quien rendirle cuentas de su desempeño o de los resultados obtenidos”. En esta misma encuesta se definen dos vocablos relevantes para este estudio: trabajador por cuenta propia y empleador. Al primero lo define: “Es el ocupado que trabaja solo o con el apoyo de integrantes de su propio hogar o ajenos, pero sin el compromiso de pagarles por sus servicios”, y al segundo como “el trabajador independiente que emplea los servicios de uno o varios trabajadores a cambio de una remuneración económica en monetario o especie” (Inegi, 2020b).

Variables independientes

A continuación, se expone la construcción de las variables independientes utilizadas para medir las probabilidades de emprender dentro y fuera de México en un contexto fronterizo.

Migración interna

Esta variable independiente de carácter no pecuniario es considerada para fines de esta investigación una *dummy*, en el sentido de que únicamente toma valores de 0 y 1. Es cero (0) cuando no ha migrado dentro del país en los últimos cinco años, es uno (1) cuando si ha migrado internamente. Para su construcción, se requiere la siguiente pregunta: “¿En qué estado de la República Mexicana o en qué país nació?”, cuyas opciones de respuesta son:

1. Aquí, en este estado
2. En otro estado
3. En los Estados Unidos de América
4. En otro país

Entonces, la opción 2 también se utiliza para la construcción de la variable migración interna a partir de identificar que, en las respuestas de esa segunda opción, esto significa haber nacido en otra entidad federativa. La utilidad de esta variable es para identificar si el haber migrado dentro del

país, le atribuye al individuo alguna probabilidad mayor de emprender respecto al que no ha migrado. La hipótesis de esto es que si fue objeto de la migración, probablemente hay un contexto de necesidad de empleo que orille, por el desconocimiento de la ciudad, a emprender antes de involucrarse en un rol laboral fijo. En los hallazgos, a esta variable se le etiquetará como migración.

Escolaridad acumulada

Una variable de interés para este estudio y una constante en las investigaciones de esta naturaleza son los años de escolaridad, que pertenece al grupo de determinantes individuales. En ese sentido, las fuentes de información referidas consideran la siguiente pregunta: “¿Cuál fue el último año o grado aprobado en la escuela?”. Con valores de 0 al 14, donde 0 es ningún grado académico cursado y 14 es doctorado, se optó por escolaridad acumulada y no grado académico, ya que con esta última variable hay menos opciones de respuesta y se pierde precisión.

Casado

La intención de usar esta variable es segmentar a la población de acuerdo con su situación conyugal, para así crear dos variables: una que contenga a los casados y que viven con su pareja, y otra que aglutina a las cuatro categorías en que no hay una relación: soltero, viudo, divorciado y separado.

Ello nos separará a nuestros sujetos de estudio en dos grupos: uno probablemente más propenso a la necesidad de empleo, como son los casados o que viven con pareja, y otro que asumirá la ausencia de un compromiso. Como es parte del rubro estado civil, esta variable se considerará para efectos de este estudio, un determinante no pecuniario.

Tamaño de localidad de residencia

Esta variable es de carácter contextual en cuanto al tipo de determinante. Se refiere a la forma de agrupar las localidades según el número de personas que las habitan, a través de una variable ordinal que puede tomar los valo-

res del 1 a 4, según la codificación correspondiente que a continuación se exhibe.

El Censo de 2020 considera las siguientes codificaciones:

1. Aquellas localidades con menos de 2 500 habitantes
2. Localidades con 2 500-14 999 habitantes
3. Localidades con 15 000-99 999 habitantes
4. Localidades con 100 000 o más habitantes

La intención de ubicar este componente territorial como variable independiente en nuestro estudio obedece a que, típicamente, es más probable que la población asentada en localidades pequeñas y rurales sea de menores ingresos. En segundo lugar, en esas mismas zonas hay una tendencia mayor a migrar internamente y también hacia Estados Unidos. Ello implica que el tipo de localidad en términos de tamaño puede capturar elementos contextuales que podrían ser importantes para la motivación a emprender o podrían afectar el nivel de emprendedurismo.

Número de hijos sobrevivientes

Las fuentes de información preguntan tanto por el número de hijos vivos que tuvo, como por los sobrevivientes en el momento del censo o encuesta. De este par de opciones se elige la segunda puesto que la primera está sujeta a que, en el transcurso del tiempo, algún o algunos hijos hayan fallecido. De tal manera, el número que interesa es el de los hijos vivos, pues éstos serán de algún modo una responsabilidad que el jefe o jefa del hogar debe atender.

Residencia en otra entidad federativa cinco años antes

Aunque pudiera confundirse con la variable de migración interna, esta variable responde a la inquietud de conocer si cinco años atrás al levantamiento de la información, la persona residía en otra entidad, pero no necesariamente nació en ella. Es decir, puede ser que en 2020 un habitante de Coahuila mencione que en 2015 vivía en Ciudad de México, más no es garantía de que haya nacido en la capital del país. Su construcción emplea la pregunta: “Hace cinco años, en junio de 2015, ¿en qué estado de la República o en qué país vivía?”

El anterior ejemplo claramente aplica para el Censo de Población y Vivienda de 2020, y su respuesta que valida el que haya sido residente en otra entidad federativa cinco años atrás es la opción: “En otro estado”. No se quiso profundizar en la siguiente cuestión dirigida al municipio de residencia en 2015, pues nuestro interés de investigación es por el estado.

Nacido en EUA

La pregunta 6 de la sección III “Características de las personas” (del Censo 2020) interroga sobre la entidad o país de nacimiento de este modo: “¿En qué estado de la República o en qué país nació?”; y las posibles respuestas son:

- a) Aquí en este estado
- b) En otro estado
- c) En los Estados Unidos de América
- d) En otro país

Entonces, si el censado(a) / encuestado(a) menciona la opción C, en automático para efectos de nuestro estudio lo convierte en ciudadano americano.

Esta variable, tanto en su diseño como en sus interpretaciones, debe hacerse con cautela por diversas cuestiones. La primera de ellas es tal vez la más relevante, porque se asume la acción racional de que el individuo hará uso de las ventajas que le confiere el haber nacido en territorio estadounidense.

La segunda es la implicación *per se* de que haber nacido en Estados Unidos de América supondrá más facilidades para emprender un negocio en dicho territorio. Este supuesto descansa en las facilidades que brinda dicho país, donde, si bien es cierto que los inmigrantes tienen ventajas para la apertura de negocios, éstas se facilitan con la ciudadanía norteamericana, pues su condición de nacionalidad le permitirá desempeñarse en un contexto más favorable.

Una vez que fueron conceptualizadas y operacionalizadas las variables dependientes e independientes, se procederá a plantear el modelo formal que se usará para la obtención de resultados. El planteamiento metodológico de la regresión se relaciona con una variable dependiente Y_i y variables

explicativas X_{ki} a través de una función de distribución, de la siguiente manera:

Valores para nuestra variable dependiente Emprendedor:

- 0 = con probabilidad P (no emprender)
- 1 = con probabilidad $1-P$ (emprender)

En la siguiente ecuación se muestra el modelo logit generalizado que define la probabilidad de que ocurra un evento, para objeto y fin de este estudio, el evento de emprender un negocio con una probabilidad $[0,1]$.

$$P_i = E \left(Y_i = \frac{1}{X_i} \right) = \frac{1}{1 + e^{-Z_i}} \quad (9)$$

Resultados

Esta sección presenta los resultados obtenidos con los datos analizados a partir de los modelos de regresión logística, para cada tipo de emprendedor (fronterizo y transfronterizo), con vistas a conocer en conjunto la probabilidad de emprender en la entidad investigada.

Tabla 2. Resultados del modelo de regresión logística sobre emprendimiento fronterizo en 2020

Variable	Razones de momio					
	BC	COAH	CHIH	NL	SON	TAM
Sexo	1.723*	1.838*	3.215*	2.122*	2.199*	2.040*
Escolaridad	1.004*	1.010*	1.018*	1.006*	1.010*	1.009*
Tamaño de localidad	1.052*	1.089*	1.168*	1.075*	1.022*	1.151*
Migrante	0.725*	1.016	1.074	0.718*	0.783*	0.802*
Edad	0.998	1.002*	1.007*	0.996*	1.003*	1.001
Casado	1.069	0.994	1.127*	0.948*	1.006	0.884*
Número de hijos	1.000	0.994	0.942*	0.994	0.976*	0.975*
Entidad foránea	1.131*	1.279*	1.097	1.113*	1.193*	1.175*
Nacido en EUA	0.57*	0.562	0.726	0.614	0.554*	0.623
Pruebas de bondad de ajuste						
Curva ROC	0.56	0.55	0.61	0.54.	0.53	0.58
Predichos +	90.67%	93.01%	93.47%	92.76%	91.48%	91.51%
Hosmer-Lemeshow	0.00	0.18	0.00	0.06	0.79	0.01
Observaciones	3,793,797	3,137,032	3,725,058	5,768,781	2,924,652	3,518,497

Nota: * p valor < 0.05.

Fuente: Elaboración propia, con datos del Censo Nacional de Población y Vivienda 2020 (Inegi, 2020a).

Los resultados del cuadro previo sugieren que el pertenecer al sexo masculino eleva hasta en tres veces la probabilidad de emprender dentro del territorio nacional. De igual forma, la escolaridad es una variable que al aumentar sus valores, es decir, al contar con más grado de estudio, también se incrementa la probabilidad de emprendimiento. Otro hallazgo resulta que el habitar en localidades tipificadas como grandes urbes se asocia con incrementos en el emprendimiento. Una de las variables que disminuye la probabilidad estudiada es la haber nacido en EUA; sin embargo, sólo es significativa en dos entidades.

En cuanto a la validación del modelo de la Tabla 2, podemos apreciar la prueba de Hosmer-Lemeshow, que es otro método para estudiar la bondad de ajuste del modelo de regresión logística. Consiste en comparar los valores previstos (esperados) por el modelo con los valores realmente observados. Ambas distribuciones, esperada y observada, se contrastan mediante una prueba ji al cuadrado. La hipótesis nula del test de Hosmer-Lemeshow es que no hay diferencias entre los valores observados y los valores pronosticados (el rechazo de este test indicaría que el modelo no está bien ajustado). En nuestro escrutinio, solo tres entidades presentaron valores de la prueba superiores a 0.05.

Tabla 3. Resultados del modelo de regresión logística sobre emprendimiento transfronterizo en 2020

Variable	Razones de momio					
	BC	COAH	CHIH	NL	SON	TAM
Sexo	1.348	9*	2.844*	5.088*	2.226*	3.252*
Escolaridad	0.992	1.025	1.027*	1.061	1.021	0.998
Tamaño de localidad	1.192	0.906	1.520*	1.060	0.974	1.340
Migrante	1.057	2.772*	1.714	7.889*	1.833	2.857
Edad	1.004	1.008*	1.007*	1.006	1.015	1.006
Casado	1.540	2.033	2.735*	1.127	1.271	0.955
Número de hijos	0.861	1.065	1.011	1.032	1.012	0.793
Entidad foránea	0.935	3.092*	3.699*	0.720	0.855	3.113
Nacido en EUA	14.577*	44.173*	16.607*	41.261*	72.863*	22.277*
Pruebas de bondad de ajuste						
Curva ROC	0.66	0.86	0.72	0.55	0.74	0.80
Predichos +	99.91%	100%	99.99%	99.99%	99.97%	99.98%
Hosmer-Lemeshow	1	0.31	0.00	0.00	1	1
Observaciones	3,793,797	3,137,032	3,725,058	5,768,781	2,924,652	3,518,497

Nota: * p valor < 0.05.

Fuente: Elaboración propia, con datos del Censo Nacional de Población y Vivienda 2020 (Inegi, 2020a).

La Tabla 3 presenta los hallazgos de la probabilidad de emprendimiento transfronterizo, lo cual ilustra ciertas diferencias con la Tabla 2. Una de las más significativas es que la variable sexo (el ser hombre) presenta razones de momio más elevadas (hasta ocho veces se eleva la probabilidad en entidades como Coahuila). Otra situación claramente diferente es esta última información de resultados en que la variable nacer en EUA incrementa la probabilidad de emprender en territorio de ese país hasta 71 veces (caso Sonora), es decir, presenta más sensibilidad en el emprendimiento transfronterizo que en el fronterizo.

Conclusiones

La incursión de los modelos de elección discreta en las investigaciones y estudios nace como necesidad de estimar efectos sobre una variable dependiente que no es continua, sino antes bien es discreta. En ese sentido, y ante las limitaciones de un modelo de mínimos cuadrados ordinarios, han surgido y cobrado fuerza modelos como la regresión logística.

Una implicación del uso de esta metodología es abordar los términos de probabilidad para cuantificar los efectos de las variables independientes sobre la dependiente. Ello nos lleva a identificar variables protectoras (las que minimizan la probabilidad de ocurrencia) y variables expositoras (las que aumentan la probabilidad de ocurrencia).

El tópico estudiado en este capítulo como caso práctico fue el emprendimiento, que no es un proceso de reciente aparición, ni se trata de un fenómeno de escaso estudio. A lo largo de los últimos años, ha habido una serie importante de esfuerzos para conceptualizarlo, entenderlo, categorizarlo, medirlo y sobre todo encontrar los determinantes asociados a su aparición, por lo que ha sido tema de análisis en muchos países del orbe. La manera en que se estimaron los efectos sobre nuestra variable dependiente, al tratarse de una probabilidad, la utilización de un modelo de elección discreta se asumió como un correcto planteamiento econométrico.

La manera de cuantificar los determinantes del emprendimiento es diversa. Sin embargo, este capítulo centró sus esfuerzos en un modelo de elección discreta, ya que éste permite identificar probabilidades de empre-

der (emprender = 1, no emprender = 0) en territorio fronterizo mexicano o en los estados vecinos al sur de Estados Unidos y no coeficientes u otros estimadores propios de modelos de regresión lineal o alguna otra metodología.

En países como México —donde persiste una notoria heterogeneidad en los 32 estados que lo conforman, pues cada localidad, municipio, entidad federativa y región responden a procesos propios no solo económicos sino sociales—, es importante el estudio de un fenómeno dinámico y evolutivo como el emprendimiento y sus determinantes, asociado no solo al país como un único ente o sitio de estudio, sino al tomar en cuenta las peculiaridades regionales que posee, en específico, la frontera norte con Estados Unidos.

Una lección de esta investigación es que las condiciones propias del sitio de estudio elegido interactúan con las variables seleccionadas, para dar lugar a un análisis segmentado en grupos de interés con hallazgos diversos. En efecto, dado el abordaje a la frontera norte de México y la manera en cómo se comportó la probabilidad de emprender en territorio nacional y en suelo estadounidense, las fuertes ligaduras y cointegración social, cultural y, desde luego, económica, propiciaron que, en muchos casos, se tuvieran odds ratio de dos dígitos.

Dentro de esta diversidad de combinaciones arrojadas por las razones de momio se identificó el perfil de individuo con más probabilidad de emprender. En términos generales, presenta las siguientes condiciones: hombre, soltero y haber nacido en otra entidad, mientras en el caso del emprendimiento transfronterizo, el haber nacido en Estados Unidos. Esto último apuntala a aceptar nuestra hipótesis secundaria, que establece: las probabilidades de emprendimiento fronterizo y transfronterizo son determinadas de manera diferenciada por los distintos factores, entre los que la condición migratoria es el principal diferenciador.

Referencias

Andréu, A. J. (2011). El análisis multinivel: Una revisión actualizada en el ámbito sociológico. *Metodología de Encuestas*, 13, 161-176.

- Field, A., Miles, J. y Field, Z. (2012). *Discovering Statistics Using R*. SAGE.
- González Ferrer, V., Alegret Rodríguez, C. M., González Ferrer, Y., y Moreno Arias, A. (2015). Validación interna de modelo predictivo creado mediante nueva metodología aplicable en la atención primaria de salud. *Medicentro Electrónica*, 19(4), 218-224. <http://scielo.sld.cu/pdf/mdc/v19n4/mdc02415.pdf>
- Gujarati, D. N., y Porter, D. C. (2010). *Econometría* (5ª ed.). McGraw-Hill.
- Hosmer Jr, D. W., Lemeshow, S., y Sturdivant, R. X. (2013). *Applied Logistic Regression* (3ª ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- Instituto Nacional de Estadística y Geografía. (2020a). *Censo de Población y Vivienda 2020*. <https://www.inegi.org.mx/programas/ccpv/2020/>
- Instituto Nacional de Estadística y Geografía. (2020b). *Encuesta Nacional de Ocupación y Empleo*. <https://www.inegi.org.mx/programas/enoe/15ymas/>
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67-82. <https://doi.org/10.1093/esr/jcp006>
- Ortega Páez, E., Ochoa Sangrador, C. y Molina Arias, M. (2022). Regresión logística binaria simple. *Evidencias en Pediatría*, 18(1), 1-9. <https://evidenciasenpediatria.es/articulo/7943/regresion-logistica-binaria-simple>
- Ucedo Silva, V. H. (2013). *Comparación de los modelos logit y probit del análisis multinivel, en el estudio del rendimiento escolar* [Tesis de grado]. Universidad Mayor de San Marcos.

IV. Modelos de ecuaciones estructurales con estimación PLS

DUNIESKY FEITÓ MADRIGAL*
MAGDELIS MORENO ORTEGA**

DOI: <https://doi.org/10.52501/cc.131.04>

Resumen

El presente capítulo tiene como objetivo revisar los antecedentes de los Modelos de Ecuaciones Estructurales (SEM), destacar sus principales características, en especial la simbología empleada para su representación, explicar los tipos de modelos reflexivos y formativos, así como presentar el procedimiento para la modelación de los SEM a partir de sus tres etapas: modelo teórico, modelo de medida y modelo de relaciones estructurales. Además, se explica cómo la aplicación de los SEM con estimación de los Mínimos Cuadrados Parciales permite la formación del binomio SEM-PLS, para convertirse en una metodología de modelación con un enfoque diferenciador y comprensible. A través del capítulo se demuestra la pertinencia de su aplicación en las investigaciones empíricas de las ciencias aplicadas, al ofrecer la posibilidad de estimar de manera simultánea las relaciones de dependencia entre variables independientes y dependientes para evaluar o testear los modelos teóricos de las investigaciones.

Palabras clave: *métodos cuantitativos, ecuaciones estructurales, estimación PLS, modelación.*

* Doctor en Ciencias Económicas. Profesor investigador de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0001-7225-2598>

** Doctorante en Ciencias Administrativas en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-4282-3577>

Introducción a los modelos de ecuaciones estructurales

Los Modelos de Ecuaciones Estructurales (Structural Equation Modeling, SEM por sus siglas en inglés) han revolucionado la aplicación de la estadística en la investigación científica. Éstos permiten probar teóricamente modelos causales (Haenlein y Kaplan, 2004) e incrementar la capacidad explicativa de la comprobación empírica de la teoría. Han demostrado su pertinencia en las investigaciones empíricas de las ciencias aplicadas, al ofrecer la posibilidad de estimar de manera simultánea las relaciones de dependencia entre variables independientes y dependientes a la hora de evaluar o testear los modelos teóricos de las investigaciones. Su aplicación con estimación de los Mínimos Cuadrados Parciales (Partial Least Squares, PLS por sus siglas en inglés) permite la formación del binomio SEM-PLS, que se basa en el análisis de la varianza, a partir de una metodología de modelación con un enfoque diferenciador, más flexible dado que, principalmente en la distribución de los datos, no demanda supuestos paramétricos rigurosos (Martínez Ávila y Fierro Moreno, 2018).

En primer lugar, es importante reconocer que los SEM pertenecen a los modelos estadísticos multivariados y que permiten estimar los efectos y las relaciones entre múltiples variables, en lo que coinciden varios autores como Esposito *et al.* (2010), Gómez Cruz (2011), Cupani (2012) y otros. Estos modelos surgen de la necesidad de otorgar mayor flexibilidad a los modelos de regresión (Bollen, 1989; Kline, 2005) y son considerados como una extensión de varias técnicas multivariadas, entre ellas la regresión múltiple, el análisis factorial y el análisis de senderos (Kahn, 2006; Ruiz *et al.*, 2010). Con la concepción de estos modelos, se realiza la evaluación de las complejas interrelaciones de dependencia entre variables. Además, la peculiaridad de que las variables definidas como dependientes en un determinado tipo de relación pueden ser independientes en otra relación dentro del propio modelo (Sáenz López y Tamez González, 2014) posibilita que el investigador pueda realizar disímiles evaluaciones a su objeto de estudio. Del mismo modo, son añadidos los efectos del error de medida sobre los coeficientes estructurales como parte de la propia definición del modelo (Cupani, 2012).

Se puede afirmar que matemáticamente son modelos complejos de estimar, por la multiplicidad de variables que intervienen a la hora de realizar un análisis.

Antecedentes de los modelos de ecuaciones estructurales

En la literatura especializada, se recogen los que pudieran ubicarse como los antecedentes del SEM, con sus inicios en 1921, cuando Sewall Wright¹ estableció las bases de los análisis de rutas o trayectoria (o *path analysis*), consistentes en una regresión múltiple expresada a través de un diagrama de flujo de la relación interdependiente entre variables (Sáenz López y Tamez González, 2014).

Posteriormente en los años sesenta, se reconoce su potencial como herramienta para analizar datos no experimentales desde las ciencias de la sociología y la psicología a través de los estudios de Blalock (1964), Boundon y Lazarsfeld (1965) y Duncan (1966) de acuerdo con Moneta Pizarro (2017). Luego en 1970, Karl Jöreskog² presentó la primera formulación de un análisis de estructura de covarianza (Covariance Structure Analysis o CSA, por sus siglas en inglés), para realizar la estimación de un sistema de ecuaciones estructurales lineales. En los años de 1970, Jöreskog, Keesling y Wiley unifican el análisis factorial, el análisis de estructuras de covarianza y el modelado de ecuaciones estructurales en un solo modelo, que resultó ser la fórmula más empleada, en este periodo, para desarrollar la estimación de parámetros a partir del ajuste de la matriz de covarianzas. Este modelo fue llamado Lisrel (Lineal Structural Relations, por sus siglas en inglés) o modelo JKW.

¹ Biólogo y estadístico (1889-1988), considerado como el padre de la síntesis neodarwiniana por sus contribuciones a la genética de poblaciones; estableció la base matemática que explica la deriva genética de las especies; desarrolló el concepto de tamaño efectivo de la población; creó la teoría del equilibrio cambiante y el sistema estadístico análisis de rutas o de trayectoria (Casasola Hernández, 2021).

² Estadístico sueco (1935-actual), miembro de la Real Academia Sueca de Ciencias y profesor emérito de la Universidad de Upsala. Es reconocido en la comunidad científica como coautor, junto con Dag Sörbom, del programa estadístico Lisrel. Es uno de los discípulos de Herman Wold y realizó sus estudios en Upsala.

Exactamente, en 1977 Herman Wold³ creó la versión definitiva del algoritmo mínimos cuadrados parciales (Geladi, 1988). La regresión por PLS se emplea para una reducción de dimensiones del conjunto de variables explicativas, de acuerdo con Mateos-Aparicio (2011). Así quedó planteado que la regresión PLS resulta ser un intento para definir variables latentes que contengan la mayor parte de la variación de las variables independientes observadas, y que de esta forma, también puedan emplearse para modelar la variable dependiente.

En la primera etapa de uso del método multivariado SEM, tiempos en los que Lisrel⁴ se convirtió en el programa por excelencia, fueron nombrados Modelos Causales que brindaron la posibilidad de identificar condiciones de causalidad entre variables a partir de las estimaciones de coeficientes de correlación de Pearson o de Spearman, según el caso (Guàrdia Olmos, 2016). Posteriormente, Jöreskog perfeccionó el *software* e incluyó métodos para trabajar con datos transversales, longitudinales, multigrupo y multinivel. De acuerdo con el estudio de Moneta Pizarro (2017), existen otros aportes realizados por investigadores seguidores de Jöreskog, entre ellos Sörbom, que extiende el modelo multigrupo con medias en las variables latentes; Muthén, quien aporta métodos para incluir variables observadas categóricas, y Häggglund, que desarrolló el método de PLS en dos etapas. Por otro lado, en 1992, Quiroga introduce estudios de robustez con correlaciones policóricas para desviaciones del supuesto de normalidad, y en el 1997, Yang-Wallentin propone métodos para estimar relaciones no lineales. Estas son algunas de las contribuciones a la evolución de la técnica en el decursar del tiempo.

En los ochenta, la aplicación de los SEM-PLS se convirtió en un ascensor de rigor científico que llevó las relaciones funcionales propias de la correlación a los niveles de las relaciones causales de las ciencias experimentales (Guàrdia Olmos, 2016). Los avances recientes comprenden extensiones

³ Investigador noruego (1908-1992), reconocido por sus aportes a la economía matemática, vinculados al análisis de series de tiempo, y a la econometría estadística. Contribuyó desde la estadística matemática con el teorema de Cramér-Wold. En cuanto a la microeconomía, realizó aportes a la teoría de la utilidad y a la teoría de la demanda de los consumidores. Una de sus grandes contribuciones fue a la estadística multivariante, con los métodos de mínimos cuadrados parciales (PLS) y modelos gráficos.

⁴ Primer *software* desarrollado para el procesamiento de los SEM.

para datos que provienen de muestras complejas, o sea, modelos lineales generalizados y series de tiempo (Moneta Pizarro, 2017).

Se coincide con autores como Hair *et al.* (2019), Nitzl *et al.* (2016) y Rasoolimanesh y Ali (2018) en que actualmente los SEM son considerados como parte del paquete de herramientas de los investigadores en varias disciplinas para la comprobación de sus análisis empíricos. Además son un método que ha estado en el debate académico durante los últimos años, cuando se ha contado con los aportes de Richter *et al.* (2016) y Rigdon *et al.* (2010).

Características de los modelos de ecuaciones estructurales

Teniendo en cuenta lo referido por Madrigal Moreno (2018), los SEM se determinan a partir de la concepción de modelos de investigación que surgen con la transformación de conceptos teóricos y derivados en variables no observables (latentes) y la conversión de conceptos empíricos en indicadores. Ambos se relacionan entonces a través de hipótesis expresadas gráficamente por diagramas de trayectoria (Haenlein y Kaplan, 2004). De acuerdo con Armijos Biutrón (2020), los SEM tienen la capacidad de estimar y evaluar la relación entre constructos no observables, denominados generalmente variables latentes. Además, se coincide con García Veiga (2011) al exponer que los SEM permiten emplear múltiples medidas que representan el constructo y controlar el error de medición específico de cada variable. Justo ésta es una necesaria aclaración para el investigador a la hora de evaluar la validez de cada constructo medido.

Otro elemento importante en el proceso de interpretación de los resultados del SEM consiste en evaluar de manera exhaustiva las pruebas estadísticas aplicadas y los índices que determinan si la estructura teórica proporciona un buen ajuste de los datos empíricos. Es aquí donde se verifica si los valores de los parámetros estimados reproducen tan estrechamente como sea posible la matriz observada de covarianza (Kahn, 2006; Cupani, 2012).

Del mismo modo, los modelos SEM proporcionan la posibilidad de representar el efecto causal entre variables, pero es importante considerar

que, aunque se puedan estimar los parámetros a estos efectos, el hecho de estimar no demuestra la causalidad. La existencia de cualquier relación causal entre variables debe venir sustentada por la articulación teórica del modelo y no por su estimación con los datos (Ponsoda, 2012). Siendo así, los SEM no prueban causalidad, solo permiten seleccionar hipótesis causales relevantes que eliminan aquéllas no sustentadas por la evidencia empírica. De ese modo, se especifican una serie de relaciones y luego se evalúa cuáles de éstas quedan representadas por los datos.

Se resume, de acuerdo con Haenlein y Kaplan (2004) y Sáenz López y Tamez González (2014), que esta técnica multivariante forma parte de una “segunda generación” y se diferencia de las basadas en regresión en las dos características que se enuncian a continuación:

1. Ofrece la posibilidad de realizar el modelado simultáneo de relaciones entre múltiples constructos independientes y dependientes, que pasan a denominarse variables latentes exógenas y endógenas.
2. Le facilita al investigador construir variables no observables medidas por indicadores, así como estimar el error de las variables observadas.

Otro de los rasgos característicos de estos modelos es que con su representación gráfica han permitido desarrollar la posibilidad de hipotetizar efectos causales entre las variables como lo describe Ruiz *et al.* (2010), al referirse al hecho de que permiten la concatenación de efectos entre variables y las relaciones recíprocas entre ellas. De hecho, son muchos los tipos de modelos que se pueden definir por esta metodología.

Simbología utilizada en los SEM

Dada la importancia de tener un dominio claro en la definición de los símbolos con los que se representan los SEM, se emplea como referencia lo expresado en la literatura especializada, particularmente por Kline (2005) y Manzano Patiño (2017), al tener en cuenta que las variables manifiestas u observables son las que se miden directamente.

Para una mejor comprensión de estos códigos, Dijkstra (2010) refiere que una variable puede ser llamada observable si y sólo si su valor es obte-

nido por medio de un experimento muestral real. Para realizar su representación se emplean las figuras geométricas de cuadrados o rectángulos, y son asignadas por lo general las letras X y Y , tal como se muestra a continuación:

Figura 1. *Representación de las variables observables*

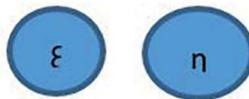


Fuente: Elaboración propia.

Las variables observables, llamadas variables indicadoras, suelen ser ítems de un cuestionario a modo de ejemplificación. En concordancia y de acuerdo con Dijkstra (2010), cualquier variable que no corresponda directamente con algo observable debe ser considerada como no observable. En este sentido, se trata de variables abstractas que representan conceptos unidimensionales, o sea, son las que no pueden ser medidas directamente.

Las variables latentes son constructos de la teoría que solo pueden ser medidos indirectamente a través de variables observables, que son sus manifestaciones. Como estas variables latentes corresponden a conceptos, son variables hipotéticas que varían en su grado de abstracción y terminan siendo variables latentes abstractas creadas en la teoría. Como se muestra en la Figura 2, son empleadas para su representación las figuras de círculos o elipses. Se ha adoptado una simbología específica determinada por las letras griegas ξ y η , según su función en el modelo, ya sea exógena o endógena, respectivamente.

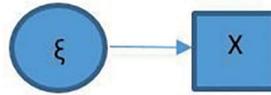
Figura 2. *Representación de variables latentes*



Fuente: Elaboración propia.

Las relaciones entre variables se determinan con flechas como parte de esta representación. Las flechas unidireccionales simbolizan la hipótesis de un efecto directo de una variable sobre otra. El origen de la flecha indica la causa y la punta señala el efecto (Relación de causalidad), como se indica en la Figura 3.

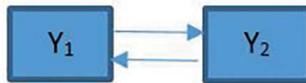
Figura 3. Representación de la relación de causalidad entre variables



Fuente: Elaboración propia.

Existen además las relaciones recíprocas entre las variables. En este caso la relación se representa con dos flechas en la Figura 4 (Relación causal recíproca).

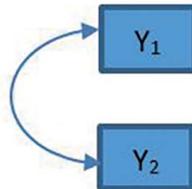
Figura 4. Representación de la relación de causal recíproca entre variables



Fuente: Elaboración propia.

La correlación entre dos variables se representa en la Figura 5 con una flecha curva bidireccional (Relación de covariación).

Figura 5. Representación de la relación de covariación entre variables

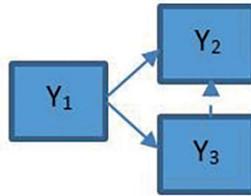


Fuente: Elaboración propia.

También están los efectos no analizados, referidos a las relaciones omitidas, dado por los siguientes motivos: error de especificación o porque sus efectos están incorporados en los términos de error y se manifiesta cuando las variables observables están correlacionadas más allá de lo que el modelo predice, por lo que aparece covarianza entre los errores de medición.

Puede existir la relación espuria entre variables, referida a dos variables que aparentemente no tienen relación y, sin embargo, están altamente relacionadas por una tercera variable que influye en esta relación (Figura 6).

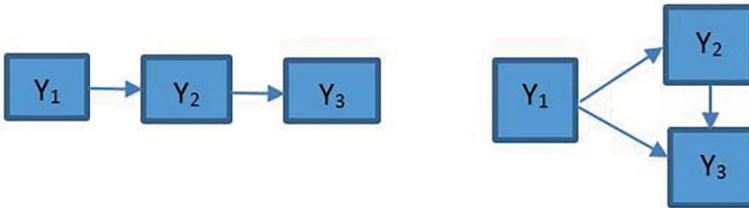
Figura 6. Representación de la relación espuria entre variables



Fuente: Elaboración propia.

Existen también las relaciones causales directas o indirectas, en que la variable tiene una influencia directa o indirecta en las demás variables que conforman el modelo (Figura 7).

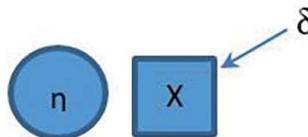
Figura 7. Representación de la relación directa o indirecta entre variables



Fuente: Elaboración propia.

Los errores se representan como variables y pueden representarse, según García Veiga (2011), tanto los errores asociados a la medición de una variable como el conjunto de variables que no han sido contempladas en el modelo y que pueden afectar la medición de una variable observada. Estas variables comparten la característica de las variables latentes, de no ser observadas directamente (Figura 8).

Figura 8. Representación del error como variable

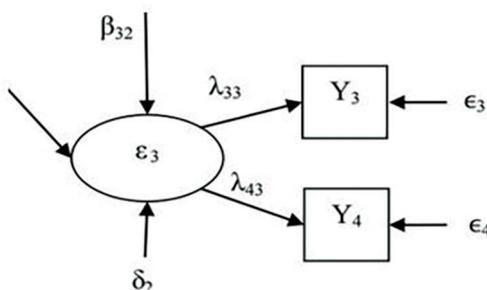


Fuente: Elaboración propia.

La variable exógena, también conocida como independiente, regresora o predictora, es aquella cuyas causas son desconocidas y afectan a otras

variables del modelo, pero no recibe efecto de ninguna otra variable. Según Gómez Cruz (2011), éstas son determinadas fuera del modelo para que influyan en el comportamiento de las variables endógenas. Se caracterizan en los SEM por no recibir impactos, de ahí que de ellas sólo salen flechas, como se representa en la Figura 9.

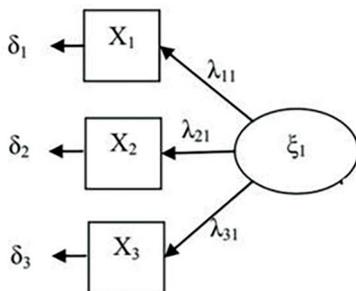
Figura 9. Representación de variable exógena



Fuente: Elaboración propia.

La variable endógena, dependiente o criterio, en general y con base en el modelo econométrico, se caracteriza por ser explicada mediante el funcionamiento del modelo. Las variables endógenas son explicadas por las variables exógenas propuestas (Gómez Cruz, 2011). Éstas se identifican en los SEM por recibir impactos de otras variables. Una regla importante es que toda variable endógena debe ir acompañada de un error, como se representa en la Figura 10.

Figura 10. Representación de variable endógena



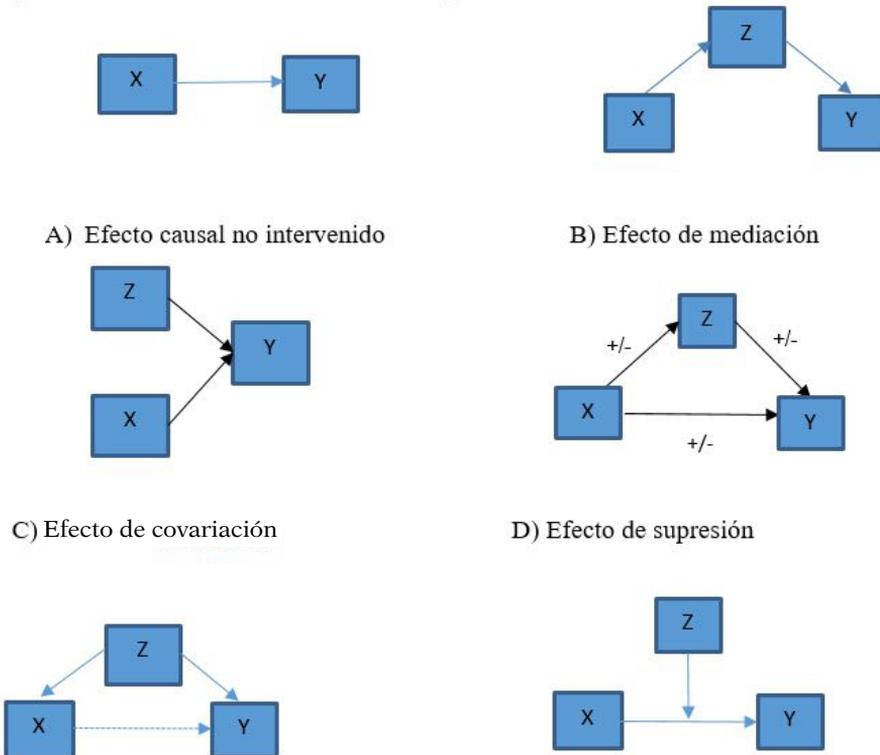
Fuente: Elaboración propia.

Las variables que, a su vez, causan efectos sobre otras endógenas se llaman mediadoras.

En la especificación de los SEM, existe el caso de que una variable tenga doble función como endógena y exógena; justamente se le denomina a esta doble función efecto indirecto o mediador (Gómez Cruz, 2011).

Según Ato y Vallejo (2011), cuando se establece una relación que se asume causal entre una variable independiente (variable X) y una variable de respuesta (variable Y), resulta interesante para el investigador suponer la posible relación con terceras variables (o variables Z). En la Figura 11 se representan a través de diagramas Path (Kenny, 1979) situaciones de la investigación que incorporan esta tercera variable para quedar definidos los efectos según el caso en cuestión.

Figura 11. Efectos de terceras variables en la investigación



Fuente: Elaboración propia.

En el caso B las *variables mediadoras* son aquéllas que explican cómo y por qué se produce un determinado efecto en la relación entre la variable independiente (X) y la dependiente (Y). Pueden ser cualquier variable relacionada con el proceso del objeto de estudio. Por ello, la teoría que sustenta la investigación aporta las variables mediadoras que se pueden incluir en el estudio.

La teoría que respalda la investigación es muy importante para saber qué variables mediadoras se pueden incluir en la investigación, y así las explicaciones teóricas ayuden a una mejor comprensión de esos efectos entre variables.

Las variables de agrupación son variables categóricas, que representan pertenencia de las observaciones a distintas subpoblaciones. Se utilizan generalmente para análisis de comparaciones de grupos. También se les llama variables moderadoras.

Una variable moderadora es la que interfiere, influye y modifica la relación entre la variable independiente y la variable dependiente. De tal modo, el signo y la fuerza de la relación entre ambas variables dependen de la influencia de dicha variable moderadora (Hayes, 2013) (*vid.* Figura 11).

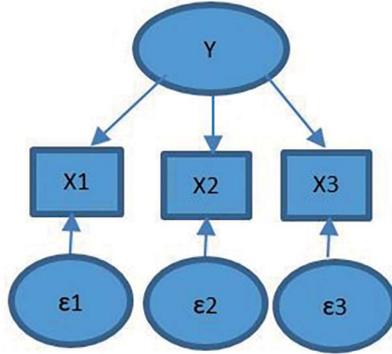
Existe también la relación de mediación moderada, que pertenece a la categoría de lo que Hayes (2013) denomina análisis de los procesos condicionales y supone una combinación de los estudios de la mediación (de los procesos) con los de la moderación. Este análisis sirve como punto de partida para revelar si un efecto mediador se produce de manera diferente en los distintos niveles (valores) de la variable moderadora. Es así como los mecanismos que explican los efectos de una variable independiente sobre una dependiente pueden ser diferentes en función de ciertas variables: las moderadoras.

Modelos reflectivos y modelos formativos

De acuerdo con Sáenz López y Tamez González (2014), se pueden determinar el tipo de variables latentes o las direcciones de las relaciones entre una variable latente y sus indicadores. A continuación, en la Figura 12 y 13 se muestran las características de los dos tipos de indicadores según su direccionalidad:

a) *Indicadores reflectivos*, que dependen de la variable latente.

Figura 12. *Indicador reflectivo de medida*



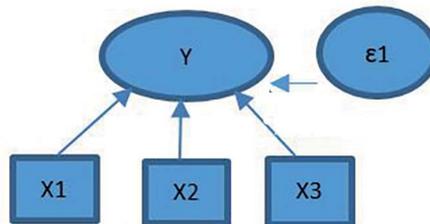
Fuente: Elaboración propia.

El modelo reflectivo se considera como un modelo de medida donde los indicadores de la variable latente son competitivos entre sí y representan manifestaciones de la variable latente. La relación causal va de la variable latente a los indicadores y un cambio en aquélla será reflejado en todos sus indicadores.

En la Figura 12 se muestra que cuando los indicadores (X_1) (X_2) (X_3) dependen de la variable latente, en función de ésta será la respuesta que se obtendrá en cada uno de los indicadores. Los indicadores (X_1) (X_2) (X_3) pasan a ser variables respuestas con un término de error asociado a ellas (ϵ_1) (ϵ_2) (ϵ_3), respectivamente. Son altamente correlacionadas en forma positiva.

b) *Indicadores formativos*, que causan la variable latente.

Figura 13. *Indicador formativo de medida*



Fuente: Elaboración propia.

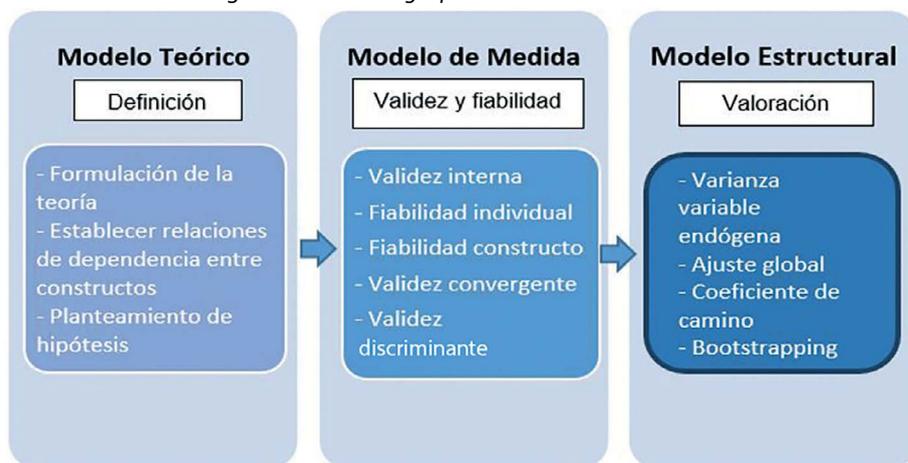
Según Martínez Ávila y Fierro Moreno (2018), las medidas formativas son constructos latentes compuestos por indicadores de medida, en el que éstos son causa o antecedente del constructo (Diamantopoulos y Winklhofer, 2001). Visto así, cada indicador en el modelo formativo representa una dimensión del significado de la variable latente; por lo cual, el eliminar un indicador significa que la variable pierde parte de su significado. Se evidencia así la relevancia de que los indicadores causen el constructo.

La Figura 13 muestra el hecho de cuando los indicadores (X_1) (X_2) (X_3) forman la variable latente; además refleja que la variable latente pasa a ser variable endógena, variable respuesta de los indicadores (X_1) (X_2) (X_3) con término de error asociado a ella (ϵ_1), donde las relaciones pueden ser positivas, negativas o no correlacionadas.

Etapas para la modelación de los SEM

Los Modelos de Ecuaciones Estructurales constan de tres componentes fundamentales: modelo teórico, modelo de medida y modelo de relaciones estructurales. Para una mejor comprensión se explicará su aplicación a través de la metodología descrita mediante el gráfico representado en la Figura 14, que toma como referencia a Ramírez *et al.* (2014).

Figura 14. Metodología para el desarrollo de los SEM



Fuente: Elaboración propia.

Descripción detallada de las fases

Fase 1: Definición del modelo teórico

Los SEM parten de la construcción del modelo teórico donde se realiza la argumentación teórico-científica que sustenta las relaciones de dependencia entre los constructos estudiados. Por lo tanto, este sustento teórico posibilita que el investigador reconozca que los SEM constituyen un método confirmatorio, guiado más por la teoría que por los resultados empíricos. Entonces debe examinarse cada relación propuesta desde esta perspectiva para garantizar que los resultados sean conceptualmente válidos (Hair *et al.*, 2019). Por lo tanto, los SEM pueden ser utilizados para probar los supuestos teóricos con datos empíricos (Haenlein y Kaplan, 2004).

El modelo teórico debe especificar las relaciones que se esperan encontrar entre las variables, para proponer estructuras causales de forma que unas variables causen un efecto sobre otras variables que, a su vez, pueden trasladar estos efectos a otras variables (correlaciones, concatenaciones, efectos directos, efectos indirectos, bucles) que definen los constructos (Ruiz *et al.*, 2010).

Fase 2: Definición del modelo de medida. Validez y fiabilidad

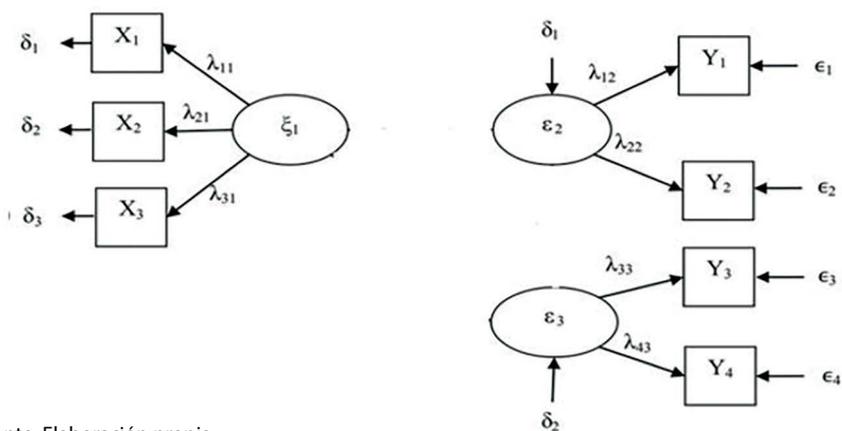
El modelo de medida muestra las relaciones entre indicadores y constructos, o sea, la forma en que cada constructo latente está medido mediante sus indicadores observables y quedan así definidos los errores que afectan a las mediciones y las relaciones de dependencia entre los constructos cuando éstos están relacionados entre sí (Kaplan, 2000). Todo esto se realiza en sintonía con la definición teórica del objeto de estudio. El objetivo fundamental del modelo de medida es corroborar la idoneidad de los indicadores seleccionados en la medición de los constructos de interés, o sea, que el investigador evalúe qué tan bien las variables observadas combinan (cavarían o correlacionan) para identificar el constructo hipotetizado. En el caso de que los indicadores propuestos correlacionen débilmente entre sí, se puede considerar que el investigador ha especificado el modelo errónea-

mente o que hay un desacuerdo en las presuntas relaciones entre las variables (Weston y Gore, 2006).

De acuerdo con Ruiz *et al.* (2010), es necesario considerar que en un modelo completo quedan definidos dos modelos de medida, uno para las variables predictoras y otro para las variables dependientes. A continuación, se exponen en detalles estas especificidades según Ramírez *et al.* (2014): El modelo de medida se define a partir de las variables latentes (VL) y cómo éstas se conforman por indicadores correspondientes a variables observables. Como ya se había visto anteriormente, existen dos tipos de VL. Una VL se define por indicadores reflectivos si éstos son manifestaciones del constructo que representan, luego se establece una relación donde la VL precede al indicador en un sentido causal, y el indicador está en función de este constructo como indicador reflejo. En cambio, la otra VL se define por indicadores formativos cuando el constructo queda expresado como una función de dichos indicadores.

Es decir, el modelo de medida refleja las relaciones existentes entre los constructos (variables latentes) y los indicadores (variables observables). Se permite, de esta forma, evaluar la contribución de cada ítem (reactivo) a la escala de medición, o sea, especificar qué indicadores definen a cada constructo (Ramírez *et al.*, 2014). Otro elemento importante es que valora la fiabilidad de constructos e indicadores. En la Figura 15, se realiza una ejemplificación de este tipo de modelo.

Figura 15. Representación del modelo de medida



Fuente: Elaboración propia.

Un elemento de necesaria consideración es que si una variable no es directamente observable, deben identificarse los indicadores que permitan medirla. Se recomienda formular el modelo en formato gráfico e identificar las ecuaciones y los parámetros. Una vez formulado el modelo, cada parámetro debe ser correctamente identificado y ser derivable de la información contenida en la matriz de varianzas-covarianzas. Existen estrategias para conseguir que todos los parámetros sean identificados (Ruiz *et al.*, 2010).

De acuerdo con Ramírez *et al.* (2014), el primer paso para estudiar las VL con indicadores formativos consiste en verificar su validez interna al descartar problemas de multicolinealidad.⁵ Se procede a calcular el factor de inflación de varianza (VIF) a partir de un análisis de regresión, al considerar que un VIF mayor que 10 revela un problema de multicolinealidad. En consecuencia, es necesario asegurar la validez y fiabilidad de las medidas de los constructos estudiados. Posteriormente, se examina la fiabilidad individual de cada uno de los ítems con el cálculo de las cargas (λ o *loading*), o correlaciones simples de los indicadores pertenecientes al constructo (Kaplan, 2000).

Resulta importante tener en cuenta que no existe consenso al evaluar la fiabilidad individual sobre el valor que debe arrojar esta prueba para ser aceptada. Hay autores como Falk y Miller (1992) que indican que se acepta $\lambda \geq 0.55$, y, por otro lado, Carmines y Zeller (1979) aseguran que se acepta $\lambda \geq 0.707$.

A la hora de evaluar la fiabilidad del constructo, el análisis se puede realizar a partir del coeficiente alfa de Cronbach (CA) y la fiabilidad compuesta del constructo (CR) como medidas de consistencia interna. De acuerdo con Nunnally (1978), para ambos índices se acepta un 0.7 con una fiabilidad modesta en las etapas tempranas de la investigación y un 0.8 para la investigación básica.

Siguiendo los pasos propuestos en la metodología, para evaluar las consistencias internas del modelo se debe analizar la validez convergente. Para realizar este proceso, hay que considerar la Varianza Extraída Media (AVE),

⁵ El problema de la multicolinealidad hace referencia, en concreto, a la existencia de relaciones aproximadamente lineales entre los regresores del modelo, cuando los estimadores obtenidos y la precisión de éstos se ven seriamente afectados. Disponible en: <https://www.uv.es/uriel/material/multicolinealidad3.pdf>.

medida sólo aplicable a las VL conformadas de indicadores reflectivos (Chin, 1998).

Como límite inferior de un AVE aceptable, sugieren Fornell y Larcker (1981) un valor de 0.5, lo que representa que más del 50% de la varianza del constructo se debe a sus indicadores. Por ello, se puede concluir que el ajuste de los indicadores será significativo y éstos estarán altamente correlacionados.

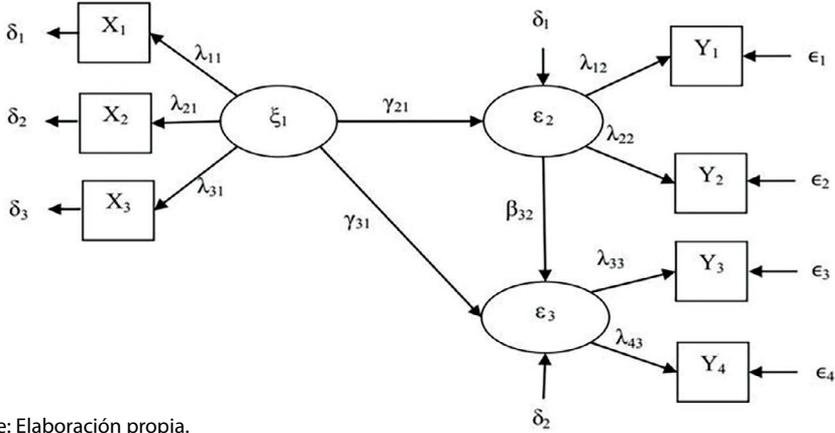
Dentro de la última prueba, para determinar el grado de diferencia de cada VL con las otras VL del modelo, se realiza la medición de la validez discriminante. Y es cuando la raíz cuadrada de la AVE de cada VL es mayor a las correlaciones (r de Pearson) del resto de VL del modelo, lo que permite afirmar que existe en el modelo validez discriminante (Salazar y Ramírez, 2014).

Fase 3: Definición del modelo estructural. Valoración

El modelo estructural es la parte del sistema en donde se definen las relaciones causales entre las variables del modelo. De acuerdo con Ruiz *et al.* (2010), en él se reflejan los efectos y relaciones entre los constructos, los cuales serán normalmente variables latentes. Para representar un modelo causal y las relaciones que se desean incluir en él, se acostumbra a utilizar diagramas similares a los diagramas de flujo, denominados diagramas causales, gráfico de rutas o diagramas estructurales. La representación gráfica del modelo estructural a través de un diagrama del mismo nombre especifica las relaciones manifiestas en el modelo y los parámetros contenidos en él (Figura 16).

En la actualidad, los *softwares* facilitan la definición del modelo al presentarlo a través del interfaz gráfico. A partir del diagrama estructural, el propio programa deriva las ecuaciones del modelo e informa de las restricciones necesarias para que sea completamente identificado. Los diagramas estructurales siguen unas convenciones particulares que son necesarias conocer para poder derivar las ecuaciones correspondientes (Ruiz *et al.*, 2010).

Figura 16. Modelo de ecuaciones estructurales



Fuente: Elaboración propia.

Para realizar la representación matemática del modelo, se tomará como referencia a Gómez Cruz (2011), por lo que la parte estructural de los modelos se denota por:

$$\eta = \beta\eta + \tau\xi + \varepsilon \tag{1}$$

Donde:

η : variables endógenas

ξ = vector de variables latentes independientes

ε = vector de variables latentes dependientes

β = matriz de coeficientes correspondientes a ε

τ = matriz de coeficientes de ξ a ε

δ = vector de errores asociado a ε

Se establecen los supuestos de γ , por no estar correlacionadas los errores con las variables.

Luego para el modelo de medida la ecuación es:

$$Y = \Lambda_y\eta + \varepsilon \tag{2}$$

Y = vector de $q \times 1$ de variables observadas

Λ_y = matriz de $q \times \eta$ de coeficientes

η : variables endógenas

ε = vector de $\eta \times 1$ de variables latentes

El planteamiento del modelo estructural, presentado en la Figura 16, con matrices es:

$$\begin{bmatrix} \eta_2 \\ \eta_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \gamma_{21} \\ \gamma_{31} \end{bmatrix} (\xi_1) + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \quad (3)$$

Y para el modelo de medida es:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} [\xi_1] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} = \begin{bmatrix} \lambda_{12} & 0 \\ \lambda_{22} & 0 \\ 0 & \lambda_{33} \\ 0 & \lambda_{43} \end{bmatrix} \begin{bmatrix} \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \quad (5)$$

Cuando el resultado de la estimación es inadmisibles, es decir, hay parámetros estimados con valores ilógicos, como varianzas negativas o correlaciones entre un factor y un indicador con valor absoluto mayor a 1, puede ser causado por:

- a) errores en la especificación del modelo.
- b) no identificación del modelo.
- c) presencia de valores atípicos o una combinación de tamaño de muestra pequeño y sólo dos indicadores por factor en un modelo de medida.
- d) valores de inicio inadecuados o correlaciones extremadamente altas o bajas que resultan empíricas sobre identificaciones (Kline, 2005).

De acuerdo con Ramírez *et al.* (2014), en esta fase, se debe analizar si ciertamente la cantidad de la varianza de la variable endógena es explicada por los constructos que la predicen en un primer momento, donde se utiliza el valor de la varianza explicada. Entonces R_2 es representativo de esta

varianza explicada. En el logro de que esta varianza sea suficientemente explicada por las variables independientes, se debe examinar que el R_2 deba ser ≥ 0.1 .

Posteriormente, para determinar el ajuste del modelo se procede a realizar el cálculo del índice de ajuste global (GoF). Para ello se multiplica la raíz cuadrada del promedio de AVE por la raíz cuadrada del promedio de R_2 . De esa manera, para que quede comprobada la confiabilidad y ajuste del modelo, el índice GoF debe ser ≥ 0.5 .

Luego, son las hipótesis las que deben ser contrastadas mediante el examen de los coeficientes de camino (β) y el análisis de bootstrapping. En el primero caso, se permite conocer si las variables predictoras favorecen a la varianza explicada de la variable endógena. Los pesos de regresión estandarizados son representados por los valores β . Los $\beta \geq 0.2$ son considerados significativos, aun cuando el valor ideal que se espera es $\beta \geq 0.3$, definido por Chin (1998).

El análisis de bootstrapping determina la estabilidad de las estimaciones ofrecidas por el análisis PLS (Chin, 1998). De acuerdo con este autor, se debe calcular la distribución T de Student de dos colas con $n-1$ grados de libertad, donde n es el número de submuestras con niveles de significación de $*p < 0.05$, $** p < 0.01$, y $*** p < 0.001$, dado que se debe aplicar un remuestreo a partir de los datos de la investigación.

Los resultados del análisis de bootstrapping deben ser comparados con el valor T de Student. Posteriormente se examina si el valor β entre ellas es mayor o igual a 0.2, para poder concluir que existe una relación causal entre dos VL del modelo y además es significativa estadísticamente.

Caso práctico. Factores que inciden en la satisfacción residencial en México: una estimación a partir de microdatos de vivienda y la modelación PLS

Fuente de información

La fuente de información provino de la Encuesta Nacional de Vivienda (ENVI), la cual fue aplicada entre los meses de octubre a diciembre del año

2020 en México. Esta encuesta tiene como objetivo generar información estadística sobre las características de las viviendas en México, así como conocer con mayor exactitud las necesidades y demandas de la población al respecto. Señalar que, a diferencia de la aplicación anterior, en 2020 se profundizan aspectos relacionados con la percepción de los integrantes con la calidad de la vivienda, la satisfacción con el entorno, los impactos del covid-19, entre otros temas de interés (Instituto Nacional de Estadística y Geografía [Inegi], 2020).

Población y muestra

Para el análisis de la población se tomó en cuenta la base de microdatos de la ENVI. En este caso se trabajaron un total de 55 147 registros, los que posterior a un proceso de depuración en función de las variables de interés arrojaron un total de 10 615 viviendas.

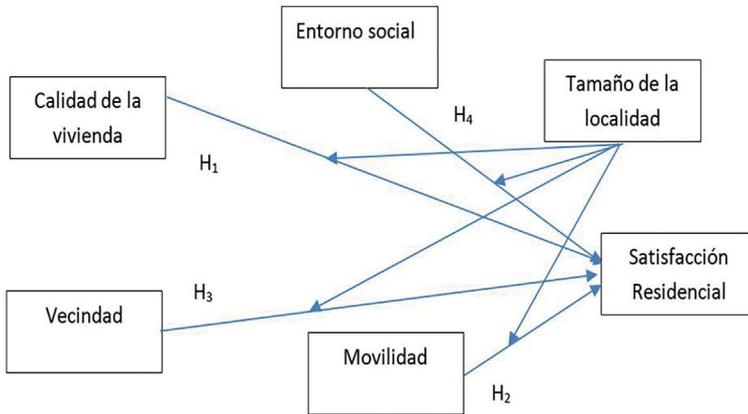
Modelo teórico y operacionalización de variables

Tal y como hemos abordado en el desarrollo de este capítulo, el primer paso en la estimación de ecuaciones estructurales es la identificación del modelo teórico. Para el estudio se retoman los planteamientos de Francescato *et al.* (1989), quienes abordan la satisfacción residencial como un fenómeno multinivel que impacta directamente en la calidad de vida de los individuos y en el cual influyen factores físicos, sociales y vecinales. Estos argumentos son reforzados por autores como Amérigo y Argonés (1997), los que enfatizan que para estudiar la satisfacción residencial no solo hay que analizar la calidad de la vivienda, sino también otros factores a los que denominan el barrio, los vecinos y el medio social. De acuerdo con Chapman y Lombard (2006), esta denominación de barrio engloba las viviendas y su conexión con sistemas físicos, sociales y económicos.

Con base en lo anterior, la propuesta de modelo teórico (Figura 17) propone que la satisfacción residencial sea explicada por el grado de satisfacción con la calidad de la vivienda (Calidad de la vivienda), la satisfacción

con el entorno social (Entorno social), la satisfacción con la movilidad social (Movilidad) y la satisfacción con la convivencia vecinal (Vecindad). Al mismo tiempo, se busca explicar si el lugar donde habitan las personas (localidades urbanas o rurales) influye en las relaciones causales incluidas en el modelo teórico.

Figura 17. Modelo teórico de investigación



Fuente: Elaboración propia.

Derivadas del modelo de investigación se plantearon las siguientes hipótesis:

- H_1 = La satisfacción con la calidad de la vivienda se relaciona positivamente con la satisfacción residencial.
- H_2 = La satisfacción con la movilidad social se relaciona positivamente con la satisfacción residencial.
- H_3 = La satisfacción con la convivencia vecinal se relaciona positivamente con la satisfacción residencial.
- H_4 = Los entornos sociales adversos se relacionan negativamente con la satisfacción residencial.

Para la definición de las variables Calidad de la vivienda, Movilidad, Entorno social, Movilidad y Satisfacción residencial fueron seleccionados en la ENVI un conjunto de indicadores que pretenden explicar los constructos en el modelo de investigación (Tabla 1).

Tabla 1. *Operacionalización de las variables latentes y sus respectivos indicadores*

VARIABLES LATENTES	DESCRIPCIÓN	ESCALA DE MEDIDA
Calidad de la vivienda (exógena)	X ₁ Pensado en esta vivienda, dígame, ¿qué tan satisfechos están con la calidad del piso?	1 - Nada satisfechos 2 - Poco satisfechos 3 - Algo satisfechos 4 - Muy satisfechos
	X ₂ Pensado en esta vivienda, dígame, ¿qué tan satisfechos están con la calidad de los muros y techos?	
	X ₃ Pensado en esta vivienda, dígame, ¿qué tan satisfechos están con el tamaño de los cuartos para dormir?	
	X ₄ Pensado en esta vivienda, dígame, ¿qué tan satisfechos están con el tamaño de los baños?	
	X ₅ Pensado en esta vivienda, dígame, ¿qué tan satisfechos están con el tamaño de la cocina?	
Movilidad social (exógena)	X ₆ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y el trabajo?	1 - Nada satisfechos 2 - Poco satisfechos 3 - Algo satisfechos 4 - Muy satisfechos
	X ₇ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y los centros escolares?	
	X ₈ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y los centros de salud?	
	X ₉ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y mercados o centros comerciales?	
	X ₁₀ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y parques o espacios deportivos?	
Vecindad (exógena)	X ₁₁ ¿Qué tan satisfechos están con la “distancia-tiempo” entre esta vivienda y los centros de recreación o instalaciones culturales?	1 - Nada 2 - Poco 3 - Algo 4 - Mucho
	X ₁₂ Con relación a sus vecinos, ¿qué tanto se conocen entre ustedes?	
	X ₁₃ Con relación a sus vecinos, ¿qué tanto cooperan entre todos?	
	X ₁₄ Con relación a sus vecinos, ¿qué tanto se puede confiar en ellos?	
	X ₁₅ Con relación a sus vecinos, ¿qué tanto se organizan para resolver asuntos de la colonia o barrio (localidad)?	
Entorno social (exógena)	X ₁₆ En su colonia o barrio (localidad), ¿qué tanto problema tienen con la basura tirada en las calles?	1 - Nada 2 - Poco 3 - Algo 4 - Mucho
	X ₁₇ En su colonia o barrio (localidad), ¿qué tanto problema tienen con el deterioro por abandono de viviendas o espacios públicos?	
	X ₁₈ En su colonia o barrio (localidad), ¿qué tanto problema tienen con el vandalismo, grafitis, pintas, vidrios rotos, entre otros?	
	X ₁₉ En su colonia o barrio (localidad), ¿qué tanto problema tienen con los robos o asaltos?	
Satisfacción residencial (endógena)	Y ₁ En una escala del 0 al 10, ¿qué tan satisfechos están con su colonia o barrio (localidad)?	00 - 10
Tamaño de la localidad (variable moderadora)	1_Localidades urbanas 2_Localidades rurales	1 - (≥ 15,999 habitantes) 2 - (< 15,999 habitantes)

Fuente: Elaboración propia.

Por último, en la Tabla 2 se establecieron las ecuaciones lineales que definen las relaciones entre las variables latentes del modelo de investigación.

Tabla 2. Ecuaciones lineales del modelo de investigación

Indicadores exógenos			
ξ 1 (Calidad de la vivienda)	ξ 2 (Movilidad)	ξ 3 (Vecindad)	ξ 4 (Entorno social)
$X_1 = \lambda^x_{11} \xi_1 + \delta_1$	$X_6 = \lambda^x_{62} \xi_2 + \delta_6$	$X_{12} = \lambda^x_{123} \xi_3 + \delta_{12}$	$X_{16} = \lambda^x_{164} \xi_4 + \delta_{16}$
$X_2 = \lambda^x_{21} \xi_1 + \delta_2$	$X_7 = \lambda^x_{72} \xi_2 + \delta_7$	$X_{13} = \lambda^x_{133} \xi_3 + \delta_{13}$	$X_{17} = \lambda^x_{174} \xi_4 + \delta_{17}$
$X_3 = \lambda^x_{31} \xi_1 + \delta_3$	$X_8 = \lambda^x_{82} \xi_2 + \delta_8$	$X_{14} = \lambda^x_{143} \xi_3 + \delta_{14}$	$X_{18} = \lambda^x_{184} \xi_4 + \delta_{18}$
$X_4 = \lambda^x_{41} \xi_1 + \delta_4$	$X_9 = \lambda^x_{92} \xi_2 + \delta_9$	$X_{15} = \lambda^x_{153} \xi_3 + \delta_{15}$	$X_{19} = \lambda^x_{194} \xi_4 + \delta_{19}$
$X_5 = \lambda^x_{51} \xi_1 + \delta_5$	$X_{10} = \lambda^x_{102} \xi_2 + \delta_{10}$		
	$X_{11} = \lambda^x_{112} \xi_2 + \delta_{11}$		
Indicadores endógenos			
η_1 (satisfacción residencial)			
$Y_1 = \lambda^y_{11} \eta_1 + \epsilon_1$			
Modelo de relaciones			
$\eta_1 = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \gamma_4 \xi_4 + \zeta_1$			

Fuente: Elaboración propia.

Resultados

El apartado de resultados se divide en dos partes. En la primera se analiza la confiabilidad y validez del modelo de medida, y en la segunda se aborda la valoración del modelo estructural y se corroboran las hipótesis planteadas con anterioridad.

Con relación a la validez convergente, la Tabla 3 muestra las estimaciones de la varianza media extraída (AVE) y la confiabilidad compuesta (CR) para las variables del modelo. Tal y como se puede apreciar, los valores de la AVE superan el umbral establecido (> 0.5), lo que corrobora que las variables latentes explican en todos los casos al menos el 50% de la varianza de los indicadores que lo conforman. Por su parte, el test de confiabilidad compuesta y las cargas factoriales de los indicadores (Tabla 4) superan el valor objetivo de 0.70, lo que garantiza la consistencia interna de los indicadores en el modelo de medida.

Tabla 3. *Confiabilidad y validez del modelo de medida*

Variables	Alfa de Cronbach	rho_A	Composite Reliability (CR)	AVE
Calidad de la vivienda	0.83	0.84	0.88	0.60
Movilidad social	0.83	0.83	0.87	0.55
Vecindad	0.76	0.79	0.85	0.58
Entorno social	0.72	0.74	0.83	0.55
Satisfacción residencial	1.00	1.00	1.00	1.00

Fuente: SmartPLS, 2023.

Tabla 4. *Carga factorial de los indicadores del modelo de medida*

Indicador	Calidad de la vivienda	Movilidad social	Vecindad	Entorno social	Satisfacción residencial
X ₁	0.723				
X ₂	0.739				
X ₃	0.806				
X ₄	0.804				
X ₅	0.798				
X ₆		0.833			
X ₇		0.882			
X ₈		0.752			
X ₉		0.793			
X ₁₀		0.799			
X ₁₁		0.757			
X ₁₂			0.747		
X ₁₃			0.810		
X ₁₄			0.831		
X ₁₅			0.754		
X ₁₆				0.884	
X ₁₇				0.734	
X ₁₈				0.830	
X ₁₉				0.792	
Y ₁					1.00

Fuente: SmartPLS, 2023.

Adicionalmente al análisis anterior, en la Tabla 5 se muestran los resultados de la validez discriminante del modelo de medida. Al respecto, los valores individuales de la raíz cuadrada de la varianza media extraída de los constructos (diagonal principal) son superiores a los valores compartidos con el resto de los constructos, lo que permite validar el efecto diferenciador de las variables del modelo.

Tabla 5. *Análisis de validez discriminante*

	Calidad de la vivienda	Movilidad social	Vecindad	Entorno social	Satisfacción residencial
Calidad de la vivienda	0.755				
Movilidad social	0.439	0.739			
Vecindad	0.115	0.063	0.764		
Entorno social	-0.122	-0.049	-0.147	0.739	
Satisfacción residencial	0.253	0.245	0.287	-0.348	1

Fuente: SmartPLS, 2023.

Asimismo, como complemento del análisis discriminante, la Tabla 6 muestra las cargas factoriales cruzadas de los indicadores en sus variables latentes y su comparativo con el resto de las variables. En este caso, se puede apreciar que las cargas factoriales tienen un mayor valor con su propia variable latente que con el resto de las variables del modelo.

Tabla 6. *Cargas factoriales cruzadas*

Indicador	Calidad de la vivienda	Movilidad social	Vecindad	Entorno social	Satisfacción residencial
X ₁	0.723	0.329	0.077	-0.083	0.180
X ₂	0.739	0.348	0.074	-0.075	0.187
X ₃	0.806	0.318	0.113	-0.119	0.212
X ₄	0.804	0.360	0.088	-0.101	0.200
X ₅	0.798	0.350	0.089	-0.093	0.200
X ₆	0.336	0.833	0.069	-0.093	0.171
X ₇	0.295	0.882	0.070	-0.048	0.173
X ₈	0.311	0.752	0.068	-0.081	0.183
X ₉	0.365	0.793	-0.008	0.020	0.168
X ₁₀	0.325	0.799	0.036	-0.030	0.202
X ₁₁	0.314	0.757	0.042	-0.019	0.185
X ₁₂	0.024	-0.024	0.747	-0.083	0.154
X ₁₃	0.089	0.050	0.810	-0.089	0.209
X ₁₄	0.149	0.113	0.831	-0.153	0.280
X ₁₅	0.056	0.014	0.754	-0.110	0.208
X ₁₆	-0.128	-0.113	-0.137	0.884	-0.260
X ₁₇	-0.088	-0.025	-0.089	0.734	-0.195
X ₁₈	-0.079	-0.006	-0.010	0.830	-0.281
X ₁₉	-0.071	-0.004	-0.098	0.792	-0.280
X ₂₀	0.253	0.245	0.280	-0.348	1.00

Fuente: SmartPLS, 2023.

Finalizada la evaluación de la validez y fiabilidad del modelo de medida, pasamos a comentar los resultados del modelo estructural. En este sentido, debemos analizar en primer lugar la magnitud, el signo algebraico y el grado de significación de los coeficientes de regresión estandarizados, también conocidos como coeficientes path.

Como se puede observar en la Tabla 7, se encontró que existe una relación positiva moderada y estadísticamente significativa entre el constructo Vecindad y la Satisfacción de la vivienda (0.220), una relación negativa moderada y significativa entre el constructo Entorno social y la Satisfacción de la vivienda (-0.292). Del mismo modo, aunque en menor magnitud, se encontró que la Movilidad social tiene impactos positivos y significativos sobre la Satisfacción residencial (0.164), al igual que con la Calidad de la vivienda (0.120). Es importante señalar que, sin el afán de realizar un abordaje a mayor profundidad en los resultados, la evidencia empírica demuestra que en el caso mexicano la Satisfacción residencial es explicada en mayor medida por factores conocidos en la literatura como subjetivos (entorno social, vecindad) y no tanto por factores objetivos como Calidad de la vivienda y Movilidad social.

Tabla 7. Modelo estructural a partir del método bootstrapping en SmartPLS

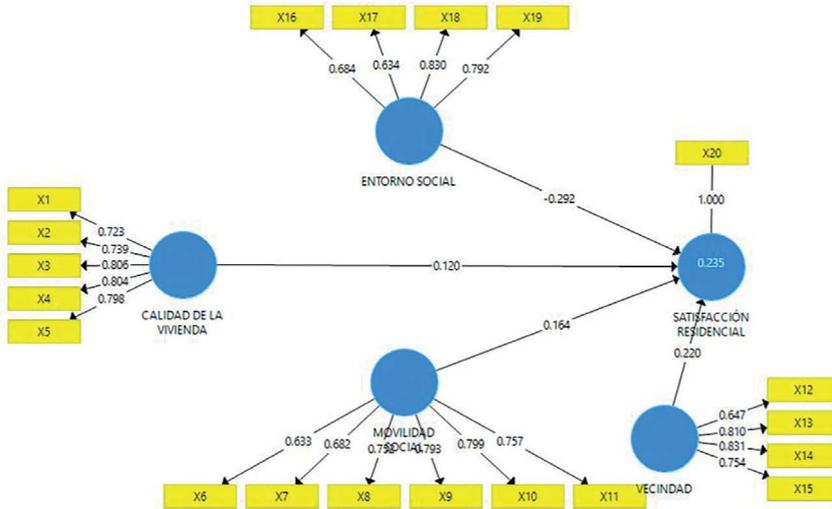
	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistic (O/STDEV)	P Values	Hipótesis
Calidad de la vivienda > Satisfacción residencial	0.120	0.120	0.011	11.363	0.000**	Aceptada
Entorno social > Satisfacción residencial	-0.292	-0.293	0.009	32.233	0.000**	Aceptada
Movilidad social > Satisfacción residencial	0.164	0.165	0.010	15.997	0.000**	Aceptada
Vecindad > Satisfacción residencial	0.220	0.220	0.009	25.451	0.000**	Aceptada

Nota: Niveles de significación: ** $P < 0.05$.

Fuente: SmartPLS, 2023.

Aunado a lo anterior, se muestra la validación del modelo estructural en la Figura 18:

Figura 18. Modelo estructural validado



Nota. Se obtuvo la validación del modelo a través del software SmartPls 3.

Fuente: Elaboración propia.

Efecto moderador de la variable tamaño de la localidad en el modelo estructural

En este apartado se comenta el efecto de la variable moderadora en el modelo estructural. A propósito, los resultados de la prueba de permutación (5 000 permutaciones) muestran diferencias significativas entre las localidades urbanas (grupo 1) y las localidades rurales (grupo 2). Como se puede apreciar en la Tabla 8, las relaciones causales Calidad de la vivienda, Entorno social y Movilidad social con la Satisfacción residencial son diferentes entre las localidades urbanas y rurales (p -Values < 0.05). En un análisis detallado de los efectos se visualiza un mayor valor de los coeficientes path en el grupo 2 relacionado con Calidad de la vivienda-Satisfacción residencial (0.173) y el Entorno social-Satisfacción residencial (-0.24). Por el contrario, los estadísticos reflejan que en el caso de la Movilidad social y la Satisfacción residencial, el coeficiente path del grupo 2 (0.107) es inferior al grupo 1 (0.179), lo que pudiera ser explicado por la carencia de accesos a los sistemas físicos, sociales y económicos que persisten en las localidades rurales de México.

Tabla 8. Resultados de la prueba permutation multigrupo

	Path Coef Original (GROUP_ TLOC (1.0)	Path Coef. Original (GROUP_ TLOC (2.0)	Path Coef. Original Difference GROUP (1.0) - (2.0)	Path Coef. Permutation Difference GROUP (1.0) - (2.0)	2.50%	97.50%	Permutation p-Values
Calidad de la vivienda > Satisfacción residencial	0.096	0.173	-0.077	0	-0.054	0.054	0.005**
Entorno social > Satisfacción residencial	-0.322	-0.24	-0.082	0	-0.047	0.046	0.00**
Movilidad social > Satisfacción residencial	0.179	0.107	0.073	-0.001	-0.049	0.048	0.002**
Vecindad > Satisfacción residencial	0.207	0.222	-0.015	0	-0.043	0.042	0.489

Nota: Niveles de significación: ** $P < 0.05$.

Fuente: SmartPLS, 2023.

Conclusiones

A modo de resumen, los Modelos de Ecuaciones Estructurales son técnicas avanzadas de segunda generación y de gran utilidad para el desarrollo de la investigación científica, en especial en la rama de las ciencias sociales. Su utilidad se centra en la construcción de modelos teóricos que posibilitan una mejor comprensión de la realidad y de los complejos fenómenos observados. Asimismo, determinan en qué medida los datos apoyan un modelo teórico de múltiples relaciones de dependencia, al comprobar posibles efectos (directos o indirectos) que no pueden ser observados ni medidos directamente. Del mismo modo, su potencia y versatilidad en su confección y validación a través de diferentes índices de ajuste incrementan sin lugar a dudas la capacidad explicativa de la comprobación empírica de la teoría.

Referencias

- Amérigo, M., y Aragonés, J. I. (1997). Residential satisfaction in council housing. *Journal of Environmental Psychology*, 10(4), 313-325. [https://doi.org/10.1016/S0272-4944\(05\)80031-3](https://doi.org/10.1016/S0272-4944(05)80031-3)
- Armijos Biutrón, V. A. (2020). *La comunicación por acción y la Responsabilidad Social Corporativa desde el enfoque de su influencia en la imagen y reputación: análisis de la banca privada ecuatoriana integrada en la RED CERES* [Tesis de doctorado]. Universidad Nacional de Rosario, México.
- Ato, M., y Vallejo, G. (2011). Los efectos de terceras variables en la investigación psicológica. *Anales de Psicología/Annals of Psychology*, 27(2), 550-561. <https://revistas.um.es/analesps/article/view/123201>
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
- Carmines, E. G., y Zeller, R. A. (1979). *Reliability and Validity Assessment*. Sage Publications.
- Casasola Hernández, L. (2021). *Sewall Wright: biografía de este genetista estadounidense*. <https://psicologiaymente.com/biografias/sewall-wright>.
- Chapman, D. W., y Lombard, J. R. (2006). Determinants of neighborhood satisfaction in fee-based gated and nongated communities. *Urban Affairs Review*, 41(6), 769-799. <https://doi.org/10.1177/1078087406287164>
- Chin, W. W. (1998). The partial least squares approach for structural equation modeling. En G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Lawrence Erlbaum Associates.
- Cupani, M. (2012). Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*, 2, 186-199. <http://hdl.handle.net/11086/22039>
- Diamantopoulos A., y Winklhofer H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38(2), 269-277. <https://doi.org/10.1509/jmkr.38.2.269.18845>
- Dijkstra, T. K. (2010). Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares. En V. Esposito Vinzi, W. W. Chin, J. Henseler y H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, vol. II)* (pp. 23-46). Springer.
- Esposito V., Trinchera L., y Amato S. (2010). PLS Path Modeling: From Foundation to Recent Developments and Open Issues for Model Assessment and Improvement. En V. Esposito Vinzi, W. W. Chin, J. Henseler y H. Wang (Eds.), *Handbook of Partial Least Squares* (pp. 47-82). Springer.
- Falk, R. F., y Miller, N. B. (1992). *A primer for soft modeling*. University of Akron Press.
- Fornell, C., y Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 18(3), 382-388. <https://doi.org/10.1177/002224378101800313>
- Francescato, G., Weidemman, S., y Anderson, J. R. (1989). Evaluating the built environ-

- ment from the users point of view: An attitudinal model. En W. F. E. Preiser (Ed.), *Building Evaluation* (pp. 181-198). Springer.
- García Veiga, M. (2011). *Análisis causal con ecuaciones estructurales de la satisfacción ciudadana con los servicios municipales* [Proyecto en opción a Máster, Universidad de Santiago de Compostela]. Minerva Repositorio Institucional DA USC.
- Geladi, P. (1988). Notes on the history and nature of partial least squares (PLS) modeling. *Journal of Chemometrics*, 2(4), 231-246. <http://doi.org/10.1002/cem.1180020403>
- Gómez Cruz, M. (2011). *Estimación de los modelos de ecuaciones estructurales, del índice mexicano de la satisfacción del usuario de programas sociales mexicanos, con la metodología de mínimos cuadrados parciales*. [Tesis de maestría, Universidad Iberoamericana]. Repositorio IBERO.
- Guàrdia-Olmos, J. (2016). Esquema y recomendaciones para el uso de los Modelos de Ecuaciones Estructurales. *Revista de Estudios e Investigación en Psicología y Educación*, 3(2), 75-80. <https://doi.org/10.17979/reipe.2016.3.2.1847>
- Haenlein, M. y Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics*, 3(4), 283-297. http://doi.org/10.1207/s15328031us0304_4
- Hair, J. F., Risher, J. J., Sarstedt, M., y Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2-24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional. A regression-based approach* (3ª ed.). Guilford Press.
- Instituto Nacional de Estadística y Geografía. (2020). *Encuesta Nacional de Vivienda*. <https://www.inegi.org.mx/programas/>
- Kahn, J. H. (2006). Factor analysis in Counseling Psychology research, training and practice: Principles, advances and applications. *The Counseling Psychologist*, 34(5), 684-718. <https://doi.org/10.1177/0011000006286347>
- Kaplan, D. (2000). Structural equation modeling: Foundations and extensions. *Journal of Educational Measurement*, 39(2), 183-186. <https://www.jstor.org/stable/1435255>
- Kenny, D. A. (1979). *Correlation and causality*. Wiley.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2ª ed.). Gilford Press.
- Madrugal Moreno, F. (2018). *Branding en el consumo de tecnología de los universitarios de la generación milenio*. [Tesis doctoral, Universidad Michoacana de San Nicolás de Hidalgo]. Repositorio UMSNH.
- Manzano Patiño, A. P. (2017). Introducción a los modelos de ecuaciones estructurales. *Investigación en Educación Médica*, 7(25), 67-72. <https://doi.org/10.1016/j.riem.2017.11.002>
- Martínez Ávila, M. y Fierro Moreno, E. (2018). Aplicación de la técnica PLS-SEM en la gestión del conocimiento: un enfoque técnico práctico. *Revista Iberoamericana para la Investigación y el desarrollo educativo*, 8(16). <https://doi.org/10.23913/ride.v8i16.336>

- Mateos-Aparicio, G. (2011). Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. *Communications in Statistics: Theory and Methods*, 40(13), 2305–2317. <https://doi.org/10.1080/03610921003778225>
- Moneta Pizarro, A. M. (2017). *Uso de ecuaciones estructurales para identificar factores determinantes del desempeño académico en Educación a Distancia*. <https://rdu.iaa.edu.ar/handle/123456789/364>
- Nitzl, C., Roldan, J.L., y Cepeda, G. (2016). Mediation analysis in partial least squares path modeling: Helping researchers discuss more sophisticated models. *Industrial Management & Data Systems*, 116(9), 1849-1864. <https://doi.org/10.1108/IMDS-07-2015-0302>
- Nunnally, J. C. (1978) *Psychometric theory* (2ª ed.). McGraw-Hill.
- Ponsoda, J. (2012). La investigación sociolingüística de la comunidad de habla. El origen inconformista de la dialectología social. *Revista de Filología de la Universidad de La Laguna*, (30), 155-176. https://riull.ull.es/xmlui/bitstream/handle/915/13486/RF_30_%282012%29_07.pdf?sequence=1&isAllowed=y
- Ramírez, P. E., Mariano, A. M., y Salazar, E. A. (2014). Propuesta metodológica para aplicar modelos de ecuaciones estructurales con PLS: El caso del uso de las bases de datos científicas en estudiantes universitarios. *Revista ADMpg Gestión Estratégica*, 7(2), 133-139. http://www.admpg.com.br/revista2014_2/Artigos/15%20%20-%20Artigo_15.pdf
- Rassolimanesh, S y Ali, F. (2018). Partial least squares – structural equation modeling in hospitality and tourism. *Journal of Hospitality and Tourism Technology*, 9(3), 238-248. <https://doi.org/10.1108/JHTT-10-2018-142>
- Richter, N. F., Sinkovics, R. R., Ringle, C. M., y Schlägel, C. (2016). A Critical Look at the Use of SEM in International Business Research. *International Marketing Review*, 33(3), 376-404. <https://doi.org/10.1108/IMR-04-2014-0148>
- Rigdon, E. E., Ringle C. M., y Sarstedt, M. (2010). Structural Modeling of Heterogeneous Data with Partial Least Squares. January 2010. *Review of Marketing Research*, 7, 255-296. [https://doi.org/10.1108/S1548-6435\(2010\)0000007011](https://doi.org/10.1108/S1548-6435(2010)0000007011)
- Ruiz, M. A., Pardo, A., y San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del Psicólogo*, 31(1), 34-45. <https://www.redalyc.org/articulo.oa?id=77812441004>
- Sáenz López, K., y Tamez González, G. (2014). *Métodos y técnicas cualitativas y cuantitativas*. Tirant Humanidades México.
- Salazar, E. A., y Ramírez, P. E. (2014). Efecto de los Talleres de Alfabetización Informacional en el uso de Bases de Datos Científicas. *Formación universitaria*, 7(3), 41-54. <https://doi.org/10.4067/S0718-50062014000300006>
- Weston, R., y Gore, P. A. (2006). A Brief Guide to Structural Equation Modeling. *The Counseling Psychologist*, 34(5), 719–751. <https://doi.org/10.1177/0011000006286345>

V. Modelos de cointegración y ciclo común

JORGE QUIROZ FÉLIX*

CARLOS ALBERTO FLORES SÁNCHEZ**

MARTHA OFELIA LOBO RODRÍGUEZ***

DOI: <https://doi.org/10.52501/cc.131.05>

Resumen

El presente capítulo tiene como fin ayudar a formar profesionales capaces de utilizar y aplicar herramientas econométricas en la investigación. La naturaleza aplicada de este estudio se alinea con el objetivo del libro de favorecer que los métodos estadísticos sean más accesibles y relevantes para las áreas económico-administrativas.

Este capítulo se adentra en el campo del turismo, al examinar los factores que impactan en la demanda de turistas hacia México y se centra principalmente en los de Estados Unidos y Canadá. Con técnicas de análisis de series de tiempo, pruebas de estacionariedad, pruebas de cointegración y de ciclo común, el estudio busca establecer el comportamiento en el corto plazo y relaciones de largo plazo entre variables en series de tiempo que influyen en la demanda turística.

Estas técnicas permiten descubrir relaciones importantes y significativas entre variables económicas, en este caso la demanda de turistas y los factores

* Doctor en Ciencias Económicas. Director de la Facultad de Comercio Internacional de la Universidad Estatal de Sonora (UES), México. ORCID: <https://orcid.org/0000-0002-2907-8771>

** Doctor en Ciencias Económicas. Profesor investigador en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-1516-166X>

*** Doctora en Ciencias Económicas. Coordinadora general de Vinculación y Cooperación Académica de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0002-9483-2216>

de precio y riqueza. Los hallazgos de este capítulo son particularmente relevantes para los responsables de la toma de decisiones, pues subrayan la importancia de mantener precios competitivos y promover eficientemente el destino turístico, para atraer turistas y enfrentar la competencia en el mercado globalizado.

En general, este capítulo demuestra la relevancia y aplicabilidad de los modelos estadísticos y las técnicas cuantitativas en el ámbito económico-administrativo, y la manera en que éstos pueden ser utilizados para abordar problemas reales y generar soluciones efectivas.

Palabras clave: *demanda y competitividad turísticas, cointegración, ciclo común.*

De la interacción entre series de tiempo existen efectos de retroalimentación en las variables. El análisis de esta relación entre series de tiempo permite identificar los posibles choques idiosincráticos que surgen a partir de su retroalimentación. Este proceso se logra mediante un vector autorregresivo (VAR), que es básicamente una generalización del análisis de procesos autorregresivos, donde en lugar de considerar una sola variable, se considera un vector de variables. Por otro lado, la especificación del VAR nos lleva a la metodología que plantea Johansen (1991) y que permite observar la presencia de cointegración.

Fundamentos del análisis VAR

Un vector autorregresivo (VAR) es un sistema que consta de dos o más series de tiempo, organizado según rezagos de las variables y la interacción dinámica que pudiera o no existir entre ellas. Básicamente, consiste en dos dimensiones: el número de variables (k) y el número de rezagos (p). A continuación, se muestra un VAR de tres variables y un rezago:

$$\begin{aligned} (y_{1t} \ y_{2t} \ y_{3t}) &= (\phi_1 \ \phi_2 \ \phi_3) \\ &+ (\pi_{11.1} \ \pi_{12.1} \ \pi_{13.1} \ \pi_{21.1} \ \pi_{22.1} \ \pi_{23.1} \ \pi_{31.1} \ \pi_{32.1} \ \pi_{33.1}) (y_{1t-1} \ y_{2t-1} \ y_{3t-1}) \\ &+ (\varepsilon_{1t} \ \varepsilon_{2t} \ \varepsilon_{3t}) \end{aligned}$$

Usualmente, un VAR de k variables y p rezagos se precisa de la siguiente manera:

$$y_t = \Phi + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots + \pi_p y_{t-p} + \varepsilon_t$$

Donde:

$$\Phi' = (\Phi_1, \Phi_2, \dots, \Phi_k)$$

$$y_t' = (y_{1t}, y_{2t}, \dots, y_{kt})$$

$$\varepsilon_t' = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt})$$

$$\pi_p = \begin{bmatrix} \pi_{11p} & \dots & \pi_{1kp} & \vdots & \vdots & \pi_{k1p} & \dots & \pi_{kkp} \end{bmatrix}$$

Con esto, es importante identificar las propiedades estocásticas del sistema, es decir, la estacionariedad del VAR.

Para mostrar el concepto de estacionariedad, se considerarán un VAR de dos variables y un rezago, que se muestran a continuación:

$$y_t = \pi_1 y_{t-1} + \varepsilon_t$$

Expresado de otra forma:

$$(y_{1t} \ y_{2t}) = [\pi_{11.1} \ \pi_{12.1} \ \pi_{21.1} \ \pi_{22.1}] (y_{1t-1} \ y_{2t-1}) + (\varepsilon_{1t} \ \varepsilon_{2t})$$

La condición necesaria y suficiente para la estabilidad del VAR es que todos los valores característicos de π_1 tengan módulos menores a uno. Dichos valores son las raíces de la matriz π_1 en $y_t = \pi_1 y_{t-1} + \varepsilon_t$. Se observa que éstas se pueden obtener al resolver la ecuación del polinomio característico:

$$|\pi_1 - vI| = 0, \text{ o bien}$$

$$|I - \rho\pi_1| = 0$$

Esta última es el inverso del polinomio. En este momento se debe precisar que la condición de estabilidad que se refiere al valor de los eigenvalues es la ecuación $|\pi_1 - \nu I| = 0$. En el caso de la expresión $|I - \rho\pi_1| = 0$, la condición se traduce a que ninguno de los eigenvalues tenga un módulo menor a 1.

Raíces unitarias

El primer paso para realizar la prueba de Johansen consiste en identificar si el sistema de series de tiempo es de grado $I(1)$, es decir, que su proceso estocástico muestra la no existencia de estacionariedad. Las condiciones que exhibe el proceso estocástico se evidencian por medio de la estimación de pruebas de hipótesis sobre los coeficientes de un PEG. Las pruebas de hipótesis se conocen comúnmente como pruebas de raíces unitarias.

Para conocer el concepto de raíz unitaria, se representa mediante un desarrollo formal. Si se considera la representación de un modelo AR (1), sería así:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

Si se utiliza el operador de diferencia, se tiene que:

$$(1 - \phi L)y_t = \varepsilon_t$$

Si se invierte la expresión, se obtiene:

$$y_t = (1 - \phi L)^{-1} \varepsilon_t$$

Si se representa como una expansión de Taylor, queda de la siguiente forma:

$$y_t = (1 + \phi L + \phi^2 L^2 + \dots) \varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

Esto muestra que el proceso AR(1) se puede representar como un proceso MA (∞). El mismo converge si $|\phi| < 1$. Observe que ésta es una condi-

ción de convergencia. Si $|\phi| = 1$, la serie no converge, y si $|\phi| > 0$, el proceso es explosivo.

También se puede reescribir $(1 - \phi L)y_t = \varepsilon_t$ como

$$\phi(L)y_t = \varepsilon_t$$

Donde $\phi(L) = (1 - \phi L)$.

Si se resuelve por las raíces del polinomio, esto es $\phi(L) = 0$, se obtiene $L = \frac{1}{\phi}$.

Considerando la condición de convergencia anterior $|\phi| < 1$, se tiene que esto implica que $L > 1$.

Es decir, se afirma que una serie presenta raíz unitaria si existen raíces del polinomio representativo dentro del círculo unitario. En tal caso, $|\phi| = 1$, $L = 1$, y el sistema sería inestable, pues la condición de convergencia no se cumpliría. La teoría econométrica señala que, si bien el análisis del correlograma es un instrumento empírico de gran utilidad, se necesitan aplicar otras pruebas, que identifiquen la magnitud del parámetro ϕ .

Tomando la expresión

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

se resta en ambos lados la variable y_{t-1} y se obtiene:

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + u_t$$

La expresión anterior se puede reescribir como:

$$\Delta y_t = \alpha y_{t-1} + u_t.$$

En ella Δ es el operador de la primera diferencia y $\alpha = (\phi - 1)$. Al realizar una prueba de hipótesis sobre el valor de ϕ , se valoraría en $y_t = \phi y_{t-1} + \varepsilon_t$ si $\phi = 1$ y en $\Delta y_t = \alpha y_{t-1} + u_t$ si $\alpha = 0$. Estas pruebas de hipótesis forman el fundamento de las diversas pruebas de raíz unitaria. Si se comprueba estadísticamente que $\phi = 1$, entonces se determina que la serie es no estacionaria, de lo contrario se dice que la serie no contiene raíz unitaria o que es estacionaria. Por otro lado, si se acepta la hipótesis nula de que el pa-

rámetro $\alpha = 0$, entonces la expresión $\Delta y_t = \alpha y_{t-1} + u_t$ se reduce a $\Delta y_t = (y_t - y_{t-1}) = u_t$, lo que significaría que la primera diferencia de la variable y_t es igual al error aleatorio u_t , y como éste es ruido blanco se deduce que la serie en primera diferencia es estacionaria.

En este escenario, de una serie que se diferencia en una observación y resulta ser estacionaria, se dice que es integrada de orden 1, o bien $I(1)$. A su vez, si un proceso en el que los datos fluctúan sin un patrón evidente se diferencia dos veces para considerar estacionaria la serie, se dice que es $I(2)$, y si se tiene que diferenciar d veces, será de orden $I(d)$. En pocas palabras, si una serie es caracterizada por ser $I(1)$, $I(2)$, ..., $I(d)$, se determina que no es estacionaria en niveles y toma (d) diferencias para que sea estacionaria.

Algunas pruebas formales para determinar la presencia de raíz unitaria en un proceso estocástico son la de Dickey y Fuller (1981), la Dickey-Fuller Aumentada que se muestra en Said y Dickey (1984), la de Phillips-Perron (1988) y la desarrollada por Kwiatkowski *et al.* (1992).

Cointegración

Se reconoce que dos o más series están cointegradas, cuando éstas se mueven simultáneamente a lo largo del tiempo y las diferencias entre ellas son constantes (es decir, estacionarias), sin importar si cada serie cuenta con una tendencia estocástica y por lo tanto sea no estacionaria. La cointegración manifiesta la presencia de un equilibrio a largo plazo hacia donde el sistema de ecuaciones converge en un espacio de largo plazo. Las discrepancias, o término de error, en la ecuación de cointegración se interpretan como el error de desequilibrio para cada punto de tiempo en particular.

En econometría, si se cuenta con dos o más series de tiempo que son no estacionarias de orden $I(1)$, están cointegradas si existe una composición lineal de dichas series que sea estacionaria o de orden $I(0)$. El vector de coeficientes que crea esta serie estacionaria es el vector cointegrante.

Metodología de cointegración de Johansen

De acuerdo con Johansen (1991), la mayor parte de las series temporales son no estacionarias, y las técnicas clásicas de regresión basadas en datos no estacionarios tienden a producir resultados espurios. Sin embargo, las series no estacionarias pueden presentar cointegración si alguna combinación lineal de las series llegara a ser estacionaria. Es decir, la serie puede deambular, pero en el largo plazo hay fuerzas que tienden a llevar a un equilibrio. Por lo tanto, las series cointegradas no se apartarán mucho unas de otras, debido a que están enlazadas a largo plazo.

La técnica de cointegración de Johansen es un proceso multivariado, aplicable a sistemas de ecuaciones. El método se basa en modelos VAR (Vectores AutoRegresivos), es una prueba de máxima verosimilitud que requiere volúmenes de datos de 100 o más; valida la existencia de múltiples vectores de cointegración entre las variables, mediante la prueba de la traza y del Eigenvalue máximo; descansa en la relación entre el rango de la matriz y sus raíces características.

La metodología de Johansen se basa esencialmente en el análisis de un VAR. Si se toma el siguiente modelo:

$$\Delta y_t = \Pi y_{t-1} + \varepsilon_t$$

donde $\Pi = \Pi_1 - I$

Se inicia por suponer que se encuentra una raíz unitaria en el VAR de interés, de tal manera que uno de los eigenvalues de Π_1 es igual a 1. Esto implicaría que el rango de Π es reducido. Esto es: 1) en una matriz de dimensión $k \times k$ existen sólo n renglones o columnas independientes, siendo $n < k$, o 2) en una matriz de dimensión $k \times k$ existen sólo n renglones o columnas independientes, siendo $n < k$. Cuando el rango de la matriz Π es reducido, entonces se sabe que existe una combinación lineal de las variables en el sistema que es estacionaria; en otras palabras, existe al menos un vector de cointegración. Así, la identificación de la existencia de cointegración se reduce a determinar el rango de la matriz Π y, en particular, si el rango es reducido. Dicha tarea se puede realizar con al menos dos criterios: uno se

refiere a identificar los eigenvalues de Π , ya que el número de eigenvalues distintos a 0 es en sí su rango. Alternativamente, se puede determinar que el rango de Π es reducido si su determinante es igual a 0.

Las pruebas de cointegración dan como resultado el número de rezagos óptimos del sistema. Éste se determinó a través de los criterios de información estándar. Se obtienen los valores probabilísticos (p -values) para los estadísticos de prueba de traza y valor característico. Considerando que la hipótesis nula es la no existencia de al menos un vector de cointegración, cualquier valor de p menor a 0.05 a 0.10 indica que la hipótesis nula es rechazada y se acepta la existencia de cointegración. Es así como los estadísticos de Johansen sugieren que existe cointegración entre las variables. Cabe especificar que la hipótesis nula de este tipo de pruebas es la no existencia de cointegración, de tal manera que cualquier valor estadístico superior al valor crítico lleva a rechazar la hipótesis.

Johansen (1991) genera estadísticos de prueba para determinar el número de vectores de cointegración. Dichos estadísticos se muestran a continuación:

1. Estadístico de la traza

$$\text{Traza } (r_0|k) = -T \sum_{i=r_0+1}^n \ln(1 - \lambda_i)$$

donde λ_i son los eigenvalues estimados, los cuales se ordenan como $\lambda_1 > \lambda_2 > \dots > \lambda_n$ y se encuentra en el rango $[0, n - 1]$.

2. Estadístico del máximo eigenvalue

$$\lambda_{\max}(n - 1) = -T \ln(1 - \lambda_n)$$

En la instrumentalización práctica de la prueba de Johansen, la prueba de hipótesis es secuencial. Se inicia por probar la existencia de al menos un vector de cointegración hasta la posible existencia de $n - 1$ vectores.

Una vez determinada la existencia de cointegración entre las variables, se procede a conocer los coeficientes de cointegración. Regresamos a la representación VEC del sistema de series:

$$\Delta y_t = \Pi y_{t-1} + \varepsilon_t$$

Y se parte de la especificación de corrección de error de la ecuación

$\Delta y_t = \alpha z_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-(p-1)} + \vartheta(L)\varepsilon_t$ y se observa que $\pi = \alpha \beta'$, por lo cual el VEC se puede reescribir así:

$$\Delta y_t = \alpha \beta' y_{t-1} + \varepsilon_t$$

A partir de esta ecuación se pueden reconocer dos elementos relevantes: la matriz de coeficientes de ajuste y la matriz de coeficientes de equilibrio (cointegración), β . Se destaca el hecho de que las columnas de β potencialmente representan un vector de cointegración. Por lo tanto, es posible determinar el vector de cointegración relevante al tomar cualquiera de las columnas y normalizar con respecto a cualquiera de los coeficientes.

Una de las ventajas de la metodología de Johansen es la libertad de normalizar el vector de cointegración con respecto a la variable de interés, sin preocuparse por la posibilidad de que los coeficientes cambien al modificar el ordenamiento de las variables.

Modelo para determinar ciclos comunes

A continuación, se expone una aplicación del análisis multivariado de series de tiempo, que ha tomado importancia en fechas recientes; y se trata sobre la identificación de ciclos comunes en un conjunto de variables de un sistema de ecuaciones. Nótese que a través del análisis de cointegración se puede explicar la existencia de tendencias comunes entre series de tiempo. Esto propone que, si los componentes tendenciales de dos series se comportan de manera similar, entonces se podría construir una tendencia que les sea común a ambas. De la misma forma, si dos series presentan un comportamiento cíclico similar, se puede entonces generar un ciclo común entre ellas. Una serie de tiempo se puede entender como la suma de dos componentes, uno de tendencia y otro estacionario; este último constituye su ciclo. El estudio sobre la descomposición de series de tiempo en sus componentes

es amplio, se presenta una discusión que puede motivar al lector a profundizar en el tema.

Beveridge y Nelson (1981) proponen que una serie de tiempo puede ser descompuesta en dos partes, una referida a una tendencia que sigue un proceso de marcha aleatoria y otra que se representa como un proceso ARMA (p,q) y que se entiende como el ciclo de la serie. Formalmente, la serie y_t se descompone como:

$$y_t = \tau_t + c_t$$

En ella $\tau_t = \varphi + \tau_{t-1} + \eta_t$ y $A(L)c_t = B(L)\varepsilon_t$, con los polinomios $A(L)$ y $B(L)$, representan los componentes p y q del proceso ARMA. La tarea de descomposición se puede realizar por medio de diferentes metodologías.

A continuación, se describe la idea de descomposición. Si existe un vector de n variables que se encuentran integradas de orden 1, y_p , y cuya transformación en primeras diferencias es estacionaria y se representa como:

$$\Delta y_t = C(L)e_t = C(1)e_t + (1 - L)C^*(L)e_t$$

Donde $C(L)$ es una matriz polinomial en el operador de rezagos L. Con la integración de ambos lados se puede reescribir como:

$$y_t = C(1)\sum e_{t-i} + C^*(L)e_t$$

Esta es esencialmente la versión multivariada de la descomposición tendencia-ciclo de Beveridge y Nelson (1981), donde el primer término del lado derecho se refiere a la tendencia y el segundo al ciclo.

Este sistema multivariado muestra la tendencia para una serie en particular, que se puede estimar por medio de un vector de corrección de errores (Vector Error Correction Model, VEC por sus siglas en inglés), como por ejemplo, cuando se normaliza el vector de cointegración con respecto a la variable de interés. De esta forma, se incluye la restricción de cointegración en el cálculo de la tendencia. Esta metodología presenta una mejora importante con respecto a la metodología del filtro, que no impone ninguna restricción sobre la dinámica de las series. Sin embargo, como argumentan Vahid y Engle (1993), resulta importante considerar no sólo la posible existencia

de relaciones estables a largo plazo entre las variables, sino también la de posibles relaciones estables a corto plazo.

Esto significa que si tomamos la expresión $y_t = C(1)_{\Sigma} e_{t-i} + C^*(L)e_t$, se observa que la existencia de cointegración implica que $\alpha' C(1) = 0$, donde se incluye los r vectores de cointegración. De igual forma, para el componente cíclico sería posible encontrar que $\alpha' C^*(L) = 0$ indicaría la existencia de relaciones estables a corto plazo en el sistema, tal como la condición $\alpha' C(1) = 0$ sugiere la existencia de relaciones estables a largo plazo.

Cuando se cumple la condición $\alpha' C^*(L) = 0$, se dice que existen s relaciones de ciclo común y α agrupa los vectores de comovimiento, con la restricción $r + s \leq n$. Explicado de otra forma, el concepto de ciclo común es el equivalente al de cointegración, pero al corto plazo. La importancia de tomar en cuenta la posible existencia de comovimientos cíclicos es no minimizar el efecto de choques transitorios, por lo tanto, es un cálculo no eficiente de la tendencia común. De tal forma, para un sistema de ecuaciones donde se encuentra cointegración, la técnica que produce estimaciones de tendencias y ciclos comunes más eficientes es la sugerida por Vahid y Engle (1993).

La metodología de Vahid y Engle (1993) considera en primera instancia el cálculo de las correlaciones canónicas entre las diferencias de las variables y su historia. Ésta se define por el rezago del término de corrección de error y el rezago de las diferencias de los elementos de la matriz de variables explicativas, X_t . Enseguida se calculan las correlaciones canónicas cuadradas ($\lambda^2 \forall i = 1, 2, \dots, n$), que se obtienen al resolver un problema de eigenvalues. Una vez realizado este ejercicio, se evalúa la posible rotación del sistema; así que las variables que se encuentran en el lado derecho del mismo, las que forman la historia, no presenten correlación entre ellas. Después se continúa con la realización de una prueba de hipótesis sobre el número de combinaciones lineales de las variables diferenciadas que cumplen con las características de que no pueden ser pronosticadas y eliminan los patrones de correlación serial en el sistema. En notación de Vahid y Engle (1993), s representa el número de combinaciones lineales y, de hecho, es igual al número de correlaciones canónicas igual a cero. Y constituyen vectores de comovimiento; de hecho es posible mostrar que $s \leq n - r$. La estadística de prueba para la hipótesis nula de que las s correlaciones canónicas cuadradas son igual a cero ($\lambda^2 = 0, \forall j = 1, 2, \dots, s$) se calcula con la siguiente expresión:

$$C(p, s) = - (T - p - 1) \sum \log(1 - \lambda^2)$$

Bajo la hipótesis nula de que la dimensión del espacio de vectores de comovimiento es al menos s , esta estadística tiene una distribución χ^2 con grados de libertad igual a $s(np + r) - s(n - s)$.

Caso práctico. Competitividad turística: análisis a corto y largo plazo

A nivel internacional, de los 95 estudios considerados en Crouch (1994), cerca del 70% utiliza como variable dependiente el número de turistas, debido, en la mayoría de los casos, a que son estudios concretos entre el país de destino y el de origen. El resto de los trabajos emplea como variable dependiente gastos o ingresos por turismo, y unos pocos, la duración de la estancia.

Con respecto a la variable de precio, la forma apropiada de determinarla no está clara, porque el turista adquiere bienes muy heterogéneos que corresponden a varios sectores económicos, pero es importante considerar el costo de vida en el país de destino para el turista. El índice de precios al consumidor¹ del país de destino, ajustado o no por el tipo de cambio, se ha utilizado como proxy para el precio del turismo. Uysal y Crompton (1985) toman en cuenta la existencia de países competidores turísticos y especifica la variable precio de turismo como el costo de vida en el país de destino, relativo a una media ponderada de un conjunto de destinos alternativos.

En este trabajo se utiliza como variable dependiente la demanda de turistas de los principales países de origen (Estados Unidos y Canadá), y se toma como dato el número de turistas por vía aérea en el periodo 1980-2020, según el Sistema Nacional de Información Estadística del Sector Turismo de México (DataTur, 2021). Las variables independientes son la riqueza relativa del país de origen de los turistas (Riqueza del turista en México con

¹ Martín y Witt (1987) compararon dos indicadores del precio de turismo: el Índice de precios al consumidor (IPC) y un costo de turismo más específico, calculado como la media diaria del costo de comer y hospedarse en un hotel de categoría media, y no obtienen mejores resultados estadísticos con el segundo indicador que con el IPC.

respecto a la riqueza en su país de origen) y los precios relativos del país de destino con respecto al país de competencia. La fuente para ingreso per cápita y el índice de precios al consumidor es el Banco Mundial. Para el ejercicio econométrico, omitimos el 2020 por tratarse del año de pandemia covid-19, y se espera retomarlo para próximos documentos.

El modelo desarrollado:

$$\text{Ln}D_{it} = a_0 + a_1 YR_{it} + a_2 \text{IPCR}_{jt} + e_{it}$$

$$i = 1,2 \text{ (países de origen)}$$

$$j = 1,2,3 \text{ (países de la competencia)}$$

Donde:

D_{it} = Número de turistas que arribaron a México del país i en el tiempo t .

YR_{it} = Riqueza relativa del país i en el tiempo t .

Éste es determinado por:

$$YR_{it} = \frac{\text{pib}_i}{\text{pib}_{mx}}$$

Pib_i = Pib per cápita del país i .

Pib_{mx} = Pib per cápita México.

IPCR_{jt} = Precios relativos del país j en el tiempo t .

Éste es determinado por:

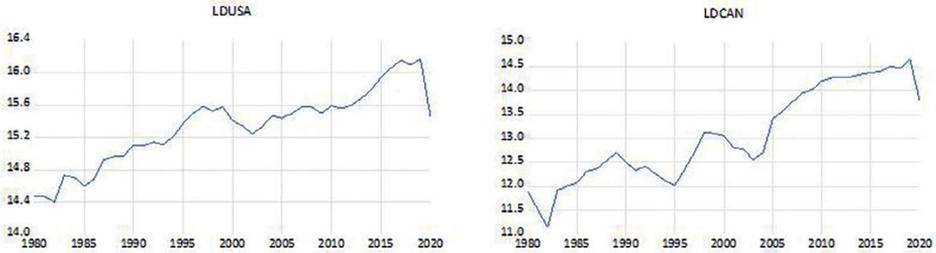
e_{it} = Error aleatorio

La relación entra las variables son expresadas en logaritmos.

Gráficas de variables

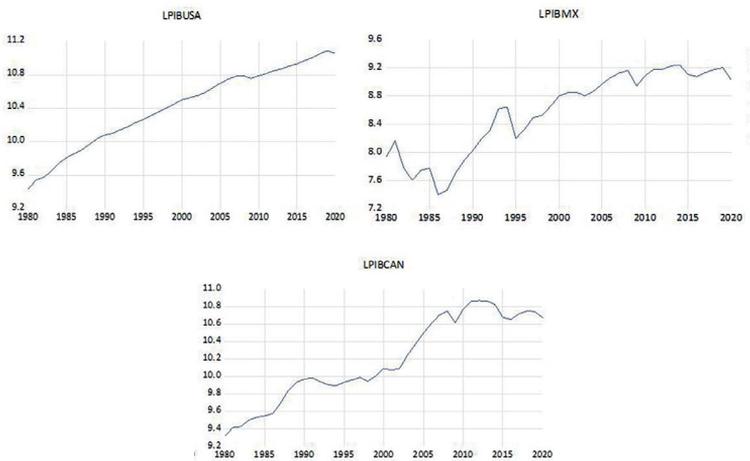
En este apartado se exploran las variables del estudio. Las gráficas ofrecen una visión panorámica de la demanda turística hacia México desde Estados Unidos y Canadá (Figura 1), el comportamiento del producto interno bruto de Estados Unidos, Canadá y México (Figura 2), y la evolución de los índices de precios al consumidor en México, República Dominicana y Costa Rica (Figura 3). Éstas aportan una base visual para entender las dinámicas turísticas y económicas, lo que ayuda a evidenciar las tendencias y relaciones clave con vistas al ejercicio econométrico.

Figura 1. Gráficas Demanda de turismo a México de Estados Unidos y Canadá



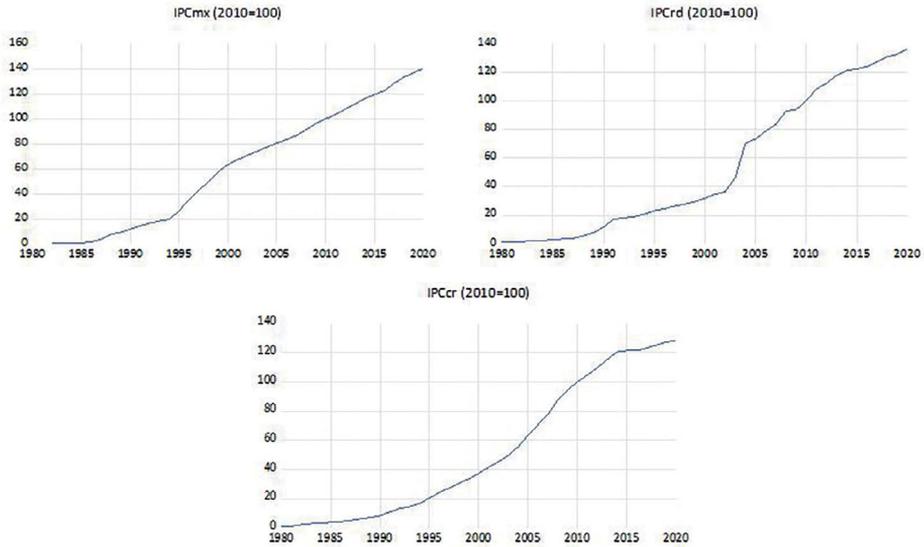
Fuente: Elaboración propia, con datos de DataTur (2021).

Figura 2. Gráficas PIB Estados Unidos, Canadá y México



Fuente: Elaboración propia, con datos del Banco Mundial.

Figura 3. Gráficas IPC México, República Dominicana y Costa Rica



Fuente: Elaboración propia, con datos del Banco Mundial.

Ejercicio econométrico

Se realizaron pruebas de raíz unitaria ADF (1981), KPSS (1992) y PP (1988). Dado que las series presentan cambios estructurales, se propone emplear la prueba de Harvey *et al.* (2013), para determinar el orden de integración, pues dicha prueba controla la particularidad de corte estructural. Esta prueba considera una serie generada por el proceso:

$$y_t = \alpha + \beta t + \gamma' DT_t(\tau_0) + u_t \text{ de } t = 1, \dots, T$$

$$\text{Con } u_t = \rho_T u_{t-1} + \varepsilon_t \text{ de } t = 2, \dots, T$$

Donde:

$DT_t(\tau_0) := [DT_t(\tau_{0,1}), \dots, DT_t(\tau_{0,m})]$ contiene los elementos que indican el periodo en el que sucede el quiebre.

Resultados

A partir de esta especificación se lleva a cabo una prueba de raíz unitaria ajustada de tipo Dickey-Fuller.

Dicha prueba produce valores estadísticos bajo la hipótesis nula de existencia de raíz unitaria. La prueba permite hasta dos cortes estructurales, por lo que se reportan dos estadísticos denotados como MDF1 y MDF2. De tal manera, si estos valores son menores en términos absolutos al valor crítico, entonces no se rechaza la hipótesis nula y se concluye que la serie exhibe raíz unitaria.

Los resultados de las pruebas de estacionariedad se muestran en la Tabla 1. Las series resultaron ser no estacionarias en niveles y estacionarias en primera diferencia, en otras palabras, series integradas de orden 1. En la mayoría de los casos, los estadísticos de la prueba ADF apuntan en esa dirección, y los resultados de la prueba de Harvey lo confirman. Las variables índices de precios al consumidor de México (IPCMx) e índice de precios al consumidor de Costa Rica (IPCcr) sugerían no estacionariedad en niveles en la prueba ADF; sin embargo, las pruebas KPPS y Harvey la rechazan.

Tabla 1. Prueba de estacionariedad para las variables analizadas

Variable	ADF		KPPS		PP		Harvey		Nivel Integración
	Nivel	1ra Dif.	Nivel	1ra Dif.	Nivel	1ra Dif.	MDF1	MDF2	
Dusa	-0.98	-4.68*	0.65	0.08	-0.91	-5.51*	-2.346**	-4.571**	I ₁
Dcan	-0.99	-4.61*	0.67 [^]	0.09	-0.44	-5.03*	-2.431**	-3.285**	I ₁
Yusa	-1.23	-2.33*	0.71 [^]	0.61 [^]	-6.81*	-4.11*	-2.374**	-3.854**	I ₁
Ycan	-1.29	-3.22*	0.68 [^]	0.13	-1.22	-3.59*	-1.155**	-2.498**	I ₁
Ymx	-0.87	-5.33*	0.66 [^]	0.08	-0.73	-5.44*	-2.376**	-4.765**	I ₁
IPCMx	-0.19	-2.31	0.68 [^]	0.31	-2.71	-2.26	-3.976**	-3.822**	I ₁
IPCrd	1.83	-4.12*	0.66 [^]	0.42 [^]	-1.52	-4.58*	-2.543**	-4.588**	I ₁
IPCjam	0.66	-4.40*	0.66 [^]	0.57 [^]	0.46	-2.11	-2.123**	-4.168**	I ₁
IPCcr	-0.61	-1.65	0.65 [^]	0.56 [^]	2.88	-1.69	-2.591**	-3.001**	I ₁

Notas: * Rechaza nula de no estacionariedad al 5%; [^] No rechaza nula de estacionariedad al 5%; ** Rechaza nula de raíz unitaria al 5%.

Fuente: Elaboración propia, con datos de ejercicio econométrico.

Posterior a las pruebas de raíz unitaria, y una vez que se comprobó que las variables son integradas de orden $I(1)$, se procede a realizar la prueba de cointegración de Johansen en cada una de las estimaciones específicas generales, para generar un total de cuatro estimaciones, presentadas en dos grupos generales; cada grupo corresponde al país emisor de turistas y los países competidores. La Tabla 2 muestra los resultados de este ejercicio.

En todos los sistemas analizados se encontró por lo menos un vector de cointegración, con coeficientes significativos y los signos esperados. Se confirma que la demanda de turistas de Estados Unidos y Canadá guarda una relación estable a largo plazo con la riqueza relativa entre los países en cuestión con México y con los precios relativos entre México y el país de competencia.

Tabla 2. Pruebas de cointegración

Sistema	Estructura de rezagos	Hipótesis	Probabilidad	Vector normalizado
Dusa, YRusa, IPCRrd	4	$r \geq 1$	0.0466	1, 0.350, -1.296
Dusa, YRusa, IPCRcr	4	$r \geq 1$	0.0244	1, 0.360, -1.742
Dcan, YRcan, IPCRrd	4	$r \geq 1$	0.0023	1, 0.032, -0.342
Dcan, YRcan, IPCRcr	4	$r \geq 1$	0.0107	1, 0.024, -0.511

Fuente: Elaboración propia, con datos de ejercicio econométrico.

Los vectores normalizados de cointegración se reportan en la última columna del cuadro. En todos los casos, la relación entre la demanda de turistas y la riqueza relativa es positiva. En efecto, cuando la riqueza relativa de los turistas aumenta, la afluencia de turistas al país crece. En el caso de los precios relativos, la relación es negativa, es decir, cuando los precios de México en comparación con el país de competencia aumentan, la demanda de turistas disminuye.

Debido a que las variables fueron tratadas en su transformación logarítmica, los coeficientes pueden ser interpretados como elasticidades a largo plazo. Respecto a la demanda de turistas de Estados Unidos, cuando México se compara con República Dominicana y Costa Rica, se encuentra que en ambos casos es de mayor peso el aumento en el costo de vida en México, el cual impacta en la disminución de demanda en este mercado. En contraste con República Dominicana, un aumento en el costo de vida en México dismi-

nuye en 1.29% la demanda de este mercado. En el caso de Costa Rica, un cambio de 1% en los precios genera una disminución de 1.74% de la demanda de turistas. Sin embargo, la riqueza relativa no genera un impacto relevante. Al ser comparado con República Dominicana, un cambio del 1% en la riqueza relativa de los americanos aumenta en 0.35% la demanda de turistas. Y cuando es comparada con Costa Rica, un aumento del 1% en la riqueza relativa aumenta en 0.36% la demanda de turistas de Estados Unidos. Esto implica que la variable más importante para este mercado es el costo de vida o los precios relativos de los servicios turísticos que ofertan los países de destino.

Respecto a la demanda del mercado de Canadá, cuando México se compara con República Dominicana, el comportamiento de la variable riqueza relativa de los turistas es positivo, pero con poco impacto. Así pues, un aumento de riqueza del 1% aumenta en 0.032% la demanda de turistas, mientras que un aumento en el costo de vida en México disminuye en 0.34% la demanda de este mercado. Cuando se contraponen con Costa Rica, es de mayor relevancia el costo de vida, debido a que un cambio de 1% en los precios, disminuye en 0.51% la demanda de turistas. En tanto, un aumento del 1% de la riqueza relativa sólo aumenta en 0.024% la demanda de turistas. Por último, se encontró que para ambos mercados la variable de precio relativo o costo de vida es la que genera mayor impacto en la demanda de turistas.

Como ejercicio adicional se realizó una prueba de ciclos comunes (Vahid y Engle, 1993) para conocer la relación estable en el corto plazo de la demanda de turistas de Canadá con la riqueza relativa de éstos. Esta estimación es diseñada para establecer las asociaciones a corto plazo. Las estimaciones de relaciones de cointegración o ciclo común pueden realizarse a través de diversas metodologías. Sin embargo, en la literatura se ha mostrado que la aplicación conjunta de las técnicas propuestas produce resultados más eficientes con relación a otras alternativas. En suma, el procedimiento de Vahid y Engle (1993) impone restricciones de cointegración sobre la estimación del componente cíclico del sistema, lo cual incrementa la eficiencia de los resultados.

Por otra parte, el estadístico de prueba sugerido por Vahid y Engle es:

$$C(p, s) = -(T - p - 1) \sum_{i=1}^s \log \log (1 - \lambda_i)$$

Donde T es el número de observaciones, p es el número de rezagos del Vector autorregresivo (VAR) en niveles, s es el número de correlaciones canónicas al cuadrado iguales a cero que se está sometiendo a prueba, y λ_i es la i -ésima correlación canónica al cuadrado más pequeña entre las variables del sistema y su historia relevante.

Tal estadístico se distribuye ji-cuadrada con $s^2 + snp + sr - sn$ grados de libertad. Siguiendo la notación convencional, n es el número de variables en el sistema, y r representa el número de vectores de cointegración.

Adicionalmente, Vahid y Engle (1993) demuestran que los vectores de cointegración y de comovimiento son ortogonales, por lo que cuando la suma del número de ambos tipos de vectores es igual al número de variables en el sistema, se puede crear una base para proyectar R^n (n = número de variables). De este modo, mediante una sencilla manipulación algebraica de la matriz que agrupa los vectores de cointegración y de comovimiento, se puede obtener tanto el componente tendencial como el cíclico del sistema (Vahid y Engle, 1993).

Para realizar la prueba antes mencionada, se construyó un sistema bivariado para el país seleccionado (Canadá). Se consideró como variable dependiente la demanda de turistas a México y como variable independiente la riqueza relativa del país de origen de los turistas, en cada uno de los sistemas.

Tabla 3. Pruebas de ciclos comunes

Sistema	β	t-est	Pvalue 1 ciclo	Pvalue 2 ciclos
Dusa, YRusa	0.32	2.00	0.06*	0.00
Dcan, YRcan	0.46	0.54	0.18*	0.00

Nota: * Acepta existencia de un ciclo común.

Fuente: Elaboración propia con datos del Banco Mundial.

Se encuentra que las series tienen ciclos económicos comunes. Es posible afirmar que en el corto plazo la riqueza de los turistas de Estados Unidos

tiene impacto directo en la demanda de turistas a México. Según el modelo, un aumento del 1% de la riqueza de Estados Unidos genera un aumento del 0.32% en la demanda de turistas. Estos resultados tienen sentido dada la cercanía geográfica y la competitividad en precios de México. Así, el aumento de turismo impacta directamente la economía, al tener un aumento de ingresos extraordinario, aunque podemos inferir que México no es un bien de lujo para el mercado americano. Aun cuando las series tienen un ciclo común, el mercado de Canadá, en el cual si bien existe un ciclo común, éste no es estadísticamente significativo. Ello implica que el ciclo existe por alguna razón fortuita y no como resultado de una relación de la riqueza de los turistas, pues una de las principales razones que consideran es la ubicación geográfica.

Conclusiones

En conclusión, si se busca conocer los factores de mayor impacto en la demanda de turistas a México, se deduce que el precio es un factor significativo para los dos mercados estudiados (Estados Unidos y Canadá), lo cual ofrece una ventaja competitiva a México, dado que somos un destino más económico que los países analizados como competencia. Tiene sentido que el aumento de riqueza no sea un factor tan relevante para la demanda de los mercados analizados, debido a que a mayor poder adquisitivo, los turistas eligen destinos ubicados geográficamente más lejos o exclusivos.

En el mercado globalizado, México se encuentra con diversos competidores, que han venido adquiriendo fuerza en los mercados estudiados. Aunque si bien México todavía tiene una mayor demanda de turistas, es necesario conocer los factores que influyen en la decisión de compra de la demanda, frente a otras opciones de destino, con el objetivo de tener información para la toma de decisiones en la elaboración de una eficiente política pública, en la promoción del destino y la generación de valor agregado por las empresas del sector.

Los resultados obtenidos en este documento muestran que a largo plazo el precio es un factor de impacto en la decisión de compra de ambos mercados analizados. Por ello se trata de una demanda elástica al precio, lo cual

implica que el mercado es muy sensible a las variaciones del precio de servicios turísticos y opta por el destino que ofrece mejores precios. Esto coincide con estudios previos realizados en otros destinos turísticos, como en el trabajo de Patsouratis *et al.* (2005). Por otro lado, el aumento de riqueza de Estados Unidos y Canadá no se considera como relevante, pues solamente tiene impacto cuando la competencia es República Dominicana o Costa Rica para la demanda de Estados Unidos y Canadá. Sin embargo, en el corto plazo, la riqueza relativa si es un factor específico para el mercado de Estados Unidos, caso contrario a lo que sucede con el mercado de Canadá.

El trabajo, meramente cuantitativo, manifiesta la necesidad de tener un acercamiento a los turistas para elaborar su perfil e identificar cuáles son los factores con impacto sobre su toma de decisión para venir al destino y, por otro lado, sobre su lealtad al destino. También deja como requisito contar con información sobre los precios que manejan otros destinos turísticos para que no sea un factor en contra sino a favor, así como de la riqueza relativa de los turistas visitantes, lo cual se identifica con el tipo de cambio y con el desempeño económico de su país de origen. Juntos los trabajos de investigación cuantitativa y cualitativa apoyarán la elaboración de políticas públicas eficientes que visualicen la diversificación del mercado. Por ello, se necesita impulsar la diversificación de productos turísticos, dado que México puede ofrecer destinos integrales (playa, historia, sitios arqueológicos, gastronomía, etc.), con lo cual podría generar productos turísticos para atraer al turismo que busca más que sol y playa, así como integrar programas de promoción en países diferentes al mercado de Norteamérica (USA y Canadá) y apoyar a las empresas que ofertan servicios turísticos tanto en su planeación como en su operación.

Referencias

- Beveridge S., y Nelson, C.R. (1981). A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of Business Cycle. *Journal of Monetary Economics*, 7(2), 151-174. [https://doi.org/10.1016/0304-3932\(81\)90040-4](https://doi.org/10.1016/0304-3932(81)90040-4)
- Crouch, G. I. (1994). The Study of International Tourism Demand: A Survey of Practice.

- Journal of Travel Research*, 32(4), 41-55. <https://doi.org/10.1177/004728759403200408>
- Dickey, D. A., y Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 49(4) 1057-1072. <https://doi.org/10.2307/1912517>
- Harvey, D. I., Leybourne, S. J., y Robert Taylor, A. M. (2013). Testing for unit roots in the possible presence of multiple trend breaks using minimum Dickey–Fuller statistics. *Journal of Econometrics*, 177(2), 265-284. <https://doi.org/10.1016/j.jeconom.2013.04.012>
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society*, 59(6), 1551-1580. <https://doi.org/10.2307/2938278>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., y Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3), 159-178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Martín, C.A., y Witt, S.F. (1987). Tourism demand forecasting models: Choice of appropriate variable to represent tourists' cost of living. *Tourism management*, 8(3), 233-246. [https://doi.org/10.1016/0261-5177\(87\)90055-0](https://doi.org/10.1016/0261-5177(87)90055-0)
- Patsouratis, V., Franfouli, Z., y Anastasopoulos, G. (2005). Competition in tourism among the Mediterranean countries. *Applied economics*, 37(16), 1865-1870. <https://doi.org/10.1080/00036840500217226>
- Phillips, P. C. B., y Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346. <https://doi.org/10.2307/2336182>
- Said E. S., y Dickey, D. A. (1984). Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika*, 71(3), 599-607. <https://doi.org/10.2307/2336570>
- Sistema Nacional de Información Estadística del Sector Turismo de México (2021). *Compendio Estadístico del Turismo en México*. Secretaría de Turismo. <https://www.datatur.sectur.gob.mx/SitePages/ActividadHotelera.aspx>
- Uysal, M., y Crompton, J. L. (1985). Deriving a relative price index for inclusion in international tourism demand estimation models. *Journal of Travel Research*, 24(1), 32-34. <https://doi.org/10.1177/004728758502400106>
- Vahid F. y Engle R. F. (1993). Common Trends and Common Cycles. *Journal of Applied Econometrics*, 8(4), 341-360. <https://www.jstor.org/stable/2285000>

VI. Diseño y validación de encuestas

MALENA PORTAL BOZA*
EDGAR JIMÉNEZ CERRA**

DOI: <https://doi.org/10.52501/cc.131.06>

Resumen

Este capítulo tiene como objetivo describir las bases teóricas y metodológicas y los procedimientos que deben ser aplicados en el momento de diseñar y validar un instrumento de recopilación de información primaria. Para ello se incluyen las principales formas de aplicación de la encuesta, así como los elementos por considerar para obtener resultados exitosos en este proceso. De forma específica, se hace énfasis en la conformación del cuestionario, las características sociodemográficas más utilizadas, los tipos de preguntas y los tipos de escala más empleados en el ámbito de las ciencias económico-administrativas. Además, se describe el proceso de validación de los instrumentos mediante el empleo de métodos de análisis factorial exploratorio y confirmatorio.

Palabras clave: *encuesta, cuestionario, validación, escalas, preguntas.*

* Doctora en Ciencias Económicas. Profesora de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0002-4237-1534>

** Doctor en Ciencias Administrativas. Profesor de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-1285-3345>

La encuesta como fuente primaria de recolección de datos

A nivel global la encuesta es una de las técnicas de recopilación de información primaria más utilizadas en el ámbito de los estudios científicos. Por eso se considera un referente obligado al abordar el proceso investigativo. Este instrumento permite obtener datos de forma indirecta según la interpretación de las personas encuestadas, lo que puede incidir en la calidad de la información resultante. Es importante destacar la masividad que permite el uso de esta herramienta, lo que asociado al empleo de las técnicas de muestreo correctas, garantiza que los resultados puedan ser extensibles a poblaciones completas.

Al emplear este tipo de herramienta, los resultados esperados no reflejan características específicas de los sujetos que contestan, sino que conforman conjuntos que permiten identificar patrones asociados a la población a la que pertenecen las personas estudiadas. En este sentido, es importante comentar que la información obtenida debe ser anónima y recogida mediante la aplicación de un único cuestionario a todos los sujetos incluidos en la muestra, de forma que permita aplicar diferentes técnicas estadísticas para la realización de estudios comparativos sobre los resultados obtenidos. Existen diversas formas de aplicar una encuesta, entre ellas las más comunes son la vía telefónica, la entrevista directa y mediante el empleo de una aplicación o formulario en línea.

Las encuestas telefónicas se utilizan en menor grado; sin embargo, aún existen empresas o investigadores que contratan los servicios de *callcenter* para aplicar el instrumento diseñado para esta vía. El encuestador transmite las preguntas al encuestado mediante la vía telefónica y emplea como guía un formulario en su equipo de cómputo o dispositivo digital, en el cual registrará las respuestas obtenidas. Las principales ventajas de este método radican en la reducción de recursos para la aplicación presencial, la verificación de las respuestas por el sistema empleado, la reducción del tiempo de entrevista y la mejora de la calidad de los datos resultantes. Además, al ingresar datos de forma sistemática a la aplicación, el sistema puede generar reportes intermedios que orienten al investigador sobre el desarrollo del

proceso. La principal desventaja estriba en la disponibilidad de las personas para responder la llamada telefónica.

Las entrevistas personales o directas son las más costosas por el elevado nivel de recursos requeridos para su aplicación: impresión de las encuestas, capacitación de los encuestadores, vestuario e implementos que identifiquen a los aplicadores, salarios de todo el personal implicado, equipos de cómputo y personal para la captura de la información recopilada, así como recursos para la divulgación del proceso y su soporte logístico. Al realizarse de forma directa en los hogares de las personas, esto puede resultar una ventaja porque permite constatar las respuestas asociadas al perfil sociodemográfico de manera directa por el entrevistador, aunque puede considerarse una desventaja por razones de seguridad. Además, existe la posibilidad de que los aplicadores adulteren la información, con la finalidad de concluir el proceso con mayor rapidez.

La variante más empleada en el presente, asociada a la digitalización de la sociedad, son las encuestas en línea, principalmente por las ventajas con respecto a las formas tradicionales (Evans y Mathur, 2005). La reducción en los costos de despliegue del instrumento, el procesamiento de la información y la disminución del tiempo para la obtención de los datos son elementos significativos que inciden en el uso de este método. Además, el entrevistado se siente mucho más cómodo al poder contestar las preguntas del cuestionario en su ambiente personal, sin importar el tipo de dispositivo utilizado, ni el momento y lugar en que lo haga. En la actualidad existen diversas aplicaciones que facilitan la realización de encuestas en línea, entre las más utilizadas se pueden encontrar SurveyMonkey, Google Forms y QuestionPro.

La posibilidad de efectuar encuestas por internet ha eliminado las barreras geográficas, idiomáticas y culturales, lo que permite a los investigadores considerar poblaciones más diversas y obtener resultados más abarcadores. Sin embargo, existen elementos que se deben tomar en consideración al utilizar este método. En primer lugar, se debe garantizar que todos los sujetos de la población estudiada tengan igual oportunidad de ser encuestados, y esto puede ser un problema de no contarse con listados de sus características y sus direcciones de correos electrónicos (Díaz de Rada, 2012).

Por otra parte, el acceso a internet y el conocimiento para utilizar aplicaciones móviles son requisitos indispensables para la aplicación de una encuesta en línea, por lo que aquellos integrantes de la población que no cuenten con este recurso o estas habilidades, automáticamente serán excluidos del estudio. Esto puede considerarse un sesgo negativo importante en cuanto a la validez de la muestra utilizada para generalizar los resultados obtenidos. Por último, se ha identificado que uno de los principales errores cometidos por parte de los investigadores es la orientación intencionada hacia ciertas respuestas, lo que incide de forma considerable en la veracidad de los resultados.

Con el objetivo de garantizar el éxito en la aplicación de este tipo de instrumento, se sugieren algunas buenas prácticas como son la personalización del marco muestral a través de un listado con las principales características de los integrantes y su dirección de correo electrónico comprobada, y el empleo de herramientas de recordatorio para la consecución de un mayor porcentaje de respuestas y de seguimiento a los individuos que abandonan la encuesta sin responder todas las preguntas. De igual forma, debe considerarse limitar a una respuesta por dirección de correo electrónico o IP del dispositivo utilizado. Además, es importante incluir la descripción del instrumento y una explicación breve pero clara de lo que se quiere alcanzar como parte del mensaje de petición de respuesta a los encuestados y, de esta forma, suplir la ausencia de un encuestador presencial.

Otra clasificación de las encuestas está relacionada con su estructura de preguntas y respuestas. En este sentido, existen las no dirigidas o libres, que proponen un tema sobre el cual el entrevistado diserta libremente sin ninguna orientación por el entrevistador y son las que se ajustan más a los estudios de corte cualitativo. En segundo lugar, se encuentran las semidirigidas o en profundidad. En las preguntas asociadas a la temática por investigar se organizan en secciones internas del cuestionario, las cuales no son rígidas u ordenadas. En este caso el encuestador utiliza preguntas indirectas establecidas en un guion preparado previamente para conducir al entrevistado, y se busca obtener respuestas en su subconsciente mediante estímulos; un clásico ejemplo de este tipo es el conocido test de Rorschach. Por último, existe la encuesta dirigida, que se compone por preguntas en las cuales existe un elevado grado de direccionalidad por el investigador, tanto en el orden

como en la forma de formular las preguntas. En éstas se podrá profundizar en la información que se busca, al no existir limitaciones en el momento de responder por los participantes.

Elementos de importancia en el diseño del cuestionario

El instrumento empleado para el despliegue del contenido de la encuesta es el cuestionario, que según Rojas *et al.* (2000) se puede definir como la estructura integrada por diferentes tipos de preguntas que permite recopilar los valores para las variables asociadas a una investigación de una forma ordenada y que responde a los objetivos de la misma. En su elaboración el investigador debe ser capaz de transformar las variables empíricas que se estudian en preguntas válidas, confiables, claras, concretas y fáciles de responder por los encuestados.

En el momento de diseñar las preguntas es importante considerar varios aspectos asociados a la población objetivo de la aplicación. Elementos como el nivel cultural, la edad, el estado de salud y las habilidades informáticas, entre otros, permiten definir aspectos como la cantidad de preguntas que formarán el cuestionario, el sistema de aplicación que se utilizará, el lenguaje, el idioma, la forma en la que se solicitarán las respuestas, y cualquier otra cuestión con incidencia en la calidad de las respuestas. Además, es importante que no incidan en el sentido de la respuesta y que no induzcan inexactitud en ésta (López-Roldán y Fachelli, 2016).

El éxito de un cuestionario estará relacionado de forma directa con las preguntas que el investigador incluya en él. La cantidad es un elemento fundamental, ya que un número excesivo puede hacer que el entrevistado se canse y desista de responder; sobre esta base es importante eliminar preguntas innecesarias y redundantes. Para las investigaciones asociadas a las ciencias administrativas, se sugiere utilizar preguntas cerradas y clasificadas en un determinado número de categorías. La forma de enunciar las preguntas debe ser sencilla, positiva, mediante el empleo de un lenguaje coloquial directo, sin ambigüedades, y no debe incluir posibles respuestas. Las palabras utilizadas no deben ser rebuscadas y debe procurarse que tengan el

mismo significado para todos los encuestados. De igual forma, se recomienda no incluir cuestiones que impliquen elevados esfuerzos de memorización y cálculo; los temas tratados deben ser de la competencia de los entrevistados y no muy íntimos, de tal forma que no provoquen una respuesta condicionada. Con respecto a las respuestas, se sugiere que éstas cubran todo el espectro posible en torno a la pregunta realizada y se garanticen los requisitos de mutua exclusión entre ellas. Además, deben considerarse respuestas como “otros”, “no contesta” y “no aplica” para alcanzar todas las posibilidades.

Como parte del diseño del cuestionario, se pueden utilizar diferentes tipos de pregunta para lograr aglutinar toda la información requerida en el proceso de investigación; es importante conocer esa tipología para crear instrumentos en función de los intereses que se persiguen.

Presentemos la clasificación según la función de la pregunta en el cuestionario. En primer lugar, debe considerarse utilizar preguntas introductorias, que permiten al encuestador romper el estado inicial de incertidumbre o desconfianza y garantizan que el sujeto entrevistado sienta mayor disposición a contestar el resto de las preguntas más enfocadas al objeto de estudio. También existen las llamadas preguntas filtro, que permiten realizar un proceso de selección entre los encuestados según sus características, y a partir de responder a ellas, el grupo seleccionado tendrá preguntas específicas. Otro reactivo muy utilizado son las preguntas de control o congruencia, que son empleadas para comprobar la consistencia de las respuestas obtenidas mediante el contraste de preguntas que significan lo mismo, pero se enuncian de manera diferente; éstas deben estar espaciadas a lo largo del cuestionario.

En correspondencia con la respuesta esperada, podemos clasificar las preguntas de la siguiente manera: las cerradas o de respuesta fija, que son aquéllas donde las alternativas del entrevistado para responder están predeterminadas. Al profundizar en esta categoría, se encuentran las dicotómicas, que solo permiten una respuesta u otra; las politómicas, que cumplen las mismas condiciones pero para un número mayor de posibilidades; las de escala, que establecen un sentido creciente o decreciente a las posibles respuestas, y las de opción múltiple, que permiten seleccionar más de una alternativa. Por otra parte, las abiertas cumplen con el criterio de que no prefijan las posibles respuestas y permiten a los encuestados discernir sobre el tema en cuestión de una forma amplia, y las semiabiertas combinan las

características de las dos anteriores (cerradas y abiertas) de manera flexible. Es importante destacar que los cuestionarios no tienen que contener un único tipo de preguntas, sino que deben integrarse por las que se ocupen para lograr el objetivo propuesto.

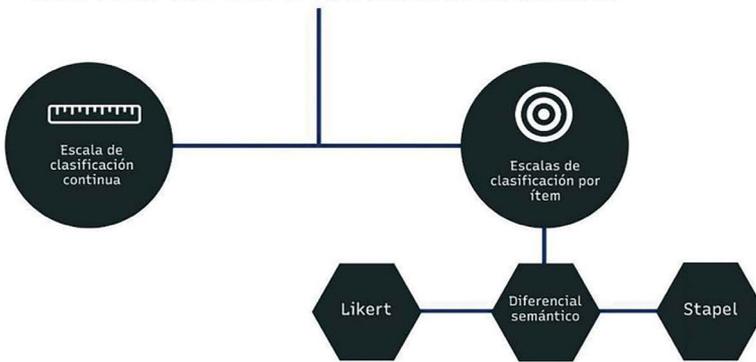
De forma específica y por ser una de las más utilizadas, haremos énfasis en las preguntas de opción múltiple, especialmente las que como respuesta incluyen elementos escalares. Las escalas se usan con el objetivo de obtener una distribución de las frecuencias como medida cuantitativa asociada a características o aspectos del sujeto de estudio. Son consideradas una extensión del proceso de medición al establecer sus categorías. Dentro de las clasificaciones para los tipos de escala, una de las más conocidas es la emitida por Stevens (1951). Este afamado autor particulariza que existen las nominales, que se caracterizan por establecer un sistema de etiquetado por categorías para los objetos incluidos en la muestra y, de esta forma, clasificarlos e identificarlos; las escalas ordinales, que permiten dar un orden a los elementos de la muestra y se utilizan principalmente para mostrar las preferencias de los encuestados; las escalas de intervalos, que permiten a partir de la definición de una unidad de medida la asignación a cada sujeto de un valor según la medición realizada y son empleadas con mayor frecuencia en la medición de la satisfacción o el grado, de acuerdo con afirmaciones planteadas por el investigador; y las escalas de razón, en las cuales el cero está definido, lo que contribuye a la realización de comparativas a partir de las mediciones realizadas; este tipo de escala es muy utilizado para la medición de variables cuantitativas como la edad y el nivel de ingresos, entre otros.

Además de las escalas básicas, existen las comparativas y las no comparativas. En las primeras se solicita al entrevistado que a partir de un punto de referencia explicado con anterioridad, compare los objetos de la muestra según un grupo de estímulos; en este caso la principal desventaja de este tipo de escalas es la imposibilidad de generalizar los resultados obtenidos debido a la particularidad de los objetos comparados. Por eso se recomienda el empleo de escalas no comparativas o métricas, donde los elementos de la muestra son evaluados de forma independiente.

Las escalas no comparativas generan resultados en forma de intervalos o razón (Amoo y Friedman, 2001). Los encuestados no comparan los obje-

tos estudiados entre sí, sino que ofrecen su valoración personal sobre el objeto en cuestión en cada una de las preguntas. Estas escalas se dividen en dos grupos (Figura 1), las de clasificación continua y las de clasificación por ítems. Las primeras se muestran como una línea continua que en sus extremos exhibe dos categorías antagónicas relacionadas con la variable estudiada, de forma que el encuestado ubique su valoración. Este tipo de respuestas solo es objetiva al aplicarla mediante el uso de herramientas digitales, pues de otra manera se complejiza su interpretación.

Figura 1. Escalas no comparativas básicas



Fuente: Elaboración propia.

Por su parte, las escalas de clasificación por ítems son las más utilizadas en el ámbito investigativo. Se presentan al encuestado en forma de categorías ordenadas mediante números o palabras que describan el comportamiento por evaluar y permitan al entrevistado expresar su parecer en este sentido. Se describen a continuación los tres tipos que se emplean con más frecuencia.

La más difundida es la conocida como escala de Likert, llamada de esa forma en reconocimiento a su creador, el afamado psicólogo Rensis Likert. Ésta es de muy fácil construcción e incluye entre 5 y 9 elementos (por lo general 5), o sea, un número impar que permite que exista una categoría central para que el entrevistado pueda mostrar su indiferencia ante el tema. Su resultado está enfocado al grado de acuerdo o desacuerdo de los entrevistados con las afirmaciones planteadas en cada uno de los reactivos.

La escala de diferencial semántico se construye con un total de siete elementos, definida en sus extremos por elementos de significado totalmente opuestos, ejemplos “frío y caliente”, “antiguo, novedoso”, “fiable, engañoso”, “satisfecho, insatisfecho”. Ésta se emplea en investigaciones de corte emocional, principalmente con respecto a situaciones, servicios o productos, lo que las hace las más utilizadas en estudios de marketing. Sus principales ventajas radican en la sencillez y que son de fácil comprensión para los encuestados. Sin embargo, al utilizar este tipo de escalas es recomendable emplear preguntas filtro en la parte introductoria del cuestionario, para realizarlas según el nivel de conocimientos sobre el tema de los encuestados.

En tanto, la escala de Stapel tiene un enfoque unipolar y se integra por un número par de categorías para las posibles respuestas entre las cuales no se incluye un término medio o neutro. Los entrevistados deben indicar su nivel de acuerdo o desacuerdo mediante un valor numérico (+5 a -5), con la precisión de la descripción del objeto evaluado según las categorías propuestas. Debe mantenerse una relación creciente positiva entre el resultado y la precisión categórica seleccionada, o sea, que una puntuación más elevada indica mayor aceptación de la descripción del objeto en cuestión.

Hasta el momento, se han descrito en el este capítulo los principales elementos de diseño y conformación de las encuestas. Sin embargo, todo lo anterior debe ser validado para garantizar la fiabilidad de la información resultante con la aplicación del instrumento y, de esa forma, poder generalizar los resultados obtenidos.

Validez y fiabilidad del cuestionario

Validez de constructo

La validez es considerada la más importante de las pruebas a la que se debe someter el instrumento desarrollado en el ámbito científico y consiste en evaluar cómo la encuesta propuesta mide y representa el modelo teórico planteado como parte de la investigación (Hernández-Sampieri *et al.*, 2006). Para la realización de este proceso de validación, se emplean métodos estadísticos mediante los cuales se explica la relación entre las variables incluidas

en el cuestionario en función de un grupo de factores determinados. El método organiza los elementos estudiados y los reduce a un grupo de categorías (constructos), que permiten hacerlos más accesibles para su análisis estadístico. Su nombre es análisis factorial. Es importante el conocimiento de los datos que se manejan para la definición de los constructos, porque a partir del análisis de éstos se busca explicar la variabilidad e interrelación de las variables observadas (Levin *et al.*, 1996). Además, estos factores deben reflejar la estructura conceptual del cuestionario.

Díaz y Morales (2012) argumentan que el análisis factorial permite establecer si las covarianzas o correlaciones observadas sobre un conjunto de variables pueden ser explicadas a partir de un número pequeño no observable de variables latentes. Estos autores presentan la estructura del modelo a partir de un vector aleatorio X de tamaño $(p \times 1)$ con media μ y matriz de covarianza Σ como sigue:

$$X = \mu + \Lambda f + U \quad (1)$$

Donde Λ es una matriz de ponderaciones, cargas o pesos de tamaño $(p \times k)$; f es un vector columna de k componentes $(k \leq p)$ y U es un vector aleatorio de tamaño $(p \times 1)$ con distribución independiente de f . Respectivamente:

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pk} \end{pmatrix}, \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}. \quad (2)$$

Las ecuaciones se representan de la siguiente manera:

$$\begin{cases} X_1 = \mu_1 + \lambda_{11}f_1 + \cdots + \lambda_{1k}f_k + u_1 \\ X_2 = \mu_2 + \lambda_{21}f_1 + \cdots + \lambda_{2k}f_k + u_2 \\ \vdots \\ X_p = \mu_p + \lambda_{p1}f_1 + \cdots + \lambda_{pk}f_k + u_p, \end{cases} \quad (3)$$

$$X_p = \mu_p + \lambda_{p1}f_1 + \dots + \lambda_{pk}f_k + u_p$$

Donde los elementos de f son los factores comunes y los de los factores únicos o específicos. Los factores comunes están a su vez estandarizados ($E(f) = 0$; $\text{Cov}(f) = I$, los factores específicos tienen media 0 y están incorrelados ($E(u) = 0$; $\text{Cov}(U) = E(UU') = \Psi$ y $\text{Cov}(f, u) = 0$).

El modelo expresado en la ecuación 4 muestra que:

$$X_i = \mu_i + \sum_{j=1}^k \lambda_{ij}f_j + u_i \quad i = 1, \dots, p \quad (4)$$

Por tanto:

$$\text{var}(X_i) = \sigma_{ii} = \sum_{j=1}^k \lambda_{ij}^2 + \psi_{ii}$$

De tal forma que la varianza puede descomponerse en dos partes, la primera es la siguiente:

$$\lambda_{i1}^2 + \dots + \lambda_{ik}^2 = \sum_{j=1}^k \lambda_{ij}^2 = h_i^2 \quad (5)$$

Esto significa que la comunalidad h^2 representa la varianza explicada por los factores comunes, y la segunda ψ_{ii} la variabilidad exclusiva de X_{ij} , es decir, la varianza que no es compartida con otras variables y es llamada especificidad o varianza única. Por lo tanto, los factores comunes explican las relaciones existentes entre las variables del problema.

Salvador Figueras y Gargallo Valero (2006) mencionan que una vez formulado el problema y obtenida la matriz de datos X , el siguiente paso es el examen de la matriz de correlaciones maestras $R = (r_{ij})$, donde r_{ij} es la correlación muestral observada entre las variables X_i y X_j . El objetivo de este paso es justificar que las características son convenientes para realizar un análisis factorial.

Estos mismos autores explican que una de las exigencias para que el análisis factorial sea viable es que las variables estén altamente intercorrelacionadas; también se espera que las variables con correlación muy alta entre sí la tengan con el mismo factor o factores.

Para comprobar que la estructura de los datos incluidos en la muestra es adecuada para la realización del análisis factorial, se emplean la media de

adecuación muestral Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett. Según De la Fuente (2011), de forma específica el estadístico de KMO permite constatar el volumen correlacional entre las variables, lo que se explica al interpretar que valores resultantes cercanos a 0 indican que el modelo no es el adecuado; pero en sentido opuesto, valores que se aproximan a 1 conducen a un mejor ajuste del modelo.

Por su parte, el test de esfericidad de Bartlett pone a prueba la hipótesis nula que plantea la posibilidad de que la matriz correlacional resultante sea una matriz de identidad. Esto implicaría la no existencia de correlaciones significativas entre las variables implicadas y, por tanto, la inutilidad del modelo propuesto. De acuerdo con López-Aguado y Gutiérrez-Provecho (2019), este test presenta una distribución asintótica según la ecuación X^2 con $\frac{p(p-1)}{2}$ grados de libertad, lo que hace que —como parte del análisis de su resultado— una relación inversa entre los valores del estadístico y la significatividad indicarían el rechazo de la hipótesis nula y permitirán concluir que la relación entre las variables implicadas en el estudio es suficiente para la realización del análisis factorial.

En el caso de que el investigador desconozca la cantidad de constructos en los que se agruparán las variables estudiadas, según la relación existente entre ellas, y la estructura mediante la cual fue concebido el cuestionario no ha sido validada con anterioridad, el análisis factorial realizado se califica como exploratorio (AFE), ya que los factores de agrupación se generan como parte de los resultados de la aplicación de esta técnica (Méndez Martínez y Rondón Sepúlveda, 2012).

La estimación de cada uno de los constructos se puede realizar mediante el uso de la técnica de análisis multivariado de componentes principales (ACP), cuyo principal objetivo es explicar en un alto porcentaje la varianza total observada en un grupo de variables con la menor cantidad de componentes posibles. Mediante este método se logra convertir, a través de una transformación lineal, los indicadores originales en otro conjunto de variables sin correlación alguna, bajo el nombre de componentes principales. Éstos permiten evitar la multicolinealidad entre los indicadores originales incluidos en el estudio y realizar una clasificación inmediata sobre la base de los factores iniciales (Méndez Martínez y Rondón Sepúlveda, 2012).

Los resultados al desarrollar el AFE por lo general resultan un tanto complejos, por lo que se recomienda (Guerrero Pulido, 2018) realizar una rotación a la versión inicial de éstos, de forma tal que se facilite su interpretación. Se puede emplear el método ortogonal, que basa su funcionamiento en la reducción de la ambigüedad en los valores del efecto del factor sobre las variables estudiadas, mientras los demás factores permanecen constantes. El resultado de la aplicación del método ortogonal en cualquiera de sus variantes (equamax, quartimax o varimax) se presenta como una matriz que muestra la correlación entre las variables y los constructos resultantes. Estos valores obtenidos deben ser evaluados para seleccionar las variables que tienen mayor peso, al explicar cada factor y eliminar las que no lo hacen. Siguiendo esta línea, Méndez Martínez y Rondón Sepúlveda (2012) proponen una escala que establece que los valores por debajo de 0.5 tienen un aporte mínimo, entre 0.5 y 0.7 significativo y por encima de 0.7 relevante.

Por su parte, el análisis factorial confirmatorio (AFC), como su nombre lo indica, comprueba y confirma a partir de una fundamentación teórica previamente definida, que los factores que agrupan las diferentes variables logran representar el modelo teórico y los datos de la muestra se ajustan a éste. Por tanto, en la medida en que el investigador tenga más información y referentes sobre la problemática de estudio, tendrá una mayor probabilidad de contrastar y comprobar hipótesis más contundentes. Debe realizarse de forma obligatoria al concluir un AFE, pues elimina las ambigüedades propias de esa técnica y permite validar los constructos resultantes mediante la asignación específica de las variables implicadas a los factores que saturan únicamente, lo que garantiza una conformación más acertada de las secciones del cuestionario.

Como parte de la validación del modelo mediante el AFC, se define si éste integra los supuestos y limitantes que se asocian a la población estudiada con la bondad de ajuste, o sea, si el modelo es adecuado para analizar la muestra en cuestión. Existen varios índices de bondad de ajuste en los que el investigador se puede apoyar mediante su contrastación para la valoración del modelo. Los principales de ellos se muestran en la Tabla 1.

Tabla 1. *Índices de bondad de ajuste*

Indicador	Descripción	Interpretación
Error de aproximación cuadrático medio (RMSEA)	Varianza no explicada en el modelo según el grado de libertad.	≤ 0.05 buen ajuste $0.05 < \leq 0.08$ ajuste razonable (Browne y Cudeck, 1993).
Chi cuadrado normalizado por los grados de libertad (CMIN/df)	Discrepancia mínima por grado de libertad.	≤ 3 Ajuste aceptable ≤ 5 ajuste razonable (Fernández García <i>et al.</i> , 2008).
Índice de ajuste incremental (IFI)	Influencia del tamaño de la muestra y parámetros añadidos.	Valores cercanos a 1 indica buen ajuste (De Souza, 2013).
Índice de ajuste no normalizado o Tucker-Lewis Index (TLI)	Discrepancia entre los modelos teórico y empírico por grados de libertad.	Valores cercanos a 1 indican ajuste aceptable (Melhado y Barroso, 2004).
Índice de ajuste comparado (CFI)	Contrasta rendimiento de los modelos empírico y teórico considerando la no existencia de correlación entre las variables.	< 0.90 Ajuste deficiente $0.90 < 0.95$ Ajuste aceptable $0.95 < \leq 1$ Ajuste muy bueno (Widaman y Thompson, 2003).

Fuente: Elaboración propia.

Luego de demostrarse la validez de constructo, existen otras pruebas a las cuales debe someterse la herramienta propuesta como son la fiabilidad, que garantiza que al aplicar cada uno de los componentes del cuestionario en diferentes momentos se obtengan resultados similares; la validación de convergencia, que evalúa el nivel de congruencia de las variables incluidas con los constructos identificados; y la validación divergente, que busca demostrar que los factores encontrados que no deben relacionarse, no se relacionan.

Validez convergente y validez discriminante

Para evaluar la correlación entre los ítems que explican de mejor manera un factor identificado, o sea, la validez convergente del instrumento, se calcula la varianza media extraída (AVE, por sus siglas en inglés). Con respecto al AVE, sus creadores (Fornell y Larcker, 1981) establecieron que los valores resultantes de este indicador se pueden interpretar de la siguiente manera: superiores a 0.5 indican un nivel aceptable de convergencia, y ya superiores a 0.7 lo consideran un nivel bueno.

Y para la validación divergente mediante la cual establecemos que los constructos difieren unos de otros de forma empírica, se pueden emplear

dos métodos: el análisis de las cargas cruzadas, en el cual se considera que cada indicador debe mostrar una carga factorial mayor al cruzarse con el factor al que pertenece, superior a 0.70; o el criterio establecido por Fornell y Larcker, quienes calculan la raíz cuadrada de AVE y establecen que el resultado debe ser superior a las correlaciones entre los demás constructos.

Fiabilidad

En lo que respecta a la fiabilidad, existen varios métodos para evaluarla. En primer lugar, se puede utilizar la técnica conocida como Test-Retest, que consiste en aplicar el cuestionario en diferentes momentos a un grupo reducido de la muestra, de tal forma que se pueda comprobar su fiabilidad mediante el contraste de las respuestas obtenidas. Otro de los métodos más empleados es el cálculo del índice de la fiabilidad compuesta (IFC) que, muy similar al Alpha de Crombach, se realiza mediante la ecuación 7.

$$IFC = \frac{(\sum_{i=1}^p \lambda_i)^2}{(\sum_{i=1}^p \lambda_i)^2 + \sum_i V(\delta)} \quad (7)$$

Donde:

λ_i = Carga completamente estandarizada para el i -ésimo indicador

$V(\delta_i)$ = Varianza del término de error para el i -ésimo indicador

p = cantidad de indicadores

La interpretación de este índice es definida por Netemeyer *et al.* (2003), quienes consideran que valores superiores a 0.8 indican una construcción confiable y bien definida.

Caso práctico. Calidad percibida y satisfacción de los servicios educativos desde la perspectiva de los estudiantes de educación superior

El caso de estudio que se presenta se llevó a cabo en el año 2022 en la Universidad Autónoma de Baja California, específicamente en la Facultad de Contaduría y Administración, campus Tijuana. El objetivo principal con-

sistió en realizar una valoración global de la calidad y la satisfacción de los servicios de apoyo al aprendizaje a partir de la perspectiva de estudiantes de la educación superior.

La investigación incluyó un total de 369 alumnos, 61.8% del género femenino y 38.2% del masculino. El tamaño de la muestra se determinó a través de un muestreo probabilístico estratificado, y para ello se estableció un nivel de confianza del 95% y un error máximo permisible del 5%. Del total de la muestra, el 31.7% corresponde a la Licenciatura de Administración de Empresas, 19.5% a la Licenciatura en Contaduría, 17.6% a la Licenciatura en Negocios Internacionales, 5.7% a la Licenciatura en Informática y el 25.5% a estudiantes que se encuentran en la etapa básica, más conocida por tronco común. Asimismo, y con el objetivo de abarcar todas las etapas de los programas educativos y la aleatoriedad del muestreo, se subdividieron las muestras según se observa en la Tabla 2.

Tabla 2. *Distribución de la muestra por etapas*

Etapa	Frecuencia	Porcentaje
Tronco común	125	33.8
Básica	60	16.3
Disciplinaria	84	22.8
Terminal	100	27.1
Total	369	100

Fuente: Elaboración propia.

De este total, el 73.2% asistía a la universidad en el turno matutino y el 26.8% restante en el turno vespertino. La recopilación de la información se llevó a cabo a través de una encuesta aplicada en el mes de mayo de 2022. El instrumento se dividió en tres secciones: la primera, referida a la recopilación de datos generales de los estudiantes, como edad, género, carrera, etapa de estudio y turno en que asiste a la universidad. La segunda sección recopila información sobre la percepción de los estudiantes en cuanto a la calidad de los servicios, basada en el modelo Servqual. A propósito, se establecieron 14 indicadores que logran representar las dimensiones de fiabilidad, capacidad de respuesta, empatía, seguridad y elementos tangibles. La escala de medida utilizada fue de tipo Likert, con cinco categorías en la siguiente forma: (1) Totalmente en desacuerdo, (2) Moderadamente en desacuerdo, (3) Ni de acuerdo ni en desacuerdo, (4) Moderadamente de acuerdo,

(5) Fuertemente de acuerdo. En la tercera y última sección, se les solicitó a los estudiantes evaluar en qué medida se encontraban satisfechos con los principales servicios ofrecidos en la institución, para ello se establecieron las siguientes escalas: (1) Totalmente insatisfecho, (2) Poco satisfecho, (3) Satisfecho, (4) Bastante satisfecho y (5) Muy satisfecho (Tabla 13).

Tabla 3. *Operacionalización de las variables*

Constructo	Indicadores
Calidad percibida (CP)	P ₁ . Los profesores de la FCA cumplen con los contenidos establecidos en la carta descriptiva.
	P ₂ . Los profesores de la FCA le comunican los criterios de evaluación a inicios del semestre y cumplen con ellos.
	P ₃ . Los profesores aclaran las dudas de los estudiantes.
	P ₄ . Los profesores usan estrategias para aplicar los conocimientos teóricos a la realidad.
	P ₅ . Cuando tengo un problema en la institución muestran interés en solucionármelo.
	P ₆ . Los procesos administrativos (inscripciones, reinscripciones, intersemestrales, servicio social, prácticas profesionales) son ágiles.
	P ₇ . Cuando necesito realizar modificaciones en algunos de mis trámites (inscripciones, reinscripciones, intersemestrales, servicio social, prácticas profesionales) recibo un trato adecuado.
	P ₈ . El personal de la FCA muestra amabilidad al atenderle.
	P ₉ . Los profesores poseen un nivel suficiente y actualizado de conocimientos teóricos y prácticos.
	P ₁₀ . El campus de la FCA es suficiente para desarrollar su función y logra un ambiente y entorno agradables.
	P ₁₁ . Existe una comunicación fluida y de confianza entre profesores y estudiantes.
	P ₁₂ . Los medios y materiales utilizados por los profesores en el proceso de enseñanza son visualmente atractivos.
	P ₁₃ . Las instalaciones físicas de la FCA son cómodas, limpias, agradables y atractivas.
	P ₁₄ . La presentación personal de los profesores se caracteriza por estar bien vestidos, limpios y aseados.
Satisfacción con los Servicios (SS)	S ₁ . Atención del personal Administrativo (secretarías, personal de limpieza).
	S ₂ . Servicios de Orientación Psicológica.
	S ₃ . Servicio de Tutorías.
	S ₄ . Servicio de Asesoría.
	S ₅ . Servicio social primera o segunda etapa.
	S ₆ . Servicios Administrativos (dirección, subdirección, coordinador de carrera, coordinador de área).
	S ₇ . Servicio de Internet.
	S ₈ . Servicios de biblioteca.
	S ₉ . Servicios de laboratorio.
	S ₁₀ . Préstamo de instrumental de enseñanza (cañón, laptop, copias).
	S ₁₁ . Servicio de limpieza.

Fuente: Elaboración propia.

Resultados

Validez de constructo

Como primer paso se procede a evaluar la validez de la escala de medida con base en la metodología de análisis factorial exploratorio (AFE). Para el ajuste del modelo se tomaron en cuenta el test de esfericidad de Bartlett y el índice KMO.

Tabla 4. *KMO y prueba de Bartlett*

Medida de adecuación muestral de Kaiser-Meyer-Olkin		0.953
	<i>Ji al cuadrado aproximado</i>	7098.73
Prueba de esfericidad de Bartlett	gl	300
	Sig.	0.00

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Como se puede apreciar en la Tabla 4, las pruebas de medida de adecuación muestral de Kaiser-Mayer-Olkin (> 0.60) y la prueba de esfericidad de Bartlett ($p < 0.05$) indican que existe una adecuada correlación entre las variables, y los estadísticos se encuentran en el rango aceptable para la aplicación del modelo. Los resultados del AFE arrojaron dos factores con autovalores superiores a la unidad y una varianza acumulada total de 62.27% (Tabla 5). En concordancia con lo anterior y partiendo de la matriz de rotación de los factores y la agrupación de las variables (Tabla 6), podemos definir el componente 1 como Calidad percibida y el componente 2 como Satisfacción con los servicios (SS).

Tabla 5. *Varianza total explicada AFE*

Componente	Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	12.024	48.09	48.09	8.432	33.726	33.726
2	3.544	14.17	62.27	7.136	28.545	62.271

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Tabla 6. *Matriz de componentes rotados*

	Componente	
	1	2
P ₉	.835	
P ₁₄	.824	
P ₁₁	.807	
P ₁₀	.802	
P ₂	.801	
P ₃	.800	
P ₁	.796	
P ₄	.762	
P ₁₂	.753	
P ₈	.714	
P ₁₃	.701	
P ₇	.662	
P ₅	.643	
P ₆	.604	
S ₆		.826
S ₈		.810
S ₄		.801
S ₉		.778
S ₂		.763
S ₁		.756
S ₃		.746
S ₅		.707
S ₁₀		.680
S ₁₁		.656
S ₇		.611

Método de extracción: Análisis de componentes principales.

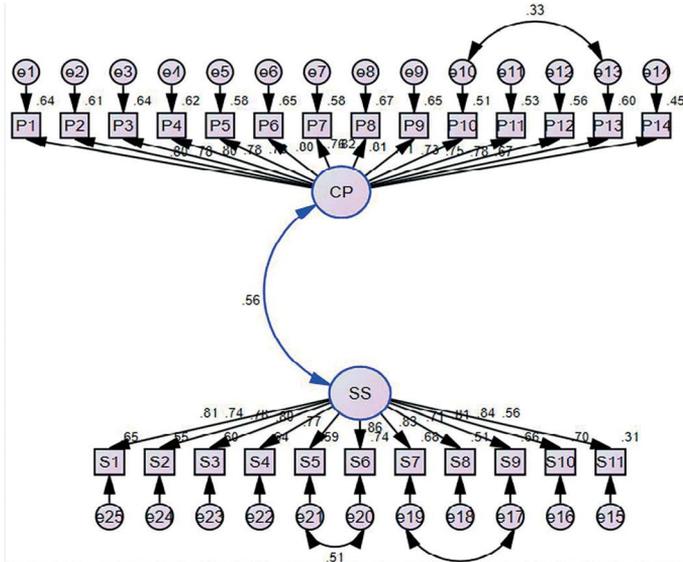
Método de rotación: Normalización varimax con Kaiser.

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Culminado el paso anterior y con el propósito de eliminar los problemas de la perspectiva exploratoria, se procedió a través del análisis factorial confirmatorio a reafirmar si el grupo de indicadores seleccionados permiten explicar las medidas propuestas en el modelo (*vid.* Figura 2).

De acuerdo con los estándares definidos para evaluar el ajuste de los modelos de análisis factorial confirmatorio: CMIN/DF (< 5), IFI (≥ 0.9), TLI (≥ 0.9), CFI (≥ 0.9) y RMSEA (≤ 0.07), podemos afirmar que el modelo propuesto logra representar la estructura de datos originales (*vid.* Tabla 7).

Figura 2. Modelo de análisis factorial confirmatorio



Fuente: Elaboración propia, con base en el sistema SPSS Amos 24.

Tabla 7. Bondad de ajuste AFC

Model	CMIN/DF	IFI Delta2	TLI rho2	CFI	RMSEA
Default model	2.797	.933	.926	.933	.069

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Fiabilidad, validez discriminante y validez convergente

Culminada la validación de los constructos del modelo, se procede a evaluar que los criterios de fiabilidad, validez discriminante y validez convergente, como se muestra en la Tabla 8. Siguiendo los planteamientos de Fornell y Lacker (1981), la medición de la confiabilidad del instrumento de recopilación de información debe de ser evaluada a través del índice de fiabilidad compuesta (IFC), que se interpreta como el alfa de Cronbach, pero tiene en cuenta las interrelaciones de los constructos extraídos. En este sentido se plantea que el valor estadístico del índice debe ser superior a 0.7.

Con relación a la validez convergente se busca evaluar el grado en que la medida de los ítems de los constructos que recogen el mismo concepto

están correlacionados. Para ello se analiza la varianza media extraída (AVE), cuyo valor recomendable debe superar el umbral de 0.5. Por último, se procede a analizar si los constructos del modelo que son diferentes no tienen ningún tipo de relación entre ellos, nos referimos a la validez discriminante. Para validar el efecto discriminador de los constructos, se utiliza la raíz cuadrada de varianza media extraída (AVE), que se ubica en la diagonal principal (0.769 -0.777) y debe ser superior a las correlaciones al cuadrado en la celda de cruce entre ambos factores (0.559).

Tabla 8. *Análisis de fiabilidad, validez convergente y validez discriminante*

	FC	AVE	MSV	MaxR(H)	CP	SS
CP	.953	0.592	0.312	0.954	0.769	
SS	.943	0.604	0.312	0.949	0.559***	0.777

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Discusión de resultados

Antes de comenzar con la discusión de resultados, se requirió realizar una agrupación en SPSS de los constructos CP y SS de la forma en que se muestra en la Tabla 9. Posterior a ello, se calcularon los estadísticos descriptivos y se analizaron posibles diferencias en los valores medios asociadas a variables independientes referidas al turno en el que asisten los estudiantes a la universidad (matutino/vespertino), a la etapa de formación profesional (tronco común, básica, disciplinaria y terminal) y la licenciatura que se encuentran cursando (tronco común, LAE, LC, LI y LNI).

Tabla 9. *Método de estimación CP-SS*

Constructo	Indicadores	Estimación
CP	$P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9,$ $P_{10}, P_{11}, P_{12}, P_{13}, P_{14}$	$CP = (\sum_{i=1}^n (Pid))/n$ $1 \leq CP \leq 5$
SS	$S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9,$ S_{10}, S_{11}	$SS = (\sum_{i=1}^n (Ss))/n$ $1 \leq SE \leq 5$

Fuente: Elaboración propia, con base en el sistema SPSS 24.

La Tabla 10 recoge la información de las medidas de tendencia central y de dispersión de los índices calculados. Como primer punto y de acuerdo

con el valor ideal (5 puntos), podemos afirmar que en general existe una buena percepción de la calidad de los servicios brindados por la facultad. En el caso de la satisfacción de los estudiantes, si bien es cierto que los alumnos en promedio se encuentran satisfechos (3.45), los estadísticos de media, mediana y moda no sobrepasan el puntaje de 3.5, lo que los ubica en el menor puntaje de los índices analizados y alejados del valor ideal (muy satisfecho - 5 puntos).

Tabla 10. *Estadísticos descriptivos CP, SS*

Estadísticos	CP	SS
Media	3.86	3.45
Mediana	4.07	3.36
Moda	5.00	3.00
Desv. típ.	0.97	0.83
Rango	4	4
Mínimo	1	1
Máximo	5	5

Fuente: Elaboración propia.

En un análisis más detallado de los ítems que obtuvieron menor calificación con relación a la Calidad percibida, se encuentran indicadores asociados a la dimensión de empatía (P5 y P7) y capacidad de respuesta (P6). Por último y de acuerdo con los datos reflejados en la Tabla 9, los tres servicios que recibieron la menor puntuación de los estudiantes fueron servicio de internet (2.29), préstamo de instrumental de enseñanza (3.46) y el servicio de orientación psicológica (3.47). En el caso del servicio de internet es importante señalar que este indicador recibió la peor puntuación de todos los analizados en el cuestionario (*vid.* Tabla 11).

Siguiendo con los objetivos planteados en la investigación, la Tabla 12 analiza posibles diferencias significativas entre los índices calculados y los turnos en los que asisten los estudiantes a la universidad. En este caso, los estadísticos para el test de prueba ($\text{sig} < 0.05$) permiten aceptar la hipótesis nula referida a la igualdad de los valores medios para los diferentes niveles de la variable independiente. Por lo tanto, podemos afirmar con 95% de confianza que las puntuaciones obtenidas en los índices CP y SS no se encuentran influenciados por el hecho de que los estudiantes asistan a la universidad en el turno matutino o vespertino.

Tabla 11. Valores medios de los indicadores

CP		SS	
Indicador	Media	Indicador	Media
P14	4.28	S8	3.72
P2	4.16	S11	3.67
P9	4.08	S9	3.66
P3	4.07	S6	3.62
P10	4.03	S1	3.59
P1	3.95	S3	3.54
P13	3.88	S5	3.55
P11	3.85	S4	3.49
P4	3.79	S2	3.47
P8	3.78	S10	3.46
P12	3.72	S7	2.29
P7	3.62		
P6	3.47		
P5	3.41		

Fuente: Elaboración propia.

Tabla 12. Anova Turno Vs CP y SS

		N	Media	F	Sig.
CP	Matutino	270	3.8807	.323	.570
	Vespertino	99	3.8153		
SS	Matutino	270	3.4370	.591	.443
	Vespertino	99	3.5124		

Notas: *significativo 0.05; ** significativo 0.10.

Fuente: Elaboración propia, con base en el sistema SPSS 24.

A diferencia del análisis anterior y de acuerdo con los estadísticos de significancia reflejados en la Tabla 13, podemos afirmar que la Calidad percibida y la Satisfacción de los servicios de apoyo muestran diferencias significativas asociadas a la etapa de estudio en la que se encuentra el estudiante.

De acuerdo con la Figura 3, es posible visualizar que los resultados obtenidos en cada una de las etapas son congruentes con el realizado a nivel agregado. Además, existe mejor valoración de la calidad percibida, seguida por el grado de satisfacción de los estudiantes con los servicios de apoyo. Un aspecto interesante es la tendencia que muestran los datos a la baja a medida que el estudiante transita por las diferentes etapas de su formación profesional. Por solo poner un ejemplo, la valoración de la calidad

disminuye en un 9.5% del tronco común a la etapa terminal y 5.14% en la Satisfacción de los estudiantes. Con relación a este último indicador y de acuerdo con el análisis gráfico, en la etapa disciplinaria existe una mejoría en el grado de satisfacción de los estudiantes. Sin embargo, ésta vuelve a caer en la etapa terminal, lo que pudiera explicarse a partir de que en este periodo se concentra la mayor parte de los servicios demandados por los estudiantes.

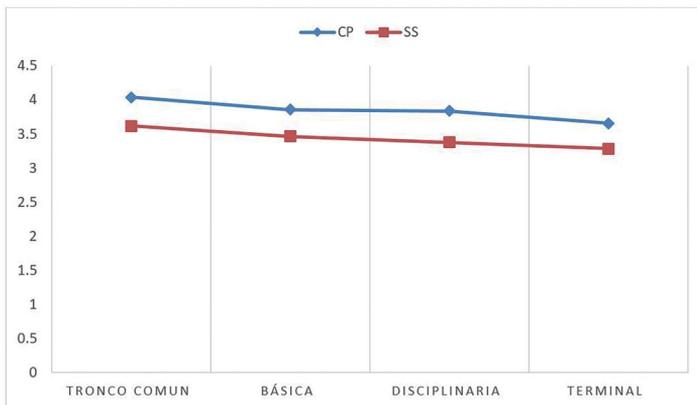
Tabla 13. Anova Etapa Vs CP y SS

		N	Media	F	Sig.
CP	Tronco Común	125	4.0480	3.028	.029*
	Básica	60	3.8571		
	Disciplinaria	84	3.8367		
	Terminal	100	3.6579		
SS	Tronco Común	125	3.6233	2.152	.025*
	Básica	60	3.4788		
	Disciplinaria	84	3.3831		
	Terminal	100	3.2991		

Notas: *significativo 0.05; ** significativo 0.10.

Fuente: Elaboración propia, con base en el sistema SPSS 24.

Figura 3. Comportamiento de CP Y SS por etapas de formación



Fuente: Elaboración propia.

Por último, se analiza si los valores medios Cp y SS difieren en función de los programas educativos que se ofrecen en la facultad. De acuerdo con

los resultados obtenidos (Tabla 14), la Calidad percibida es valorada de manera diferente en cada una de las carreras ($\text{sig}=0.013$). Caso contrario es en lo referido a la satisfacción de los servicios de apoyo, en donde no se encontraron diferencias significativas ($\text{sig} = 0.308$).

Tabla 14. *Anova Carrera VS CP y SS*

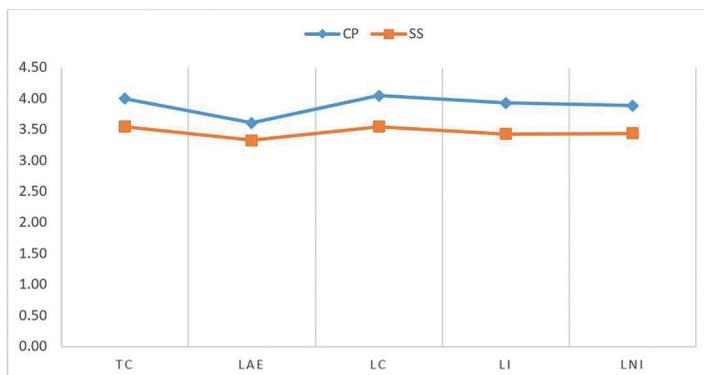
		N	Media	F	Sig.
CP	TC	94	4.000	1.164	.013
	LAE	117	3.608		
	LC	72	4.051		
	LI	21	3.925		
	LNI	65	3.893		
SS	TC	94	3.555	2.594	.308
	LAE	117	3.333		
	LC	72	3.550		
	LI	21	3.432		
	LNI	65	3.443		

Notas: *significativo 0.05; ** significativo 0.10.

Fuente: Elaboración propia, con base en el sistema SPSS 24.

En términos generales, se destaca que los estudiantes de la carrera de Licenciatura en Administración de Empresas son los que perciben en menor medida la Calidad de los servicios y, por ende, los que menos satisfechos se encuentran en la actualidad (3.33) (Figura 4).

Figura 4. *Comportamiento de la CP y SS por programa de estudio*



Fuente: Elaboración propia.

Las implicaciones de la investigación se centran en la elaboración y validación de un instrumento de investigación que, teniendo en cuenta la literatura sobre la Calidad de los servicios, permite evaluar en un contexto educativo la percepción de los estudiantes sobre la calidad de los diversos procesos que contribuyen a su formación profesional. Los resultados obtenidos demuestran que los estudiantes de la Facultad de Contaduría y Administración, campus Tijuana, tienen una percepción positiva sobre la calidad docente y los diversos servicios que la institución ofrece como soporte del proceso de aprendizaje, lo que se ve reflejado en su grado de satisfacción.

Como parte de los hallazgos encontrados, existen interrogantes que requieren ser estudiadas con mayor profundidad, como por ejemplo: ¿Qué factores determinan que a medida que los estudiantes transitan de etapas inferiores a etapas superiores de su formación profesional, la percepción de la calidad de los servicios y su nivel de satisfacción disminuyan? ¿Por qué los estudiantes de la Licenciatura de Administración de Empresas, que además es la de mayor matrícula en la FCA, son los que en promedio se encuentran menos satisfechos con las variables estudiadas? El estudio también marca pautas concretas que pueden tomarse en cuenta para elevar la calidad de los servicios asociados al empleo de la tecnología en la práctica docente y en la mejora de la infraestructura (conexión a internet). Asimismo, evidencia la necesidad de mejorar la capacidad de respuesta de los diversos procesos que funcionan en la institución.

Conclusiones

El propósito de este capítulo consistió en describir las bases teóricas y metodológicas, así como los procedimientos que deben ser aplicados en el momento de diseñar y validar un instrumento de recopilación de información primaria. Se incluyeron las principales formas de aplicación de la encuesta —en especial las más utilizadas, que son las realizadas mediante el empleo del internet y herramientas digitales—, así como los elementos de consideración para obtener resultados exitosos en este proceso. De forma específica, se hizo énfasis en la conformación del cuestionario, las características sociodemográficas de interés, los distintos tipos de preguntas y sus fun-

cionalidades, los tipos de escala que se emplean para la elaboración de los reactivos, entre las cuales sobresalen las no comparativas básicas, en particular la de Likert, por ser de las más utilizadas en el ámbito de las ciencias económico-administrativas.

En cuanto a la validación de los instrumentos, se describe el empleo de métodos de análisis factorial exploratorio y confirmatorio, que permiten establecer una estructura de factores a partir de un conjunto de variables previamente definidas. Al mismo tiempo, determinan la contribución que hace cada factor a la variabilidad total, con la finalidad de corroborar las variables que quedarán definitivamente incluidas en el estudio. De igual forma, permiten analizar el grado de aportación de las mismas, al determinar si se consideran significativas o relevantes para el estudio. Es importante destacar que esta metodología requiere de una revisión previa de la literatura teórica y empírica sobre la cual el investigador propone un modelo de relaciones esperado, el que se intentará validar en un momento posterior.

Referencias

- Amoo, T., y Friedman, H. H. (2001). Do numeric values influence subjects' responses to rating scales? *Journal of International Marketing and Marketing Research*, 26, 41-46. <https://ssrn.com/abstract=2423064>
- Browne, M. W., y Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Bollen y J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- De la Fuente Fernández, S. (2011). *Análisis factorial*. Universidad Autónoma de Madrid. <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>
- De Souza, R. (2013). *Contribuições à análise de outliers em modelos de equações estruturais*. [Tesis de maestría, Instituto de Matemática e Estatística]. Biblioteca Digital USP. <https://teses.usp.br/teses/disponiveis/45/45133/tde-19062013-135858/es.php>
- Díaz de Rada, V. (2012). Ventajas e inconvenientes de encuestas por Internet. *Papers*, 97(1), 193-223. <https://doi.org/10.5565/rev/papers/v97n1.71>
- Díaz Monroy, L. G., y Morales Rivera, M. A. (2012). *Análisis estadístico de datos multivariados*. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/79916>
- Evans, J. R., y Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195-219. <https://doi.org/10.1108/10662240510590360>
- Fernández García, E. F., Sánchez Bañuelos, F., y Salinero Martín, J. J. (2008). Validación y

- adaptación de la escala PACES de disfrute con la práctica de la actividad física para adolescentes españolas. *Psicothema*, 20(4), 890-895. <https://www.redalyc.org/pdf/727/72720457.pdf>
- Fornell, C., y Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 18(3), 382-388. <https://doi.org/10.2307/3150980>
- Guerrero Pulido, J. F. (2018). Validación de constructo a un cuestionario relacionado con el diagnóstico estratégico de las TIC en la Educación Superior. Caso de estudio. *Acción Pedagógica*, 27(1), 22-33. <http://www.saber.ula.ve/handle/123456789/46624>
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2006). Análisis de los datos cuantitativos. En *Metodología de la investigación* (4a ed., pp. 407-499). McGraw-Hill / Interamericana.
- Levin, R., Rubin, D., y Samaniego, A. (1996). *Estadística para administradores*. Prentice-Hall Hispanoamericana.
- López-Aguado, M., y Gutiérrez-Provecho, L. (2019). Com dur a terme i interpretar una anàlisi factorial exploratòria utilitzant SPSS. *REIRE, Revista d'Innovació i Recerca en Educació*, 12(2), 1-14. <http://doi.org/10.1344/reire2019.12.227057>
- López-Roldán, P., y Fachelli, S. (2016). *Metodología de la investigación social cuantitativa*. Dipòsit Digital de Documents Universitat Autònoma de Barcelona.
- Melhado, T. T., y Barroso, L. P. (2004). *Medidas de ajuste de modelos de equações estruturais*. [Tesis de maestría, Universidade de São Paulo, São Paulo]. Repositorio USP. <https://teses.usp.br/teses/disponiveis/45/45133/tde-20210729-134614>
- Méndez Martínez, C., y Rondón Sepúlveda, M. A. (2012). Introducción al análisis factorial exploratorio (AFE). *Revista Colombiana de Psiquiatría*, 41(1), 197-207. <http://www.scielo.org.co/pdf/rcp/v41n1/v41n1a14.pdf>
- Netemeyer, R. G., Bearden, W.O., y Sharma, S. (2003). *Scaling Procedures: Issues and Applications*. Sage Publications.
- Rojas, A. J., Fernández, J. S., y Pérez, C. (2000). Investigar mediante encuestas. Fundamentos teóricos y aspectos prácticos. *Psicothema*, 12(2), 320-323. <https://reunido.uniovi.es/index.php/PST/article/view/7598/7462>
- Salvador Figueras, M., y Gargallo Valero, P. (2006). *Análisis factorial*. <https://www.ciberconta.unizar.es/LECCION/Factorial/>
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. En S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). Wiley.
- Widaman, K. F y Thompson J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, 8(1), 16-37.

VII. Sistemas complejos aplicados a la investigación. El enfoque de redes sociales

NADXILLI LÓPEZ PÉREZ*
MARTHA CECILIA JARAMILLO CARDONA**

DOI: <https://doi.org/10.52501/cc.131.07>

Resumen

Este capítulo tiene por objetivo demostrar el uso del enfoque de redes, con la finalidad de conocer la transversalidad con perspectiva de género en las políticas públicas para la violencia familiar en Baja California. Se utilizó una metodología cualitativa y cuantitativa, en la cual se construyeron indicadores propuestos por la Comisión Europea (2006), con atributos transversales en las políticas públicas, tales como Liderazgo, Estructura, Capacidad y Habilidades, Participación y compromiso de la comunidad, Investigación y Evaluación, plasmados en un cuestionario de cinco preguntas abiertas, no estructuradas, con una muestra de 65 actores que conforman la red del sistema integral para prevenir y atender la violencia familiar. Se utilizaron los sistemas Nvivo y Ucinet para realizar el análisis de la información cualitativa. Los resultados arrojan la relación entre los diversos actores que conforman la red, la influencia y acción, la estructura gubernamental e interacciones para contrastar la transversalidad con perspectiva de género en las políticas públicas.

Palabras clave: *enfoque de redes, cuestionario, análisis cualitativo.*

* Doctora en Estudios del Desarrollo Global. Profesora investigadora de tiempo completo en la Facultad de Turismo y Mercadotecnia de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0002-9054-9552>

** Doctora en Ciencias Sociales con especialidad en Estudios Regionales. Profesora investigadora de tiempo completo en la Facultad de Economía y Relaciones Internacionales de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0002-8091-0551>

Introducción al enfoque de redes sociales

En la actualidad, el término red forma parte de un nuevo paradigma complejo. Diversas disciplinas utilizan el término para hacer referencia a sistemas, o células de información como los microbiólogos, al medio ambiente como sistema de redes en el caso de los ecologistas, o bien a las redes neuronales con capacidades de autoorganización y aprendizaje en la disciplina de la informática. Para las ciencias sociales contemporáneas, las redes se estudian como nuevas formas de organización social, ya sea en la economía de industrias y tecnología de redes, en la administración de los negocios y en las políticas públicas (Borzel, 1997).

En la perspectiva del análisis de la política pública se entienden las redes como la interacción entre los actores públicos y privados, al aceptar que el Estado no es el actor dominante en el diseño e implementación de la política pública. Y desde la perspectiva de las organizaciones de intereses, se utiliza el término para reconocer la existencia de dependencias mutuas entre lo privado y lo público, que afectan la dinámica interna y los mecanismos de actuación. Para realizar el análisis de las redes en la política pública, se establecen una serie de características como el número de actores, el ámbito de actuación, las funciones básicas, la estructura de la relación y estabilidad entre los actores, el grado de institucionalidad, las reglas de conducta, distribución de poder y estrategias de los actores (Jordana, 1995).

De acuerdo con Adam y Kriesi (2019), las redes de políticas constituyen una nueva forma de gobernanza que se caracteriza por las relaciones horizontales informales y descentralizadas, por lo que el enfoque de redes para el análisis de la política pública proviene de la teoría de redes sociales, que se centra en una visión de la estructura social como un conjunto de vínculos que une a los individuos y a los colectivos. De acuerdo con Santos (2003), tiene sus raíces en la sociometría, la teoría matemática de los grafos, las teorías del equilibrio estructural y la psicología social. Este autor menciona que su fundamento está basado en la sociología estructural, donde la estructura social visualiza los efectos de la conducta, las limitaciones y oportunidades de los actores, las normas culturales y otros elementos subjetivos.

Siguiendo al autor, menciona que la teoría de redes sociales permite observar una realidad social, simultánea a nivel micro a través de las relaciones de los actores individuales y a nivel macro con las relaciones entre los actores colectivos. También explica que la operatividad de ésta se centra en tres dimensiones: la primera en el efecto de la conducta de la posición en la red del actor, la segunda identifica a los subgrupos en la estructura de la red, y finalmente una tercera hace alusión a la naturaleza de las relaciones entre los actores individuales o colectivos.

De acuerdo con Adam y Kriesi (2019), se establece una tipología de redes bidimensional, caracterizada por dos tipos de variables: las de composición, referida a los atributos del actor, y la de las estructuras,¹ que aluden a los tipos específicos de relaciones entre los actores. En la dimensión referida a los atributos de los actores, se tienen las capacidades, percepciones y preferencias específicas. Por otro lado, Jordana (1995) establece una tipología que se caracteriza por número de participantes, tipo de interés, frecuencia de interacción, continuidad, consenso, distribución del recurso en la red, distribución interna, que se refiere a la capacidad y habilidades de los miembros para liderar o regular a otros miembros y el poder.

En otro sentido, Lozares Colina (1996) realiza un análisis y evoca el origen de la teoría de las redes sociales hasta llegar a formar parte de una metodología conocida el día de hoy como análisis de las redes o enfoque de redes. En su texto, menciona como a partir de los años cincuenta los sociólogos se dedican al estudio de las redes sociales y se hacen progresos en el campo de las matemáticas en la realización de investigaciones sobre urbanización y organizaciones sociales, y es partir de los años setenta cuando se establece el análisis de las redes sociales.

Los autores Brand Monsalve y Gómez (2006) argumentan que el análisis de las redes sociales es un paradigma teórico y un conjunto de técnicas de investigación, que desarrollan conceptos matemáticos con múltiples aplicaciones desde la psicología hasta la sociología, la economía y otras dis-

¹ Estas estructuras de red en la política pública están determinadas por contextos transnacionales, nacionales y el contexto específico. Estas estructuras, a su vez, están conformadas por la distribución del poder, que puede ser de concentración o fragmentación, y arrojan los resultados de interacción de los actores, ya sea de conflicto, negociación o cooperación, lo que resulta un potencial para el cambio de las políticas o un tipo de cambio de política.

ciplinas, y abarca diversos objetos de estudio como son las redes de política pública, gestión del conocimiento, relaciones comunitarias, relaciones inter-organizacionales, circuitos y relaciones de productividad, capital social, relaciones para el desarrollo y efectividad empresarial, entre otros.

Continuando con los autores, definen el análisis de redes o enfoque de redes como una metodología de investigación que permite la construcción del conocimiento sobre las múltiples relaciones existentes y los mecanismos de intercambio y poder que se conjugan, denominado análisis estructural. Según ellos, esta metodología estudia la conducta de los individuos a nivel micro, los patrones de relaciones (la estructura de la red) a nivel marco y las interrelaciones entre los dos niveles.

Ledesma (2003) menciona diversos enfoques para explicar el análisis de redes. El primero es el conocido como *networking analysis*, del cual se ha especificado en el párrafo anterior. El segundo enfoque son las redes de políticas, que estudia la centralidad de las relaciones entre el conjunto de relaciones que involucra de manera específica a las redes de organización, es decir, los roles, que combinan competencia, colaboración y las estructuras verticales y horizontales. El tercer enfoque es el relacionado con las redes de acción pública o *policy networks*, basado en la teoría del comportamiento identifica las relaciones estructurales entre las instituciones y las organizaciones, cercanas a la teoría de coordinación.

De acuerdo con la autora, el enfoque de redes aplicado al análisis del Estado se centra en la fragmentación y la diversidad que operan tanto dentro de los grupos de interés como entre los grupos y el Estado, para ocuparse de la composición de las redes y las decisiones políticas. En resumen, son el resultado de la cooperación más o menos estable no jerárquica entre individuos y organizaciones que se conocen y reconocen, al intercambiar recursos y compartir normas e intereses. Finalmente, las redes como matriz de relaciones de intercambio —que parten de la teoría de la coordinación social y la gobernanza— plantean que el mercado no es la única forma de coordinación social, sino que existen arreglos institucionales que se combinan, complementan o entran en conflicto con diferentes objetivos y mecanismos, cuando las instituciones imponen restricciones a las elecciones y preferencias de los actores.

En sintonía con ello, Casas (2003) menciona que el enfoque basado en las instituciones es fundamental para conocer los procesos de interacción, pues sostiene que las instituciones son los nodos o elementos centrales de la estructura de las redes, donde tiene lugar el desarrollo de proyectos específicos de colaboración y flujo de conocimientos, para generar marcos de referencia que dependen de su capacidad, recursos, políticas y capital institucional para la vinculación. De tal manera, estos marcos influyen en la construcción, caracterización y dinámica de redes de conocimientos, por lo que en el desarrollo de las colaboraciones un conjunto de instituciones tiene peso importante en la construcción de redes como universidades, centros de salud, instituciones, gobierno, etc.

El análisis de las políticas públicas a través del enfoque de redes se utiliza para conocer como el Estado se vincula con otros actores sociales, con la movilización de los recursos humanos, financieros y tecnológicos, y con la organización y comunicación, ya que está fuertemente influenciado por la teoría interorganizacional, que define a la red en dos vertientes: por un lado, el vehículo de comunicación entre el Estado y la sociedad y, por el otro, la herramienta de procesos de la elaboración de la política pública. De esa manera, se tiene una comprensión de los vínculos entre los actores que conforman la red (Nogueira, 2015).

Por otro lado, la red de políticas posee varias raíces, entre ellas la influencia de la teoría interorganizativa, que pone énfasis en el hecho de que los actores dependen unos de otros, ya que necesitan de los recursos del otro para poder alcanzar sus objetivos. El enfoque de redes puede utilizarse como la estructura del gobierno nueva y bien definida, que distinga las formas jerárquicas organizadas en forma vertical y las estructuras del mercado organizadas en forma horizontal, para enfatizar la coordinación horizontal y autoorganizativa entre actores privados y públicos involucrados, en este caso autónomos y autogobernantes. También se utiliza para conocer los patrones de interacción entre los actores y finalmente el enfoque formalizado cuantitativo del análisis de redes sociales, que permite la caracterización de los aspectos clave de los actores como el grado de centralización, conectividad, densidad, etc. (Adam y Kriesi, 2019).

En conclusión, y de acuerdo con los autores, la teoría de redes sociales es un paradigma compuesto por diversas ciencias como la política, la socio-

logía, la economía y la psicología. Por otro lado, la teoría de grafos permite brindar una explicación a las diferentes interacciones entre los individuos y las organizaciones, ya sean de forma vertical u horizontal. A su vez, esta teoría se fragmenta en una metodología conocida como enfoque de redes. Utilizado en esta investigación, ese enfoque es cuantitativo y permite el análisis de la política pública en su diseño, implementación y ejecución a través del examen de la coordinación e intercambio para conocer las conexiones entre las instituciones y otros actores. Además, permite determinar las preferencias y elecciones de éstos, con base en las características propias de la red, lo que facilita el análisis.

Modelo matemático de las redes

Para la correcta interpretación de los indicadores utilizados en la teoría de redes, es importante definir cada uno de ellos. Velázquez y Aguilar (2005) describen cuatro indicadores que se desarrollan para cada nodo y para toda la red:

- a) *Densidad*: Indicador para un nodo y para toda la red, que muestra alta o baja conectividad en la red. Se expresa como porcentaje del cociente entre el número de relaciones existentes con las aristas posibles.
- b) *Centralidad y Centralización*: La primera es un indicador para el nodo y no para la red completa; representa el número de actores a los cuales un actor está directamente unido. La segunda se utiliza para la red completa y no para el nodo, y muestra la condición especial en la que un actor ejerce un papel claramente central, al estar altamente conectado en la red.
- c) *Intermediación*: Indicador utilizado para el nodo y la red completa; muestra la posibilidad de un nodo para intermediar en las comunicaciones entre pares de nodos, conocidos como actores puente.
- d) *Cercanía*: Al igual que el indicador anterior, es utilizado para un nodo y para la toda la red, e indica la capacidad del actor para alcanzar a todos los nodos de la red.

Wasserman y Faust (2013) mencionan que la densidad de un grafo oscila entre 0 y 1, y está dada por:

$$\Delta = \frac{d}{(g-1)} \quad (1)$$

Donde:

Δ = Densidad del grafo

d = grado modal²

$(g-1)$ = número de líneas presentes menos uno

Y la densidad de toda la red es entendida como la densidad de un grafo y es igual a la proporción de los actos presentes en el grafo dirigido, calculado como el número de arcos, L , y dividido por el número de posibles arcos³ $g(g-1)$, por lo que en matemática se obtiene:

$$\Delta = \frac{L}{g(g-1)} \quad (2)$$

Borgatti *et al.* (2018) mencionan que en una matriz de una red la centralidad de grados es simplemente la suma de la fila o columna de la matriz de adyacencia.⁴ Si es el grado de centralidad del actor i y x_{ij} , la entrada (i, j) de la matriz de adyacencia entonces matemáticamente se representa:

$$d_i = \sum_j x_{ij} \quad (3)$$

Dependiendo de la naturaleza de los vínculos en la red, existen variedad de interpretaciones del grado de centralidad.

Por otro lado, la Centralización muestra que mientras más grande sea, es más probable que exista solo un actor central y los demás actores residan en la periferia, por lo que mide la variabilidad o lo heterogénea que son las centralidades de los actores. Matemáticamente se expresa:

² Es un estadístico que informa el grado promedio de los nodos del grafo.

³ Un arco es un par ordenados de nodos (Wasserman y Faust, 2013).

⁴ Una matriz de adyacencia es una matriz en la cual las filas y columnas representan nodos, y una entrada en la fila i y la columna j representan la línea o vínculo que conecta a partir de i a j , es decir, la matriz A de un gráfico no valorado es definida como la matriz en la cual $a_{ij} = 1$, donde hay un vínculo de i a j , y $a_{ij} = 0$, a la inversa (Borgatti *et al.*, 2018).

$$C_A = \frac{\sum_{i=j} g[C_A(n') - C_A(n_i)]}{\max \sum_{i=j} g[C_A(n') - C_A(n_i)]} \quad (4)$$

Donde:

C_A = Centralidad de la red

$g = a$ los actores

$C_A(n') = \max_i C_A(n_i)$ = al mayor valor de índice particular que se tiene entre los actores.

$\sum_{i=j} g[C_A(n') - C_A(n_i)]$ = a la suma de las diferencias entre el valor máximo y los valores observados.

$\max \sum_{i=j} g[C_A(n') - C_A(n_i)]$ = suma teórica posible de las diferencias en la centralidad de actor, que se lleva a cabo sobre todos los grafos posibles con g actores.

Los autores mencionan que el índice oscilará entre 0 y 1. C_A es igual a cero cuando todos los actores tienen el mismo índice de centralidad y es igual a 1 si existe un actor que domine o sea sombra de los demás actores.

Por otro lado, Borgatti *et al.* (2018) denotan la intermediación como:

$$b_j = \sum_{i < k} \frac{g_{ijk}}{g_{ik}} \quad (5)$$

Donde:

g_{ijk} = al número de rutas geodésicas que conectan i y k a través de j .

g_{ik} = total de rutas geodésicas que conectan i y k .

Por lo que la fórmula representa, la suma de estas proporciones en todos los pares y el resultado son un valor único para cada nodo en la red. Una intermediación es cero, cuando nunca está en el camino más corto entre otros dos, que ocurre cuando se encuentra aislado o alterado con un nodo que está conectado a otro. La intermediación tiene su máximo valor cuando el nodo se sitúa en una ruta más corta entre cada par de otros nodos.

Finalmente, siguiendo con Borgatti *et al.* (2018), la cercanía es definida como la suma de la distancia de las geodésicas de otro nodo a todos los otros nodos. La cercanía es lo inverso que la centralidad, en el sentido que un número largo indica que un nodo es altamente periférico, mientras que los números pequeños indican que el nodo es más central y se denota como el cálculo de la varianza de los índices estandarizados de cercanía del actor dada por Wasserman y Faust (2013):

$$S_c^2 = \left[\sum_{i=j}^g (C'_c(n_i) - \underline{C}_c)^2 \right] / g \quad (6)$$

Donde:

$C'_c(n_i)$ = la cercanía a nivel de actor g .

$\sum_{i=j}^g C'_c(n_i) / g = \underline{C}_c$ = cercanía normalizada, es decir, la media de las centralidades de cercanía a nivel de actor.

La varianza alcanza su valor mínimo de 0 en una red con índices de actores iguales, y este índice crece a medida que la red se haga menos homogénea con respecto a las distancias.

Caso práctico. La transversalidad de las políticas públicas con perspectiva de género para la violencia intrafamiliar: un análisis de los avances del estado de Baja California a la luz de América Latina

Recolección de la información

El enfoque de redes se aplicó para conocer la transversalidad con perspectiva de género en las políticas públicas del Estado de Baja California. Con ese propósito, se construyeron indicadores propuestos por la Comisión Europea (2006), en el informe final "Promoting the Mainstreaming of Social Inclusion" de acuerdo con los atributos que deben contener las políticas públicas para ser efectivamente transversales, tales como Liderazgo, Estructura,

tura, Capacidad y habilidades, Participación y compromiso de la comunidad, Investigación y evaluación.

Para ello, se realizó un cuestionario de cinco preguntas abiertas, no estructuradas, el cual se envió a un total de 65 actores, determinados por la red de actores que conforman el sistema integral para prevenir y atender la violencia intrafamiliar. Éstos fueron contactados vía telefónica y por correo electrónico, de ellos se obtuvo una respuesta de 30 actores. El cuestionario está estructurado de la siguiente manera:

Tabla 1. *Atributos que miden la transversalidad de las políticas públicas de acuerdo con el informe final de la Comisión Europea (2006), "Promoting the Mainstreaming of Social Inclusion"*

Pregunta	Atributo de la transversalidad que mide
<ul style="list-style-type: none"> • En su opinión, ¿cómo considera que ha impactado en su localidad la transversalidad de las políticas públicas para la violencia familiar desde las convenciones internacionales o algunos países de América Latina? 	<ul style="list-style-type: none"> • Liderazgo: Refiere el compromiso entre los países y la sensibilización del tema a través de la agenda pública. Esta pregunta considera el conocimiento sobre la transversalidad y busca conocer la influencia de los tratados internacionales, convenciones de los que México forma parte, y algunos países que pudieran ser referentes en el diseño y la implementación de la política pública en la localidad.
<ul style="list-style-type: none"> • En su opinión, ¿cómo considera las políticas (legislaciones federales, estatales, locales, planes de desarrollo federales, locales y estatales) con perspectiva de género y transversales para la atención de la violencia familiar? ¿Por qué? • Del sistema integral para la atención, prevención, sanción y erradicación de la violencia familiar, ¿cómo considera la interacción entre las dependencias (federales, estatales y locales) que conforman el sistema para abordar la problemática de la violencia familiar en el estado? ¿Considera alguna con liderazgo? ¿Por qué? 	<ul style="list-style-type: none"> • Estructura gubernamental nacional. Mide la unidad centralizada y centrada en la promoción de la política pública. En el caso de México, la promoción y el proceso de la perspectiva de género en las políticas públicas para la prevención, atención y sanción de la violencia familiar a través del conocimiento de las entidades estatales y locales.
<ul style="list-style-type: none"> • En su opinión, ¿considera que los recursos económicos (presupuestos), humanos (expertos en la materia, capacidades y habilidades) y tecnológicos (acceso y registro de información) son suficientes para la atención de la violencia familiar? ¿Por qué? 	<ul style="list-style-type: none"> • Capacidad y habilidades. Mide las habilidades y recursos humanos, técnicos y materiales que posee la administración pública para llevar a cabo el diseño o mejora de la política pública.
<ul style="list-style-type: none"> • En su opinión, ¿considera que se toma en cuenta a la población civil y a las organizaciones no gubernamentales para el diseño e implementación de las políticas públicas? ¿Por qué? 	<ul style="list-style-type: none"> • Contribución de las partes y compromiso. Mide la contribución de las partes interesadas a través del intercambio de información recabada entre los distintos departamentos, agentes gubernamentales y participación de las organizaciones y sociedad civil.

Fuente: Elaboración propia, a partir de Comisión Europea (2006).

Análisis de resultados

Una vez recabada la información con los actores descritos anteriormente, se utilizó el sistema Nvivo y Ucinet para realizar el análisis de la información cualitativa y contrastar la hipótesis:

- *Ho*: La transversalidad de las políticas públicas con perspectiva de género para la violencia familiar en América Latina impacta positivamente en las decisiones de política pública para atender, prevenir, sancionar y erradicar la violencia familiar contra las mujeres en el estado de Baja California.

Cada pregunta del cuestionario se constituyó como un nodo,⁵ lo que permitió clasificar la información y generar codificaciones o agrupaciones que aportan a la hipótesis de la investigación. El análisis de cada nodo arrojó resultados acerca del conocimiento sobre los tratados internacionales, de la percepción del impacto de la transversalidad en la ejecución de las políticas, la noción de la perspectiva de género, así como obtener información sobre los recursos de que disponen y la participación de la sociedad civil y las organizaciones no gubernamentales.

Con base en la información del análisis anterior, se procedió a construir una matriz conformada por 20 actores. En la matriz se abrevian las filas y columnas con letras que corresponden a cada uno de los actores y se describen en el documento. Los datos son binarios, donde 1 corresponde a si emite y ejecuta los lineamientos trazados por las convenciones y el Instituto Nacional de las Mujeres, como eje rector y supervisor de la implementación de la política para el tratamiento de la violencia familiar en los estados de la República Mexicana, en relación con los demás actores locales que conforman la red para el caso de Baja California, y 0 de no ser de esta manera.

- A. Convenciones y tratados internacionales (Actor construido a través de las respuestas brindadas por los entrevistados).
- B. Secretaría Distrital de la Mujer en Bogotá, Colombia.
- C. Programa para la Prevención de la Violencia Intrafamiliar (Propevi), Guatemala.

⁵ Un nodo es una colección de referencias vinculadas al tema que se está analizando (Castleberry, 2014).

- D. Instituto Nacional de las Mujeres (Inmujer), México (Actor construido a través de las respuestas brindadas por los entrevistados).
- E. Instituto de la Mujer en el estado de Baja California (Imujer BC).
- F. Instituto Chihuahuense de la Mujer (ICH Mujer).
- G. Secretaría Estatal de la Mujer en Guerrero (Ser Mujer).
- H. Instituto Estatal de las Mujeres, Tabasco (IEM).
- I. Congreso del poder legislativo del estado de Baja California.
- J. Sistema para el Desarrollo Integral de la Familia estatal (DIF estatal BC).
- K. Procuraduría General de Justicia del estado de Baja California (PGJE).
- L. Secretaría de Seguridad Pública del estado de Baja California (SSP).
- M. Sistema Estatal de Educación (SEE) del estado de Baja California.
- N. Poder Judicial del estado de Baja California.
- O. Institutos municipales de las mujeres (Integran los institutos de Ensenada, Tijuana, Tecate y Playas de Rosarito).
- P. Sistema de Desarrollo Integral para la Familia en el municipio de... (Constituye la suma de las respuestas que integran el DIF municipal de Ensenada, Playas de Rosarito, Tecate y Tijuana).
- Q. Secretaría de Desarrollo Municipal (Sedom). Comprende las respuestas de las autoridades de los municipios de Ensenada, Playas de Rosarito y Tijuana.
- R. Secretaría de Seguridad Pública Municipal. Compuesta por las respuestas de Ensenada y Playas de Rosarito.
- S. Instituto Municipal de la Juventud (Imjuve). Comprende a las autoridades de Playas de Rosarito y Tijuana.
- T. Comisión de Género del municipio. Realizada por la regidora de Ensenada.

También se creó otra matriz para conocer el grado de coordinación interinstitucional de los actores locales, conformada únicamente por ocho dependencias, de acuerdo con la pregunta número 3 sobre las dependencias que se consideraban con mayor liderazgo. En la que 1 representa la interacción entre las dependencias y 0 nula interacción, de tal manera que corroboré con las respuestas emitidas por los actores. Además, en esta matriz se agrega a la Sociedad Civil y a la Secretaría de Salud, las que comentaron que

tenían liderazgo o mayor actuación para atender, prevenir y sancionar la violencia familiar.

Matriz 1. Red de actores que implementan y ejecutan las políticas públicas para la violencia familiar

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	1	0	0		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	1		0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
F	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	1	0	0	0		1	1	1	1	1	1	1	1	1	1	1
J	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0
O	0	0	0	0	1	0	0	0	1	0	1	1	1	0		1	1	1	1	1
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
T	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	

Nota: Cada letra en la matriz corresponde a un actor.

Fuente: Elaboración propia.

En contraste con la matriz anterior, las letras de los actores son diferentes:

A. Instituto Nacional de las Mujeres.

B. Instituto Estatal de la Mujer del estado de Baja California.

C. Instituto Municipal de la Mujer (Ensenada, Mexicali, Playas de Rosarito, Tijuana y Tecate).

D. Secretaría de Salud Estatal.

E. Procuraduría General de Justicia del Estado (PGJE).

F. Sistema de Desarrollo Integral para la Familia del estado de Baja California (DIF estatal).

G. Sistema de Desarrollo Integral para la Familia Municipal (Ensenada, Mexicali, Playas de Rosarito, Tijuana y Tecate).

H. Las organizaciones de la sociedad civil (OSC).

Matriz 2. Muestra la coordinación interinstitucional de los actores locales para el tratamiento de la violencia familiar en BC

	A	B	C	D	E	F	G	H
A		1	0	0	0	0	0	0
B	1		1	0	1	1	0	0
C	0	1		0	1	0	1	1
D	0	0	0		1	0	0	0
E	0	1	1	1		1	1	0
F	0	0	0	0	1		1	0
G	0	0	1	0	1	0		1
H	0	1	1	0	0	0	1	

Nota: Las letras representan un actor.

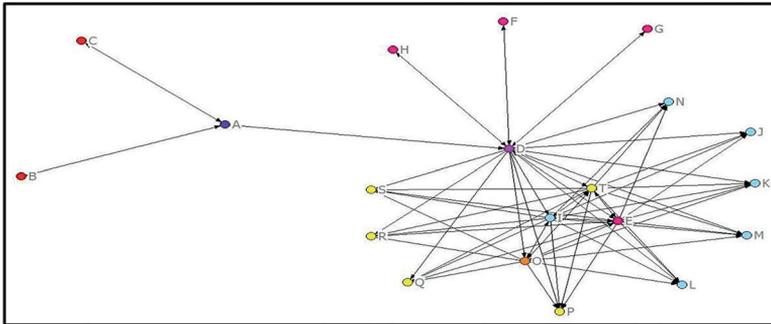
Fuente: Elaboración propia.

Con base en la información recabada, es importante conocer los resultados del análisis de la red, lo que permitirá sustentar la información cualitativa de las entrevistas con los diversos actores que abordan el tema de la violencia familiar hacia las mujeres, no solo a nivel local, estatal o federal, sino también a nivel internacional. De ese modo, es posible observar el impacto de las políticas públicas en la localidad desde las convenciones internacionales, tratados y legislaciones federales y otra red que permita conocer la coordinación interinstitucional para atender, prevenir, sancionar y erradicar la VIF.

Red de actores que explican la transversalidad de las políticas públicas para la violencia familiar

La Figura 1 muestra en color morado las convenciones internacionales que interactúan con las instituciones encargadas de la problemática de Colombia, Guatemala (color rojo, letras C y B) y con la institución en México (color rosa, letra D). A su vez, esta información es transmitida a las dependencias estatales de la República Mexicana, promotoras y ejecutoras de acuerdo con la legislación nacional, que corresponde a las instituciones estatales de las mujeres con las letras: E, F, G, H, y es la letra E la que representa al Instituto de la Mujer en el estado de Baja California.

Figura 1. Red de transversalidad de la política pública para la violencia familiar en el estado de Baja California



Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

En este sentido, siguiendo el modelo de red de Adam y Kriesi (2019), las instituciones que conforman el contexto transnacional son los organismos internacionales y los países de América Latina, mientras que en el contexto nacional se encuentran el Instituto Nacional de las Mujeres y los institutos de las mujeres en los estados de Baja California, Chihuahua, Guerrero y Tabasco. Finalmente, en el contexto específico del ámbito de las políticas, se ubican las instituciones municipales del estado de Baja California. A través de la red es posible observar la interacción.

A su vez, se observa que la red tiene forma parecida a una estrella, en cuyo centro aparecen el Instituto de la Mujer del estado, los institutos municipales de las mujeres (color naranja) y dos institutos relevantes municipales (color azul) y estatales (color amarillo). Estos actores interactúan entre sí, mientras quedan aislados las convenciones internacionales, las instituciones de otros países y los institutos estatales de las mujeres de otros estados de México.

Indicadores por nodo de la red de actores que ejecutan e implementan la política pública contra la violencia familiar en el estado de Baja California

Como se mencionó antes en el modelo matemático de redes, existen cuatro indicadores clave para cada nodo y para toda la red: densidad, centraliza-

ción, intermediación y cercanía. En la Tabla 2 se muestran los resultados para dichos parámetros en los actores ejecutores de la política pública para combatir la violencia familiar en Baja California.

Tabla 2. *Indicadores de los actores que ejecutan e implementan la transversalidad de las políticas públicas en el estado de Baja California*

Densidad ^a	0.189
Centralización ^b	Grado de centralización de salida en la red: 67% ^b Grado de centralización de entrada en la red: 7%
Intermediación ^c	35.65%
Cercanía ^d	9.958

- a. Con una desviación estándar de 0.392 y un promedio de grado por actor 3.6, número de vínculos 72 y un α de 0.824
- b. Posee una varianza de 30 en grados de salida y de 2.6 en grados de entrada, lo que indica una red centralizada y heterogénea.
- c. Posee una media de 10.050, una desviación estándar igual a 29.181, un rango máximo de 126 y mínimo de cero y una varianza de 851.514, esto indica que cada actor se encuentra muy alejado uno del otro y no tan cercanos.
- d. En un rango mínimo de 9.091 y máximo de 10.615, con una desviación estándar de 0.603 y una varianza de 0.363, con un total de observaciones para 20 actores, esto indica que los actores no son tan cercanos unos entre otros en toda la red.

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Densidad de los actores de la red

En la red se observa que los actores con mayor densidad son D, E, T, I y O, es decir, con mayor nivel de conectividad en toda la red al interactuar con otros actores (Tabla 3). El Instituto Nacional de las Mujeres es el encargado de implementar o ejecutar la política pública con perspectiva de género en todas las entidades, provenientes de los tratados internacionales, los que se encuentran distantes a esa ejecución en los países que los firmaron. Posteriormente, el Instituto de la Mujer en el estado es el encargado de ejecutar esa transversalidad con quienes interactúa (Figura 2).

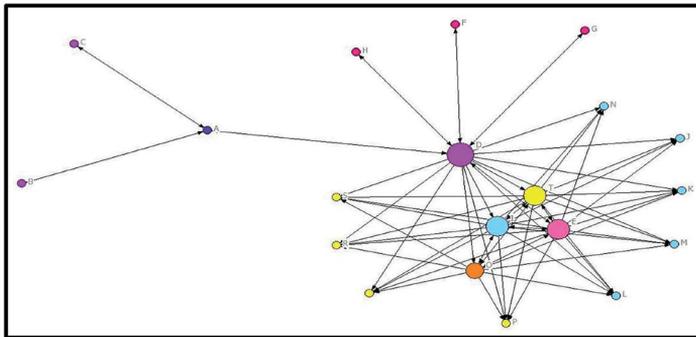
Tabla 3. *Densidad de cada uno de los actores que conforman la red en el estado de Baja California*

Actor	No. de vínculos	% de densidad
A	6	8.33
B	2	2.78
C	2	2.78
D	22	30.56

E	17	23.61
F	2	2.78
G	2	2.78
H	2	2.78
I	16	22.22
J	4	5.56
K	5	6.94
L	5	6.94
M	5	6.94
N	4	5.56
O	14	19.44
P	5	6.94
Q	5	6.94
R	5	6.94
S	5	6.94
T	16	22.22

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Figura 2. Red que muestra la densidad de los nodos para la transversalidad de la política pública para la violencia familiar en BC



Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

En los grados de centralidad se observa que los actores con mayor porcentaje de salida son D, E, I, O y T, es decir, aquellos actores que sirven de emisores en el caso de que se cumplan los mandatos establecidos en los convenios internacionales en cuanto a las políticas públicas para la violencia familiar. También es posible observar que el actor D funge como receptor, es decir, que se retroalimenta o vigila a los estados en el cumplimiento de los mandatos establecidos. Al igual que el actor anterior, el actor E vigila que se cumplan y ejecuten los procedimientos establecidos en el estado a

través de los institutos municipales de la mujer, mientras quedan en la periferia las convenciones internacionales y otros actores.

Grados de centralidad y centralización

Tabla 4. *Grados de centralidad entre los actores de la red*

Actor	% Grados de salida	% Grados de entrada
A	15	15
B	5	5
C	5	5
D	85	25
E	65	20
F	5	5
G	5	5
H	5	5
I	60	20
J	0	20
K	0	25
L	0	25
M	0	25
N	0	20
O	50	20
P	0	25
Q	0	25
R	0	25
S	0	25
T	60	20

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

El grado de centralización de los actores está dado por los resultados de las estadísticas descriptivas, el cual corresponde al promedio con un 18% en grados de entrada y salida para todos y cada uno de los actores, para los 72 posibles vínculos y con datos muy dispersos respecto a la media. La desviación estándar es de 5.481 para los grados de salida y de 1.59 para los grados de entrada. Los rangos máximos para los grados de salida oscilan entre 0 a 17, mientras que los rangos de grados de entrada oscilan entre 1 y 5.

Esto refleja que la centralización está dada por los mismos actores descritos con anterioridad, es decir, que se concentra en dos instituciones para las mujeres, la Nacional en México y la Estatal correspondiente a Baja California, y se refleja mediante la implementación de acciones y programas en

los ejes de atención, prevención y sanción para la violencia familiar y a través de los planes de desarrollo federales, estatales, y a través de los planes operativos anuales en el caso de los municipios.

Intermediación de los nodos

Los grados de intermediación están dados por los actores A, D, E, I y T. El actor con mayor grado de intermediación o puente es el D, que como se ha explicado con anterioridad corresponde al Instituto Nacional de las Mujeres. Este grado de intermediación se explica a través de los programas que promueven la transversalidad con perspectiva de género como el Programa de Apoyo a Instancias de Mujeres de las Entidades Federativas (Paimef), enfocados en prevenir y erradicar la violencia contra las mujeres y por ende la violencia familiar (Tabla 5).

Tabla 5. *Grado de intermediación de los actores en la red*

Actor	% Grado de intermediación
A	15,205
B	0
C	0
D	36,842
E	15,205
F	0
G	0
H	0
I	0,195
J	0
K	0
L	0
M	0
N	0
O	0
P	0
Q	0
R	0
S	0
T	0,195

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Los actores A y E son los segundos en grados de intermediación con un porcentaje del 15.205%. El primero corresponde a las convenciones internacionales, y este grado se explica a través de los convenios que firman los Estados como el tratado del Comité para la Eliminación de la Discriminación contra la Mujer o la Convención de Belém do Pará, entre otros. El actor E le corresponde al Instituto de la Mujer en Baja California y es intermediario no solo en la promoción de la perspectiva de género o prevención de la violencia, sino a través de las mujeres víctimas que acuden a la institución y son canalizadas a otras dependencias para su atención con servicios que no puede brindar en ese momento y debe gestionar con otros actores de la red.

Finalmente, aunque con un mejor porcentaje, los actores correspondientes a la Comisión de Género de las Municipalidades (T) y al Congreso del estado de Baja California (I) son intermediarios de otros actores, al generar políticas estatales y municipales para la prevención, atención y sanción de la violencia familiar a través de la legalidad, como la Ley General de Acceso para una Vida Libre de Violencia en el Estado, y su reglamento, o bien la creación de los institutos de las mujeres, así como la implementación de nuevas políticas a través de los consejos consultivos en la materia para ayudar a las mujeres víctimas de violencia.

Cercanía

Como se explicó en el modelo matemático de las redes, la cercanía permite conocer que tan próximo está un actor de otro que pueda ejercer cierta influencia o para el objeto de investigación colaborar de forma interinstitucional para el cumplimiento y ejecución de la política pública para la problemática que se aborda en este documento.

En la Tabla 6 se observa que los actores más cercanos entre sí son J, K, L, M, N, P, Q y R, correspondientes a autoridades estatales y municipales como la Procuraduría General de Justicia, el Instituto Municipal de la Juventud, el Sistema de Educación Estatal, el DIF estatal y municipal, la Secretaría de Desarrollo Social Municipal, la Secretaría de Seguridad Pública y el Poder de Justicia del Estado, lo que explica que conforman el Sistema Integral para Prevenir, Atender, Sancionar y Erradicar la Violencia Familiar, en menor densidad, pero cercanos al Instituto Estatal y Municipal de la

Mujer y al Instituto Nacional de las Mujeres. Esto significa que los actores con mayor cercanía son quienes ejecutan la política pública, es decir, son quienes atienden a las víctimas a través de los servicios de denuncia o bien otras acciones como refugios, con el apoyo del DIF estatal y municipal, así como becas económicas gestionadas por la Secretaría de Desarrollo Social para las mujeres que padecen la violencia familiar.

Tabla 6. Cercanía de actores que conforman la red

Actor	Cercanía
A	9,5
B	9,091
C	9,091
D	9,744
E	9,596
F	9,314
G	9,5
H	9,314
I	9,596
J	10,615
K	10,615
L	10,615
M	10,615
N	10,615
O	9,596
P	10,615
Q	10,615
R	10,615
S	10,615
T	9,596

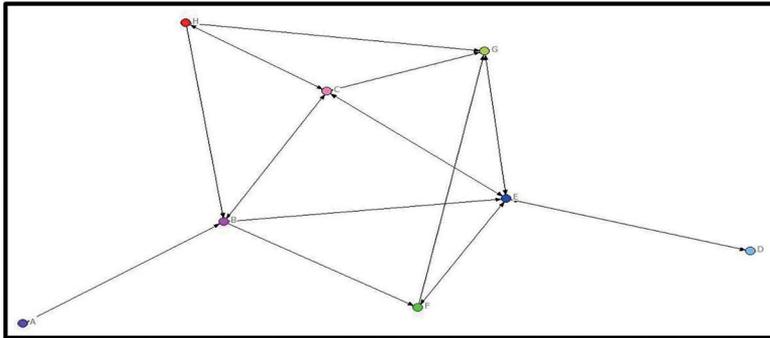
Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Red de interinstitucionalidad de los actores locales para el tratamiento de la violencia familiar en Baja California

En la Figura 3 se muestra a los actores que mencionaron las encuestas como mayores colaboradores para la atención de la violencia familiar en el estado de Baja California. Se observa en el centro de la red al actor C, que corresponde a los institutos municipales de las mujeres, y en la periferia se observa

al Instituto Nacional de las Mujeres y a la Secretaría de Salud Estatal, denotados con la letra A de color morado y D de color azul claro. Por otro lado, los indicadores generales de esta red de coordinación interinstitucional se registran en la Tabla 7.

Figura 3. Red de coordinación interinstitucional en el estado de Baja California para la Violencia Familiar



Fuente: Elaboración propia.

Tabla 7. Indicadores generales de la red

Densidad ^a	0.411
Centralización ^b	Grado de centralización de salida en la red: 26.563% Grado de centralización de entrada en la red es de 26.563%
Intermediación ^c	33.22%
Cercanía ^d	43.97%

- a. Cuenta con 23 vínculos, una desviación estándar de 0.492 y un α de 0.848.
- b. La varianza es de 1.859 en grados de salida y de 2.109 en grados de entrada, lo que significa que en los grados de salida los datos son bastante dispersos, ya que la media es de 2.875 tanto para los grados de salida y de entrada. En caso contrario, en los grados de entrada los datos son menos dispersos, más cercanos a la media, por lo que se considera una red homogénea y sin grado de centralización en los actores.
- c. La media es de 5.125, la desviación estándar es de 6.393. Los rangos mínimo y máximo son 0 a 17, y la varianza es 40.886, lo que significa que los actores se encuentran muy alejados unos de los otros, por lo que la intermediación se da en un porcentaje mínimo.
- d. Con un mínimo de 43.750 y un máximo de 77.77.

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Indicadores por nodo de la red de coordinación interinstitucional en el estado de Baja California para el tratamiento de la violencia familiar

En este apartado se muestran los indicadores de la red para cada uno de los actores que la conforman. Consiste en conocer la densidad, centralización, centralidad y cercanía de los actores con vistas a lograr la coordinación interinstitucional para prevenir, atender y sancionar la violencia familiar en el estado de Baja California.

Densidad de los actores de la red

En la Tabla 8, se muestran los resultados de la densidad de cada uno de los actores. El 43% lo obtiene el actor E, que corresponde a la Procuraduría General de Justicia del Estado (PGJE). Esto indica que tiene un rol importante en el tratamiento de la violencia familiar en la localidad, ya sea porque es el que recibe a la mayor proporción de víctimas y son registradas a través de las denuncias que llegan a cada una de las procuradurías.

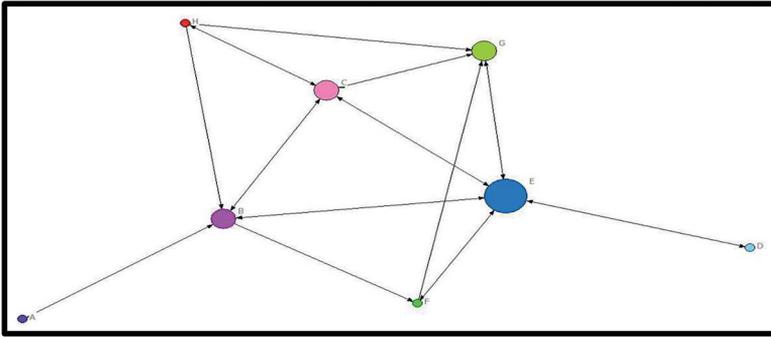
Los siguientes actores con 35% de densidad son B y C, que representan a las instituciones de las mujeres estatales y municipales, las que además de promover las políticas públicas de igualdad o de libre acceso a una vida sin violencia, son prestadoras de servicios para las víctimas y gestionan con otras dependencias programas que ayuden a las mujeres que padecen violencia familiar. Se puede observar la Figura 4 para una mejor comprensión de los resultados obtenidos en la Tabla 8.

Tabla 8. *Porcentaje de grados de densidad de los actores en coordinación interinstitucional*

Actor	% de densidad
A	9%
B	35%
C	35%
D	9%
E	43%
F	17%
G	30%
H	22%

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Figura 4. Densidad de actores en la red de coordinación interinstitucional para la violencia familiar en el estado de Baja California



Fuente: Elaboración propia,

Grados de centralidad y centralización

Los datos de la Tabla 9 muestran que el actor central es la PGJE, es decir, es tiene con mayor influencia y coordinación con otras instituciones para el tratamiento de la violencia familiar. Esto puede observarse a través de los reportes de denuncia en las procuradurías y el registro de los datos en la página web Seguridad en Baja California, que registra, entre otros delitos, la violencia de género. También es visible a través de las respuestas brindadas por los actores, en las que mencionaban mayormente a la Procuraduría General en la solución de la problemática, más allá de la prevención y atención, sino como la entidad donde se pretende erradicar la violencia familiar.

En segunda instancia, se encuentran nuevamente los institutos de las mujeres estatales y municipales, que como se ha mencionado con anterioridad no sólo ejecutan y promueven las leyes para atender, prevenir, sancionar y erradicar la violencia contra las mujeres, sino también como prestadoras de servicios y canalizadoras a otras instituciones que intervienen en la problemática. Otra institución con un rol relevante es el DIF municipal con 50% de grados de entrada, es decir, que con base en sus funciones, también atiende la problemática sin canalizarlo a otras dependencias.

Tabla 9. *Grados de centralidad y centralización para cada actor en la red de coordinación interinstitucional para la violencia familiar*

Actor	% Grados de salida	% Grados de entrada
A	12,5	12,5
B	50	50
C	50	50
D	12,5	12,5
E	65,5	62,5
F	25	25
G	37,5	50
H	37,5	25

Nota. La media para los grados de entrada y salida es de 2.875, mientras que la desviación estándar varía, con 1.364 para los grados de salida y 1.452 para los grados de entrada. La varianza es de 1.859 para los grados de salida y 2.109 para los grados de entrada. La red muestra ocho actores con 23 posibles conexiones entre cada uno de ellos.

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Intermediación de los nodos

Los resultados arrojan que los actores con mayor porcentaje de intermediación son la Procuraduría General de Justicia del estado (E) y el Inmujer BC (B). Esto significa que transfieren a otras dependencias a las víctimas para la atención psicológica o legal. También indica que son los actores con mayor actividad para atender el problema de la violencia familiar (Tabla 10).

Tabla 10. *Intermediación de los nodos en la red de coordinación interinstitucional*

Actor	% Intermediación
A	0
B	33,73
C	11,905
D	0
E	41,27
F	1,587
G	7,54
H	1,587

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Igualmente, se observa que el Inmujer Nacional y la Secretaría de Salud Estatal (SSE) quedan fuera de la línea de intermediación, es decir, si se conectan con otros actores para atender a las víctimas o bien recibir alguna información relevante sobre cambios en las legislaciones o rendición de cuentas para el caso del Inmujer Nacional, a través de los programas que se implementan por medio de la federación, y en el caso de la SSE, a través de los servicios de atención médica para las víctimas de violencia; pero no fungen como intermediarios con otros actores dentro de la red.

Cercanía

Al observar la red, es posible visualizar a los actores más lejanos, tal es el caso del Inmujer Nacional y la Secretaría de Salud Estatal, aunque trabajen en colaboración. Asimismo, los actores más cercanos y con mayor comunicación entre ellos son la PGJE y los institutos estatales y municipales de la mujer. Esto se relaciona con las respuestas de los actores en cuanto a la institución con mayor liderazgo en el tema. Cabe señalar que la PGJE como último eslabón en el tratamiento de la violencia familiar, es el primero en enterarse de los hechos, según se menciona a través de las denuncias o reportes por la Secretaría de Seguridad Pública para dar seguimiento a los casos. Finalmente, los institutos de las mujeres estatales y municipales juegan un rol muy importante en la transferencia de la información, promoción y ejecución, pero también en la atención a las víctimas (Tabla 11).

Tabla 11. *Porcentaje de grados de cercanía entre los actores para la atención de la violencia familiar*

Actor	% de grado de cercanía
A	43.75
B	70
C	70
D	46.66
E	77.77
F	65.63
G	63.63
H	58.33

Fuente: Elaboración propia, con resultados obtenidos en sistema Ucinet.

Asimismo, las otras dependencias realizan sus actividades y colaboran entre sí, aunque en menor comunicación unas con otras, como el DIF estatal y municipal, y las organizaciones civiles. Sin embargo, es notorio que la cercanía supera el 50%, lo que confirma lo dicho por las autoridades entrevistadas de que existe una buena colaboración y coordinación entre las autoridades en atención a este problema social.

Conclusiones

En este capítulo se abordan las conclusiones generales de los hallazgos encontrados y finalmente la aprobación de la hipótesis, que la transversalidad de las políticas públicas sí impacta de forma positiva en el estado de Baja California, aunque no de la forma esperada, es decir, solo con pequeños avances a corto plazo. Es importante recordar que la transversalidad está compuesta por seis atributos fundamentales de acuerdo con la Comisión Europea (2006), que son liderazgo, estructura gubernamental, capacidad y actitudes, participación y compromiso de la comunidad, investigación y evaluación, mismos que se contemplaron en el instrumento de investigación para cada uno de los actores.

Con el uso del enfoque de redes sociales, se realizó la medición de atributos que podrían ser considerados cualitativos, si observamos la relación entre los diversos actores. De esa manera, se demostró la influencia y acción de los actores que conforman la red para la prevención, atención y sanción de la violencia de género. Asimismo, permitió demostrar la centralización en la estructura gubernamental para la promoción de los programas contra la violencia hacia la mujer con perspectiva de género. De igual manera, corroboró el intento de las instituciones mexicanas, a través de los resultados de las interacciones, por armonizar las leyes actuales en materia de igualdad y para el tratamiento de la violencia familiar.

Finalmente, los indicadores utilizados en la teoría de redes por Velázquez y Aguilar (2005) permitieron conocer a través de la densidad el número de actores relacionados entre sí, para dar seguimiento a los marcos legales, lo que es característico de la definición de transversalidad. Además, posibilitaron identificar la cantidad de actores fuertemente relacionados a través

de la centralidad, mientras que la centralización permitió observar la institución que juega un papel predominante de influencia y poder sobre las demás entidades. Por otro lado, la intermediación contribuyó a comprender la transmisión de la información entre pares, conocidos como actores puente, y la cercanía, la capacidad que tiene el actor para alcanzar todos los nodos de la red, es decir, a todas las instituciones que conforman el sistema integral para prevenir, atender y sancionar la violencia hacia la mujer.

Referencias

- Adam, S., y Kriesi, H. (2019). The Network Approach. En P. A. Sabatier (Ed.), *Theories of the policy process* (2da ed., pp. 129-154). Westview.
- Borgatti, S. P., Everett, M. G., y Jhonson, J. C. (2018). *Analyzing social network*. SAGE.
- Borzel, T. A. (1997). ¿Qué tienen de especial los policy networks?: Explorando el concepto y utilidad para el estudio de la gobernación europea. *Redes verdes de transformación territorial y social*. <http://revista-redes.rediris.es/webredes/textos/policy-net.pdf>
- Brand Monsalve, E. G., y Gómez, H. (2006). Análisis de redes sociales como metodología de investigación. Elementos básicos y aplicación. *La Sociología en sus Escenarios*, (13), 1–28. http://bibliotecadigital.udea.edu.co/bitstream/10495/2542/1/BrandEdinson_analisisredesmetodologiainvestigacion.pdf
- Casas, R. (2003). Enfoque para el análisis de redes y flujos de conocimiento. En M. Luna (Coord.), *Itinerarios del conocimiento: formas dinámicas y contenido: un enfoque de redes* (pp. 19-50). Anthropos.
- Castleberry, A. (2014). NVivo 10 [software program]. *American Journal of Pharmaceutical Education*, 78(1), 25. <https://doi.org/10.5688/ajpe78125>
- Comisión Europea. (2006). Better policies Better outcomes, *Promoting the Mainstreaming of Social Inclusion*. MSI.Final Report, Combat Poverty Agency-European Comission Project Promoting the Mainstreaming of Social Inclusion, Dublín.
- Jordana, J. (1995). El análisis de los policy networks: ¿una nueva perspectiva sobre la relación entre políticas públicas y Estado? *Gestión y Análisis de Políticas Públicas*, (3), 77-89. <http://maxicamposrios.com.ar/wp-content/uploads/2014/03/JORDANA-J.-El-analisis-de-los-policy-networks.pdf>
- Ledesma, M. L. (2003). La red como mecanismo de coordinación y las redes de conocimiento. En M. Luna (Coord.), *Itinerarios del conocimiento: formas dinámicas y contenido: Un enfoque de redes* (pp. 51-78). Anthropos.
- Lozares Colina, C. (1996). La teoría de redes sociales. *Papers: Revista de Sociología*, (48), 103-126. <https://raco.cat/index.php/Papers/article/view/25386>
- Nogueira, M. E. (2015). Las redes de políticas como herramientas de análisis: notas a partir de intervenciones recientes en materia de desarrollo rural en Argentina 2008-

2014. *Cuadernos de Gobierno y Administración Pública*, 2(1). https://doi.org/10.5209/rev_CGAP.2015.v2.n1.49429
- Santos, F. R. (2003). *Análisis de redes sociales: orígenes, teorías y aplicaciones (No. 198)*. Centro de Investigaciones Sociológicas.
- Velázquez, A., y Aguilar, N. (2005). Manual introductorio al análisis de redes sociales. *Ejemplos Prácticos con UCINET*, 6. https://www.researchgate.net/publication/264311672_Manual_introductorio_al_analisis_de_redes_sociales_Medidas_de_centralidad
- Wasserman, S., y Faust, K. (2013). *Análisis de redes sociales. Métodos y aplicaciones*. Centro de Investigaciones Sociológicas.

VIII. Índice de confianza ciudadana. Percepción sobre economía y seguridad en la Zona Metropolitana de Tijuana

ISMAEL PLASCENCIA LÓPEZ*

OSWALDO PLASCENCIA LÓPEZ**

ABRAHAM NEFTALÍ SODI DÍAZ***

DOI: <https://doi.org/10.52501/cc.131.08>

Resumen

Este capítulo describe el desarrollo del índice de confianza ciudadana (ICC), parámetro que surge por la necesidad de ejemplificar herramientas de generación de información para toma de decisiones en los cursos del Doctorado en Ciencias Administrativas. El ICC busca capturar el sentir de los habitantes con respecto a su situación económica y la perspectiva de seguridad en la Zona Metropolitana de Tijuana, compuesta por los municipios de Tijuana, Rosarito y Tecate. Para su desarrollo, se tomaron como base dos instrumentos del Inegi: la Encuesta Nacional sobre la Confianza del Consumidor y la Encuesta de Seguridad Pública Urbana. Con reactivos puntuales de cada encuesta, se elaboró un nuevo instrumento que consta de 80 reactivos sobre temáticas socioeconómicas, económicas y de seguridad. La encuesta se aplica cada mes a jefes de familia, y los resultados se agregan cada

* Doctor en Ciencias Económicas. Director de la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-2860-1417>

** Maestro en Administración. Alumno del Doctorado en Ciencia Administrativas de la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0009-0002-5467-6850>

*** Maestro en Mercadotecnia. Profesor de tiempo completo en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC), México. ORCID: <https://orcid.org/0000-0003-0865-3594>

vez para realizar análisis temporal sobre el comportamiento de las expectativas en la comunidad.

Palabras clave: *índice, confianza, habitantes, economía, seguridad, Zona Metropolitana de Tijuana.*

Introducción al índice de confianza ciudadana

¿Por qué un índice de confianza ciudadana? Éste fue creado durante el curso de Estadística del Doctorado en Ciencias Administrativas en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California (UABC) en enero de 2017, como un ejercicio para explicar la importancia de la estadística en la toma de decisiones. Se diseñó una encuesta a partir de dos cuestionarios del Instituto Nacional de Estadística, Geografía e Informática (Inegi). Con apoyo de los estudiantes del doctorado, la mayoría profesores de la facultad, se comenzó a pilotear para ir afinando el instrumento.

En este capítulo se explica paso a paso cómo se construye un índice a partir de indicadores o variables que nos permitan medir la percepción ciudadana con respecto a dos temas muy importantes: la situación económica, medida por el índice de confianza del consumidor (ICC), y la percepción de inseguridad, medida por la Encuesta Nacional de Seguridad Urbana (ENSU), ambos instrumentos del Instituto Nacional de Estadística, Geografía e Informática. Con la adaptación de ambos instrumentos en uno solo, se aplica una encuesta mensual en 400 hogares de la Zona Metropolitana de Tijuana (ZMT), que comprende los municipios de Playas de Rosarito, Tecate y Tijuana.

Este índice surge de la necesidad de medir con mayor nivel de precisión cómo están viviendo los habitantes de la ZMT la situación económica y la inseguridad. Con la ponderación de ambos conceptos, se construye el índice de confianza ciudadana a partir de las siguientes premisas: Para que exista inversión y generación de empleo debe garantizarse la seguridad. A pesar de que existan niveles de inseguridad, la dinámica de la ciudad fron-

teriza sigue desarrollándose; ¿pero de qué sirve que a un ciudadano le vaya bien en su trabajo o negocio si no se siente seguro?

El objetivo de este índice es integrar dos instrumentos que se abordan por separado y adaptarlo a las condiciones y necesidades de la ZMT. Esta metodología puede ser replicada por otras ciudades o regiones del país, ya que las publicaciones del Inegi generalmente son nacionales y cada entidad, ciudad o región tiene dinámicas económicas y de seguridad con fuertes diferencias entre ellas, por lo que es muy importante que éstas no se pierdan o se diluyan en datos agregados y estadísticos promedio.

Antecedentes

En la introducción se explicó que el índice de confianza ciudadana surge de combinar indicadores de dos instrumentos o encuestas aplicados por el Inegi: el índice de confianza del consumidor (ICC) y la Encuesta Nacional de Seguridad Urbana (ENSU). En este apartado se realiza una breve explicación de cada instrumento.

El índice de confianza del consumidor (ICC)

Existen muchos índices e indicadores que pueden apoyar a los inversionistas y economistas a predecir hacia dónde se dirige la actividad económica, como son el índice de precios al consumidor (IPC), el índice de precios al productor (IPP) y el producto interno bruto (PIB). El índice de confianza del consumidor (ICC) es otro indicador clave diseñado para ilustrar el sentimiento promedio del nivel de confianza de los consumidores.

El ICC tiene sus orígenes en los años cuarenta cuando el profesor de la Universidad de Michigan George Katona creó el índice del sentimiento del consumidor de Michigan, en el Instituto de Investigación Social de esa universidad. Desde la hipótesis de las expectativas racionales, un índice de confianza debe ser el valor esperado de las variables macroeconómicas. Katona (1951) argumenta que la confianza del consumidor no sólo refleja condiciones económicas, sino también estados subjetivos de la mente de éste.

Para Katona, el nivel de confianza sobre las condiciones financieras futuras de los consumidores es el principal determinante de sus decisiones de consumo actuales. En su estudio de 1975 demuestra que la confianza del consumidor es afectada por factores psicológicos durante eventos inusuales que provocan que las actitudes de las personas cambien repentinamente (Kilic y Cankaya, 2016). Cuando las personas consideran que vienen condiciones económicas adversas y ven a algunos conocidos perder sus empleos, se preocupan y dejan de consumir para ahorrar. Si pocas personas piensan así, no hay problema; pero si muchas lo piensan, el consumo y las ventas caen y, con ello, puede existir un aumento en el desempleo por lo que la pesadilla se convierte en realidad.

Actualmente, el índice de confianza del consumidor ha sido adoptado por muchos países, y México no ha sido la excepción. Es el Inegi el encargado de aplicar la encuesta, que explica en su documento metodológico llamado Encuesta Nacional sobre la Confianza del Consumidor 2015. En 1993, el Inegi levantó la primera encuesta en las 36 principales ciudades. En 2007 se hizo un rediseño estratégico y se distribuyó una muestra total de 2 336 viviendas repartidas en las 32 ciudades más grandes de cada entidad federativa, no necesariamente una capital (Inegi, 2016).¹

El cuestionario se divide en dos apartados: el socioeconómico y el básico, con cinco preguntas principales o preguntas base de donde se construye el índice de confianza del consumidor:

1. Comparada con la situación económica que los miembros de este hogar tenían hace 12 meses, ¿cómo cree que es su situación en este momento?
2. ¿Cómo considera usted que será la situación económica de los miembros de este hogar dentro de 12 meses, respecto a la actual?
3. ¿Cómo considera usted la situación económica del país hoy en día comparada con la de hace 12 meses?
4. ¿Cómo considera usted que será la condición económica del país dentro de 12 meses respecto a la actual situación?
5. Comparando la situación económica actual con la de hace un año, ¿cómo considera en el momento actual las posibilidades de que us-

¹ Para una mejor explicación metodológica sobre la determinación del tamaño de muestra, el objeto y sujeto de estudio, consultar el documento metodológico del Inegi.

ted o alguno de los integrantes de este hogar realice compras tales como muebles, televisor, lavadora, otros aparatos electrodomésticos, etcétera?

Las respuestas utilizan escala de Likert: mucho mejor, mejor, igual, peor y mucho peor, con ponderadores de 1, 0.75, 0.50, 0.25 y 0, respectivamente. La pregunta 8 tiene cuatro opciones: mayores (1.0), iguales (0.50), menores (0.0) y no sabe, que se prorratea entre las demás. En el apartado metodológico se continúa con el desarrollo para la obtención del índice de confianza. Ahora pasamos a una breve descripción de la ENSU.

La Encuesta de Seguridad Pública Urbana (ENSU)

El crimen y la percepción de seguridad son dos conceptos que se entrelazan y que afectan la calidad de vida y el desarrollo económico de una sociedad. Sin seguridad pública garantizada, los costos de transacción de las empresas se elevan y, con ello, disminuye la inversión y el empleo. Por esa razón, no se puede desligar la inseguridad de la economía.

En el ámbito internacional, un artículo de Zhang *et al.* (2021) demuestra estadísticamente y de forma georreferenciada como las medidas de percepción de seguridad están significativamente asociadas con los crímenes reportados en ciudades urbanas. No obstante, señalan los autores que pueden existir “sesgos de percepción” entre percepción de seguridad y crímenes reportados; así, zonas que se perciben más seguras pueden ser inseguras, y zonas y lugares que se perciben inseguras pueden ser más seguras (Zhang *et al.*, 2021).

Al igual que sucede con la confianza del consumidor sobre las perspectivas de la economía, la percepción sobre la seguridad o inseguridad pública refleja el sentir de la población. En el ámbito nacional, de acuerdo con Triana (2021), los factores determinantes en la construcción subjetiva de la inseguridad y el temor ante la delincuencia surgen de la victimización, la vulnerabilidad física y social, incivildades y redes sociales en entornos cotidianos. Este autor demuestra que la prevalencia de conductas delictivas en el entorno es el principal predictor de la percepción de inseguridad y el

temor al delito, entre otros factores, en la ciudad de Acapulco, Guerrero (Triana, 2021).

Lo anterior es importante destacarlo, porque en ocasiones se quieren desestimar los resultados de estos ejercicios, al aducir que son percepciones y no realidades. Sin embargo, no necesitamos que nos roben nuestro vehículo o nos asalten personalmente, con que le suceda a un familiar o vecino es motivo suficiente para sentirnos inseguros, pensando que seremos la próxima víctima.

La Encuesta Nacional de Seguridad Pública Urbana (ENSU) comienza su levantamiento en septiembre de 2013, con el objetivo de generar estimaciones acerca de la percepción de la población sobre la seguridad pública. El objetivo general es obtener información que permita realizar estimaciones con representatividad a nivel nacional urbano sobre la percepción de la población en cuanto a la seguridad pública en su ciudad (Inegi, 2023).

La periodicidad es trimestral, y la cobertura geográfica es nacional urbano. El tamaño de la muestra nacional llega a 25 500 hogares, la unidad de muestreo es la vivienda, y el esquema de muestreo es probabilístico, trietápico, estratificado y por conglomerados. De la información recabada, el Inegi publica tabulados, microdatos y datos abiertos, así como presentaciones ejecutivas trimestrales.² Los conceptos más relevantes que incluye la ENSU (Inegi, 2023) son:

- Sensación de inseguridad por temor al delito (percepción)
- Expectativa social sobre tendencia del delito (percepción)
- Atestiguación de conductas delictivas o antisociales (conocimiento)
- Cambio de rutinas por temor a ser víctima del delito (experiencias)
- Percepción del desempeño de las autoridades sobre seguridad pública (percepción)
- Desempeño gubernamental (percepción)
- Movilidad (experiencias)

En el siguiente apartado explicamos los pasos para llegar a la construcción del índice de confianza ciudadana aplicado en la Zona Metropolitana de Tijuana (ZMT), que incluye los municipios de Playas de Rosarito, Tecate y el municipio más poblado de México: Tijuana.

²Para una mejor explicación metodológica, consultar los documentos metodológicos disponibles en la página del Inegi: <https://www.inegi.org.mx/programas/ensu/>

Metodología

La idea original fue construir un índice que considerara dos aspectos fundamentales de la vida urbana: el desempeño de la economía local y la seguridad pública. Para ello, se tomaron dos instrumentos ya existentes que se describieron en el apartado anterior: la Encuesta Nacional sobre Confianza del Consumidor (ENCO) y la Encuesta Nacional de Seguridad Pública Urbana (ENSU), ambos aplicados en el ámbito nacional y regional por el Inegi.

Cuestionario híbrido de la ENCO y la ENSU

Se realizó un análisis minucioso para discriminar preguntas menos importantes y dejar aquéllas que más servían a nuestro propósito, ya que el reto era tener una combinación de ambos cuestionarios, ENCO y ENSU, sin que el nuevo instrumento híbrido resultara extenso. Este cuestionario habría de aplicarse por nuestros alumnos de la Facultad de Contaduría y Administración de las materias de Fundamentos de Investigación Metodológica y Economía a jefe o jefa de familia de tres hogares distintos.

Al cuestionario final se le agregaron preguntas para evaluar a los mandatarios ejecutivos municipal, estatal y federal. El cuestionario final, que se aplica de manera formal desde marzo de 2017, se puede ver en el Anexo A.³

Entre las principales ventajas que podemos destacar de la aplicación de nuestro propio cuestionario aparecen:

- Los resultados nacionales publicados por el Inegi no necesariamente reflejan los resultados regionales y locales, pues al ser promedios, las diferencias se pierden. La dinámica de una ciudad fronteriza como Tijuana quedó evidente durante la pandemia. Fue la primera en ser afectada y la primera en recuperarse por muchos meses, en comparación con el resto de las ciudades o zonas metropolitanas del país.

³ El instrumento se piloteó en enero y febrero de 2017 a cien hogares de la Ciudad de Tijuana, para ajustar, calibrar y hacerlo extensivo a Tecate y Playas de Rosarito.

- Los resultados del Inegi tienen significancia estadística a nivel nacional (por la suma de todas las ciudades o zonas urbanas), pero no son estadísticamente significativos a nivel local.
- La ENSU se aplica trimestralmente, mientras que nuestro cuestionario se aplica mensualmente, lo que permite un mejor pulso de cambios a corto plazo, además de que se empata con la ENCO que es mensual, para la formación del índice.
- Mientras el Inegi aplica cuestionarios en menos de cien hogares de Tijuana, nuestra cobertura va de entre 350 a 550 cuestionarios por mes en hogares en la ZMT.

Construcción del ICC

Un índice es una medida compuesta de variables, o una forma de medir un constructo mediante más de un dato. El índice es la acumulación de marcadores de una variedad de datos individuales. Para crear un índice, debemos seleccionar los ítems del constructo o variables posibles, examinar su relación empírica, marcar el índice y validarlo (Boslaugh, 2015; Lind *et al.*, 2015; Watters, 2008).

Figura 1. Construcción de un índice



Fuente: Elaboración propia.

Para contrastar la hipótesis y cumplir con los objetivos planteados, se desarrolla esta investigación, que tiene un alcance descriptivo-correlacional con un diseño no experimental, ya que no se manipuló ninguna de las variables planteadas de acuerdo con lo establecido por Hernández Sampieri *et al.* (2010). La información fue recolectada a través del instrumento de investigación en un solo momento, es decir, en un tiempo único, lo cual se denomina corte transversal; pero el instrumento se aplicará de manera mensual en la ciudad, con el fin de poder realizar mediciones en series de tiempo y poder efectuar simulaciones del comportamiento del índice basado en la percepción de los ciudadanos.

Se aplicó un instrumento que conjunta variables de la ENCO y ENSU, y a partir de ello se determinó un índice, por lo cual la investigación se considera como de tipo cuantitativo. El índice de confianza ciudadana (ICC) se centra en medir la percepción de los habitantes de la ciudad en dos dimensiones principalmente, la seguridad y la situación económica. La Tabla 1 muestra la descripción de las variables utilizadas para medir la dimensión de seguridad y la descripción de las variables utilizadas para medir la dimensión de situación económica (ENCO, 2017; ENSU, 2017). Se estudió a los habitantes de Ciudad Obregón en Sonora, México, al considerar una muestra no probabilística y a conveniencia de 200 personas, quienes decidieron contestar la encuesta y cumplían con las características de ser mayores de edad y habitar en la ciudad objeto de estudio.

Tabla 1. Descripción de las variables

Dimensión	Variables	Pregunta	Opciones de respuesta
Seguridad	Delincuencia	En términos de delincuencia, ¿considera que vivir actualmente en su ciudad es...?	<ul style="list-style-type: none"> • Seguro • Inseguro • No sabe / No responde
Situación económica	Situación económica actual del hogar	Comparada con la situación económica que los miembros de este hogar tenían hace 12 meses, ¿cómo cree que es su situación en este momento?	<ul style="list-style-type: none"> • Mucho mejor • Mejor • Igual • Peor • Mucho peor • No sabe
Situación económica	Situación económica futura del hogar	¿Cómo considera usted que será la situación económica de los miembros de este hogar dentro de 12 meses, respecto a la actual?	<ul style="list-style-type: none"> • Mucho mejor • Mejor • Igual • Peor • Mucho peor • No sabe
Situación económica	Situación económica actual de la ciudad	¿Cómo considera la situación económica de la ciudad hoy en día comparada con la de hace 12 meses?	<ul style="list-style-type: none"> • Mucho mejor • Mejor • Igual • Peor • Mucho peor • No sabe
Situación económica	Situación económica futura de la ciudad	¿Cómo considera usted que será la condición económica de su ciudad dentro de 12 meses respecto a la actual situación?	<ul style="list-style-type: none"> • Mucho mejor • Mejor • Igual • Peor • Mucho peor • No sabe

Situación económica	Posibilidad actual de realizar compras	Comparando la situación económica actual con la de hace un año, ¿cómo considera en el momento actual las posibilidades de que usted o alguno de los integrantes de este hogar realice compras como muebles, televisor, lavadoras, otros aparatos electrodomésticos, etc.?	<ul style="list-style-type: none"> • Mayores • Iguales • Menores • No sabe
---------------------	--	---	--

Fuente: Elaboración propia basada en ENCO (2017) y ENSU (2017).

Se adaptó un instrumento para medir la percepción acerca de la seguridad y la situación económica de la ciudad, el cual consta de 80 reactivos que miden aspectos generales de los encuestados, como dirección, correo electrónico, sexo, edad, ocupación, entre otros, y aspectos de su percepción con relación a la situación económica y la seguridad de su ciudad. La muestra está representada por 200 jefes de hogar, mayores de 18 años, por lo cual el objeto de estudio no son las personas, sino los hogares representados por quienes contestaron el instrumento.

Se aplicó una prueba piloto para medir la fiabilidad del instrumento, y se obtuvo un Alpha de Cronbach de 0.87. Para calcular el ICC se tomaron siete preguntas, una para medir la dimensión de seguridad, cinco para medir la situación económica y una más como ponderador con vistas a determinar el peso entre situación económica y seguridad. Para determinar la percepción con relación a la dimensión de seguridad, se utilizó la variable y la pregunta descritas en la Tabla 1, cuyas posibles respuestas y ponderaciones se presentan en la Tabla 2.

Tabla 2. *Ponderador dimensión seguridad*

Opciones de respuesta	Ponderador
Seguro	1.00
Inseguro	0.00
No sabe	NA

Fuente: Elaboración propia basada en ENSU (2017).

A partir del ponderador se obtiene un índice para la dimensión seguridad. En el caso de la dimensión situación económica, el índice resulta de

promediar los índices de las cinco variables que miden dicha dimensión. A su vez, cada componente se calcula como el promedio ponderado de respuestas expandidas de cada una de las preguntas referentes a cada una de las variables que miden la dimensión, y cuya descripción se presenta en la Tabla 1. En cada una de las cuatro primeras preguntas, los entrevistados tienen cinco opciones de respuesta: mucho mejor, mejor, igual, peor y mucho peor. En la quinta pregunta las opciones son: mayores, iguales y menores. Los ponderadores utilizados para cada opción de respuesta se presentan en la Tabla 3.

Tabla 3. Ponderador dimensión situación económica

Opciones de respuesta	Ponderador	
	Primeras 4 preguntas	Última pregunta
Mucho mejores	1.00	
Mejores (mayores)	0.75	1.00
Igual (iguales)	0.50	0.50
Peor (menores)	0.25	0.50
Mucho peor	0.00	

Fuente: Elaboración propia basada en ENCO (2017).

Para obtener el ICC debe sumarse el resultado de multiplicar el índice de seguridad y el índice de situación económica por el ponderador de preocupación, el cual se determina a través de preguntar a los encuestados qué les preocupa más, si la situación económica o la inseguridad; la especificación del cálculo se presenta en la fórmula 1.

$$ICC = (IS * \alpha) + (ISE * \beta) \tag{1}$$

Donde:

ICC = Índice de confianza ciudadana.

IS = Índice de seguridad.

α = Porcentaje de personas a las cuales les preocupa más la inseguridad que la situación económica.

ISE = Índice de situación económica.

β = Porcentaje de personas a las cuales les preocupa más la situación económica que la inseguridad.

Para medir el desempeño y efectividad de los diferentes niveles de gobierno con vistas a resolver los distintos problemas en la ciudad, se utilizaron diferentes indicadores, que se describen en la Tabla 4.

Tabla 4. *Variables de desempeño y efectividad*

Dimensión	Variables	Pregunta
Desempeño y efectividad	Capacidad para solucionar problemas	En su opinión, hoy en día su ciudad enfrenta los siguientes problemas: <ul style="list-style-type: none"> • Fallas y fugas en el suministro de agua potable • Deficiencias en la red pública de drenaje • Coladeras tapadas por acumulación de desechos • Falta de tratamiento de aguas residuales • Alumbrado público insuficiente. • Ineficiencia en el servicio de limpieza y recolección de basura • Calles y avenidas con embotellamientos frecuentes • Baches en calles y avenidas • Parques y jardines descuidados • Delincuencia (robos, extorsiones, secuestros, fraudes, etc. • Servicio de transporte público deficiente
	Desempeño de las instituciones de seguridad	Qué tan efectivo considera el desempeño de: <ul style="list-style-type: none"> • Policía Preventiva Municipal • Policía Estatal • Policía Federal y/o Gendarmería Nacional • Ejército y/o Marina
	Desempeño del gobierno	Qué tan efectivo para resolver los problemas ha sido: <ul style="list-style-type: none"> • El gobierno municipal • El gobierno estatal • El gobierno federal
	Confianza en las instituciones de seguridad	Cuánta confianza le inspira: <ul style="list-style-type: none"> • Policía Preventiva Municipal • Policía Estatal • Policía Federal y/o Gendarmería Nacional • Ejército y/o Marina

Fuente: Elaboración propia basada en ENCO (2017).

Las variables descritas en la tabla anterior sirven para medir el desempeño de los gobiernos y sus instituciones para resolver los diferentes problemas que enfrenta la ciudad. Cada variable tiene una serie de preguntas y cada pregunta una serie de respuestas, que se ponderan para calcular un índice por cada una de las variables. Dichas ponderaciones se describen en la Tabla 5. Para calcular la probabilidad de que el índice de confianza ciudadana aumente en función del desempeño de las instituciones de gobierno, se utiliza un modelo lineal de probabilidad (modelo Probit), en donde Y es cuantitativa; el objetivo consiste en estimar su valor o media esperados, dados los valores de las regresoras.

Tabla 5. Ponderador dimensión desempeño

Opciones de respuesta a variable Capacidad para resolver problemas	Ponderador	Opciones de respuesta a variable Desempeño de las instituciones de seguridad. Y variable Desempeño del gobierno	Ponderador	Opciones de respuesta a variable Confianza en las instituciones de seguridad	Ponderador
No	1.00	Muy efectivo	1.00	Mucha confianza	1.00
		Algo efectivo	0.75	Algo de confianza	0.75
Sí	0.00	Poco efectivo	0.50	Algo de desconfianza	0.50
		Nada efectivo	0.00	Mucha desconfianza	0.00

Fuente: Elaboración propia basada en ENCO (2017): ENSU (2017).

En los modelos donde Y es cualitativa (dicotómica), el objetivo consiste encontrar la probabilidad de que un acontecimiento suceda. Los modelos de regresión con respuesta cualitativa se conocen a menudo como modelos de probabilidad. Un modelo lineal de probabilidad puede escribirse de la siguiente manera:

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \tag{2}$$

Como se parte del supuesto de que $E(\epsilon_i) = 0$, entonces el valor esperado de la variable dependiente será:

$$E(Y_i) = 1 \cdot (\pi_i) + 0 \cdot (1 - \pi_i) = \pi_i \tag{3}$$

Ello implica que:

$$E(Y_i | X_i) = E(x' i \beta) = \pi_i \tag{4}$$

Esto significa que la respuesta esperada es la probabilidad de que la variable dependiente tome el valor 1, es decir, la esperanza condicional del modelo en realidad se interpreta como la probabilidad condicional de Y_i . Como la probabilidad π_i debe encontrarse entre 0 y 1, tenemos la restricción: $0 \leq (E(Y_i | X_i)) \leq 1$. O sea, la esperanza o probabilidad condicional debe encontrarse entre 0 y 1. Al ser el índice de confianza ciudadana un indicador que puede tomar valores entre 0 y 100, y al ser el modelo probabilístico una

técnica estadística que requiere una variable dicotómica como variable independiente, se considera un umbral de ICC ≥ 50 . En otras palabras, se trata de medir la probabilidad de que el ICC sea mayor que la media en función del desempeño de las instituciones de gobierno.

Resultados

Independientemente de todas las gráficas descriptivas de cada pregunta que se puedan generar como los ejemplos a continuación, aquí nos centraremos en la tendencia a largo plazo con respecto al índice de confianza ciudadana (ICC), que como ya dijimos se compone del índice de confianza del consumidor, que recoge la percepción de la economía, y el indicador de Seguridad tomado de la Encuesta Nacional de Seguridad Urbana (ENSU).

La Tabla 6 es un ejemplo del nivel de ingreso de los hogares participantes en la encuesta durante el año 2021, que fue cuando se realizó el cálculo durante la pandemia.

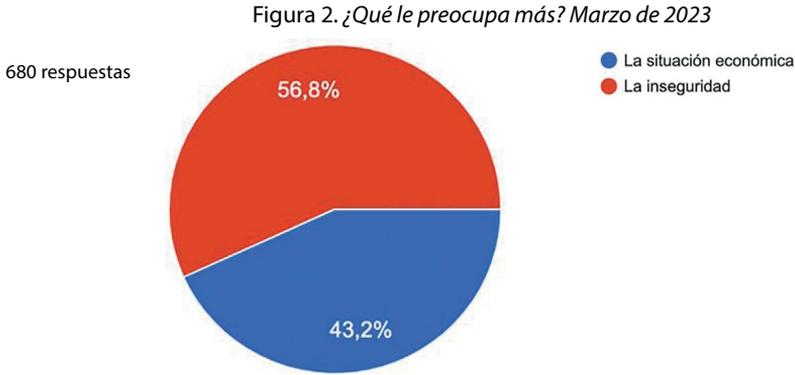
Tabla 6. Nivel de ingreso de los hogares en la ZMT, 2021 (n= 572 hogares)

Ingreso	% de los hogares	Clasificación hogar
Hasta \$6 mil pesos	14.2	Pobres
De \$6 mil a \$10 mil pesos	24.4	Clase media baja
De \$10 mil a \$20 mil pesos	27.0	Clase media media
De \$20 mil a \$30 mil pesos	18.1	Clase media alta
Más de \$30 mil pesos	16.3	Clase alta
	100.0	

Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2021.

La pregunta de la Figura 2 resulta fundamental porque es la que determina el ponderador que se dará a las preguntas de situación económica o inseguridad. Por tanto, el ponderador para las preguntas de situación económica para marzo de 2023 será de 43.2 por ciento, mientras que el ponderador para las preguntas de inseguridad será de 56.8 por ciento. Podemos decir que históricamente desde que se empezó a aplicar esta encuesta, lo que más preocupa a los tijuanaenses es la inseguridad. Únicamente durante el cierre

total de la economía a partir de marzo de 2020 hasta noviembre del mismo año, fue mayor la preocupación económica.



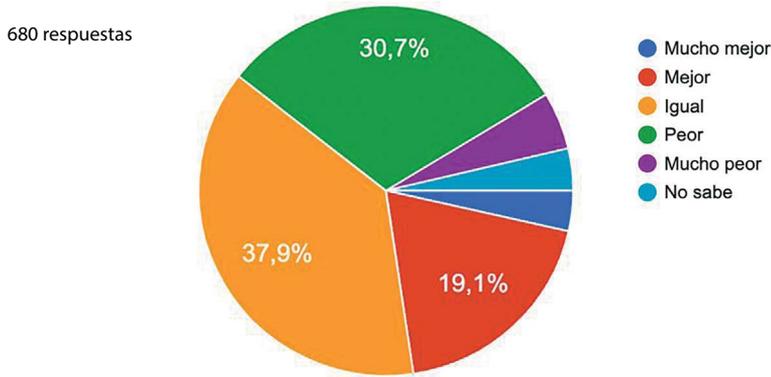
Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

En la Figura 3 se aprecia una de las cinco preguntas utilizadas para construir el indicador de confianza del consumidor. Ante la pregunta siguiente: ¿Cómo considera usted la situación económica de SU CIUDAD HOY EN DÍA comparada con la de HACE 12 MESES?, el 38 por ciento de los habitantes de la ZMT considera que está igual, el 31 por ciento considera que está peor, el 19 por ciento considera que mejor, el 5 por ciento considera que mucho peor, y el 3.5 por ciento mucho mejor.

Es decir, solo 1 de cada 5 tijuanaenses considera que está mejor o mucho mejor, 2 de cada 5 consideran que la situación económica es igual a la de hace un año, y otros 2 de cada 5 consideran que la situación es peor o mucho peor a la de hace un año. La muestra fue de 680 hogares.

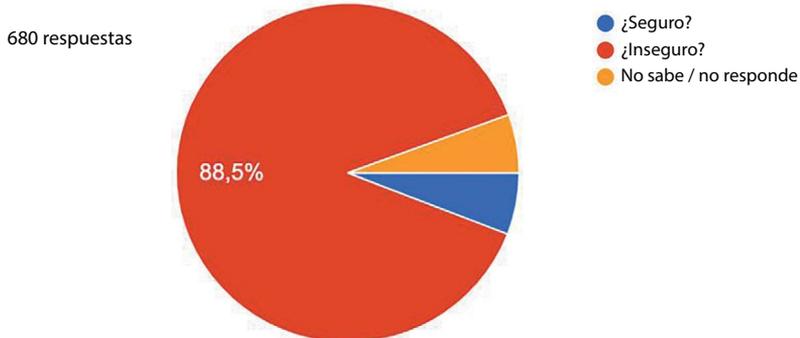
Otro ejemplo descriptivo es el de la Figura 4, donde se presenta la pregunta inicial del apartado que tiene que ver con seguridad urbana. El 88.5 por ciento de los tijuanaenses se sienten inseguros, el 5.9 por ciento se sienten seguros y un 5.6 por ciento no sabe/no responde. Como se explica en la metodología del Inegi, el último porcentaje se divide entre dos, y se obtiene que el 91.3 por ciento se siente inseguro y el 8.7 por ciento se siente seguro en la Zona Metropolitana de Tijuana.

Figura 3. ¿Cómo considera usted la situación económica de su ciudad hoy en día, comparada con la de hace 12 meses? Marzo de 2023



Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

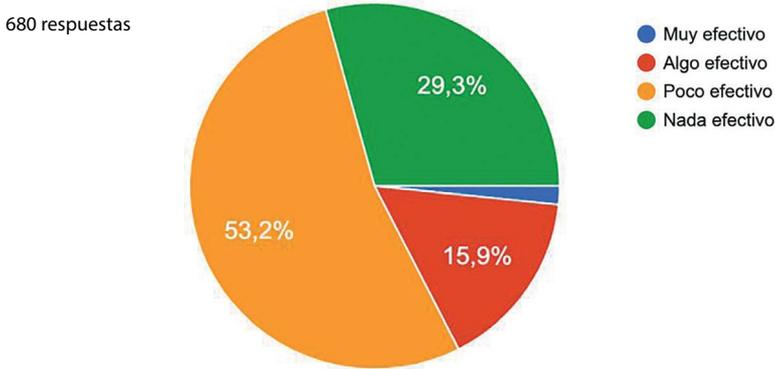
Figura 4. En términos de delincuencia, ¿considera que vivir actualmente en SU CIUDAD es? Marzo de 2023



Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

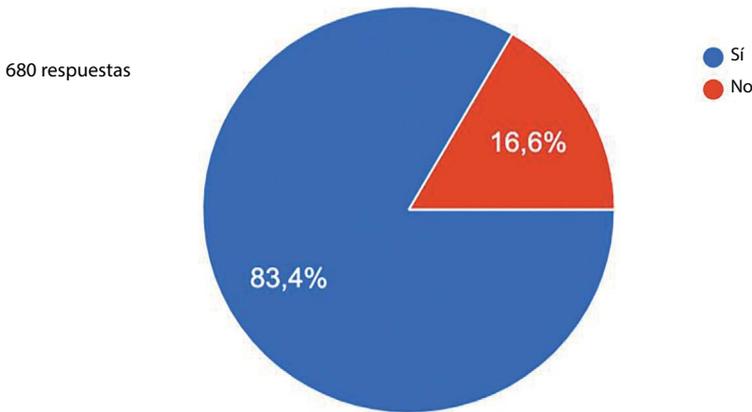
Se agregaron preguntas al cuestionario que no se contemplan en el del Inegi, como la de la Figura 5. El 53 por ciento considera que el gobierno ha sido poco efectivo, el 29 por ciento considera que nada efectivo, el 16 por ciento considera que algo efectivo y sólo el 2 por ciento considera que muy efectivo.

Figura 5. *¿Qué tan efectivo ha sido el gobierno de SU CIUDAD para resolver estos problemas?* Marzo de 2023



Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

Figura 6. *¿Para usted es una preocupación la escasez de agua en su colonia?* Marzo de 2023



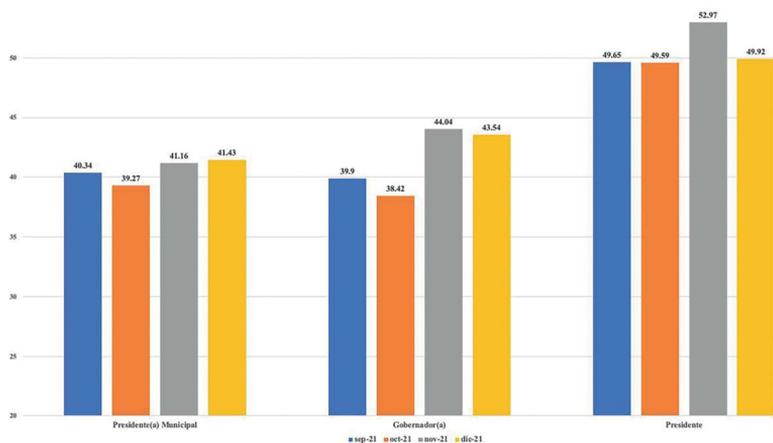
Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

La pregunta de la Figura 6 responde a un problema específico del norte del país y apremiante en la ZMT: el agua. Este tema cuenta con una batería de cuatro preguntas para monitorear el comportamiento. A 8 de cada 10 tijuanaenses les preocupa la escasez de agua en la ciudad. En la Figura 7 se muestra otro tipo de preguntas, las que se agregaron y se enuncian a continuación:

- En una escala de 0 a 100, ¿qué calificación le pondría a la presidenta municipal Montserrat Caballero Ramírez por su desempeño?
- En una escala de 0 a 100, ¿qué calificación le pondría a la gobernadora Marina del Pilar Ávila Olmeda por su desempeño?
- En una escala de 0 a 100, ¿qué calificación le pondría presidente Andrés Manuel López Obrador por su desempeño?

En la figura se aprecia la evolución de cómo califica la población a su presidenta municipal, gobernadora y presidente, con una calificación en diciembre de 2021 de 41, 43 y 50, respectivamente.

Figura 7. ¿Qué calificación les pone a sus gobernantes? Septiembre a diciembre de 2021



Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

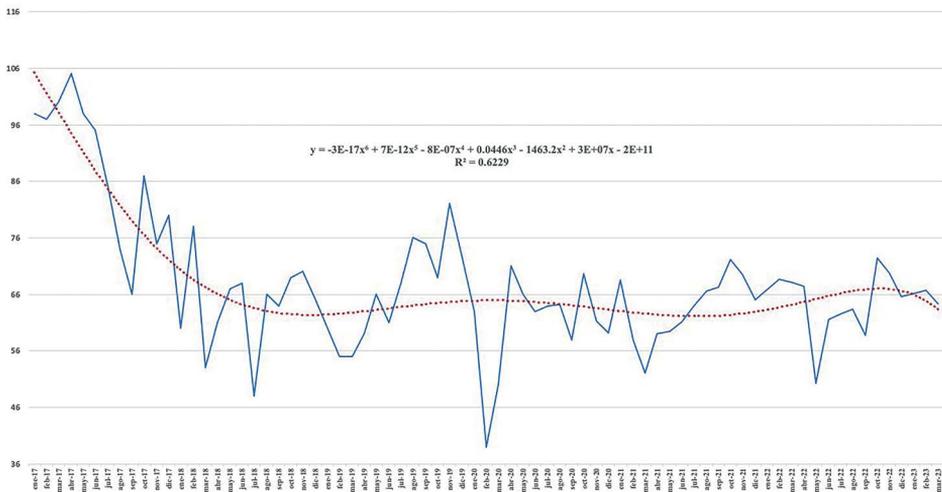
Ahora pasamos directamente al índice de confianza ciudadana, que como mencionamos se conforma de dos indicadores: el de confianza del consumidor (ICC) y el de seguridad urbana (ISU). La Figura 8 muestra la evolución del índice de confianza ciudadana de la Zona Metropolitana de Tijuana desde enero de 2017 hasta marzo de 2023, es decir, 6 años y 3 meses con un total de 75 observaciones mensuales.

En la Figura 8 se puede apreciar como hay una pérdida de confianza de 2017 a la fecha. Es importante destacar que al ser un índice mensual hay fuertes variaciones, por lo que se saca una línea de tendencia multinomial.

El indicador pasó de casi 106 en mayo de 2017 a poco menos de 66 en marzo de 2023, lo que representa una caída de 46.5 por ciento.

Surge entonces la siguiente pregunta: ¿Qué es lo que más ha contribuido a la pérdida de confianza de los tijuanaenses de enero de 2017 a marzo de 2023, el panorama económico o los niveles de inseguridad que imperan en la ciudad? En la Figura 9 se presentan las líneas de tiempo de la evolución de la percepción en la economía (línea azul), la percepción en la inseguridad (línea naranja) y la combinación de ambas en el ICC (línea gris).

Figura 8. Evolución del índice de confianza ciudadana de la ZMT, enero de 2017 a marzo de 2023



Fuente: x Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

El ICC de la Figura 9 es la contribución de la percepción de la economía y la percepción de la inseguridad en donde las preguntas de los indicadores están ponderadas por la pregunta de la Figura 2. Para suavizar las fuertes variaciones mensuales se calcula la línea de tendencia, que son las líneas punteadas de color rojo para cada indicador.

En la Figura 9 se pueden apreciar los dos indicadores: Economía y Seguridad urbana. La línea de tendencia polinomial se ajusta más a la tendencia en inseguridad con un 90 por ciento de bondad de ajuste, medido por la R2 de la ecuación, mientras que para la economía la bondad del ajuste es de sólo 34 por ciento.

Entre los eventos más relevantes de la economía local podemos destacar que a principios de 2017 se presentaba una recuperación en la percepción, después de que Donald Trump ganó la presidencia y no cumplió sus amenazas de cancelar el Tratado de Libre Comercio entre México, Canadá y Estados Unidos. Se presenta un bache a finales de 2017 y los primeros meses de 2018 por el nulo crecimiento de la economía a nivel nacional, explicado por la falta de experiencia del gobierno federal, el subejercicio de recursos y el tratar de detener el robo de combustible a Pémex sin un plan y estrategia, lo que ocasionó que se pararan sectores de la economía.

Posteriormente, tenemos la caída libre en febrero y marzo de 2020, debido al cierre productivo por la pandemia. Comienza la recuperación de la confianza en la economía sin llegar al punto más alto, que fue agosto de 2019 con 141 unidades en el índice, que tiene su dato base en marzo de 2017. En marzo de 2023 se ubica en 124 puntos, que se traduce en una confianza 32 por ciento superior a la de enero de 2017, aunque estancada durante todo 2022 y con tendencia a la baja en 2023, debido a las presiones inflacionarias derivadas del gasto público por EE. UU. para enfrentar la pandemia.

Figura 9. Evolución del índice de confianza ciudadana de la ZMT, el índice de confianza del consumidor e índice de seguridad urbana, enero de 2017 a marzo de 2023



Fuente: Elaboración propia basada en Encuesta del índice de confianza ciudadana, 2023.

Con respecto al tema de seguridad urbana, se planteó que durante todo el periodo los tijuanaenses siempre estuvieron más preocupados por la inseguridad que por la economía. Al inicio del análisis, 7 de cada 10 tijuanaenses se sentían inseguros de vivir en la ciudad, indicador ya de por sí bastante alarmante. Sin embargo, desde enero de 2017 hasta marzo de 2018 el índice de seguridad pasó de 108 a 15 puntos. Ahora, 9 de cada 10 tijuanaenses se sienten inseguros de vivir en la ciudad. Desafortunadamente, ya pasaron más de 6 años y el indicador no ha mejorado: 9 de cada 10 tijuanaenses se siguen sintiendo inseguros.

Debido a que los dos indicadores de economía y seguridad urbana alimentan el índice de confianza ciudadana, en donde el de economía, pese a la crisis económica por el covid-19 ha sido bueno, debido a la dinámica fronteriza en la que Tijuana se inserta al mercado norteamericano, el ICC no ha sido peor. El ICC tuvo su posición más alta en abril de 2017 con 105 puntos y en marzo de 2023 obtuvo 64. La caída en la confianza ciudadana fue de 48.5 por ciento.

El ICC fue favorecido por la percepción en la economía, ya que fue de las primeras en caer durante la pandemia, pero también de las primeras en recuperarse, debido a que muchas de las empresas que operan bajo el régimen IMMEX⁴ (antes Maquila) se consideran esenciales, pues se dedican a la producción de dispositivos médicos, altamente demandados durante la crisis sanitaria.

Desafortunadamente, la caída en el ICC es explicada principalmente por la caída en la percepción sobre la inseguridad que impera en la ZMT, donde como decíamos, el 90 por ciento de los habitantes nos sentimos inseguros. Con estos resultados podemos concluir que la ZMT es a pesar de sus gobiernos, una ciudad resiliente donde la población se esfuerza por salir adelante. Y si el tema de inseguridad no fuera tan grave, las condiciones económicas serían mejores, ya que se estima que a nivel nacional la inseguri-

⁴IMMEX significa Industria Manufacturera, Maquiladora y de Servicio de Exportación. Permite a los fabricantes extranjeros importar materias primas y componentes a México, para que éstos sean elaborados, transformados o arreglados. El beneficio del programa IMMEX a los fabricantes extranjeros es el de libre pago de impuestos generales a la importación, libre pago de impuestos al valor agregado y de las cuotas compensatorias (<https://www.comercioyaduanas.com.mx/comercioexterior/immex-que-es/>).

ridad representa un costo del 1.85 por ciento del PIB.⁵ Sabemos que los niveles de inseguridad y violencia en Tijuana están muy por encima de la media nacional, por lo que el costo económico es aún mayor.

Conclusiones

México es un país muy peligroso en términos de inseguridad. En el momento de terminar este capítulo, se registran 10 feminicidios y matan a 86 mexicanos diariamente. México es el país más peligroso para los periodistas, incluso por encima de Ucrania, un país en guerra. En 2022 hubo en Estados Unidos 20 200 asesinatos; en México 30 968 de acuerdo con el archivo de “Gun Violence” de Amnistía Internacional (<https://www.amnesty.org/en/what-we-do/arms-control/gun-violence>). La diferencia, además de los diez mil asesinatos, es que mientras Estados Unidos tiene una población de 336 millones de habitantes, México tiene una población de 136 millones de habitantes.

Es muy importante que desde la academia seamos críticos con situaciones de deterioro social y que hagamos contrapeso a las declaraciones oficiales triunfalistas y engañosas que señalan que vamos mejor, cuando los datos incluso oficiales nos dicen otra cosa. El bienestar y la confianza ciudadanos no sólo dependen de su seguridad económica (medida por perspectivas de ingresos presentes y futuros), sino de su seguridad en su persona, la de su familia y patrimonio (medida por su percepción de inseguridad).

El considerado padre de la economía y promotor del libre mercado, Adam Smith (1958), señaló en su famosa obra de 1976 *Investigación sobre la naturaleza y causa de la riqueza de las naciones*, que el gobierno debe encargarse de garantizar lo siguiente:

- *Soberanía nacional*: Capacidad de defender el territorio y las fronteras territoriales, aéreas y marítimas de un ejército invasor.

⁵García, A. K. (22 de septiembre de 2021). La inseguridad y violencia en México cuesta a los mexicanos un monto igual al 1.85% del PIB. *El Economista*. <https://www.economista.com.mx/politica/La-inseguridad-y-violencia-en-Mexico-cuesta-a-los-mexicanos-un-monto-igual-al-1.85-del-PIB-20210922-0063.html>

- *Seguridad nacional*: Capacidad de aplicar las normas de derecho para garantizar el respeto a la integridad de las personas y la propiedad privada. Cumplir las leyes y hacerlas cumplir, nadie por encima de ellas.
- *Infraestructura*: Invertir en las obras como puertos, vías de trenes, carreteras, puentes, aeropuertos, que garanticen una mejor movilidad de personas y un comercio que impulse la actividad y el empleo a través del crecimiento económico.
- *Educación y entrenamiento*: Invertir en educación es la mejor forma de garantizar un futuro más productivo y con menores niveles de violencia en la población, como se ha demostrado en todos los países desarrollados.

Es obvio que nuestras instituciones mexicanas no han hecho un buen trabajo en términos de garantizar la seguridad de todas y todos los mexicanos. Esto ocasiona que la confianza ciudadana disminuya afectada por la percepción de inseguridad, falta de oportunidades y el crecimiento económico.

Las teorías sociales respaldadas por datos duros y análisis estadísticos con la construcción y estudio de indicadores e índices resultan fundamentales para quitar la paja de la sustancia y la demagogia de los hechos. Las sociedades avanzadas utilizan el método científico para buscar la verdad y, con ello, detectar problemas y posibles soluciones.

Esperamos contribuir desde la academia con una mayor discusión y seriedad para analizar los problemas subyacentes que nos permitan dialogar en la búsqueda de soluciones al problema social y estructural de la inseguridad, que se traduce en problemas económicos, psicológicos y de otro tipo. Hay que evitar los círculos viciosos en donde las grandes ciudades o zonas metropolitanas transitan de economías de escala y aglomeración cuando van en ascenso, a deseconomías de escala y aglomeración cuando los costos de lo malo, como la falta de infraestructura, agua e inseguridad, superan los beneficios de lo bueno, entendido como posición geográfica y oportunidades económicas.

Referencias

- Boslaugh, S., y Watters, A. P. (2008). *Statistics in a Nutshell: A desktop Quick Reference*. O'Reilly Media.
- Encuesta Nacional sobre Confianza del Consumidor. (2017). *Principales resultados*. Instituto Nacional de Estadística y Geografía. <https://www.inegi.org.mx/programas/enco/#Tabulados>
- Encuesta Nacional de Seguridad Pública Urbana. (2017). *Principales resultados*. Instituto Nacional de Estadística y Geografía. https://www.inegi.org.mx/contenidos/programas/ensu/doc/ensu2017_trim2_presentacion_ejecutiva.pdf
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2010). *Metodología de la investigación* (5ta. ed.). McGraw-Hill/Interamericana.
- Instituto Nacional de Estadística y Geografía. (2016). *Encuesta Nacional sobre Confianza del Consumidor 2015. ENCO Documento Metodológico*. https://en.www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825078058.pdf
- Instituto Nacional de Estadística y Geografía. (2023). *Encuesta Nacional de Seguridad Pública Urbana 2022. ENSU Diseño Muestral*. <https://www.inegi.org.mx/programas/ensu/>
- Katona, G. (1951). *Psychological Analysis of Economic Behavior*. McGraw Hill.
- Kilic, E., y Cankaya, S. (2016). Consumer confidence and economic activity: a factor augmented VAR approach. *Applied Economics*, 48(32), 3062-3080. <https://doi.org/10.1080/00036846.2015.1133902>
- Lind, D. A., Marchal, W. G., y Wathen, S. A. (2015). *Statistical Techniques in Business & Economics* (6ta ed.). McGraw Hill Education.
- Smith, A. (1958). *Investigación sobre la naturaleza y causa de la riqueza de las naciones*. Fondo de Cultura Económica.
- Triana, J. L. (2021). Percepción de inseguridad, temor al delito y medidas de autoprotección: el caso de Acapulco, Guerrero. *Noésis: Revista de Ciencias Sociales y Humanidades*, 30(60), 166-190. <https://doi.org/10.20983/noesis.2021.2.9>
- Zhang, F., Fan Z., Kang, Y., Hu, Y., y Ratti, C. (2021). "Perception bias": Deciphering a mismatch between urban crime and perception of safety. *Landscape and Urban Planning*, 207, 1-14.

Anexos

Anexo 1. Preguntas del cuestionario del índice de confianza ciudadana: apartado socioeconómico

1. *Supervisor*

2. *Nombre del encuestador*

3. *¿Cuál es su ciudad?*

- Ensenada
- Mexicali
- Playas de Rosarito
- Tecate
- Tijuana

4. *Si vive en Tijuana, ¿en cuál delegación está su domicilio?*

- Centro
- Otay
- Playas de Tijuana
- La Mesa
- San Antonio de los Buenos
- Sánchez Taboada
- Cerro Colorado
- La Presa
- La Presa Este
- Fuera de Tijuana

5. *Sexo*

- Masculino
- Femenino

6. *Edad*

- 18-25

- 26-34
- 35-44
- 45-54
- 55-64
- 65 y más

7. *Escolaridad*

- Ninguna
- Primaria
- Secundaria
- Preparatoria o bachillerato
- Carrera técnica
- Profesional y posgrado

8. *¿Cuántas personas viven normalmente en esta vivienda, contando a los niños y a los ancianos?*

- 1
- 2
- 3
- 4
- 5 o más

9. *Actualmente está...*

- Empleado(a)
- Desempleado(a)
- Es estudiante o se dedica a los quehaceres de su hogar
- Otra situación/No especificado

10. *¿Cuál fue el ingreso total de este hogar el último mes?*

- \$0-\$6 000
- \$6 001-\$10 000
- \$10 001-\$20 000
- \$20 001-\$30 000
- \$30 001-\$40 000
- \$40 001-\$50 000

- \$50 001 o más
- No contestó

11. *¿Cuál es el nombre del oficio, puesto o cargo que desempeñó... en su trabajo principal la semana pasada?*

- Profesionistas y técnicos
- Funcionarios de los sectores público y privado
- Personal administrativo
- Comerciantes, vendedores y similares
- Trabajadores en servicios personales y conductores de vehículos
- Trabajadores en labores agropecuarias
- Trabajadores industriales
- No especificado/No aplica

12. *¿A qué se dedica la empresa, negocio o institución para la que trabajó la semana pasada?*

- Agropecuaria
- Industria
- Construcción
- Comercio
- Servicios
- Comunicaciones y transportes
- Administración pública y defensa
- No especificado/No aplica

13. *El jefe(a) de este hogar es...*

- Trabajador(a) a sueldo fijo o asalariado(a)
- Patrón(a)
- Trabajador(a) por su cuenta
- Desempleado(a)

14. *¿Qué le preocupa más?*

- La situación económica
- La inseguridad

Anexo 2. Encuesta sobre confianza del consumidor

1. *¿Cómo describe usted SU SITUACIÓN económica comparada con la de HACE 12 MESES?*

- Mucho mejor
- Mejor
- Igual
- Peor
- Mucho peor
- No sabe

2. *¿Y cómo cree usted que será SU SITUACIÓN económica DENTRO DE 12 MESES respecto a la situación actual?*

- Mucho mejor
- Mejor
- Igual
- Peor
- Mucho peor
- No sabe

3. *Comparada con la situación económica que los miembros de este hogar TENÍAN HACE 12 MESES, ¿cómo cree que es su situación en este momento?*

- Mucho mejor
- Mejor
- Igual
- Peor
- Mucho peor
- No sabe

4. *¿Cómo considera usted que será la situación económica de los miembros de este hogar DENTRO DE 12 MESES, respecto a la actual?*

- Mucho mejor
- Mejor
- Igual

- Peor
- Mucho peor
- No sabe

5. *¿Cómo considera usted la situación económica de SU CIUDAD HOY EN DÍA comparada con la de HACE 12 MESES?*

- Mucho mejor
- Mejor
- Igual
- Peor
- Mucho peor
- No sabe

6. *¿Cómo considera usted que será la condición económica de SU CIUDAD DENTRO DE 12 MESES respecto a la actual situación?*

- Mucho mejor
- Mejor
- Igual
- Peor
- Mucho peor
- No sabe

7. *¿En ESTE MOMENTO tiene usted mayores posibilidades de comprar ropa, zapatos, alimentos, etc. que hace un AÑO?*

- Sí
- No
- Igual
- No sabe

8. *Comparando la SITUACIÓN ECONÓMICA ACTUAL con la de HACE UN AÑO, ¿cómo considera en el MOMENTO ACTUAL las posibilidades de que usted o alguno de los integrantes de este hogar realice compras tales como muebles, televisor, lavadoras, otros aparatos electrodomésticos, etc.?*

- Mayores
- Iguales

- Menores
 - No sabe
9. *¿Considera usted que durante los próximos 12 meses usted o alguno de los integrantes de este hogar tendrán posibilidades económicas para salir de vacaciones?*
- Sí
 - No
 - No sabe
10. *¿ACTUALMENTE USTED tiene posibilidades de AHORRAR alguna parte de sus ingresos?*
- Sí
 - No
 - No sabe
 - No tiene ingresos
11. *¿Cómo considera usted que serán sus condiciones económicas para AHORRAR DENTRO DE 12 MESES, comparadas con las actuales?*
- Muy buenas
- Buenas
 - Iguales
 - Malas
 - Muy malas
 - No sabe
12. *¿Qué porcentaje aproximado del ingreso familiar destina a la alimentación?*
13. *¿Qué porcentaje aproximado del ingreso familiar destina al pago de servicios? (agua, luz, internet, teléfono, celular, gas)*
14. *¿Qué porcentaje AHORRA?*
- Menos del 5%
 - Del 5% al 10%
 - Del 10% al 15%

- Del 15% al 20%
- Más del 20%

15. *¿A qué piensa destinar su ahorro?*

a) Para enfrentar una emergencia

- Sí
- No

b) Para el retiro en la vejez

- Sí
- No

c) Para vacaciones

- Sí
- No

d) Para comprar casa

- Sí
- No

e) Para comprar carro

- Sí
- No

f) Otro

- Sí
- No

16. *Comparado con los 12 meses anteriores, ¿cómo cree usted que se comporten los precios en SU CIUDAD en los siguientes 12 meses?*

- Disminuirán mucho
- Disminuirán poco
- Permanecerán igual
- Aumentarán poco
- Aumentarán igual

- Aumentarán mucho
- No sabe

17. *¿Cree usted que el EMPLEO EN SU CIUDAD EN LOS PRÓXIMOS 12 MESES?*

- Aumentará mucho
- Aumentará poco
- Permanecerá igual
- Disminuirá poco
- Disminuirá mucho
- No sabe

18. *¿Algún miembro de este hogar o usted están planeando COMPRAR un AUTOMÓVIL nuevo o usado en los PRÓXIMOS 2 AÑOS?*

- Sí
- Probablemente
- No
- No sabe

19. *¿Algún miembro de este hogar o usted están planeando COMPRAR, CONSTRUIR O REMODELAR UNA CASA en los PRÓXIMOS 2 AÑOS?*

- Sí
- Probablemente
- No
- No sabe

Anexo 3. Encuesta metropolitana de seguridad pública urbana

1. *En términos de delincuencia, considera que vivir actualmente en SU CIUDAD es...*

- Seguro
- Inseguro
- No sabe / no responde

2. *En términos de delincuencia, dígame ¿se siente seguro(a) o inseguro(a) en...?*

a) Su casa

- Seguro
- Inseguro

b) Su trabajo

- Seguro
- Inseguro

c) Las calles que habitualmente usa

- Seguro
- Inseguro

d) La escuela

- Seguro
- Inseguro

e) El mercado

- Seguro
- Inseguro

f) El centro comercial

- Seguro
- Inseguro

g) El banco

- Seguro
- Inseguro

h) El cajero automático localizado en la vía pública

- Seguro
- Inseguro

i) El transporte público

- Seguro
- Inseguro

j) El automóvil

- Seguro
- Inseguro

k) La carretera

- Seguro
- Inseguro

l) El parque o centro recreativos

- Seguro
- Inseguro

3. *Pensando en las condiciones de delincuencia en SU CIUDAD, ¿considera que en los próximos 12 meses...?*

- Mejorará
- Seguirá igual de bien
- Seguirá igual de mal
- Empeorará
- No sabe/No responde

4. *En los últimos tres meses, ¿ha escuchado o ha visto en los alrededores de su vivienda situaciones como...?*

a) Vandalismo (grafitis, pintas, vidrios quebrados o daños) en las viviendas o negocios

- Sí
- No

b) Consumo de alcohol en las calles

- Sí
- No

c) Robos o asaltos

- Sí
- No

d) Bandas violentas o pandillerismo

- Sí
- No

e) Venta o consumo de droga

- Sí
- No

f) Disparos frecuentes con armas

- Sí
- No

5. *En este mismo periodo de tres meses, por temor a sufrir algún delito (robo, asalto, secuestro, entre otros), ¿usted cambió sus hábitos respecto a...?*

a) Llevar cosas de valor como joyas, dinero o tarjetas de crédito

- Sí
- No

b) Caminar por los alrededores de su vivienda, pasadas las ocho de la noche

- Sí
- No

c) Visitar a parientes o amigos

- Sí
- No

d) Permitir que salgan de su vivienda a sus hijos menores

- Sí
- No

6. *¿Qué tan efectivo considera el desempeño de...?*

a) Policía Preventiva Municipal

- Muy efectivo
- Algo efectivo
- Poco efectivo
- Nada efectivo

b) Policía Estatal

- Muy efectivo
- Algo efectivo
- Poco efectivo
- Nada efectivo

c) Policía Federal y/o Guardia Nacional

- Muy efectivo
- Algo efectivo
- Poco efectivo
- Nada efectivo

d) Ejército y/o Marina

- Muy efectivo
- Algo efectivo
- Poco efectivo
- Nada efectivo

7. *¿Cuánta confianza le inspira...?*

a) Policía Preventiva Municipal

- Mucha confianza
- Algo de confianza
- Algo de desconfianza
- Mucha desconfianza

b) Policía Estatal

- Mucha confianza
- Algo de confianza

- Algo de desconfianza
- Mucha desconfianza

c) Policía Federal y/o Guardia Nacional

- Mucha confianza
- Algo de confianza
- Algo de desconfianza
- Mucha desconfianza

d) Ejército y/o Marina

- Mucha confianza
- Algo de confianza
- Algo de desconfianza
- Mucha desconfianza

8. *En su opinión, ¿cuáles son los problemas más importantes que se enfrentan hoy en día EN SU CIUDAD?*

a) Fallas y fugas en el suministro de agua potable

- Sí
- No

b) Escasez de agua potable

- Sí
- No

c) Deficiencias en la red pública de drenaje

- Sí
- No

d) Coladeras tapadas por acumulación de desechos

- Sí
- No

- e) Falta de tratamiento de aguas residuales
 - Sí
 - No

- f) Alumbrado público insuficiente
 - Sí
 - No

- g) Ineficiencia en el servicio de limpia y recolección de basura
 - Sí
 - No

- h) Calles y avenidas con embotellamientos frecuentes
 - Sí
 - No

- i) Baches en calles y avenidas
 - Sí
 - No

- j) Parques y jardines descuidados
 - Sí
 - No

- k) Delincuencia (robos, extorsiones, secuestros, fraudes, etc.)
 - Sí
 - No

- l) Servicio de transporte público deficiente
 - Sí
 - No

9. *¿Qué tan efectivo ha sido el gobierno de SU CIUDAD para resolver estos problemas?*

- Muy efectivo

- Algo efectivo
- Poco efectivo
- Nada efectivo

10. *Actualmente cuenta con agua:*

- Potable
- De pipa

11. *¿Para usted es una preocupación la escasez de agua en su colonia?*

- Sí
- No

12. *¿En su colonia hay problemas de escasez de agua?*

- De acuerdo
- En desacuerdo

13. *¿En su colonia se va el agua...?*

- Nunca
- Un día a la semana
- De dos a tres días a la semana
- Más de cuatro días a la semana
- Solo cuando lo anuncia la CESPT

14. *En una escala de 0 a 100, ¿qué calificación le pondría a la presidenta municipal Montserrat Caballero Ramírez por su desempeño?*

15. *En una escala de 0 a 100, ¿qué calificación le pondría a la gobernadora Marina del Pilar Ávila Olmeda por su desempeño?*

16. *En una escala de 0 a 100, ¿qué calificación le pondría al presidente Andrés Manuel López Obrador por su desempeño?*

Sobre los coordinadores y autores

Duniesky Feitó Madrigal

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California (México). Obtuvo la maestría en Administración de Negocios y la licenciatura en Ingeniería Industrial en la Universidad Central “Marta Abreu” de Las Villas (Cuba). Se ha desempeñado como coordinador del programa educativo Licenciatura en Administración en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California. En la actualidad es coordinador de la Maestría en Administración y del Doctorado en Ciencias Administrativas, en dicha facultad. También es miembro del Sistema Nacional de Investigadores del Conahcyt y de la Asociación Nacional de Facultades y Escuelas de Contaduría y Administración. Ha publicado artículos como “El ahorro para el retiro en México: Un análisis generacional y de género” en la revista *Región y Sociedad* del Colegio de Sonora en 2023; asimismo, ha participado en la coordinación de libros como *Casos prácticos de investigación financiera* (Centro de Estudios e Investigaciones para el Desarrollo Docente [Cenid], 2022).

ORCID: <https://orcid.org/0000-0001-7225-2598>

GOOGLE SCHOLAR: <https://scholar.google.com/citations?user=XIvh-tV8AAAAJ&hl=es>

RESEARCHGATE: <https://www.researchgate.net/profile/Duniesky-Madrigal>

ACADEMIA: <https://independent.academia.edu/DunieskyFeitoMadrigal>

Malena Portal Boza

Doctora en Ciencias Económicas por la Universidad Autónoma de Baja California (México). Obtuvo la maestría en Dirección y la licenciatura en Contabilidad y Finanzas en la Universidad Central “Marta Abreu” de Las Villas (Cuba). Se ha desempeñado como coordinadora del área de Finanzas en soporte a los programas educativos y como profesora de tiempo completo en materias de licenciatura y posgrado en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California. En la actualidad es coordinadora de la Maestría en Inteligencia Financiera y de la Especialidad en Dirección Financiera, ambos programas de posgrado

ofertados en dicha facultad. También es miembro del Sistema Nacional de Investigadores del Conahcyt y de la Asociación Nacional de Facultades y Escuelas de Contaduría y Administración. Ha publicado artículos como “El ahorro para el retiro en México: Un análisis generacional y de género” en la revista *Región y Sociedad* del Colegio de Sonora en 2023. Asimismo, ha participado en la coordinación de libros como *Casos prácticos de investigación financiera* con la editorial del Centro de Estudios e Investigaciones para el Desarrollo Docente (Cenid), en 2022.

ORCID: <https://orcid.org/0000-0002-4237-1534>

RESEARCHGATE: <https://www.researchgate.net/profile/Malena-Portal-Boza>

Ismael Plascencia López

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California (UABC), maestro en Desarrollo Regional por el Colegio de la Frontera Norte (Colef) y licenciado en Economía por UABC. Es miembro del Sistema Nacional de Investigadores desde 2007; ha liderado y participado en más de 10 proyectos de investigación con financiamiento de Conahcyt y la Unión Europea. Preside el Colegio de Economistas de Baja California, es director de la Facultad de Contaduría y Administración de la UABC desde 2019; también catedrático y asesor sobre temas de economía, estadística y metodología de la investigación. Ha publicado artículos como “Evaluation of the efficacy and safety of silver nanoparticles in the treatment of non-neurological and neurological distemper in dogs: A randomized clinical trial” (Viruses, 2022) y “Knowledge transfers and absorptive capacities as determinants of competitive advantages: The Mexican case” (IEEE Engineering Management Review, 2022). Entre sus libros se encuentra *Shenzhen: Lecciones de competitividad para las zonas metropolitanas* (Porrúa, 2021).

ORCID: <https://orcid.org/0000-0003-2860-1417>

GOOGLE SCHOLAR: https://scholar.google.com.mx/citations?user=32ob_xsAAAAJ&hl=es

RESEARCHGATE: <https://www.researchgate.net/profile/Ismael-Lopez-5>

ACADEMIA: <https://uabc.academia.edu/IsmaelPlascencia>

FLORES SÁNCHEZ, Carlos Alberto

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California (México), máster en Administración e ingeniero en Computación. Se ha desempeñado en coordinaciones de área y de servicio social profesional, así como profesor de asignaturas en licenciatura y posgrado en la Unidad Académica de Contaduría y Administración de la Universidad Autónoma de Baja California. En la actualidad es profesor investigador de la Facultad de Contaduría y Administración, donde realiza investigación sobre transformación digital y economía de la innovación. También es colaborador del Grupo de Trabajo Gobierno de TI de la Corporación Universitaria para el Desarrollo de Internet (CUDI), en la Red Nacional de Educación e Investigación de México. Ha publicado capítulos de libro y artículos sobre transformación digital, uso de las TIC, clústeres y ejercicios econométricos sobre la demanda del turismo.

ORCID: <https://orcid.org/0000-0003-1516-166X>

GOOGLE SCHOLAR: <https://scholar.google.com/citations?user=BRKM-s64AAAAJ&hl=es>

RESEARCHGATE: <https://www.researchgate.net/profile/Carlos-Flores-34>

Jaramillo Cardona, Martha Cecilia

Doctora en Ciencias Sociales, con especialidad en estudios regionales, por el Colegio de la Frontera Norte (Colef). Obtuvo la maestría en Planeación Urbana y Regional por la Pontificia Universidad Javeriana y la licenciatura en Trabajo Social en la Fundación Universitaria Monserrate, ambas en Bogotá, Colombia. Ha coordinado más de 20 proyectos de investigación y se desempeña como profesora investigadora de tiempo completo en la Facultad de Economía y Relaciones Internacionales de la Universidad Autónoma de Baja California. En la actualidad es miembro del Sistema Nacional de Investigadores Nivel I y miembro del Registro Conahcyt de Evaluadores Certificados (RCEA). Pertenece a los núcleos básicos del programa de Maestría y Doctorado en Estudios del Desarrollo Global y del programa de Maestría y Doctorado en Negocios Sociales. Sus líneas de investigación son descentralización, salud y políticas públicas.

Ha publicado el libro *Seguridad alimentaria y construcción de capacidades en México y Guatemala* (Ediciones del Lirio / UABC, 2023); es auto-

ra de los artículos “Equiparar poderes de las mujeres en la esfera pública y privada: Clave para romper el techo de cristal” en la revista *Pares* (2021), y “Políticas migratorias en un contexto de globalización: El caso de México y Estados Unidos” en la revista *Población y Desarrollo-Argonautas y Caminantes* (2021).

ORCID: <https://orcid.org/0000-0002-8091-0551>

GOOGLE SCHOLAR: https://scholar.google.com/citations?user=pMO3_dsAAAAJ&hl=en

RESEARCHGATE: <https://www.researchgate.net/profile/Martha-Jaramillo-Cardona>

ACADEMIA: <https://uabc.academia.edu/MarthaCeciliaJaramilloCardona>

Jiménez Cerra, Edgar

Doctor en Ciencias Administrativas y maestro en Gestión de las Tecnologías de la Información y la Comunicación por la Universidad Autónoma de Baja California (México), y licenciado en Ciencias Computacionales por la Universidad Central “Marta Abreu” de Las Villas (Cuba). Se ha desempeñado como coordinador del área de Acreditación en soporte a los programas educativos y como profesor de tiempo completo en materias de licenciatura y posgrado en la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California. Ha publicado varios capítulos de libro como “La profesión financiera en la nueva era digital. Un análisis desde el mercado laboral y la oferta educativa en México” y “Los desafíos de la inclusión en México: educación financiera y su relación con el acceso al mercado de crédito”, en el libro *Finanzas 4.0 e inclusión financiera en México: Oportunidades y retos en la nueva era digital* (Porrúa, 2021); y “El impacto de la pandemia covid-19 en los grupos vulnerables del mercado laboral en México”, en el libro *Acciones de vinculación en sectores vulnerables* (Centro de Estudios e Investigaciones para el Desarrollo Docente [Cenid], 2021). Publicó recientemente el artículo “Inclusão financeira e sua relação com o desenvolvimento sustentável em países emergentes: O caso do Brasil e do México” en la revista colombiana *Perspectivas de las Ciencias Contables*.

ORCID: <https://orcid.org/0000-0003-1285-3345>

RESEARCHGATE: <https://www.researchgate.net/profile/Edgar-Jimenez-14>

Lobo Rodríguez, Martha Ofelia

Doctora en Ciencias Económicas por la Universidad Autónoma de Baja California. Obtuvo la maestría en Mercadotecnia en la Universidad Autónoma de Aguascalientes y la Universidad Autónoma de Baja California. Es licenciada en Turismo por la Universidad Autónoma de Baja California. Se ha desempeñado como coordinadora de la maestría de Administración y de la carrera de Gestión turística, así como directora de la Facultad de Turismo y Mercadotecnia de la Universidad Autónoma de Baja California. Actualmente es coordinadora general de vinculación y cooperación académica de la Universidad Autónoma de Baja California. También es miembro y fundadora del Colegio de Profesionistas en Turismo, de Baja California. Ha sido coordinadora de dos libros sobre el sector turístico, tiene publicaciones sobre el uso de las TIC en empresas de servicios y ejercicios econométricos sobre la demanda de turismo.

ORCID: <https://orcid.org/0000-0002-9483-2216>

GOOGLE SCHOLAR: https://scholar.google.com/citations?user=zt_OqWoAAAAJ&hl=es

RESEARCHGATE: <https://www.researchgate.net/profile/Martha-Lobo-2>

López Leyva, Josué Aarón

Doctor en Ciencias en Electrónica y Telecomunicaciones por el Centro de Investigación Científica y de Educación Superior de Ensenada. Licenciado en Ingeniería en Electrónica en Telemática por el Instituto Tecnológico Superior de Cajeme. Ha participado como evaluador en diversos programas y convocatorias del Conahcyt. Ha sido coordinador académico de la Ingeniería en Energías Renovables, Ingeniería Cibernética en Electrónica e Ingeniería Mecatrónica en CETyS Universidad. Ha participado en 11 registros en el Instituto Nacional del Derecho de Autor (Indautor), tres procesos en el Instituto Mexicano de la Propiedad Industrial (IMPI), ocho proyectos externos y ocho proyectos internos de Investigación, Desarrollo e Innovación y 28 proyectos vinculados al ID&i, 66 estudiantes son sus coautores en publicaciones técnicas y científicas, y 95 estudiantes han colaborado en proyectos liderados por el Dr. López. Ha publicado diversos artículos científicos y técnicos, así como capítulos de libro. Ha participado en congresos naciona-

les e internacionales, siempre procurando la colaboración entre instituciones y alumnos de licenciatura y posgrado. Hasta el momento, ha impartido diversos módulos en diplomados del sector técnico, 27 asignaturas en posgrado, 40 asignaturas en licenciatura. Ha sido director de 21 tesis de maestría. Además, obtuvo las siguientes distinciones institucionales: Reconocimiento a la Excelencia Educativa en Publicaciones 2013, Reconocimiento a la Excelencia Educativa en Investigación 2017, Reconocimiento a la Excelencia Educativa en Investigación 2018, Reconocimiento a la Excelencia Educativa en Vinculación 2019 y Reconocimiento a la Excelencia Educativa en Investigación 2020. Es miembro del Sistema Nacional de Investigadores Nivel 1 del Conahcyt, en el área de investigación interdisciplinaria. Ha publicado el capítulo de libro “Short- and medium-term entrepreneurship intention analysis of university students based on the theory of multiple intelligences using artificial neural networks” (Springer, 2023) y el artículo “Energy- and water-integrated management system to promote the low-carbon manufacturing industry: An interdisciplinary Mexican case study” (*Environment, Development and Sustainability*, 2023).

ORCID: <https://orcid.org/0000-0002-3004-5686>

GOOGLE SCHOLAR: <https://scholar.google.es/citations?user=mkdc-FZUAAA&hl=es>

RESEARCHGATE: <https://www.researchgate.net/profile/Josue-Lopez>

López Pérez, Nadxilli

Doctora en Estudios del Desarrollo Global y maestra en Ciencias Económicas por la Facultad de Economía y Relaciones Internacionales de la Universidad Autónoma de Baja California (México). Es licenciada en Administración por el Tecnológico Nacional de México, campus Los Mochis, Sinaloa. Cuenta con experiencia profesional en la iniciativa privada y como docente desde 2017. Actualmente, es profesora de tiempo completo en la Facultad de Turismo y Mercadotecnia, UABC, campus Tijuana. Sus principales líneas de investigación son género, violencia de género, políticas públicas con perspectiva de género y emprendimiento femenino. Ha publicado los artículos “La transversalidad de la perspectiva de género en las políticas públicas para la violencia intrafamiliar en México” en la revista *GénEroos* (2023), “Conciliación trabajo-familia: Una comparación de las decisiones de empleo entre

las mujeres de México y España” (Ciencia UAT, 2016) y “El conflicto entre el empleo extradoméstico y la dedicación al hogar en la mujer unida” (2016).

ORCID: <https://orcid.org/0000-0002-9054-9552>

GOOGLE SCHOLAR: <https://scholar.google.com/citations?user=QsUS8O-gAAAAJ&hl=es>

RESEARCHGATE: <https://www.researchgate.net/profile/Nadxilli-Perez>

Moreno Ortega, Magdelis

Máster y licenciada en Ingeniería Industrial por la Universidad Central “Marta Abreu” de Las Villas, Cuba. Profesora adjunta del Centro de Estudios de Dirección de la Facultad de Ciencias Económicas de la Universidad Central “Marta Abreu” de Las Villas, Cuba. Ha sido coordinadora de proyectos nacionales en atención ciudadana y gobierno electrónico adscriptos al Ministerio de Comunicaciones de Cuba. Actualmente cursa el Doctorado en Ciencias Administrativas en la Universidad Autónoma de Baja California.

ORCID: <https://orcid.org/0000-0003-4282-3577>

GOOGLE SCHOLAR: <https://scholar.google.com/citations?user=opeZIXkAAAAJ&hl=es>

RESEARCHGATE: <https://www.researchgate.net/profile/Guillermo-Moreno-Ortega>

Plascencia López, Oswaldo

Maestro en Administración por la Universidad Autónoma de Baja California (UABC); licenciado en Negocios Internacionales en UABC. Colaborador del cuerpo académico de Innovación y desarrollo regional de la Facultad de Contaduría y Administración de la Universidad Autónoma de Baja California. Actualmente estudia Doctorado en Ciencias Administrativas con las líneas de generación de conocimiento en Desarrollo regional y Sistemas de innovación, y es responsable del mantenimiento para las encuestas y la base de datos del ICC.

ORCID: <https://orcid.org/0009-0002-5467-6850>

ResearchGate: <https://www.researchgate.net/profile/Oswaldo-Plascencia>

Quiroz Félix, Jorge

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California (UABC), maestro en Desarrollo Regional por el Colegio de la Frontera Norte y licenciado en Economía por la Universidad de Sonora. Ha sido director de la Facultad de Comercio Internacional de la Universidad Estatal de Sonora (UES), así como también coordinador del Centro de Negocios de la UES, Unidad Académica San Luis Río Colorado. Actualmente es profesor de tiempo completo en UES y profesor de asignatura en la Facultad de Turismo y Mercadotecnia de UABC, campus Tijuana. Cuenta con el reconocimiento perfil PRODEP de la Secretaría de Educación Pública y es colaborador del Cuerpo Académico de Desarrollo Turístico (nivel: en consolidación). Ha publicado artículos en revistas nacionales e internacionales sobre economía turística y crecimiento económico.

ORCID: <https://orcid.org/0000-0002-2907-8771>

GOOGLE SCHOLAR: <https://scholar.google.com.mx/citations?user=CXU2M2gAAAAJ&hl=es>

ResearchGate: <https://www.researchgate.net/profile/Jorge-Felix-5>

Academia: <https://uabc.academia.edu/JorgeQuirozFelix>

Ramírez Urquidy, Martín Arturo

Doctor en Ciencias Económicas en el área de capital humano y desarrollo microempresarial por la Universidad Autónoma de Baja California (UABC). Maestro en Economía con énfasis en análisis económico por California State Polytechnic University y maestro en Economía con especialidad en instituciones y organización industrial por Claramont Graduate University. Licenciado en Economía por la UABC. Además, realizó estudios posdoctorales sobre instituciones y emprendimiento en la Universidad Estatal de California-Domínguez Hills, Los Ángeles. Es profesor de tiempo completo titular C en la Facultad de Economía y Relaciones Internacionales en la UABC, investigador nacional nivel II del Consejo Nacional de Humanidades, Ciencias y Tecnologías y miembro regular de la Academia Mexicana de Ciencias. Sus líneas de investigación son economía del emprendimiento y las pequeñas empresas, capital humano, emprendimiento y desarrollo, organización industrial, y ecosistemas de emprendimiento.

Su más reciente producción académica en redes de colaboración se enmarca en los libros *Ecosistemas de emprendimiento de las entidades federativas de México: una propuesta de medición y evaluación para orientar la política pública* (2022) y *Emprendimiento y educación superior en México: diversos enfoques en la construcción de ecosistemas* (2021). Asimismo, publicó recientemente los artículos “Economic downturns and regional entrepreneurship dynamics in Mexico: trends and policy implications”, en la revista *Problemas del Desarrollo* (2022), y “Entrepreneurial outcomes and the role of higher education and R&D: evidence from Mexico”, en *Journal of Developmental Entrepreneurship* (2021).

ORCID: <https://orcid.org/0000-0003-0571-0477>

Google Scholar: <https://scholar.google.com.mx/citations?user=K0d4N-HYAAAAJ&hl=en>

ResearchGate: <https://www.researchgate.net/profile/Martin-Ramirez-Urquidy>

Sodi Díaz, Abraham Neftalí

Maestro en Mercadotecnia por el Centro de Enseñanza Técnica y Superior (Cetys Universidad), maestro en Administración con especialidad en Finanzas también por Cetys Universidad, y licenciado en Economía por UABC.

Ha publicado “Análisis exploratorio sobre fuentes de información para el estudio de la movilidad urbana empleando técnicas bibliométricas” (INNOVAITESCYT, 2023).

ORCID: <https://orcid.org/0000-0003-0865-3594>

Valadez García, Alfredo

Doctor en Ciencias Económicas por la Universidad Autónoma de Baja California (UABC). Es maestro en Economía Aplicada por el Colegio de la Frontera Norte (Cofe) y obtuvo su licenciatura en Economía por la UABC. Es profesor de tiempo completo en la Escuela de Administración y Negocios en Cetys Universidad, campus Tijuana, donde también es investigador asociado del Centro de Investigación Económica del Noroeste (CIEN). Es miembro del Sistema Nacional de Investigadores, nivel Candidato, del Conahcyt. Ha participado con ponencias en congresos celebrados en el Colegio de Sonora, el Instituto Nacional de Salud Pública y otras instituciones. Sus lí-

neas de investigación son economía de la educación, economía del emprendimiento y dinámica de la economía transfronteriza. Participa en el proyecto de investigación “Condiciones académicas y laborales de la población transfronteriza estudiantil en Cetys Universidad”. Recientemente ha publicado los capítulos de libro “Short- and medium-term entrepreneurship intention analysis of university students based on the theory of multiple intelligences using artificial neural networks” (Springer, 2023), y “Virtual and hybrid classes as the challenge assumed during the SARS-CoV-2 pandemic: An interdisciplinary qualitative approach” en *IEEE German Education Conference*; así como el artículo “Impacto del covid-19 en la inserción laboral de los egresados de Cetys universidad, México”, en la revista *Diálogos sobre Educación* (2023).

ORCID: <https://orcid.org/0000-0002-2841-4972>

ResearchGate: <https://www.researchgate.net/profile/Alfredo-Garcia-29>

Índice general

<i>Prólogo</i>	11
I. Introducción a la estadística	
<i>Alfredo Valadez García, Josué Aarón López Leyva</i>	11
Resumen	11
Antecedentes e importancia general de la estadística	12
Aportes de los modelos estadísticos a las ciencias económico- administrativas.	16
Principales técnicas estadísticas aplicadas a la investigación económico-administrativa	20
Referencias	27
II. Estadística descriptiva y modelos de tabulación cruzada	
<i>Edgar Jiménez Cerra</i>	29
Resumen	29
Medidas de tendencia central de los datos.	30
Medidas de dispersión.	33
Herramientas gráficas	35
Modelos de tabulación cruzada de datos	40
Modelos de tabulación cruzada multidimensional	46
Caso práctico. Factores sociodemográficos que inciden en el uso del crédito hipotecario en México.	47
Resultados	50
Conclusiones	68
Referencias	68
III. Modelos estadísticos de elección discreta	
<i>Alfredo Valadez García, Martín Arturo Ramírez Urquidy</i>	70
Resumen	70
Introducción a los modelos de elección discreta	71
Desarrollo del modelo matemático de elección discreta.	72

Supuestos de los modelos de elección discreta	76
Caso práctico. Análisis de la probabilidad de emprendimiento fronterizo y transfronterizo: el caso de México-Estados Unidos 2020	78
Resultados	85
Conclusiones	87
Referencias	88
IV. Modelos de ecuaciones estructurales con estimación PLS	
<i>Duniesky Feitó Madrigal, Magdelis Moreno Ortega</i>	90
Resumen	90
Introducción a los modelos de ecuaciones estructurales.	91
Antecedentes de los modelos de ecuaciones estructurales.	92
Características de los modelos de ecuaciones estructurales	94
Modelos reflectivos y modelos formativos	101
Etapas para la modelación de los SEM.	103
Caso práctico. Factores que inciden en la satisfacción residencial en México: una estimación a partir de microdatos de vivienda y la modelación PLS.	110
Resultados	114
Conclusiones	119
Referencias	120
V. Modelos de cointegración y ciclo común	
<i>Jorge Quiroz Félix, Carlos Alberto Flores Sánchez</i>	
<i>Martha Ofelia Lobo Rodríguez</i>	123
Resumen	123
Fundamentos del análisis VAR	124
Metodología de cointegración de Johansen	129
Modelo para determinar ciclos comunes	131
Caso práctico. Competitividad turística: análisis a corto y largo plazo	134
Resultados	138
Conclusiones	142
Referencias	143

VI. Diseño y validación de encuestas	
<i>Malena Portal Boza, Edgar Jiménez Cerra</i>	145
Resumen	145
La encuesta como fuente primaria de recolección de datos . .	146
Elementos de importancia en el diseño del cuestionario . . .	149
Validez y fiabilidad del cuestionario	153
Caso práctico. Calidad percibida y satisfacción de los servicios educativos desde la perspectiva de los estudiantes de educación superior	159
Resultados	162
Discusión de resultados	165
Conclusiones	170
Referencias	171
VII. Sistemas complejos aplicados a la investigación. El enfoque de redes sociales	
<i>Nadxilli López Pérez, Martha Cecilia Jaramillo Cardona</i>	173
Resumen	173
Introducción al enfoque de redes sociales	174
Caso práctico. La transversalidad de las políticas públicas con perspectiva de género para la violencia intrafamiliar: un análisis de los avances del estado de Baja California a la luz de América Latina	181
Análisis de resultados	183
Conclusiones	199
Referencias	200
VIII. Índice de confianza ciudadana. Percepción sobre economía y seguridad en la Zona Metropolitana de Tijuana	
<i>Ismael Plascencia López, Oswaldo Plascencia López</i> <i>Abraham Neftalí Sodi Díaz</i>	202
Resumen	202
Introducción al índice de confianza ciudadana	203
Antecedentes	204
Metodología	208

Resultados 215
Conclusiones 223
Referencias 225

Anexos 226
Notas sobre los coordinadores y autores 241

*Modelos estadísticos de investigación
científica. Aplicaciones en las áreas económico
administrativas*, Duniesky Feitó Madrigal, Malena
Portal Boza, Ismael Plascencia López (coordinadores),
publicado por Ediciones Comunicación Científica, S. A. de C.
V., se terminó de imprimir en septiembre de 2023, en Litográfica
Ingramex S.A. de C.V., Centeno 162-1, Granjas Esmeralda, 09810,
Ciudad de México. El tiraje fue de 100 ejemplares impresos y en versión
digital para acceso libre en los formatos PDF, EPUB y HTML.

La presente obra tiene por objetivo exponer de manera práctica y en un lenguaje sencillo y amigable para el lector, la aplicación de la estadística como herramienta de investigación en las áreas económico-administrativas. A diferencia de otros libros de esta misma rama, *Modelos estadísticos para la investigación científica* explica mediante ejemplos descriptivos una variedad de metodologías estadísticas que incluyen tanto el trabajo con fuentes de información secundaria, como la elaboración, validación y aplicación de instrumentos para la recolección de datos a través de encuestas. Está dirigido fundamentalmente a estudiantes, docentes, investigadores y profesionistas que estén interesados en desarrollar o fortalecer sus capacidades de investigación a través de la aplicación y modelación de métodos cuantitativos. Para este propósito, se incluyen casos prácticos para modelos de tabulación cruzada, de elección discreta, ecuaciones estructurales, cointegración y ciclo común, la elaboración de índices y el enfoque de redes sociales, donde se integra el planteamiento metodológico acompañado con el análisis y la interpretación de métricos. Este trabajo se enriquece con la participación de autores de diversas unidades académicas adscritos a instituciones públicas de educación superior que cuentan con una vasta experiencia en la investigación científica y cuya motivación es compartir con las futuras generaciones los beneficios del manejo de bases de datos y su procesamiento estadístico, lo cual, de igual forma, contribuya al entendimiento de los fenómenos complejos del mundo actual y que derive en propuestas y soluciones a las problemáticas reales de nuestra sociedad.



Duniesky Feitó Madrigal es doctor en Ciencias Económicas; miembro del Sistema Nacional de Investigadores de México; autor de libros, capítulos de libros y artículos de investigación de reconocimiento científico, y académico certificado por la Asociación Nacional de Facultades y Escuelas de Contaduría y Administración (ANFECA).



Malena Portal Boza es doctora en Ciencias Económicas y miembro del Sistema Nacional de Investigadores de México. Forma parte del comité de revisores de artículos en revistas científicas latinoamericanas, y ha sido parte de la Academia de Ciencias Económico-Administrativas de la Universidad Autónoma de Baja California (UABC).



Ismael Plascencia López es doctor en Ciencias Económicas y miembro del Sistema Nacional de Investigadores de México. Preside el Colegio Estatal de Economistas de Baja California (CEEBC) y fue fundador de la Red Internacional de Bionanotecnología. Cuenta con una amplia productividad científica y de innovación en temas de desarrollo económico y regional.



Dimensions



[DOI.ORG/10.52501/CC.131](https://doi.org/10.52501/CC.131)



**COMUNICACIÓN
CIENTÍFICA** PUBLICACIONES
ARBITRADAS
HUMANIDADES, SOCIALES Y CIENCIAS

www.comunicacion-cientifica.com

ISBN: 978-607-59749-9-6



9 786075 974996