

**Trayectorias laborales y cambio tecnológico sesgado por tareas:
análisis del caso mexicano**

Liliana Meza González y Jael Pérez Sánchez

**Seguimiento de la distribución del ingreso en México a lo largo
del tiempo y de la geografía**

Víctor Alfredo Bustos y de la Tijera y Ana Miriam Romo Anaya

**Toma de decisiones en ambientes de salud: modelos de
jerarquización analítica de alternativas para la vacunación
en México**

Mario Gutiérrez-Lagunes

**Uso de los Censos Económicos para la detección de factores de éxito
en las empresas de la manufactura automotriz mediante modelos
de clúster**

Edgar Manuel Matus Carballo y Carlos Alberto Jiménez-Bandala

**Índices de precios de renta de la vivienda en México basados
en la ENIGH**

Jesús López-Pérez y Francisco de Jesús Corona Villavicencio

**Indicadores de agua limpia y saneamiento en la Ciudad de México y
el reto ante un desarrollo sostenible de los recursos hídricos**

Berenice García Cabrera

**Identificación de usuarios ansiosos para detectar condiciones
de salud mental en la población tuitera de México**

Ana Miriam Romo Anaya, Víctor Silva Cuevas e Irving Gibran Cabrera Zamora

Contenido

Trayectorias laborales y cambio tecnológico sesgado por tareas: análisis del caso mexicano <i>Labor Trajectories and Task-Biased Technological Change: the Mexican Case</i> Liliana Meza González y Jael Pérez Sánchez	4
Seguimiento de la distribución del ingreso en México a lo largo del tiempo y de la geografía <i>Monitoring the Income Distribution in Mexico over Time and Geography</i> Victor Alfredo Bustos y de la Tijera y Ana Miriam Romo Anaya	28
Toma de decisiones en ambientes de salud: modelos de jerarquización analítica de alternativas para la vacunación en México <i>Decision-making in Health Care Settings: Analytical Ranking Models for Vaccination Alternatives in Mexico</i> Mario Gutiérrez-Lagunes	50
Uso de los Censos Económicos para la detección de factores de éxito en las empresas de la manufactura automotriz mediante modelos de clúster <i>Use of Economic Censuses for the Detection of Success Factors in Automotive Manufacturing Companies through Cluster Models</i> Edgar Manuel Matus Carballo y Carlos Alberto Jiménez-Bandala	64
Índices de precios de renta de la vivienda en México basados en la ENIGH <i>Rental Housing Price Indexes in Mexico based on ENIGH</i> Jesús López-Pérez y Francisco de Jesús Corona Villavicencio	82
Indicadores de agua limpia y saneamiento en la Ciudad de México y el reto ante un desarrollo sostenible de los recursos hídricos <i>Clean Water and Sanitation Indicators in Mexico City and the Challenge of Sustainable Development of Water Resources</i> Berenice García Cabrera	108
Identificación de usuarios ansiosos para detectar condiciones de salud mental en la población tuitera de México <i>Identification of Anxious Users to Detect Mental Health Conditions in the Twitter Population of Mexico</i> Ana Miriam Romo Anaya, Víctor Silva Cuevas e Irving Gibran Cabrera Zamora	122
Colaboran en este número	142

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Presidencia del Instituto

Graciela Márquez Colín

Vicepresidencias

Enrique de Alba Guerra

Paloma Merodio Gómez

Adrián Franco Barrios

Mauricio Márquez Corona

Direcciones generales de:

Estadísticas Sociodemográficas

Edgar Vielma Orozco

Estadísticas de Gobierno, Seguridad Pública y Justicia

Óscar Jaimes Bello

Estadísticas Económicas

José Arturo Blancas Espejo

Geografía y Medio Ambiente

Luis Gerardo Esparza Ríos

Integración, Análisis e Investigación

Sergio Carrera Riva Palacio

Coordinación del Sistema Nacional de Información Estadística y Geográfica

Silvia Elena Meza González

Comunicación, Servicio Público de Información y Relaciones Institucionales

Julieta Alejandra Brambila Ramírez

Administración

Ricardo Miranda Burgos

Contraloría Interna

Manuel Rodríguez Murillo

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Editor responsable

Sergio Carrera Riva Palacio

Editor técnico

Gerardo Leyva Parra

Coordinación editorial

Virginia Abrin Batule y Mercedes Pedrosa Islas

Corrección de estilo

José Pablo Covarrubias Ordiales

Corrección de textos en inglés

Gerardo Hazael Piña Méndez

Diseño y formación

Eduardo Javier Ramírez Espino

Edición para internet

Ana Victoria Flores Flores y José Andrés Ortiz Domínguez

Indizada en: Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal *Latindex Catálogo*; Citas Latinoamericanas en Ciencias Sociales y Humanidades (*CLASE*) y en la Red Iberoamericana de Innovación y Conocimiento (REDIB).

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, Vol. 13, Núm. 3, septiembre-diciembre, 2022, es una publicación cuatrimestral editada por el Instituto Nacional de Estadística y Geografía, Avenida Héroe de Nacozari Sur 2301, Fraccionamiento Jardines del Parque, 20276, Aguascalientes, Aguascalientes, Aguascalientes, entre la calle INEGI, Avenida del Lago y Avenida Paseo de las Garzas, México. Teléfono 55 52781069. Toda correspondencia deberá dirigirse al correo: rde@inegi.org.mx

Editor responsable: Sergio Carrera Riva Palacio. Reserva de Derechos al Uso Exclusivo del Título Núm. 04-2014-101310563200-203, e-ISSN Núm. 2395-8537, ambos otorgados por el Instituto Nacional del Derecho de Autor. Domicilio de la publicación: Avenida Héroe de Nacozari Sur 2301, Fraccionamiento Jardines del Parque, 20276, Aguascalientes, Aguascalientes, Aguascalientes, entre la calle INEGI, Avenida del Lago y Avenida Paseo de las Garzas, México.

El contenido de los artículos, así como sus títulos y, en su caso, fotografías y gráficos utilizados son responsabilidad del autor, lo cual no refleja necesariamente el criterio editorial institucional.

Asimismo, la Revista se reserva el derecho de modificar los títulos de los artículos, previo acuerdo con los autores. La mención de empresas o productos específicos en las páginas de la Revista no implica el respaldo por el Instituto Nacional de Estadística y Geografía.

Se permite la reproducción total o parcial del material incluido en la Revista, sujeto a citar la fuente.

Página electrónica: <http://rde.inegi.org.mx>

CONSEJO EDITORIAL

Dr. Enrique de Alba Guerra

Presidente del Consejo

Mtra. Claudia Aburto Rancaño

Instituto Tecnológico Autónomo de México

México

Dr. Clemente Ruiz Durán

Universidad Nacional Autónoma de México

México

Dr. Emilio Cunjamá López

Instituto Nacional de Ciencias Penales

México

Dr. Fernando Cortés Cáceres

Profesor emérito de FLACSO PUED de la UNAM

México

Dra. Graciela Teruel Belismelis

Universidad Iberoamericana

Ciudad de México

México

Dra. Landy Sánchez Peña

El Colegio de México

México

Dra. María Martha Téllez Rojo Solís

Instituto Nacional de Salud Pública

México

Dr. Víctor Manuel Guerrero Guzmán

Instituto Tecnológico Autónomo de México

México



Offline Work

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Editorial

Concluimos este 2022 de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía* con el siguiente contenido.

Trayectorias laborales y cambio tecnológico sesgado por tareas: análisis del caso mexicano (Labor Trajectories and Task-Biased Technological Change: the Mexican Case). A partir de los datos del Módulo de Trayectorias Laborales 2015 del Instituto Nacional de Estadística y Geografía (INEGI), los autores analizan del 2010 al 2015 el número de trabajos, el acceso permanente a la seguridad social, la falta de seguridad social, la transición del trabajo asalariado al cuentapropismo y el despido en el primer empleo de los trabajadores definidos por el tipo de tareas que desarrollan: no rutinarias, cognitivas y manuales, así como rutinarias, cognitivas y manuales. De este artículo se desprende que el avance tecnológico actual sigue beneficiando, sobre todo, a los trabajadores con mayores niveles de capital humano, independientemente del tipo de tareas que se desarrollan.

Seguimiento de la distribución del ingreso en México a lo largo del tiempo y de la geografía (Monitoring the Income Distribution in Mexico over Time and Geography). El análisis se basa, principalmente, en mediciones de desigualdad derivadas de estimaciones para dichas distribuciones y su comparación tanto entre sí como con mediciones semejantes derivadas de la Encuesta Nacional de Ingresos y Gastos de los Hogares o su Módulo de Condiciones Socioeconómicas del INEGI. Los autores se dieron a la tarea de establecer nuevas comparaciones con los resultados muestrales para ratificar o rectificar el comportamiento de diversas tendencias aparentes para los años 2010, 2012, 2014 y 2016. El resultado parece sugerir que lo que ha impulsado el crecimiento de la desigualdad en el país es el aumento de esta entre los estados.

Toma de decisiones en ambientes de salud: modelos de jerarquización analítica de alternativas para la vacunación en México (Decision-making in Health Care Settings: Analytical Ranking Models for Vaccination Alternatives in Mexico). El objetivo de este trabajo es, aplicando el esquema del Proceso Analítico Jerárquico, evaluar el uso y la pertinencia de vacunación masiva en un ambiente de urgencia por combatir el COVID-19. Muestran que el análisis y la jerarquización de los criterios para vacunar y los estados de vulnerabilidad con respecto a la salud con estas variables se utilizaron para crear matrices de comparación pareadas. La construcción de las matrices para analizar y evaluar objetivamente la jerarquización presentada fue consistente.

Uso de los Censos Económicos para la detección de factores de éxito en las empresas de la manufactura automotriz mediante modelos de clúster (Use of Economic Censuses for the Detection of Success Factors in Automotive Manufacturing Companies through Cluster Models). Esta

investigación tiene por objetivo identificar las prácticas exitosas entre empresas de la industria automotriz a partir del uso de la información de ese programa estadístico del INEGI. El trabajo demuestra la importancia de los datos censales, así como la inclusión del *Big Data* y la inteligencia artificial para tomar mejores decisiones.

Índices de precios de renta de la vivienda en México basados en la ENIGH (Rental Housing Price Indexes in Mexico based on ENIGH). Aborda el problema de la construcción de estos para el mercado de vivienda mexicano, tanto para las rentadas como aquellas ocupadas por sus dueños. En esta investigación se cuantificaron las variaciones en el alquiler por entidad federativa para el periodo 2010-2018; para ello, se utilizaron tres metodologías de índices de precios (Hedónico, Estratificado y Espacial) para abordar la problemática sobre la evolución de los precios de renta en el país desde diferentes perspectivas, ya sea ajustando por la calidad de los inmuebles o al separar por tamaños de localidad o emparejando viviendas en entidades federativas que pueden estar muy distantes.

Indicadores de agua limpia y saneamiento en la Ciudad de México y el reto ante un desarrollo sostenible de los recursos hídricos (Clean Water and Sanitation Indicators in Mexico City and the Challenge of Sustainable Development of Water Resources). En este trabajo se discute la importancia de la responsabilidad social para un desarrollo sostenible y el rol del administrador y se analizan los datos e indicadores actuales en cuanto al acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente en la entidad. Se pone de manifiesto la necesidad de contar con indicadores y datos, lo cual se fundamenta en este artículo debido a que a nivel local hay carencia de estos que posibiliten medir el grado de avance ante un desarrollo sostenible en la urbe.

Identificación de usuarios ansiosos para detectar condiciones de salud mental en la población tuitera de México (Identification of Anxious Users to Detect Mental Health Conditions in the Twitter Population of Mexico). A partir de múltiples estudios que muestran que las palabras que usa una persona en sus escritos transmiten información sobre su personalidad, motivaciones, estados de ánimo y emocional e incluso nivel económico, la investigación presentada en este artículo pretende generar información relevante acerca de la emoción negativa *ansiedad* en los mensajes de los tuiteros en México, *identificando el grado de ansiedad y en qué momentos se incrementa* como indicadores en el tema de salud pública y bienestar social utilizando datos no tradicionales (los que no son obtenidos por censos o encuestas).

<http://rde.inegi.org.mx>

Trayectorias laborales y cambio tecnológico sesgado por tareas: análisis del caso mexicano

Labor Trajectories and Task-Biased Technological Change: the Mexican Case

Liliana Meza González y Jael Pérez Sánchez*

* Instituto Nacional de Estadística y Geografía, liliana.meza@ibero.mx y jael.perez@inegi.org.mx, respectivamente.



Little Latin girl with her basket full of local vegetables and fruit / iStock

Usando datos del Módulo de Trayectorias Laborales 2015 del Instituto Nacional de Estadística y Geografía, este trabajo analiza, para el periodo 2010-2015, el número de trabajos, el acceso permanente a la seguridad social, la falta de seguridad social, la transición del trabajo asalariado al cuentapropismo y el despido en el primer empleo de los trabajadores definidos por el tipo de tareas que desarrollan: no rutinarias, cognitivas y manuales, y rutinarias, cognitivas y manuales. A través de estimaciones de modelos de riesgo, esta investigación encuentra que los trabajadores que realizan tareas de naturaleza no rutinaria cognitiva son los que presentan un mejor desempeño laboral, cuando se comparan con sus contrapartes, mientras que los que hacen tareas de naturaleza no rutinaria manual son los que muestran un peor desempeño laboral. Esto sugiere que el avance tecnológico actual sigue beneficiando, sobre todo, a los trabajadores con mayores niveles de capital humano, independientemente del tipo de tareas que se desarrollan.

Clasificación JEL: J28, J21, O33.

Palabras clave: trayectorias laborales; cambio tecnológico sesgado por tareas; mercado laboral; México.

Recibido: 7 de septiembre de 2020.
Aceptado: 3 de diciembre de 2021.

1. Introducción

A lo largo de la historia se han suscitado diferentes tipos de cambio tecnológico. Por ejemplo, Acemoglu (2002) argumenta que la tecnología utilizada en la Revolución Industrial del siglo XVIII sustituyó al trabajo calificado, lo que incrementó la demanda por el de baja calificación. Por su parte, Acemoglu (2002) y Autor *et al.* (2003) mencionan que, hacia principios del siglo XXI, la creciente utilización de sistemas informáticos y de computadoras incrementó la demanda por trabajadores con mayor nivel educativo.

En la actualidad, el cambio tecnológico parece distinto, pues está sustituyendo a los trabajadores que realizan tareas rutinarias, independientemente de su nivel de calificación. Las máquinas actuales

Using data from INEGI's 2015 Labor Trajectory Module (MOTRAL), this work analyzes, for the period 2010-2015, the number of jobs, permanent access to social security, lack of social security, transition from salaried work to self-employment and dismissal in the first job of workers defined by the type of tasks they perform: non-routine, cognitive and manual, and routine, cognitive and manual. Through estimates of risk models, this work finds that workers who perform tasks of a non-routine cognitive nature are those who present a better job performance, when compared to their counterparts, while those who perform tasks of a non-routine manual nature are those with the worst job performance. This suggests that current technological advance continues to benefit workers with higher levels of human capital above all, regardless of the type of tasks performed.

JEL classification: J28, J21, O33.

Key words: labor trajectories; task-biased technological change; labor market; Mexico.

son capaces de reemplazar a quienes desempeñan tareas que siguen reglas claras de programación. En cambio, las máquinas complementan a los empleados que hacen labores en las que las reglas de programación no están bien establecidas y no han podido ser automatizadas por computadoras o ejecutadas de manera sistemática. A este cambio tecnológico se le ha denominado *sesgado por tareas* (*task-biased technological change* o *TBTC*) y es, precisamente, este el que parece estar detrás de la polarización en el empleo que se observa en los Estados Unidos de América (EE. UU.), Alemania y Reino Unido, entre otros países avanzados (Rodríguez y Meza, 2021).

Con respecto a la polarización laboral en la Unión Americana, Autor (2010) argumenta que el empleo y las remuneraciones relativas aumenta-

ron a inicios del siglo XXI en las ocupaciones de alta calificación, como las directivas y las profesionales, así como en las de baja calificación, por ejemplo, la preparación de alimentos, el cuidado de niños y ancianos y las labores de limpieza. Nótese que estas desarrollan tareas de naturaleza no rutinaria. Este aumento simultáneo en empleo y salarios relativos de los trabajadores en los extremos de las distribuciones de salarios y calificaciones se ha interpretado como un incremento sostenido en la demanda relativa por este tipo de empleados. Por otro lado, las oportunidades de empleo y las remuneraciones relativas se han contraído para quienes realizan labores administrativas, técnicas, operativas y artesanales, las cuales se caracterizan por ser fácilmente automatizables con la tecnología actual. Esta reducción se ha interpretado como una clara disminución en la demanda relativa por este tipo de trabajadores.

Claramente, el tipo de tareas que los trabajadores desarrollan puede definir el tipo de trayectorias laborales que estos experimentan. Por ejemplo, si la demanda por empleados que realizan labores no rutinarias está al alza, es probable que estos trabajadores no solo reciban mayores salarios, sino que también cuenten con mejores y más prestaciones, que tengan menor rotación laboral o que se mantengan en el sector formal por más tiempo. Meza (2019) y Rodríguez y Meza (2021) han argumentado que en el mercado laboral femenino mexicano se observa un proceso de polarización laboral similar al que se ha reportado en países desarrollados, mientras que en el masculino está sucediendo lo opuesto. Si esto es cierto, entonces deberíamos observar que la mayor demanda por trabajadores que realizan tareas de naturaleza no rutinaria, sobre todo en el caso de las mujeres, se acompaña de mejores trayectorias laborales; es decir, de mayor acceso a trabajos de calidad, con menor rotación y probabilidad de despido e informalidad.

El objetivo de esta investigación es entender cómo difieren las trayectorias laborales de los trabajadores cuando estos se clasifican de acuerdo con el tipo de tareas que desarrollan; es decir, busca analizar aquellas de quienes realizan tareas

de naturaleza no rutinaria, tanto cognitiva como manual, y las de aquellos que hacen labores de índole rutinaria, tanto cognitiva como manual. La hipótesis del estudio es que, si los trabajadores que realizan tareas no rutinarias, ya sean cognitivas o manuales, tienen mejores trayectorias laborales que sus contrapartes que hacen labores de carácter rutinario, entonces se puede argumentar que el cambio tecnológico sesgado por tareas está permeando en el país. Por el contrario, si los trabajadores que realizan tareas de índole cognitiva, ya sea rutinaria o no rutinaria, son los que presentan mejores trayectorias laborales, entonces se podría argumentar que el cambio tecnológico sesgado por calificación es el que está permeando. El análisis distingue las trayectorias laborales de hombres y mujeres, pues reconocemos que existe evidencia suficiente para decir que ellas están siendo afectadas por el cambio tecnológico más que ellos debido a que presentan complementariedad con la tecnología (Yamaguchi, 2018).

Se debe reconocer, sin embargo, que los años comprendidos entre el 2015 y 2021 han sido testigos de un gran avance de la digitalización, lo cual se considera cambio tecnológico sesgado por tareas, por lo que una limitante del estudio es que los resultados solo se pueden asociar al periodo 2010-2015 y no más adelante. Asimismo, es importante decir que los trabajadores se autoseleccionan en las ocupaciones y los empleos que mejor se adaptan tanto a sus características observables (sexo, edad y nivel educativo) como a las no observables (preferencias respecto al ocio-trabajo, aspiraciones, creencias, aversión al riesgo o motivación para el trabajo, entre otras), por lo cual es probable que las estimaciones tengan bajos coeficientes de ajuste. Esta autoselección y los bajos niveles de ajuste de estas se consideran limitaciones adicionales del estudio.

Los resultados del análisis muestran que los trabajadores no rutinarios cognitivos son los que menos trabajos realizaron entre el 2010 y 2015 o, en otras palabras, los que tuvieron una menor rotación laboral; también, son los que en mayor proporción conservaron su seguridad social a lo

largo del periodo analizado; los que en menores proporciones se mantuvieron en el sector informal y, también, los que menos fueron despedidos de su primer trabajo. Sin embargo, el estudio también muestra que los empleados con trayectorias laborales más precarias¹ fueron los que realizan tareas de naturaleza no rutinaria manual, lo cual sugiere que, en México, este tipo de cambio tecnológico (el sesgado por tareas) no ha permeado como en otros países, y que el sesgado por calificación sigue siendo el que caracteriza al sector productivo nacional.

El estudio está organizado de la siguiente manera: el apartado 2 incluye una revisión de literatura sobre las trayectorias laborales diferenciadas por grupos de trabajadores; el 3 trata de los datos y la metodología de análisis, mientras que el 4 incluye las estadísticas descriptivas y los principales resultados; finalmente, el 5 presenta las conclusiones.

2. Revisión de la literatura

En general, la que trata sobre trayectorias laborales argumenta que las características observables de los trabajadores, como su sexo o edad, determinan las sendas laborales que cada uno experimenta. Así, las mujeres tienen una mayor propensión a ser trabajadoras sin pago, mientras que los jóvenes tienden a combinar los estudios con el trabajo (Moen, 2001). Esto implica que en el análisis de las trayectorias laborales se deben incluir como variables explicativas precisamente estas características observables de los distintos tipos de trabajadores. En el caso de las mujeres, Damaske y Frech (2016) argumentan que las obligaciones familiares determinan en gran medida sus transiciones del empleo al desempleo o a la informalidad, pero también reconocen que la pobreza, la raza, el nivel educativo y ciertas características de sus familias (como la edad de los hijos) se consideran determinantes de sus experiencias laborales. Las autoras indican, sin

¹ Aquí, la precariedad se define por mayor inseguridad en el trabajo y menor acceso a la seguridad social y otras prestaciones asociadas al trabajo formal. Ver Guadarrama, Hualde y López (2014); Rubio (2010); Blanco y Julián (2019); Pérez y Ceballos (2019); Román (2013); entre otros.

embargo, que las expectativas de vida y las creencias también definen el curso de la vida laboral de las mujeres, lo cual nos sugiere que hay características no observables de los trabajadores que también influyen en el camino laboral que seguirán en un determinado periodo. Con datos longitudinales que abarcan más de 30 años, su investigación analiza la variación en las horas trabajadas de las mujeres estadounidenses, tanto en la juventud como en la edad adulta, para probar tres teorías sociológicas sobre decisiones laborales por género: la de ventajas/desventajas acumulativas, la de la de creencias de género y la de restricciones trabajo-familia. Concluyen que las mujeres más aventajadas, es decir, las que provienen de familias más acomodadas, tienen trayectorias laborales más estables y continuas. Además, demuestran que las creencias sobre el papel de la mujer en la familia y en la sociedad determinan en gran medida sus trayectorias laborales.

Este análisis de las decisiones laborales de las mujeres en función de la estructura de su familia también lo abordan Lu, Shu-Huah y Han (2017) para el caso específico de aquellas que son madres. Con base en análisis secuencial, establecen las trayectorias laborales de las madres desagregando la muestra por raza, etnia y lugar de nacimiento. Los resultados muestran que antes de ser madres, las mujeres en EE. UU. presentan una enorme continuidad laboral, pero que sus trayectorias posteriores son más inestables. Un hallazgo interesante de ese trabajo es que sus trayectorias laborales después de la maternidad difieren por raza y nacionalidad. Las de color en general muestran mayor continuidad laboral que las mujeres blancas después de tener hijos. Además, las inmigrantes con poco tiempo de residencia tienden a acortar sus periodos laborales cuando se comparan con las nativas, pero las que tienen más tiempo de residencia tienden a estar más comprometidas con el mercado laboral.

Vale la pena mencionar que la sociología es la disciplina que más se ha abocado a estudiar historias laborales, y lo ha hecho ya sea con un enfoque teórico o uno empírico. Los estudios empíricos usan datos ya sea longitudinales o de historias de

vida. Por ejemplo, la investigación de Halperc *et al.* (2015), de corte sociológico, analiza el bienestar en la edad adulta en función de las historias laborales y familiares. Ese estudio también usa datos longitudinales de EE. UU. para estimar trayectorias latentes de trabajo y familia entre la adolescencia y la edad de 65 años. Después, modela el bienestar económico a la edad de 65 en función de esas trayectorias y de indicadores de salud, familia y empleo actuales. Su hallazgo central es que las trayectorias laborales y familiares tienen un efecto directo en el bienestar económico tardío de las personas.

En México, como en otros países, también se han analizado las trayectorias laborales de distintos tipos de trabajadores para entender qué factores están detrás, por ejemplo, del número de empleos que tiene una persona en un determinado lapso, de su ascenso o descenso en la empresa, de su transición del empleo al desempleo y viceversa o de su transición del sector formal al informal. Así, el análisis de estas ha comprendido el estudio de las transiciones laborales, aunque abarca más que eso. Los trabajos que abordan estas temáticas son tanto cuantitativos como cualitativos.

Una veta interesante en su estudio es la de los trabajadores migrantes. Las investigaciones que se orientan en el análisis de las trayectorias laborales y migratorias tratan de entender, por ejemplo, por qué los migrantes de retorno y los inmigrantes tienen una mayor propensión a autoemplearse cuando se les compara con los trabajadores nativos. También, es probable que problemas económicos coyunturales, como una crisis financiera, modifiquen la forma en la que estas trayectorias laborales se están desarrollando. Asimismo, es posible que situaciones más estructurales, como un cambio en el régimen de pensiones, modifiquen la manera en la que diferentes trabajadores transitan de una situación laboral a otra.

La literatura sobre este tema en México no es muy extensa, y esta es, sobre todo, de carácter cualitativo. Los estudios cuantitativos son pocos, seguramente porque los datos necesarios para llevar a cabo este tipo de investigaciones no son muy co-

munes. El trabajo de Calderón (2018) estima el efecto de los cambios en el régimen de pensiones que tuvieron lugar en el país en 1997 (de un sistema contributivo a otro de cuentas individualizadas de ahorros) sobre las transiciones de los trabajadores del sector formal al informal. Con datos del Módulo de Trayectorias Laborales (MOTRAL) de la Encuesta Nacional de Ocupación y Empleo (ENOE) 2015,² el estudio encuentra que la reforma de pensiones de 1997 en México provocó, en el mercado laboral masculino, un aumento en el riesgo de transitar del sector formal al informal —es decir, un decremento en el tiempo promedio laborando en el sector formal—, independientemente del tipo de contrato bajo el cual estaba contratado el trabajador. Por otro lado, también detalla que la reforma no tuvo efectos en las transiciones de las mujeres hacia la informalidad. Las transiciones del mercado laboral formal al informal son importantes no solo por el deterioro que implican en la calidad del empleo actual, sino por las repercusiones que esto tiene sobre la probabilidad de recibir una pensión digna al momento de la jubilación. Calderón asume que las transiciones de los individuos entre el sector formal y el informal son voluntarias, lo cual puede ser cuestionable.

En la literatura sobre este tema es común encontrar un debate respecto de si los trabajadores en la informalidad están ahí de manera voluntaria o si el mercado laboral está segmentado de tal manera que algunos grupos específicos de empleados no pueden transitar a la formalidad debido a sus características tanto observables como no observables (por ejemplo, por su nivel educativo, color de piel o etnia). La evaluación que hace Calderón (2018) del impacto de la reforma de pensiones sobre la participación en el sector formal por parte de los trabajadores requirió de la definición de dos grupos de individuos: los de control (quienes no sufrieron cambio en su sistema de pensiones) y los de tratamiento. Asimismo, y para evitar errores de medición, fue necesario establecer una estrategia de identificación para el parámetro de interés, que en ese trabajo es la interacción entre

² Programa estadístico del Instituto Nacional de Estadística y Geografía (INEGI).

el tiempo de permanencia en el sector formal y el indicador del esquema de pensiones contributivo en el que el individuo está registrado.

Este análisis de las transiciones laborales del sector formal al informal ha sido estudiado con datos de México para entender qué factores determinan el que los trabajadores cambien de sector en economías en desarrollo. Una discusión relevante en este tema es si los individuos transitan hacia la informalidad de manera forzada o voluntaria. El estudio de Maloney (1999) aborda la pregunta de si las transiciones laborales desde el trabajo asalariado formal hacia el sector informal, ya sea como asalariado (con y sin contrato) o cuentapropista, son resultado de las decisiones racionales de los propios trabajadores (enfoque de oferta) o si estos se ven obligados a transitar del sector formal al informal debido a la segmentación del propio mercado laboral (enfoque de demanda).

En la literatura sobre esto en México es común encontrar trabajos que centren su análisis en las diferencias entre las trayectorias laborales de los trabajadores migrantes y los nativos. Un estudio que analiza las de los primeros es el de Wassink y Hagan (2018), quienes hacen uso de datos longitudinales levantados mediante entrevistas semiestructuradas en León, Guanajuato, en el 2010 y 2015. La muestra está compuesta por 200 migrantes de retorno y estas tuvieron lugar tanto en sus hogares como en sus lugares de trabajo. La investigación incluyó, también, la observación del desempeño laboral de los migrantes en empresas de tamaño grande y mediano, así como en las familiares pequeñas. Las entrevistas incluyeron una combinación de preguntas cerradas y abiertas, que capturaron las historias migratorias y laborales completas, además de las motivaciones para la migración de los encuestados. El estudio, por lo tanto, se puede considerar multimétodo, pues es posible realizar análisis tanto cuantitativo como cualitativo. La reconstrucción de las historias laborales de los migrantes de retorno permitió a los investigadores indagar temas como la entrada y salida del autoempleo de los encuestados, y averiguar si el autoempleo de los migrantes es una

opción que se facilita con los ahorros obtenidos en la estancia en el extranjero o si este es una alternativa a la persistente marginación económica que enfrentan los migrantes de retorno en México. La naturaleza longitudinal de los datos permite tomar en cuenta la heterogeneidad individual no observada en las estimaciones.

En otro tipo de estudios, Parra-Sandoval *et al.* (2019) analizan las trayectorias laborales de la gente de origen mexicano que estudia en el extranjero. El objetivo es probar la hipótesis de que la obtención de un certificado de estudios en una universidad prestigiosa de otra nación provee mejores oportunidades para seguir una trayectoria laboral exitosa, tanto en el país de origen como en otros. Las autoras caracterizan las condiciones sociales, económicas y políticas de los estudiantes y académicos mexicanos que estudiaron algún programa de prestigio internacional en seis países desarrollados, tanto de Norteamérica como de Europa: Canadá, Estados Unidos de América, Alemania, España, Francia y Reino Unido. Los datos se obtuvieron de 2 228 encuestados en línea y de 273 entrevistas a becarios del Consejo Nacional de Ciencia y Tecnología (CONACYT) de México. La información de las trayectorias laborales se obtuvo del Currículum Vitae Único (CVU) que los becarios del CONACYT y miembros del Sistema Nacional de Investigadores deben reportar al Consejo de manera periódica. Las trayectorias laborales son reconstruidas en términos de cómo, cuándo y dónde sucedieron los eventos previos, simultáneos y posteriores a los estudios en el extranjero. Asimismo, analizan cómo las trayectorias escolares impactaron las laborales, dependiendo de la generación a la que pertenece el estudiante (antes de 1984, 1985-2002 o posterior al 2002) y el campo de estudio.

Otro estudio que aborda el tema de las trayectorias laborales de los migrantes es el de Blume *et al.* (2004), aunque este utiliza datos de panel levantados en Dinamarca. La investigación hace un seguimiento en el tiempo de trabajadores tanto migrantes como nativos. Las transiciones que se analizan son entre tres situaciones laborales:

el empleo asalariado, el autoempleo y el desempleo. Las probabilidades de las transiciones se estiman usando modelos de riesgo discretos que controlan por la heterogeneidad no observada y por la dependencia de la duración. El trabajo encuentra que, para ciertos grupos de migrantes, el autoempleo es un último recurso y no una opción de primera instancia. El análisis empírico revela que los inmigrantes tienen patrones de transición laboral muy diferentes a los de los nativos. Por ejemplo, en el grupo de inmigrantes procedentes de países desarrollados, las transiciones se explican principalmente por sus características individuales. Una vez que se controla por las características individuales, los autores encuentran que los inmigrantes procedentes de naciones desarrolladas no se autoemplean de manera marginalizada. Sin embargo, para los que vienen de países emergentes, el panorama es diferente, pues ellos si se autoemplean de forma marginalizada, aun después de que en la estimación se controla tanto por características observables como no observables. El estudio es interesante, pues muestra que las trayectorias laborales en Dinamarca son muy diferentes entre los trabajadores nativos, los inmigrantes procedentes de países desarrollados y los de naciones en desarrollo, como Turquía, Irán y Vietnam.

En resumen, la literatura empírica sobre trayectorias laborales en México es fundamentalmente cualitativa, pues no es común tener datos de muestras representativas de grandes poblaciones que sigan a los encuestados en el tiempo (longitudinales) o los que pregunten sobre las historias laborales. El MOTRAL que el INEGI levantó en el 2015 junto con la ENOE permite reconstruir las trayectorias laborales para distintos grupos de trabajadores, definidos tanto por sus características sociodemográficas como laborales. El uso más común de la información sobre trayectorias es para analizar transiciones entre diferentes situaciones laborales, aunque los datos permiten estudiar otros fenómenos laborales.

A continuación, se presentan los datos usados en este trabajo y la metodología de análisis.

3. Datos y metodología

3.1. Datos

Los utilizados en esta investigación provienen del MOTRAL que se aplicó a una submuestra de viviendas de las áreas autorrepresentadas en la ENOE, en junio del 2015. Es importante enfatizar que el Módulo se levantó en una sola ocasión, por lo que no se trata de una base de datos longitudinal. Lo que sí hace es recuperar la historia laboral de los encuestados. El MOTRAL 2015 representa un esfuerzo conjunto de la Comisión Nacional del Sistema de Ahorro para el Retiro (CONSAR) y del INEGI. Su objetivo es generar información en materia de trayectorias laborales y seguridad social para el periodo que va de enero del 2010 a junio del 2015. La población objeto de estudio son las personas de 18 a 54 años de edad con experiencia laboral identificadas por la ENOE en marzo del 2015; es decir, que estaban en las siguientes situaciones:

- Ocupados, independientemente de que hayan tenido o no un trabajo anterior.
- Desocupados con experiencia laboral.
- No económicamente activos con experiencia laboral.

El MOTRAL 2015 tiene una cobertura nacional urbana por ser aplicado en una submuestra de viviendas de la ENOE correspondiente al dominio urbano alto de las 32 ciudades autorrepresentadas, por lo que tiene representatividad estadística nacional para dicho segmento. La muestra estimada fue de 7 mil viviendas.

Para clasificar todas las ocupaciones a cuatro dígitos de las bases de datos de la ENOE en las categorías: no rutinaria cognitiva, rutinaria cognitiva, rutinaria manual y no rutinaria manual, cuando se usó el Sistema Nacional de Clasificación de Ocupaciones (SINCO, vigente a partir del 2012), se consideraron las tareas principales de las ocupaciones en cada grupo unitario. Es importante mencionar que, en la mayoría de los casos, cada grupo del SINCO contiene más de una ocupación. A cada una se le clasificó en los grupos menciona-

Cuadro 1

Grupos de ocupaciones

Grupos	Definición	Ejemplo
1. No rutinaria cognitiva	<p>Para realizar tareas no rutinarias cognitivas, se necesitan habilidades de gestión, razonamiento analítico y habilidades cuantitativas, como aritmética y matemáticas avanzadas, con las cuales las computadoras complementan el trabajo humano, pero no lo sustituyen.</p> <p>Incluye tareas que requieren capacidades de resolución de problemas, intuición, creatividad y persuasión.</p>	<p>Ocupaciones que implican dirección, control y planificación de actividades que involucren un alto nivel gerencial e interpersonal. (diagnóstico médico, vender, escritura legal).</p> <p>Ocupaciones profesionales, técnicas y gerenciales, como: directivos de empresas y altos funcionarios en el sector público.</p>
2. Rutinaria cognitiva	Las ocupaciones rutinarias cognitivas requieren de establecer límites, tolerancias o normas, como un indicador, ya que siguen procedimientos precisos y bien entendidos.	Operar una máquina de facturación para transcribir datos de registros de la oficina (teneduría de libros, archivar/recuperar datos textuales, procesar interacciones/transacciones de procedimiento (por ejemplo, cajero de banco).
3. Rutinaria manual	Las ocupaciones rutinarias manuales requieren de facilidad de manipular y organizar de forma sistematizada.	Selección y clasificación de objetos de ingeniería en una línea de ensamblaje, reconfigurar líneas de producción para permitir tiradas cortas, como: obreros en plantas industriales.
4. No rutinaria manual	<p>Las tareas no rutinarias manuales no son susceptibles de informatización; esto quiere decir que no podrían ser sustituidas fácilmente por un robot.</p> <p>Requieren adaptabilidad situacional, visual y reconocimiento de idiomas e interacciones en persona y presentan desafíos desalentadores para la automatización.</p>	Ocupaciones que implican facilidad para las relaciones interpersonales y la toma de decisiones a nivel básico, como: trabajadores en servicios de limpieza, conductor de camiones, cuidado de otras personas, atención personalizada en locales de servicios.

Fuente: elaboración propia con base en Acemoglu y Autor (2011).

dos con base en el tipo de tareas que desarrollan. Para poder hacer la clasificación, se tomó como referencia lo identificado en el artículo de Acemoglu y Autor (2011), donde los grupos se construyeron con base en los criterios que se muestran en el cuadro 1.

Ya descrita la clasificación de ocupaciones y la base de datos utilizada, a continuación, se expone la metodología de análisis.

3.2. Metodología³

En esta investigación se estiman modelos discretos para analizar los determinantes de cinco variables dependientes: a) el número de trabajos en los que la persona encuestada estuvo ocupada entre enero del 2010 y junio del 2015, b) si la persona se mantuvo con seguridad social ampliada durante todo el

³ El material de este apartado se obtuvo del libro de Wooldridge (2002).

periodo analizado, c) si la persona se mantuvo fuera de la seguridad social ampliada durante todo el periodo analizado, d) si la persona experimentó una transición del trabajo asalariado al cuentapropismo y e) si la persona fue despedida de su primer trabajo.

En general, la estabilidad en el trabajo se considera una característica favorable tanto para el trabajador como para el empleador. Mientras más tiempo permanece una persona en su empleo, su productividad aumenta y, con ello, la probabilidad de que su remuneración presente una tendencia ascendente. El aprender haciendo (*learning by doing*) se considera un tipo de capital humano que se asocia con ingresos y productividad mayores (Solow, 1997; Arrow, 1971, entre otros), lo que contribuye al crecimiento económico de las naciones.

Por otro lado, la Organización Internacional del Trabajo (OIT) argumenta que un trabajo *decente* es aquel que para las personas significa la oportunidad de acceder a un empleo productivo, que genere un ingreso justo, la seguridad en el lugar de trabajo y la protección social para las familias, mejores perspectivas de desarrollo personal e integración social (OIT, 2021). En este sentido, el contar con seguridad social ampliada durante todo el periodo de empleo, el tener un trabajo formal con prestaciones de ley y el ser asalariado se consideran condiciones benéficas para los trabajadores, mientras que la transición hacia el cuentapropismo, el no contar con prestaciones de ley o el ser despedido se consideran situaciones perjudiciales.

Para conocer si el número de trabajos que desarrolló un individuo está correlacionado con el tipo de tareas que esta persona desarrolla, en primer lugar, se estima un modelo de respuesta multinomial ordenada. Para esto, se utiliza un modelo probit ordenado que parte de una variable dependiente y que representa una respuesta ordenada; en este caso, el número de trabajos que la persona tuvo en el periodo 2010-2015; y puede tomar cualquier valor entre 0 y $J \{0, 1, 2, \dots, J\}$, donde J es un número entero. El modelo probit ordenado para y (condicional en las variables explicativas x) puede ser derivado de un modelo de variable latente.

Los parámetros α y β de la regresión, es decir, los coeficientes asociados a cada una de las variables independientes, se pueden estimar por máxima verosimilitud. Para cada i , la función de log-likelihood (máxima verosimilitud) es:

$$\begin{aligned} \ell_i(\alpha, \beta) = & 1[y_i = 0] \log[\Theta(\alpha_1 - x_i\beta)] + 1[y_i = 1] \\ & \log[\Theta(\alpha_2 - x_i\beta)] - \Theta(\alpha_1 - x_i\beta) + \\ & + \dots + 1[y_i = J] \log[1 - \Theta(\alpha_J - x_i\beta)] \quad (1) \end{aligned}$$

Ahora, el resto de las variables dependientes (si la persona se mantuvo con seguridad social ampliada durante todo el periodo analizado, si estuvo fuera de esta durante todo el periodo analizado, si experimentó una transición del trabajo asalariado al cuentapropismo y si fue despedida de su primer trabajo) son de naturaleza binaria, por lo cual se estiman a través de modelos de respuesta binaria. En estos casos, también partimos de un modelo de variable latente, de tal manera que:

$$y^* = x\beta + e, \quad y = 1 [y^* > 0] \quad (2)$$

donde e es una variable distribuida de manera continua independiente de x , donde la distribución de e es simétrica en 0; por lo tanto, la probabilidad de respuesta se puede expresar como:

$$P(y = 1 | x) = P(y^* > 0 | x) = P(e > -x\beta | x) = 1 - G(-x\beta) = G(x\beta) \quad (3)$$

No hay una razón particular para requerir que e se distribuya de forma simétrica en el modelo de variable latente, pero esto es precisamente lo que sucede en el caso de los modelos de respuesta binaria que se aplican de manera más común.

4. Estadística descriptiva y resultados

4.1. Estadística descriptiva

El cuadro 2 incluye promedios y desviaciones estándar para cada una de las ocupaciones que se crearon para este trabajo: no rutinarias, cognitivas y

manuales, y rutinarias, cognitivas y manuales. Las variables por analizar son: sexo, edad, ingreso mensual y número de trabajos. La literatura revisada muestra que las trayectorias laborales de una persona se asocian tanto a sus características observables (sexo, edad, etnia, nivel educativo o región de residencia, entre otras variables) como a no observables (motivación en el trabajo, actitud hacia el riesgo, facilidad para establecer relaciones sociales, entre otras). Es por esto que se analizan características sociodemográficas y laborales de cada una de las ocupaciones.

Un primer hallazgo interesante es que las mujeres se concentran, sobre todo, en las ocupaciones que realizan tareas tanto rutinarias cognitivas como no rutinarias manuales, mientras que los hombres están sobrerrepresentados en las que hacen labores no rutinarias cognitivas y rutinarias manuales. La edad promedio de los trabajadores es muy pare-

cida en las cuatro categorías de ocupación, aunque los de mayor edad, y muy probablemente de más experiencia laboral, se encuentran en las ocupaciones no rutinarias cognitivas. El ingreso mensual más alto se devenga en las no rutinarias cognitivas, que son por lo general las de dirección y liderazgo, seguido de las rutinarias cognitivas y de las rutinarias manuales. Claramente, los ingresos más bajos se asocian a los trabajos que realizan tareas no rutinarias manuales, donde se requieren niveles mínimos de calificación. Por último, el número promedio de trabajos en el periodo 2010-2015 es ligeramente mayor en las no rutinarias manuales, lo que sugiere mayor rotación laboral para los trabajadores en este tipo de ocupaciones.

El cuadro 2 también desagrega las estadísticas descriptivas básicas según sexo. Llama la atención que los hombres de mayor edad están, sobre todo,

Cuadro 2

Estadísticas descriptivas, MOTRAL 2015

	% mujeres	Desv. est.	Edad promedio	Desv. est.	Ingreso mensual promedio	Desv. est.	Número promedio de trabajos	Desv. est.
Ocupaciones no rutinarias cognitivas	42.4	0.49	37.26	10.50	16 466.66	36 160.28	2.11	1.24
Ocupaciones rutinarias cognitivas	62.6	0.48	35.65	10.34	12 565.75	32 426.24	2.28	1.43
Ocupaciones rutinarias manuales	45.3	0.49	36.68	10.94	11 281.19	31 004.07	2.15	1.22
Ocupaciones no rutinarias manuales	53.1	0.50	36.66	11.06	10 238.71	29 789.92	2.35	1.55
Hombres								
Ocupaciones no rutinarias cognitivas	--	--	37.84	10.79	16 779.05	36 403.44	2.16	1.26
Ocupaciones rutinarias cognitivas	--	--	33.90	10.08	10 625.82	29 849.50	2.54	1.60
Ocupaciones rutinarias manuales	--	--	35.96	11.49	13 504.05	33 700.08	2.29	1.35
Ocupaciones no rutinarias manuales	--	--	36.16	11.26	10 393.41	29 893.86	2.40	1.69
Mujeres								
Ocupaciones no rutinarias cognitivas	--	--	36.47	10.03	16 040.81	35 821.77	2.05	1.21
Ocupaciones rutinarias cognitivas	--	--	36.69	10.35	13 727.37	33 822.28	2.12	1.29
Ocupaciones rutinarias manuales	--	--	37.56	10.17	8 267.94	27 134.27	1.97	1.01
Ocupaciones no rutinarias manuales	--	--	37.11	10.85	10 102.75	29 697.60	2.31	1.40

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

en las ocupaciones no rutinarias cognitivas, mientras que las mujeres de mayor edad se hallan en las rutinarias manuales. Un dato que vale la pena resaltar es que las mujeres más jóvenes se ubican en las ocupaciones que desarrollan tareas cognitivas, lo que se asocia con los mayores niveles educativos de la población más joven del país y con la situación familiar de las mujeres de mayor edad (posiblemente casadas y con hijos).

Nótese cómo, en el caso de los hombres, se devengan salarios más altos en las ocupaciones rutinarias manuales que en las rutinarias cognitivas. Esto es algo que llama la atención, pues las rutinarias manuales parecen estar dominadas por ellos, ya que hay una sobrerrepresentación de estos en esta categoría. Nótese cómo ellas reciben los sa-

larios más bajos precisamente en las ocupaciones que realizan tareas de naturaleza rutinaria manual (artesanos y obreros). Obsérvese, también, cómo el salario promedio de las mujeres es más alto en las de índole rutinaria cognitiva, donde además hay una sobrerrepresentación femenina. Estos trabajos son, sobre todo, de oficina. Este hecho sugiere que ellas son preferidas para ocupar este tipo de puestos, seguramente por la ventaja comparativa que tienen en el uso de las computadoras y en la realización de trabajos relacionados con aplicaciones digitales (Rodríguez y Meza, 2021).

El cuadro 3 presenta los niveles educativos por categoría de ocupación. Llama la atención que la gran mayoría de los que se insertan en las ocupaciones no rutinarias cognitivas cuentan con

Cuadro 3

Continúa

Niveles educativos por categoría de ocupación, MOTRAL 2015

Totales y porcentajes por nivel

Nivel educativo	Ocupaciones no rutinarias cognitivas		Ocupaciones rutinarias cognitivas		Ocupaciones rutinarias manuales		Ocupaciones no rutinarias manuales	
	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%
Sin instrucción	2 038	0.02	0	0.00	71 160	0.96	465 447	1.93
Primaria	339 448	3.13	145 746	2.11	1 897 961	25.58	5 103 863	21.16
Secundaria	1 914 863	17.65	2 148 041	31.05	3 637 288	49.01	9 059 574	37.57
Media superior	1 747 457	16.11	2 535 489	36.66	1 348 070	18.17	6 788 100	28.15
Superior	6 842 788	63.09	2 087 617	30.18	466 468	6.29	2 699 407	11.19
Total	10 846 594	100.00	6 916 893	100.00	7 420 947	100.00	24 116 391	100.00
Hombres								
Sin instrucción	0	0.00	0	0.00	49 158	1.21	134 628	1.19
Primaria	299 215	4.79	75 222	2.91	945 800	23.31	2 238 198	19.82
Secundaria	1 163 985	18.62	525 562	20.31	1 804 890	44.49	3 731 429	33.04
Media superior	1 249 968	20.00	1 018 642	39.36	975 478	24.04	3 661 612	32.43
Superior	3 536 594	56.59	968 770	37.43	281 619	6.94	1 526 465	13.52
Total	6 249 762	100.00	2 588 196	100.00	4 056 945	100.00	11 292 332	100.00

Niveles educativos por categoría de ocupación, MOTRAL 2015

Totales y porcentajes por nivel

Nivel educativo	Ocupaciones no rutinarias cognitivas		Ocupaciones rutinarias cognitivas		Ocupaciones rutinarias manuales		Ocupaciones no rutinarias manuales	
	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%
Mujeres								
Sin instrucción	2 038	0.04	0	0.00	22 002	0.65	330 819	2.58
Primaria	40 233	0.88	70 524	1.63	952 161	28.30	2 865 665	22.35
Secundaria	750 878	16.33	1 622 479	37.48	1 832 398	54.47	5 328 145	41.55
Media superior	497 489	10.82	1 516 847	35.04	372 592	11.08	3 126 488	24.38
Superior	3 306 194	71.92	1 118 847	25.85	184 849	5.49	1 172 942	9.15
Total	4 596 832	100.00	4 328 697	100.00	3 364 002	100.00	12 824 059	100.00

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

educación superior, sobre todo en el caso de las mujeres, mientras que la gran mayoría de los que realizan tareas de naturaleza rutinaria manual tienen secundaria. En cambio, los que hacen labores no rutinarias manuales pueden ostentar niveles educativos básicos o medios de manera casi indistinta. Cuando analizamos esta categoría diferenciada según sexo se observa que un porcentaje relativamente alto de hombres con educación superior (13.52) realiza este tipo de trabajos, que como vimos con anterioridad, son mayormente femeninos, lo cual, junto con los datos de los salarios, nos sugiere que los varones ocupan en esta categoría posiciones de liderazgo. Un dato que llama la atención es que, al parecer, los hombres no requieren necesariamente de educación superior para ocupar puestos de liderazgo y dirección, pues la proporción de estos que ostentan educación universitaria es significativamente menor que la que presentan las mujeres.

El cuadro 4 muestra los sectores donde se insertan los trabajadores definidos por el tipo de tareas que realizan. Primero, es notable cómo la gran mayoría de los empleados en ocupaciones de naturaleza no rutinaria cognitiva (70.23 %) encajan en los servicios, seguramente en las áreas de servi-

cios profesionales, mientras que gran parte de los que se insertan en las de índole rutinaria manual (85.85 %) se encuentran ubicados en el sector industrial. Los que desarrollan tareas de carácter no rutinario manual se ubican, sobre todo, en los servicios, con seguridad en las áreas de servicios no profesionales.

Cuando analizamos estas estadísticas según sexo observamos que más de 90 % de las mujeres que realizan tareas de naturaleza rutinaria manual se ubican en el sector industrial, donde habíamos visto que los hombres recibían un mejor trato en materia salarial.

Finalmente, el cuadro 5 presenta la estadística descriptiva de las trayectorias laborales por ocupación. Nótese cómo las ocupaciones que desarrollan tareas de índole cognitiva son las que, con mayor probabilidad, cuentan con seguridad social a lo largo de todo el periodo 2010-2015. Esto es en especial notorio en el caso de las mujeres que desarrollan tareas no rutinarias cognitivas. En cambio, las personas que hacen labores de naturaleza no rutinaria manual son las que en su mayoría carecen de seguridad social a lo largo de todo el periodo analizado. Esto también es especialmente notorio

Cuadro 4

Sector de empleo por categoría de ocupación, MOTRAL 2015

Totales y porcentajes por sector

	Ocupaciones rutinarias cognitivas		Ocupaciones no rutinarias cognitivas		Ocupaciones no rutinarias manuales		Ocupaciones no rutinarias manuales	
	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%	Valores absolutos	%
No especificado	61 353	0.57	57 111	0.83	25 052	0.34	101 933	0.42
Sector primario	21 871	0.20	5 609	0.08	11 995	0.16	651 872	2.70
Sector industrial	1 414 102	13.04	1 561 539	22.58	6 370 515	85.85	3 800 944	15.75
Sector comercio	895 654	8.26	2 125 672	30.73	245 316	3.31	7 183 647	29.78
Sector servicios	7 618 016	70.23	2 759 588	39.90	759 937	10.24	11 824 241	49.01
Gobierno	835 598	7.70	407 374	5.89	8 132	0.11	563 330	2.33
Total	10 846 594	100.00	6 916 893	100.00	7 420 947	100.00	24 125 967	100.00
Hombres								
No especificado	56 074	0.90	19 097	0.74	23 022	0.57	84 918	0.75
Sector primario	12 851	0.21	5 609	0.22	2 338	0.06	478 961	4.24
Sector industrial	1 043 868	16.70	861 464	33.28	3 216 096	79.27	3 130 747	27.70
Sector comercio	489 269	7.83	562 737	21.74	184 621	4.55	2 893 849	25.60
Sector servicios	4 130 223	66.09	1 011 272	39.07	622 736	15.35	4 294 767	38.00
Gobierno	517 477	8.28	128 017	4.95	8 132	0.20	418 666	3.70
Total	6 249 762	100.00	2 588 196	100.00	4 056 945	100.00	11 301 908	100.00
Mujeres								
No especificado	5 279	0.11	38 014	0.88	2 030	0.06	17 015	0.13
Sector primario	9 020	0.20	0	0.00	9 657	0.29	172 911	1.35
Sector industrial	370 234	8.05	700 075	16.17	3 154 419	93.77	670 197	5.23
Sector comercio	406 385	8.84	1 562 935	36.11	60 695	1.80	4 289 798	33.45
Sector servicios	3 487 793	75.87	1 748 316	40.39	137 201	4.08	7 529 474	58.71
Gobierno	318 121	6.92	279 357	6.45	0	0.00	144 664	1.13
Total	4 596 832	100.00	4 328 697	100.00	3 364 002	100.00	12 824 059	100.00

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

en el caso de ellas. Los que presentan una mayor probabilidad de experimentar una transición del trabajo asalariado al cuentapropismo son precisamente los que realizan tareas de índole no rutinaria manual. Por último, este cuadro muestra que los hombres que presentan la mayor probabilidad de ser despedidos de su primer trabajo son los que desarrollan tareas no rutinarias cognitivas, y las mujeres con mayor probabilidad de ser despedidas son las que hacen labores de carácter rutinario manual.

Llama la atención que las mujeres, en la categoría de trabajo no rutinario cognitivo, presentan un porcentaje especialmente bajo de despido en el primer empleo, mientras que los hombres, un bajo de despido del primer empleo en la categoría de trabajo no rutinario manual.

A continuación, se presenta el trabajo econométrico para entender las trayectorias laborales de las diferentes categorías ocupacionales que estamos analizando.

Cuadro 5

Estadísticas de trayectorias laborales

	% con acceso a seguridad social durante toda la trayectoria laboral	% sin acceso a seguridad social durante toda la trayectoria laboral	% que experimentó transición de empleado a cuentapropia	% que fue despedido de su primer trabajo
Ocupaciones no rutinarias cognitivas	48.47	26.68	9.92	5.61
Ocupaciones rutinarias cognitivas	49.37	19.01	5.77	5.68
Ocupaciones rutinarias manuales	38.07	33.30	11.81	6.92
Ocupaciones no rutinarias manuales	23.19	46.98	13.57	4.27
Hombres				
Ocupaciones no rutinarias cognitivas	45.93	26.72	12.06	8.49
Ocupaciones rutinarias cognitivas	49.21	24.92	5.03	6.93
Ocupaciones rutinarias manuales	37.18	31.51	11.67	5.89
Ocupaciones no rutinarias manuales	25.79	41.59	13.23	3.85
Mujeres				
Ocupaciones no rutinarias cognitivas	51.92	26.63	7.01	1.68
Ocupaciones rutinarias cognitivas	49.47	15.48	6.21	4.94
Ocupaciones rutinarias manuales	39.14	35.46	11.97	8.16
Ocupaciones no rutinarias manuales	20.91	51.73	13.87	4.65

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

4.2. Resultados

El cuadro 6 presenta los resultados del modelo probit ordenado para la variable dependiente: número de trabajos. Dado que en las ocupaciones analizadas el número de trabajos realizados en el periodo 2010-2015 es muy cercano a 1, se muestran los resultados para la probabilidad de tener un solo trabajo en el lapso analizado. Las variables explicativas incluyen, además de las categorías de ocupación que nos interesan, la edad de los trabajadores, su nivel educativo, el sexo, la industria o sector en el que se insertan y las regiones en las que se emplean, definidas estas por el grado de marginación que cada región presenta.⁴

El cuadro 6 incluye tres especificaciones distintas del modelo probit ordenado. En el primero, las variables explicativas son únicamente las cuatro

ocupaciones definidas por el tipo de tareas que los trabajadores desarrollan. La segunda contiene, además de las ocupaciones, las variables sociodemográficas de edad, sexo y nivel educativo. Nótese que también se comprende la edad al cuadrado para captar los rendimientos laborales decrecientes con la edad. En la tercera se añaden, además de las ocupaciones y las variables sociodemográficas, controles por industria y región. El modelo 3 es el que se considera más robusto. La categoría de ocupación no rutinaria manual es la omitida.

En primer lugar, la última columna del cuadro 6 muestra que los trabajadores que realizan tareas no rutinarias manuales (la categoría omitida) son los que tienen la menor probabilidad de tener un solo trabajo durante la trayectoria laboral del periodo 2010-2015. Esto implica que, una vez que controlamos por algunas características observables de estos, los que presentan mayor rotación laboral son los que hacen labores no rutinarias manuales. Los que presentan la mayor probabili-

⁴ Los grados de marginación son calculados por el Consejo Nacional de Población (CONAPO) por estado y municipio a través de datos censales e intercensales.

Cuadro 6

Continúa

Regresión de número de trabajos. Probit ordenado. Probabilidad de tener un solo trabajo. Efectos marginales

Variables independientes	Modelo 1	Modelo 2	Modelo 3
Ocupación no rutinaria cognitiva	-0.0594 *** (-397.75)	0.0505 *** (315.62)	0.0503 *** (309.15)
Ocupación rutinaria cognitiva	-0.0200 *** (-112.79)	0.0288 *** (167.96)	0.0256 *** (150.01)
Ocupación rutinaria manual	-0.0446 *** (-259.72)	0.0425 *** (262.11)	0.0670 *** (350.75)
Edad	--	0.0203 *** (485.86)	0.0200 *** (481.80)
Edad al cuadrado	--	-0.0001 *** (-182.91)	-0.0001 *** (-185.72)
Educación primaria	--	0.1365 *** (243.77)	0.1259 *** (224.25)
Educación secundaria	--	0.1876 *** (339.43)	0.1721 *** (309.89)
Educación media superior	--	0.1371 *** (245.36)	0.1142 *** (203.20)

Regresión de número de trabajos. Probit ordenado. Probabilidad de tener un solo trabajo. Efectos marginales

Variabes independientes	Modelo 1	Modelo 2	Modelo 3
Educación superior	--	0.1600 *** (283.67)	0.1440 *** (254.26)
Mujer	--	0.0254 *** (230.49)	0.0262 *** (233.46)
Efectos fijos de industrias	no	no	sí
Efectos fijos de grados de marginación estatal	no	no	sí
Número de observaciones	48 016 157	48 016 157	47 961 294
Pseudo R ²	0.001	0.045	0.050
Prob > chi ²	0.000	0.000	0.000

*** Estadísticamente significativo a 99 % de confianza.

Estadísticas z entre paréntesis.

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

dad de mantenerse en el trabajo inicial son los que realizan tareas rutinarias manuales, seguidos por los que desarrollan tareas no rutinarias cognitivas. En este caso, los empleados no rutinarios manuales son los que tendrían una trayectoria laboral más precaria y vulnerable. Nótese cómo los trabajadores con educación superior y las mujeres son quienes presentan mayor probabilidad de mantenerse en su trabajo inicial, cuando se les compara con los de otros niveles educativos y con los hombres. Esto implica que los trabajadores rutinarios manuales, los de mayor educación y las mujeres son los que presentan menor rotación e inestabilidad laboral. En esta regresión, las variables de edad y edad al cuadrado muestran los signos esperados, lo cual implica que los de mayor edad presentan una trayectoria laboral más estable, aunque la probabilidad de mantenerse en un mismo trabajo decrece con la edad.

La bondad de ajuste del modelo es bastante baja, lo cual implica que se están dejando fuera variables relevantes en la definición del número de empleos de un trabajador en un periodo determinado. Es probable que estas variables tengan que ver con las características no observables o con las

de las empresas contratantes. Sin embargo, esta información no está contenida en la Encuesta. El bajo ajuste de las regresiones se puede considerar una limitación de esta investigación, aunque este resultado se atribuye a la falta de información de las empresas en el MOTRAL.

El cuadro 7 presenta los resultados del modelo probit para la variable dependiente: acceso permanente a seguridad social ampliada. Esta toma el valor de 1 si el trabajador reportó haber sido derechohabiente de la seguridad social ampliada durante todo el periodo de análisis y el de 0 si experimentó algún lapso sin acceso a seguridad social. Como en el caso anterior, el cuadro muestra tres especificaciones distintas para el modelo, donde la más robusta es aquella que incluye como variables explicativas no solo las ocupaciones que nos interesan y las características sociodemográficas, sino también efectos fijos por industria y región.

La columna 3 del cuadro 7 revela que los trabajadores con mayor probabilidad de mantenerse con seguridad social ampliada durante toda su trayectoria laboral son aquellos que realizan tareas de índole rutinaria manual. En este caso, trabajadores

Cuadro 7

Regresión acceso total a seguridad social. Modelo probit. Efectos marginales

Variables independientes	Modelo 1	Modelo 2	Modelo 3
Ocupación no rutinaria cognitiva	0.2431 *** (1568.76)	0.1672 *** (920.19)	0.1630 *** (878.88)
Ocupación rutinaria cognitiva	0.2510 *** (1364.01)	0.2191 *** (1153.80)	0.1987 *** (1042.26)
Ocupación rutinaria manual	0.1502 *** (799.28)	0.1495 *** (801.78)	0.0943 *** (424.93)
Edad	--	0.0360 *** (714.12)	0.0342 *** (684.87)
Edad al cuadrado	--	-0.0004 *** (-622.21)	-0.0004 *** (-599.34)
Educación primaria	--	0.2878 *** (306.73)	0.2568 *** (276.13)
Educación secundaria	--	0.3481 *** (373.98)	0.3162 *** (342.28)
Educación media superior	--	0.3269 *** (348.93)	0.2998 *** (322.24)
Educación superior	--	0.4370 *** (465.83)	0.4169 *** (447.91)
Mujer	--	-0.0170 *** (-129.64)	-0.0035 *** (-26.81)
Efectos fijos de industrias	no	no	sí
Efectos fijos de grados de marginación estatal	no	no	sí
Número de observaciones (ponderadas)	49 310 401	49 310 401	49 255 538
Pseudo R ²	0.047	0.073	0.089
Prob > chi ²	0.000	0.000	0.000

*** Estadísticamente significativo a 99 % de confianza.

Estadísticas z; entre paréntesis.

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

como los artesanos o los obreros, que mayormente se desarrollan en el sector industrial, que cuentan con educación secundaria y donde los de sexo masculino están sobrerrepresentados, son los que disfrutan de condiciones laborales más aptas. En segundo lugar, se encuentran los trabajadores no rutinarios cognitivos, donde también ellos están sobrerrepresentados. Esto indica que las mujeres tienden a desempeñarse en empleos de menor ca-

lidad, donde la permanencia en la seguridad social ampliada no está asegurada. De hecho, aquellos que realizan trabajo de índole no rutinaria manual son los que, de nuevo, aparecen con las condiciones laborales menos adecuadas, seguidos de los trabajadores que desarrollan tareas rutinarias cognitivas. Por lo que respecta a otras variables, este cuadro muestra que los que cuentan con educación superior y los varones son los que ostentan

las mayores probabilidades de desempeñarse con seguridad social ampliada a lo largo de toda su trayectoria laboral. La edad y la edad al cuadrado muestran los signos esperados, lo cual implica que los trabajadores de más edad tienen mayor probabilidad de contar con seguridad social ampliada durante todo su periodo laboral,⁵ aunque esta probabilidad crece a una tasa decreciente.

Nuevamente, la bondad de ajuste es baja, lo cual implica que se están dejando fuera del modelo variables importantes para explicar el acceso permanente a la seguridad social ampliada. Es probable, como en el caso anterior, que estas sean, sobre todo, no observables o referentes a las empresas, las cuales, por desgracia, no se incluyen en la base de datos. Esto, de nuevo, se considera una limitación del estudio.

5 El periodo laboral analizado es de cinco años y va del 2010 al 2015. No se refiere a toda la vida laboral del trabajador.

El cuadro 8 presenta los resultados del modelo probit para la variable dependiente: sin acceso a seguridad social ampliada. Esta toma el valor de 1 si el trabajador reportó no haber sido derechohabiente de la seguridad social ampliada durante todo el periodo de análisis y el de 0 si experimentó algún lapso con acceso a seguridad social. Como en el caso anterior, el cuadro muestra tres especificaciones distintas para el modelo, donde la más robusta es aquella que incluye como variables explicativas no solo las ocupaciones que nos interesan y las características sociodemográficas, sino también efectos fijos por industria y región.

La columna 3 del cuadro 8 sugiere que los trabajadores con mayor probabilidad de mantenerse sin seguridad social ampliada durante toda su trayectoria laboral son aquellos que realizan tareas de índole no rutinaria manual. En este caso, trabajadores como los cuidadores o los de limpieza en viviendas y establecimientos, que mayormente se desarrollan en el sector de servicios no profesiona-

Cuadro 8

Continúa

Regresión sin acceso a seguridad social. Modelo probit. Efectos marginales

VARIABLES INDEPENDIENTES	Modelo 1	Modelo 2	Modelo 3
Ocupación no rutinaria cognitiva	-0.1967 *** (-1181.61)	-0.1173 *** (-607.36)	-0.1222 *** (-619.72)
Ocupación rutinaria cognitiva	-0.2884 *** (-1394.35)	-0.2405 *** (-1135.65)	-0.2246 *** (-1055.72)
Ocupación rutinaria manual	-0.1280 *** (-668.11)	-0.1348 *** (-711.10)	-0.0766 *** (-336.23)
Edad	--	0.0029 *** (59.00)	0.0041 *** (82.66)
Edad al cuadrado	--	0.000 *** (26.23)	0.0000 *** (7.50)
Educación primaria	--	-0.1570 *** (-326.66)	-0.1320 *** (-202.75)
Educación secundaria	--	-0.2117 *** (-326.66)	-0.1853 *** (-287.45)
Educación media superior	--	-0.2523 *** (-385.11)	-0.2260 *** (-346.27)

Cuadro 8

Concluye

Regresión sin acceso a seguridad social. Modelo probit. Efectos marginales

Variables independientes	Modelo 1	Modelo 2	Modelo 3
Educación superior	--	-0.3406 *** (-515.04)	-0.3229 *** (-490.91)
Mujer	--	0.0326 *** (244.98)	0.0199 *** (147.60)
Efectos fijos de industrias	no	no	sí
Efectos fijos de grados de marginación estatal	no	no	sí
Número de observaciones	49 310 401	49 310 401	49 255 538
Pseudo R ²	0.040	0.065	0.077
Prob > chi ²	0.000	0.000	0.000

*** Estadísticamente significativo a 99 % de confianza.

Estadísticas z; entre paréntesis.

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

les, donde las trabajadoras de sexo femenino están sobrerrepresentadas, son los que disfrutan de condiciones laborales más precarias. En segundo lugar, se encuentran los rutinarios manuales, donde los hombres están sobrerrepresentados. Los que presentan la menor probabilidad de carecer de acceso a seguridad social durante toda la trayectoria laboral son los trabajadores rutinarios cognitivos, seguidos de los no rutinarios cognitivos. Esto sugiere, de nuevo, que el trabajo mental se premia en el mercado laboral con condiciones más aptas y de mejor calidad. Los resultados de este ejercicio muestran, nuevamente, que las mujeres tienden a desempeñarse en empleos de menor calidad, donde la permanencia en la seguridad social ampliada no está asegurada. De hecho, quienes realizan trabajo de índole no rutinario manual son los que aparecen, una vez más, con las condiciones laborales menos adecuadas, seguidos de los trabajadores que hacen tareas rutinarias manuales. Por lo que respecta a otras variables, este cuadro muestra que los que no cuentan con educación y las mujeres son los que ostentan las mayores probabilidades de desempeñarse sin seguridad social ampliada a lo largo de toda su trayectoria laboral. La edad y la edad al cuadrado muestran signo positivo, lo cual implica que la falta de seguridad social crece con la edad a una tasa creciente.

Una vez más, la bondad de ajuste es baja, lo cual implica que se están dejando fuera del modelo variables importantes para explicar la falta de acceso permanente a la seguridad social ampliada. Es probable, como en el caso anterior, que estas sean, sobre todo, no observables o referentes a las empresas, las cuales, desafortunadamente, no se incluyen en la base de datos. Esto se vuelve a considerar una limitante del estudio.

El cuadro 9 presenta los resultados del modelo probit para la variable dependiente: transición del trabajo asalariado al cuentapropismo. Esta toma el valor de 1 si el trabajador reportó haber iniciado el periodo de análisis en una situación de trabajo asalariado y terminado el mismo lapso como cuentapropista, y adquiere el de 0 si no experimentó este tipo de transición laboral. Como en los casos anteriores, el cuadro muestra tres especificaciones distintas para el modelo, donde la más robusta es aquella que incluye como variables explicativas no solo las ocupaciones que nos interesan y las características sociodemográficas, sino también efectos fijos por industria y región.

La columna 3 del cuadro 9 sugiere que los trabajadores con mayor probabilidad de transitar del trabajo asalariado al cuentapropismo —es decir,

de pasar de una situación laboral con ingreso seguro y, en algunos casos con prestaciones, a otra con ingreso fluctuante y sin prestaciones— son de nuevo aquellos que realizan tareas de índole no rutinario manual. En este caso, como en el anterior, trabajadores como los cuidadores, los que preparan alimentos o los de limpieza en viviendas y establecimientos, que mayormente se desarrollan en el sector de servicios no profesionales, donde las

trabajadoras de sexo femenino están sobrerrepresentadas, son los que experimentan condiciones laborales más precarias. En segundo lugar, se encuentran los rutinarios manuales, donde los hombres están sobrerrepresentados. Los que presentan la menor probabilidad de transitar del trabajo asalariado al cuentapropismo son los que desarrollan tareas de índole rutinaria cognitiva (oficinistas), seguidos de los no rutinarios cognitivos.

Cuadro 9

Regresión de transición de empleado a cuentapropista. Modelo probit. Efectos marginales

Variables independientes	Modelo 1	Modelo 2	Modelo 3
Ocupación no rutinaria cognitiva	-0.0355 *** (-304.88)	-0.0509 *** (-375.74)	-0.0372 *** (-268.52)
Ocupación rutinaria cognitiva	-0.0906 *** (-569.54)	-0.0921 *** (-576.07)	-0.0895 *** (-556.28)
Ocupación rutinaria manual	-0.0161 *** (-124.79)	-0.0124 *** (-97.08)	-0.0191 *** (-128.40)
Edad	--	0.0090 *** (248.71)	0.0094 *** (261.25)
Edad al cuadrado	--	-0.0000 *** (-95.22)	-0.0000 *** (-102.90)
Educación primaria	--	0.0738 *** (165.90)	0.0732 *** (164.71)
Educación secundaria	--	0.0739 *** (167.54)	0.0731 *** (165.29)
Educación media superior	--	0.1111 *** (249.25)	0.1110 *** (248.15)
Educación superior	--	0.0984 *** (217.80)	0.0977 *** (215.97)
Mujer	--	-0.0052 *** (-58.79)	-0.0079 *** (-87.14)
Efectos fijos de industrias	no	no	sí
Efectos fijos de grados de marginación estatal	no	no	sí
Número de observaciones	49 310 401	49 310 401	49 255 538
Pseudo R ²	0.011	0.061	0.071
Prob > chi ²	0.000	0.000	0.000

*** Estadísticamente significativo a 99% de confianza.

Estadísticas z entre paréntesis.

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015.

Esto sugiere, una vez más, que el trabajo mental se premia en el mercado laboral con condiciones más aptas y de mejor calidad. Los resultados de este ejercicio muestran, nuevamente, que las mujeres tienden a desempeñarse en empleos de menor calidad, donde la permanencia en un trabajo asalariado no está asegurada. De hecho, los trabajadores que realizan trabajo de índole no rutinario manual son los que aparecen, una vez más, con las condiciones laborales menos adecuadas, seguidos de los que hacen tareas rutinarias manuales. Por lo que respecta a otras variables, este cuadro muestra que los que cuentan con educación media superior y los hombres son los que presentan las mayores probabilidades de transitar del trabajo asalariado al cuentapropismo. La edad y la edad al cuadrado muestran que la probabilidad de transitar del trabajo asalariado al cuentapropismo aumenta con la edad, pero a una tasa decreciente. Este resultado es consistente con la literatura que argumenta que las personas, con la edad, tienden a moverse al trabajo independiente.

Nuevamente, la bondad de ajuste es baja, lo cual implica que se están dejando fuera del modelo variables importantes para explicar la transición del trabajo asalariado al cuentapropismo. Como en los casos anteriores, es probable que estas sean, sobre todo, no observables o referentes a las empresas. Esto se vuelve a considerar una limitación del análisis.

Por último, el cuadro 10 presenta los resultados del modelo probit para la variable dependiente: despido del primer trabajo. Esta toma el valor de 1 si el trabajador reportó haber sido despedido de su primer empleo y el de 0 si no experimentó esta situación. Como en los casos anteriores, el cuadro muestra tres especificaciones distintas para el modelo, donde la más robusta es aquella que incluye como variables explicativas no solo las ocupaciones que nos interesan y las características sociodemográficas, sino también efectos fijos por industria y región.

La columna 3 del cuadro 10 sugiere que los trabajadores con mayor probabilidad de ser despedidos de su primer trabajo, una vez que controlamos

por diversas características sociodemográficas y laborales de estos, son aquellos que desarrollan trabajo de naturaleza rutinaria manual; es decir, artesanos y obreros. Cabe recordar que, en este tipo de empleos, los varones están sobrerrepresentados.

En segundo lugar, se encuentran los trabajadores rutinarios manuales, donde los hombres están sobrerrepresentados. Los que presentan la menor probabilidad de ser despedidos de su primer empleo son los que desarrollan tareas no rutinarias manuales, lo que, de alguna manera, contradice los hallazgos anteriores. Es de llamar la atención que los trabajadores que menor probabilidad tienen de ser despedidos de su primer trabajo son los que realizan tareas de índole no rutinaria. Esto sugiere, de alguna manera, que los empleados no rutinarios experimentan mejores condiciones laborales que sus contrapartes que hacen tareas de naturaleza rutinaria. Respecto de las mujeres, se tiene que ellas presentan menor probabilidad de ser despedidas de su primer trabajo cuando se les compara con sus contrapartes masculinas. En cuanto a los niveles de educación, los resultados muestran que los trabajadores sin educación son los que presentan la probabilidad mayor de ser despedidos de su primer empleo. La variable edad muestra un signo positivo, lo cual indica que la práctica del despido del primer empleo era más común en el pasado.

Nuevamente, la bondad de ajuste es baja, lo cual implica que se están dejando fuera del modelo variables importantes para explicar el despido del primer empleo. Esto vuelve a considerarse una limitante del análisis.

5. Conclusiones

Trabajos como los de Acemoglu (2002), Autor *et al.* (2003) y Autor (2010) argumentan que el cambio tecnológico en la actualidad está sesgado por tareas, más que por calificaciones; es decir, ellos sugieren que las computadoras, los robots y las aplicaciones digitales están sustituyendo trabajadores que llevan a cabo trabajos de naturaleza rutinaria, mientras que estos complementan a los

Cuadro 10

Regresión despido del primer trabajo. Modelo probit. Efectos marginales

VARIABLES INDEPENDIENTES	Modelo 1	Modelo 2	Modelo 3
Ocupación no rutinaria cognitiva	0.0138 *** (170.47)	0.0145 *** (153.08)	0.0147 *** (152.24)
Ocupación rutinaria cognitiva	0.0145 *** (153.93)	0.0199 *** (203.31)	0.0212 *** (214.78)
Ocupación rutinaria manual	0.0250 *** (283.69)	0.0253 *** (285.11)	0.0267 *** (252.34)
Edad	--	0.0031 *** (124.16)	0.0033 *** (131.40)
Edad al cuadrado	--	-0.0000 *** (-79.89)	-0.0000 *** (-85.64)
Educación primaria	--	-0.0367 *** (-145.62)	-0.0408 *** (-162.99)
Educación secundaria	--	-0.0376 *** (-151.83)	-0.0419 *** (-169.95)
Educación media superior	--	-0.0350 *** (-138.54)	-0.0394 *** (-155.83)
Educación superior	--	-0.0421 *** (-163.54)	-0.0467 *** (-182.04)
Mujer	--	-0.0114 *** (-177.40)	-0.0127 *** (-192.59)
Efectos fijos de industrias	no	no	sí
Efectos fijos de grados de marginación estatal	no	no	sí
Número de observaciones	49 310 401	49 310 401	49 255 538
Pseudo R ²	0.004	0.015	0.023
Prob > chi ²	0.000	0.000	0.000

*** Estadísticamente significativo a 99 % de confianza.

Estadísticas z entre paréntesis.

Fuente: estimaciones propias con datos del MOTRAL de la ENOE 2015

que desarrollan tareas de índole no rutinaria, ya sea cognitiva o manual.

En el caso de México, Meza (2019) y Rodríguez y Meza (2021) han argumentado que el mercado laboral femenino muestra una polarización similar a la que caracteriza a los mercados laborales de países desarrollados como Estados Unidos de América, Reino Unido y Alemania, entre otros, mientras que el masculino se caracteriza por un aumento en

la demanda por los trabajadores que realizan tareas de naturaleza rutinaria cognitiva.

El presente estudio trató de mostrar algunas características de las trayectorias laborales de los trabajadores definidos por el tipo de tareas que desarrollan; a saber, no rutinarias, cognitivas y manuales, y rutinarias, cognitivas y manuales. Si de verdad el cambio tecnológico que está permeando en México está sesgado por tareas,

entonces esperaríamos encontrar que las condiciones laborales de los trabajadores que realizan tareas de índole no rutinaria son significativamente mejores que las de aquellos que hacen labores de carácter rutinario.

Las características de las trayectorias laborales que se analizan en esta investigación son, para el periodo 2010-2015: el número total de trabajos, el acceso permanente a la seguridad social, la falta de esta durante todo el lapso analizado, la transición del trabajo asalariado al cuentapropismo y el despido en el primer empleo. Estas trayectorias se analizan en función de varias características sociodemográficas de los individuos, como su edad, sexo y nivel de escolaridad, pero de especial importancia son el tipo de tareas que desarrollan en su ocupación. Las estimaciones controlan por el sector donde el trabajador se inserta y por la región donde reside.

Los resultados son consistentes y muestran que las condiciones de trabajo más precarias son las que experimentan los trabajadores que desarrollan tareas de índole no rutinaria manual, mientras que los que presentan las condiciones laborales más adecuadas son los que hacen trabajo de naturaleza cognitiva, ya sea rutinaria o no rutinaria. El único resultado que no confirma esta hipótesis es la del despido del primer empleo, pues muestra que los trabajadores en ocupaciones no rutinarias manuales son los que presentan la menor probabilidad de ser despedidos de su primer trabajo.

Los resultados que ofrece el estudio son consistentes con un avance tecnológico sesgado por calificación, que es el que prevaleció en las economías más desarrolladas a finales del siglo XX, el cual se asocia con la generalización del uso de las computadoras en el trabajo. Este resultado es relevante, pues contradice de alguna manera los hallazgos anteriores respecto de la polarización del mercado laboral femenino. De hecho, muestra que las mujeres están sobrerrepresentadas precisamente en los trabajos que presentan las trayectorias laborales más precarias y con mayores niveles de vulnerabilidad.

Aunque las estimaciones econométricas según sexo nos darían mayores elementos para saber si, efectivamente, el mercado laboral femenino se está polarizando o no, los resultados de esta investigación muestran, sobre todo, la precariedad en la que desarrollan los trabajadores que hacen tareas de naturaleza no rutinaria manual, entre los que están los cuidadores, los encargados de la limpieza o los de la preparación de alimentos.

El estudio tiene varias limitaciones que es importante reconocer. En primer lugar, la base de datos utilizada hace referencia a las trayectorias laborales en el periodo 2010-2015, y los años más recientes son, seguramente, los que han presenciado un avance más decidido del cambio tecnológico sesgado por tareas, sobre todo por el avance de la digitalización. Es probable que una base de datos con información más reciente nos mostraría un panorama distinto al que ha sido discutido en esta investigación. Asimismo, la bondad de ajuste de los modelos es muy baja, lo que se atribuye a la falta de información sobre los propios empleados y también de las empresas, negocios y establecimientos donde se emplean estos trabajadores. Finalmente, para probar discriminación laboral hacia algún grupo poblacional, se requerirían métodos econométricos más sofisticados, como descomposiciones tipo Blinder-Oaxaca. Aun así, creemos que la evidencia presentada sugiere que el cambio tecnológico sesgado por tareas no está avanzando de manera contundente en México, sino que el relevante es el sesgado por calificación, el cual implica condiciones laborales más favorables para los trabajadores con mayores niveles de capital humano.

Dado este contexto, la primera recomendación de política laboral para mejorar las condiciones de trabajo en México sería seguir avanzando en los niveles de escolaridad y, sobre todo, trabajar en favor de la mejora de la calidad de la educación, de tal manera que los trabajadores sean capaces de tomar decisiones complejas en cualquier entorno de trabajo. Otra recomendación que se desprende del estudio es la de promover mayor igualdad laboral entre mujeres y hombres. Claramente, ellas se insertan en condiciones de des-

ventaja respecto de sus contrapartes masculinas, y políticas de acción afirmativa podrían evitar esta desigualdad. Por último, nos parece que el cuentapropismo no debería asociarse con condiciones laborales más precarias, sino con ingresos mayores por el riesgo que corren los emprendedores, así como con acceso a la seguridad social a través de esquemas públicos y privados ofrecidos a precios accesibles.

Fuentes

- Acemoglu, D. "Technical Change, Inequality, and the Labor Market", en: *Journal of Economic Literature*. Vol. 40, No.1, 2002, pp. 7-72.
- Acemoglu, D., and D. Autor. "Skills, tasks and technologies: implications for employment and earnings", en: Ashenfelter O. and D. Card (eds.). *Handbook of Labor Economics*. 4B. Elsevier, 2011, pp. 1043-1771.
- Arrow, K. J. *The economic implications of learning by doing*. Readings in the theory of growth. Springer, 1971.
- Autor, David, Frank Levy and Richard Murnane. "The skill content of recent technological change: an empirical exploration", en: *Quarterly Journal of Economics*. 116 (4), 2003, pp. 1279-1333.
- Autor, David. *The Polarization of Job Opportunities in the U.S. Labor Market. Implications for Employment and Earnings*. Center for American Progress, The Hamilton Project, April, 2010, pp. 1-40.
- Blanco, O. y D. Julián. "Una tipología de precariedad laboral para Chile: la precariedad como fenómeno transclasista", en: *Revista de la CEPAL*. Núm. 129, 2019, pp. 99-137.
- Blume, K., M. Ejrnaes, H. Skyt Nielsen and A. Würtz. *Self-employment among immigrants: a last resort?* CAM Working papers 2003-08. University of Copenhagen, Department of Economics, Centre for Applied Microeconometrics, 2004.
- Calderón, A. *Effects of Contributory Pension System Reforms on Segmented Labor Markets: The Mexican Experience on Worker's Transitions out of the Formal Sector*. Mimeo. El Colegio de México, 2018.
- Damaske, S. y A. Frech. "Women's work pathways across the life course", en: *Demography*. 53, 2016, pp. 365-391.
- Guadarrama, R., A. Hualde y S. López. "Precariedad laboral y heterogeneidad ocupacional: una propuesta metodológica", en: *Revista Mexicana de Sociología*. 74(2), 2012, pp. 213-243.
- Halperc, A., J. R. Warren, J. M. Raymo y D. A. Nicholson. "The impact of work and family life histories on economic well-being at older ages", en: *Social Forces*. 93(4), 2015, pp. 1369-1396.
- Lu, Y., J. Shu-Huah Wang y W. Han. "Women's short-term employment trajectories following birth: patterns, determinants and variations by race/ethnicity and nativity", en: *Demography*. 54, 2017, pp. 93-118.
- Maloney, W. F. "Does informality imply segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico", en: *The World Bank Economic Review*. Vol. 13, No. 2, 1999, pp. 275-302.
- Meza, L. "El cambio tecnológico y el mercado de trabajo en México", en: Castro Lugo, David y Reyna Elizabeth Rodríguez Pérez (eds.). *El mercado de trabajo en México: tendencias en el siglo XXI*. Editorial Fontamara y Universidad Autónoma de Coahuila, 2019, pp.19-50.
- Moen, P. "The gendered life course", en: George, L. and R. Binstock (eds.). *Handbook of ageing and the social sciences*. San Diego, California, Academic Press, 2001, pp. 179-196.
- Organización Internacional del Trabajo (OIT). *Trabajo decente*. 2021 (DE) <https://oit.org/global/topics/decent-work/lang—es/index.htm>
- Parra-Sandoval, M. C., M. F. Orozco y R. Grediaga-Kuri. "Labor trajectories and studies abroad", en: *Sociology Insights*. Vol 3(1), 2019, pp. 1-14.
- Pérez, J. y G. Ceballos. "Dimensionando la precariedad laboral en México de 2005 a 2015, a través del modelo logístico ordinal generalizado", en: *Nósis. Revista de Ciencias Sociales y Humanidades*. 28(55), 2019, pp. 109-135.
- Rodríguez, R. E. y L. Meza. "Employment and earnings by gender in Mexico: its relationship with task biased technological change", en: *Estudios Económicos*. Julio-diciembre, 2021.
- Román, Y. "Impactos sociodemográficos y económicos en la precariedad laboral de los jóvenes en México", en: *Región y Sociedad*. XXV(58), 2013, pp. 165-202.
- Rubio, J. "Precariedad laboral en México. Una propuesta de medición integral", en: *Revista Enfoques*. VIII(13), 2010, pp. 77-87.
- Solow, R. M. *Learning form "learning by doing": Lessons for Economic Growth*. Stanford University Press, 1997.
- Wassink, J. T. y J. M. Hagan. "A Dynamic Model of Self-Employment and Socio-economic Mobility among Return Migrants: The case of urban Mexico", en: *Social Forces*. 96(3), March, 2018, pp. 1069-1096.
- Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge Massachusetts and London, England, The MIT Press, 2002.
- Yamaguchi, S. "Changes in returns to task-specific skills and gender wage gap", en: *The Journal of Human Resources*. 53(1), 2018, pp. 32-70.

Seguimiento de la distribución del ingreso en México a lo largo del tiempo y de la geografía

Monitoring the Income Distribution in Mexico over Time and Geography

Víctor Alfredo Bustos y de la Tijera y Ana Miriam Romo Anaya*

Este trabajo discute la evolución de la distribución del ingreso en México para el 2010, 2012, 2014 y 2016. Adicionalmente, se hace una desagregación geográfica al nivel estatal para los mismos años. El análisis se basa, principalmente, en mediciones de desigualdad derivadas de estimaciones para dichas distribuciones y su comparación tanto entre sí como con mediciones semejantes derivadas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) o su Módulo de Condiciones Socioeconómicas del Instituto Nacional de Estadística y Geografía. Las estimaciones son obtenidas aplicando el criterio de máxima pseudo-verosimilitud restringida (MPVR) en un afán de hacer uso simultáneo de varias fuentes de datos para mejorar la estimación de la distribución del ingreso. Nos proponemos establecer nuevas comparaciones con los resultados muestrales para ratificar o rectificar el comportamiento de diversas tendencias aparentes durante el periodo. Mostraremos que los ajustes realizados dan lugar a una tendencia creciente en el Gini a nivel na-

This paper discusses the evolution of income distribution in Mexico for 2010, 2012, 2014 and 2016. Additionally, a geographic disaggregation is made at the state level for the same years. The analysis is mainly based on inequality measurements derived from estimates for these distributions and their comparison both with each other and with similar measurements derived from the National Household Income and Expenditure Survey (ENIGH) or its Socioeconomic Conditions Module of the National Institute of Statistics and Geography (INEGI). The estimates are obtained by applying the restricted maximum pseudo-likelihood criterion (MPVR) in an effort to make simultaneous use of several data sources to improve the estimation of income distribution. We propose to establish new comparisons with the sample results to ratify or rectify the behavior of various apparent trends during the period. We will show that the adjustments made give rise to an increasing trend in the Gini at the national level, which contrasts with the decrease in the value of the same indicator for the ENIGH. In other words, our results

* Instituto Nacional de Estadística y Geografía (INEGI), alfredo.bustos@inegi.org.mx y miriam.romo@inegi.org.mx, respectivamente.

Nota: las opiniones contenidas en esta investigación son responsabilidad exclusiva de sus autores y pueden no coincidir con las de su institución de adscripción, el INEGI; se agradecen las aportaciones de árbitros anónimos que han contribuido a mejorar el presente trabajo; los errores que permanezcan son total responsabilidad de quienes lo desarrollaron.

cional, lo que contrasta con el decremento en el valor del mismo indicador para la ENIGH. En otras palabras, nuestros resultados indican que, durante el periodo, la desigualdad se incrementó. La que se presenta al interior de las entidades federativas parece en cambio no haber experimentado una evolución significativa. Lo anterior parece sugerir que lo que ha impulsado el crecimiento de la desigualdad en el país es el aumento de esta entre los estados.

Palabras clave: ingreso; máxima pseudo-verosimilitud restringida (MPVR); desigualdad.

Recibido: 8 de octubre de 2020.
Aceptado: 10 de diciembre de 2021.

indicate that, during the period, inequality increased. On the other hand, inequality within states does not seem to have evolved significantly. This seems to suggest that what has driven the growth of inequality in the country is the increase in inequality between states.

Key words: income; maximum constrained pseudo-likelihood (MCPL); inequality.



Cancun, Mexico. May 30, 2021. Skeleton couple sitting on outdoor bench at roadside against entrance of restaurant/ Ingus Krukltitis / iStock

Introducción

Las fuentes de información estadística sobre el ingreso de los hogares son diversas. Todas ellas muestran ventajas y desventajas en cuanto a la estimación de la distribución del ingreso.¹ Sin duda, la mejor conocida es la representada por los censos de población y vivienda, que son levantados con diversas frecuencias, usando distintas metodologías, por casi todos los países del planeta. Su cobertura universal da lugar a un alto costo de recolección y, en consecuencia, a un intervalo que comprende varios años entre un levantamiento y el siguiente. Dicha cobertura lleva habitualmente a pensar que la información recolectada durante esos ejercicios es más precisa. Sin embargo, ello es más o menos cierto dependiendo del tema o de la variable de que se trate. En el caso particular del ingreso se presentan limitaciones que han llevado a la División de Estadística de la Organización de las Naciones Unidas a excluirla del grupo de variables censales importantes a ser recolectadas (ver DESA, 2017, tabla 3). Por principio de cuentas, no es del todo claro que la persona que proporciona los datos conozca con precisión los ingresos de cada uno de los perceptores en la vivienda, o sus fuentes. Tampoco es claro si la declaración se refiere a ingresos brutos o netos, pues no es fácil que conozca las deducciones experimentadas por los ingresos por trabajo, sean estas fiscales, de seguridad social o de alguna otra naturaleza. Lo anterior es agravado cuando la declaración del monto del ingreso es falseada o de plano omitida, por ejemplo, bajo circunstancias de alta inseguridad.

La recolección de información que se lleva a cabo a través de encuestas no especializadas en la medición del ingreso de los hogares enfrenta limitaciones similares a las censales. Una excepción a lo anterior la representa el caso de aquellas para las que una de cuyas temáticas principales es el estudio del ingreso de los hogares.² En ellas se busca,

por lo general, que cada uno de los perceptores aporte datos acerca de su propio ingreso. Adicionalmente, para cada uno de ellos, se indaga sobre el ingreso percibido por distintas fuentes, entre las que se encuentran las provenientes de un trabajo, ya sea dependiente de terceros o no, las transferencias gubernamentales o de parientes, así como los apoyos y préstamos de familiares o amigos, por mencionar algunas de ellas. En otras palabras, se contempla un mayor desglose temático a costa de uno geográfico menor. Además de esta desventaja, otras limitaciones de este enfoque se refieren a la representatividad de la muestra, pregunta importante en la producción de estadística oficial. En general, es muy difícil que la pequeñísima fracción de la población cuyos ingresos son extraordinariamente altos esté representada en una muestra aleatoria. Dado su pequeño tamaño, puede pensarse que es despreciable el error cometido al considerar que esta sí representa a la población, pero cuando se considera el tamaño extraordinario de sus ingresos dicha idea debe ser revisada. Asimismo, como en el caso censal, tampoco es claro si el monto del ingreso es subdeclarado cuando no, de plano, omitido.

Una fuente que podría representar una cobertura universal en relación con los ingresos está constituida por los registros fiscales. En teoría, cada uno de los perceptores reportaría de manera periódica, por lo menos una vez cada año, a la autoridad hacendaria tanto el monto de sus ingresos como las aportaciones fiscales realizadas en un periodo determinado. En adición a la declaración del total percibido, las declaraciones fiscales por lo general permiten hacer referencia a cada una de las fuentes que componen dicho ingreso. Sin embargo, con frecuencia, los registros administrativos nacen con la intención de atender necesidades específicas, sin tomar en cuenta usos estadísticos alternativos que es deseable dar a la información. Por supuesto, los fiscales no son la excepción. En general, no cumplen con la cobertura universal por diversas razones, entre las que destaca la situación de informalidad laboral en la que se desempeña una fracción más o menos importante de la población económicamente activa; asimismo, no

1 El lector interesado podrá extender esta información en UNECE (2011).

2 A manera de ejemplo, la información en este apartado puede ser ampliada, para el caso mexicano, en la sección *Documentación* de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), en la siguiente URL: <https://www.inegi.org.mx/programas/enigh/nc/2016/default.html>

todos los trabajadores formales están obligados a presentar una declaración anual. Por lo anterior, el registro que recoge las declaraciones fiscales anuales de las personas físicas solo se refiere a aquella parte de la población que, o bien está obligada a presentarla por satisfacer una o más condiciones reglamentarias, o que lo hace voluntariamente para obtener una devolución parcial de impuestos. Un registro fiscal complementario es aquel que concentra los reportes patronales sobre los pagos de sueldos y salarios a sus trabajadores, así como sobre las retenciones fiscales o de otra naturaleza que les realizan. Esta información se actualiza mes con mes, pero debe enfatizarse que cubre solo a los empleados formales.

A su vez, el Sistema de Cuentas Nacionales (SCN) representa un esfuerzo internacional por hacer disponibles grandes resúmenes que reflejen la situación económica de los países. Entre sus mayores virtudes está la de basarse en una metodología coherente que les otorga gran consistencia y que está en constante actualización para adaptarse a nuevas circunstancias (INEGI, 2017). Las cuentas nacionales proporcionan una descripción integral de toda la actividad económica de una nación, incluyendo aquellas que involucran unidades domésticas (es decir, los individuos y las entidades residentes en el territorio) y las externas (los residentes en otros países).

Existen dos grandes tipos de sectores institucionales.³ De un lado se encuentran las entidades jurídicas o sociales reconocidas por ley que llevan a cabo actividades y operaciones en nombre propio, como las empresas, unidades gubernamentales e instituciones sin fines de lucro que sirven a los hogares, para las cuales se posee un conjunto completo de cuentas y balances de activos y pasivos. El otro reúne a las personas y a los hogares o grupos de individuos, así como a las empresas individuales sin personalidad jurídica, que pueden ser propietarios de activos y contraer pasivos, pero no tienen obligación legal de contabilizar sus actividades.

3 INEGI. *Sistema de Cuentas Nacionales de México. Fuentes y metodologías. Año base 2013. Cuentas por Sectores Institucionales*, https://www.inegi.org.mx/contenidos/programas/si/2013/metodologias/SCNM_Metodo_CSI_B2013.pdf

Para fines de la determinación de la distribución del ingreso, el SCN solo aporta agregados a nivel nacional y, en algunos casos, al estatal. Sabemos que se realizan esfuerzos por lograr lo que se denomina Cuentas Nacionales Distributivas⁴ (DINA, por sus siglas en inglés). Dichos esfuerzos destacadamente hacen uso de series de datos fiscales y de cuentas nacionales armonizados, complementadas con encuestas de ingresos y de riqueza, así como con datos de pago de impuestos por herencias y propiedades.

Inconsistencia entre los agregados muestrales y los de la contabilidad nacional

La estimación del ingreso corriente total (ICT) de los hogares aportada por las encuestas en diversas naciones muestra una discrepancia más o menos importante con el total determinado por los SCN para el sector institucional de hogares. En el caso mexicano, la información muestral estima un total para el ingreso corriente de estos que alcanza consistentemente un valor menor a la mitad del que consigna el Sistema de Cuentas Nacionales de México (SCNM). Ha sido usual, en este como en muchos otros países, asignar mayor credibilidad a esta última cifra. En consecuencia, se ha buscado determinar las causas que dan lugar a la aparente subestimación. Hoy en día se identifican dos principales, ya mencionadas, e identificadas como subdeclaración y truncamiento.

Leyva-Parra (2004) analiza diversas propuestas cuyo fin es el de reducir la discrepancia mencionada, desde Martínez (1970) hasta el 2004, sobre el ajuste de los ingresos de los hogares obtenidos de encuestas especializadas a resultados provenientes del SCNM para, de esta manera, hacer explícitos los supuestos de los que depende la validez de tales ajustes y, con ello, mostrar las limitaciones de los procedimientos analizados. De paso, cuestiona la validez de las cifras provenientes del SCNM como

4 OCDE. "For Good Measure: Advancing Research on Well-Being Metrics Beyond GDP", en *Chap. 6: Distributional national accounts*.

referente para la corrección de los datos de las encuestas de ingresos de los hogares.

De acuerdo con Leyva, la experiencia mexicana en cuanto al "... ajuste a Cuentas Nacionales, ya sea por niveles, fuentes o alguna combinación de ambos, descansa en los siguientes supuestos básicos:

1. Los conceptos de ingresos manejados en ambas fuentes son equiparables.
2. Los ingresos captados en las Cuentas Nacionales son al menos tan verosímiles como los de las encuestas de ingresos de los hogares.
3. Las diferencias entre ambas fuentes se deben fundamentalmente a problemas de subreporte y no a problemas de truncamiento.
4. Existe una regla de asignación óptima que permite distribuir el ingreso de los hogares, a nivel macroeconómico, al ingreso (expandido) de cada uno de los hogares en la muestra de la encuesta de ingresos (nivel microeconómico). Es una regla óptima en el sentido de que maximiza la verosimilitud de la distribución del ingreso resultante..."

Desde luego, algunos de los anteriores supuestos son más sensatos que otros, en tanto que los de un tercer grupo no se sostienen ante la evidencia. Por ejemplo, si un grupo de población pequeño, pero con ingresos muy importantes, no está representado en la muestra de la encuesta, "... el valor total del ingreso expandido de la misma, aún sin la presencia del subreporte, debe ser inferior al correspondiente con la Contabilidad Nacional, que por su metodología y cobertura incluye en principio los ingresos de todos los perceptores, sin excepción...". La ausencia de tal grupo de hogares en la muestra da lugar a una censura por truncamiento que puede llegar a ser sustancial. En consecuencia, el tercer supuesto no corresponde con la realidad. Si en el ajuste al SCNM se ignora al mencionado truncamiento, el resultado será la redistribución de una cantidad desconocida, y tal vez importante, entre los hogares incluidos en la muestra. En los hechos, se incrementarán los ingresos observados por la encuesta en diversas proporciones, lo que dará lugar a consecuencias impredecibles en la

aplicación de la política social del Estado mexicano; por ejemplo, al reducir en gabinete el número de hogares en condiciones de pobreza. Concluye Leyva que la distinción entre subreporte y truncamiento no es trivial y que a la sazón no existía ningún procedimiento sólido que permitiera hacerla.

Altimir (1987) propone un método "... basado en la asignación de las discrepancias con las cuentas nacionales según el tamaño del ingreso para cada tipo de ingreso (sueldos y salarios, empresariales, renta de la propiedad, transferencias, ingresos en especie, etc.) en vez de tomar solo en cuenta el ingreso total...". Entre sus cinco supuestos principales, que se refieren exclusivamente al subreporte, destacamos el segundo: "La magnitud total del subreporte de los ingresos por tipo es aproximada mediante la discrepancia entre los ingresos muestrales y la estimación correspondiente para ese tipo de ingreso proveniente de las cuentas nacionales, ya corregidos por diferencias conceptuales, si aquel es menor que este...". Y continúa diciendo que el "... procedimiento de ajuste consistió, por tanto, en ajustar separadamente los ingresos en la distribución de los hogares de cada grupo socioeconómico, mediante el coeficiente ponderado correspondiente al grupo asumiendo una elasticidad ingreso unitaria de subregistro, excepto los ingresos monetarios de la propiedad..."; suponiendo que "... el subregistro de los ingresos monetarios de la propiedad se concentra en el quintil más alto de hogares...".

Para el caso en el que solo se dispone de la "... distribución general de los hogares por tamaño de ingreso, pero existe información sobre la composición del ingreso en cada segmento i , solo sería posible ajustar los ingresos en cada intervalo de acuerdo con la composición por tipo de ingreso, ya que los factores de ajuste en cada intervalo serían promedios ponderados de manera diferente de los factores de ajuste para cada tipo de ingreso..."; como se muestra en (1), donde Y_j^u representa el monto de la discrepancia para el tipo de ingreso j . Entonces, el ingreso total ajustado Y_i^A se calcula como la suma ponderada de los totales muestrales para cada tipo de ingreso Y_{ij}^0 , dentro de cada segmento de ingreso i :

$$Y_i^A = \sum_j c_j Y_{ij}^0 = \sum_j \left(1 + \frac{y_j^H}{y_j^0}\right) Y_{ij}^0 = Y_i^0 + \sum_j \left(\frac{y_j^H}{y_j^0}\right) Y_{ij}^0 \quad (1)$$

Esta expresión más simple permite observar con mayor claridad que los ingresos del mismo tipo son *ajustados* de manera proporcional a lo largo de todos los intervalos de ingreso, tal como ocurre en el caso general para el cual se da un tratamiento especial a los ingresos provenientes de la renta del capital/propiedad y en el que se consideran diversos *grupos socioeconómicos*. Hace evidente, asimismo, que Altimir decide ignorar el hecho de que la muestra solo es representativa de una subpoblación cuyos ingresos no llegan a ser desmesuradamente grandes.

A pesar de lo anterior, este es un procedimiento citado con frecuencia. La afirmación que Altimir hace en el sentido de que "... aunque inevitablemente arbitrario en sus supuestos, y limitado en sus aplicaciones por la desagregación de los datos disponibles, este método fue diseñado con el propósito de ajustar datos de ingreso de diferentes tipos de encuestas de una manera uniforme..."; parece indicar que detrás del método está la intención de llevar a cabo comparaciones entre los países de la región. Aunque su propuesta consigue el único propósito claramente establecido de reducir las discrepancias entre los resultados muestrales y los provenientes de las cuentas nacionales, no queda claro que el resultado sea óptimo en algún sentido. En efecto, hay un método, pero no un criterio. Tal vez por ello tiene que reducirse a una alternativa arbitraria en sus supuestos y limitada en sus aplicaciones.⁵

Por lo que toca a las metodologías analizadas, Leyva establece que "... gravitan críticamente en los supuestos de cada autor acerca de las reglas de

5 La propuesta de Altimir parece haber sido usada por la Comisión Económica para América Latina y el Caribe (CEPAL) para el cálculo de incidencia de la pobreza en los países de esa región. Recientemente, la CEPAL (2019) ha eliminado este procedimiento y trabaja con los datos de las encuestas sin ajustarlos. Al recurrir a la identificación de una población de referencia mediante el método que denominan de *quintiles móviles*, los cuales parecen surgir de la propia encuesta, argumenta que "... el ajuste de ingresos no mejora la plausibilidad, comparabilidad o confiabilidad de las mediciones de pobreza...". Por supuesto, ello ocurre ya que el método de ajuste no modifica las posiciones relativas de los hogares.

asignación que permiten distribuir entre los hogares o grupos de hogares, el valor de las diferencias entre los ingresos totales de los hogares reportados por las encuestas de ingresos y la Contabilidad Nacional...". Afirma, además, que "... sin excepción, todas las reglas de asignación disponibles hasta el momento resultan arbitrarias...".

Aunque en otro contexto, la polémica suscitada por el libro *El capital en el siglo XXI*, del economista francés Thomas Piketty (2014), ilustra que el problema comentado líneas arriba para México ha estado, y continúa, presente también en otras latitudes. Una de las críticas⁶ más difundidas a los resultados presentados en la publicación provino de Chris Giles, editor económico de *The Financial Times* (FT). En ella, ataca la credibilidad de Piketty argumentando que cometió una serie de errores de naturaleza diversa que sesgan sus hallazgos. De hecho, Giles afirma que, después de corregir los aparentes errores, no encuentra evidencia de que la concentración de la riqueza haya aumentado en los 30 años más recientes. En su respuesta,⁷ Piketty demuestra que algunas de las correcciones practicadas por el FT son menores y no alteran sus conclusiones; otras se basan en elecciones metodológicas que son debatibles. Encuentra Piketty que, entre las opciones metodológicas del FT, es particularmente problemático su uso de fuentes fiscales para las décadas iniciales y el cambio a estimadores muestrales para periodos más recientes. Y continúa, "... es problemático ya que sabemos que en todo país las encuestas sobre riqueza tienden a subestimar la participación de los más ricos en contraste con estimadores basados en datos administrativos fiscales...". Por ello, dichas opciones metodológicas pueden sesgar los resultados hacia la declinación de la desigualdad.

Paul Krugman participa en el mismo debate.⁸ Sugiere que los supuestos errores señalados por Giles son, en esencia, "... los tipos de ajustes de datos que son normales en cualquier investigación que

6 <http://piketty.pse.ens.fr/files/capital21c/en/media/FT23052014c.pdf>

7 <http://www.voxeu.org/article/factual-response-ft-s-fact-checking>

8 http://www.nytimes.com/2014/06/02/opinion/krugman-on-inequality-denial.html?_r=0

se apoya en una variedad de fuentes...". Continúa diciendo, "... a riesgo de aportar demasiada información, este es el punto. Tenemos dos fuentes de evidencia tanto sobre ingreso como sobre riqueza: encuestas por muestreo, en las que se pregunta a las personas acerca de sus finanzas, y datos fiscales. Los datos muestrales, aunque útiles para dar seguimiento a las clases pobres y medias, subestiman notoriamente la riqueza y los ingresos más altos —en términos llanos— porque es difícil entrevistar suficientes billonarios. De este modo, los estudios sobre el 1%, el 0.1%, y otros semejantes se apoyan principalmente en datos fiscales...".

Leyva-Parra (2004) concluye: "... cabe señalar que incluso si efectivamente las cifras de ingresos de los hogares de la Contabilidad Nacional pudieran considerarse como más verosímiles que las de la ENIGH, persistiría el problema de que aún no se ha desarrollado un criterio de asignación óptimo que permita establecer el puente para distribuir las diferencias macroeconómicas a nivel microeconómico. Este problema se complica cuando se considera el hecho de que las Cuentas Nacionales y la ENIGH hacen referencia a universos de hogares distintos, en virtud de que hay una fracción que se sospecha importante de los ingresos que se reportan en las estadísticas macroeconómicas (que se supone incluyen a todos los grupos de hogares), las cuales no se captan en las encuestas tipo ENIGH, lo que implica la existencia de un truncamiento (al menos) en la parte alta de la distribución del ingreso. Así, incluso en el caso de que se dispusiera de una regla de asignación óptima para pasar de lo macro a lo micro, sería necesario saber qué parte de la discrepancia de la ENIGH con las Cuentas Nacionales corresponde al subreporte y cuál, al truncamiento...".

Ante la dificultad de asignar la proporción del déficit representada por cada una de las anteriores causas, ha sido usual actuar como si ninguna o solamente una de las dos estuviera presente. Por supuesto, ello puede llevarnos a tomar decisiones contra cuyos sesgos habrá que estar prevenidos. Con fines de ejemplificación, consideren el uso de los valores del ingreso sin modificación alguna, se-

gún fueron declarados en la encuesta. Bajo estas condiciones, al no tomar en cuenta la subdeclaración, se exagerará la proporción de hogares que exhibirá un ingreso inferior a una cantidad dada.⁹ Es decir, esta proporción resultará ser mayor que la que se alcanzaría si los ingresos hubieran sido ajustados por subdeclaración. No es difícil pensar que, en estas condiciones, medidas tales como la prevalencia de pobreza por ingresos lleguen a ser sobreestimadas. De manera similar, la ausencia de ingresos extraordinariamente grandes entre los datos muestrales podría también llevarnos a una subestimación de la desigualdad de ingresos.

La distribución del ingreso es una de las estadísticas más referidas tanto en revistas especializadas como en diarios y noticieros. Sus aplicaciones son múltiples en estudios de pobreza, de desigualdad y aún de recaudación fiscal. Por otro lado, si se supone que la discrepancia es debida exclusivamente a la subdeclaración del ingreso en la encuesta, puede sugerirse que, de acuerdo con algún criterio, los valores declarados sean incrementados de modo que la suma expandida de los nuevos valores coincida con el total de cuentas nacionales. Aunque lo anterior puede resolver el problema de la discrepancia, añade incertidumbre para fines de política pública, pues la proporción del ingreso que sería explicado por truncamiento va a ser distribuido entre los hogares en la muestra, lo que puede dar lugar a un incremento excesivo de sus respectivos ingresos. Considérese entonces el caso de aquellos hogares cuyos ingresos declarados se encuentran cercanos, pero por debajo de una línea de pobreza por ingresos. Es posible pensar que, en algunos de estos casos, el nuevo ingreso resulte mayor a la mencionada línea de pobreza. Los hogares en estas condiciones no serán incluidos en la estimación del porcentaje de los que se encuentran en condiciones de pobreza. Es claro que proceder de este modo puede dar lugar a una subestimación de magnitud desconocida de las cifras de pobreza. Lo que ocurra con la medición de la desigualdad de ingresos dependerá de la forma o el método uti-

⁹ Como los 2 dólares diarios, según la paridad de poder de compra, usados por el Banco Mundial.

lizado para realizar el ajuste, pero probablemente decrecerá. La gráfica 1 ilustra el efecto de realizar un ajuste deficiente a los datos del ingreso según las encuestas en hogares. Supongamos por un momento que la mediana de la distribución del ingreso obtenida a partir de los datos muestrales aporta una estimación razonable a lo que consideraríamos una línea de pobreza por ingresos. De manera gruesa, esto implicaría que la mitad de los hogares del país, cuyos ingresos se encuentran por debajo de dicha línea, son considerados en pobreza.

Más aún, supongamos que, ya que para nuestros fines basta con que la suma expandida de los datos ajustados $Y_{(i)}^{(a)}$, $i = 1, \dots, n$, coincida con el total estimado por el Sistema de Cuentas Nacionales, consideramos adecuado el ajuste que multiplica a las cifras observadas $Y_{(i)}$ por 2.43, como sería el caso de México para el 2012. Obsérvese que la proporción de hogares considerados en pobreza se reduciría a alrededor de 12 % si la línea de pobreza no presentara desplazamiento alguno después de llevar a cabo el ajuste. Debe ser claro que semejante éxito de la política social no es sino el resultado de un

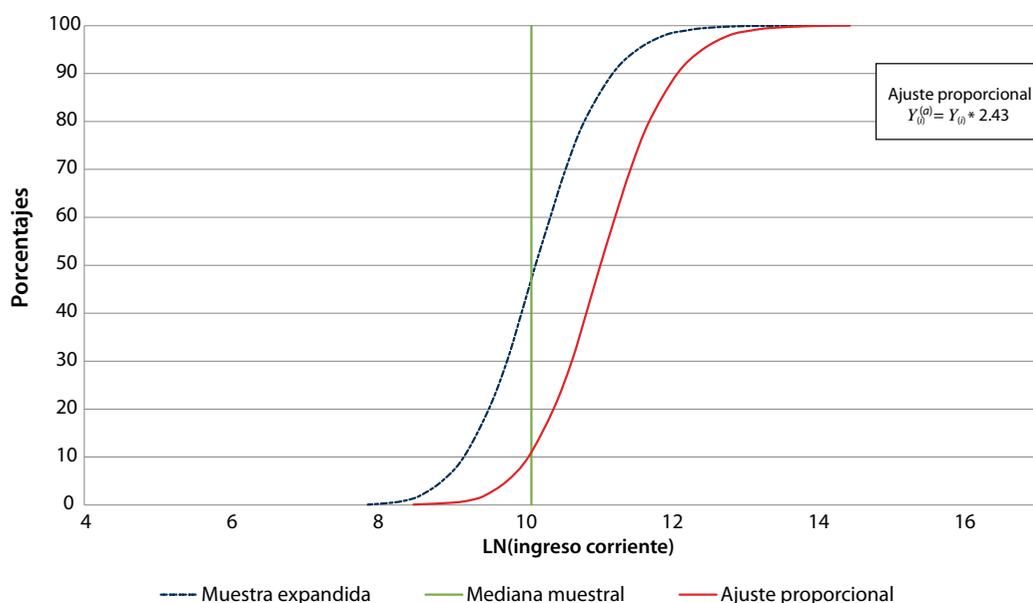
truco aritmético, que satisface la única condición impuesta sobre la suma expandida de los valores ajustados.

Por último, también se corren riesgos cuando se supone que el truncamiento explica totalmente la discrepancia entre ambas estimaciones. En este caso, autores como Campos *et al.* (2014), siguiendo a Lakner y Milanovich¹⁰ (2013), han seguido el curso de no aplicar modificación alguna a los ingresos de los hogares que se encuentran en los primeros ocho deciles de la encuesta de ingresos, de modo que la discrepancia entre las fuentes es distribuida en diferentes proporciones entre los que ocupan los dos deciles mayores. Por lo que toca a los ingresos menores, aquellos asociados con condiciones de pobreza, regresamos a la situación discutida párrafos arriba en la que los ingresos no son sometidos a ajuste alguno. En otras palabras, es probable que la magnitud de la pobreza sea sobreestimada sin que se pueda determinar la dimensión de dicha sobreestimación causada por

¹⁰ Estos autores ajustan por el consumo, no el ingreso, a diferencia de Campos y coautores.

Gráfica 1

Ajuste proporcional de los valores observados del ingreso, 2012



Fuente: cálculos propios a partir de la base de datos de la ENIGH 2012.

ignorar la presencia de subdeclaración para esos niveles de ingreso. Por supuesto, mediciones de la desigualdad por ingresos crecerán más allá de lo que parecería un límite razonable.

En el trabajo de Campos *et al.*, la distribución del ingreso que obtienen siguiendo el anterior procedimiento se vincula con la tributación óptima y con la captación posible. Es claro que la imagen posiblemente distorsionada¹¹ de dicha distribución puede conducir a la conclusión de que la captación de ingresos fiscales obtenida hoy en día es muy inferior a la que tendría lugar si sus supuestos se cumplieran. Si el encargado de elaborar la política de ingresos del gobierno federal tomara estos resultados de forma acrítica, no es difícil pensar que llegaría a la conclusión de que la magnitud de la evasión fiscal alcanzaría niveles importantes. Las mismas decisiones que tomaría con base en resultados tan inciertos podrían conducirle a, por ejemplo, instrumentar ajustes fiscales que podrían conducir a mayores déficits gubernamentales, pues los ingresos efectivamente tributados no alcanzarán el nivel esperado, según estas estimaciones.

En la siguiente sección se hará un breve recuento de la metodología desarrollada para obtener los resultados que se presentan más adelante. A lo largo de esta investigación se ha recurrido a nuestra propuesta original que se ha denominado criterio de máxima pseudo-verosimilitud restringida (MPVR). También, se hace referencia a algunos resultados publicados con anterioridad con el fin de facilitar la comprensión del material. El lector interesado podrá recurrir a los trabajos de Bustos (2015 a, b) y Bustos y Leyva (2017). La segunda sección exhibe el comportamiento de los resultados de aplicar dicha metodología a datos mexicanos entre los años 2010 y 2016, tanto a nivel nacional como desagregándolos por entidad federativa. Finalmente, la última incluye consideraciones acerca de la metodología, sobre sus extensiones, así como algunas sugerencias para trabajo futuro de investigación.

11 "... tal ajuste puede parecer excesivo [...] cuando la media muestral es igual a solo 50% del consumo privado (NT: como llega a ocurrir en México con el ingreso) entonces simplemente asignar este 50% al decil más alto es probablemente excesivo..."; Lakner *et al.* (2013, p. 38).

Metodología

Ante todo, es necesario dejar claro que nuestra investigación pretende encontrar una aproximación a la distribución del ingreso de los hogares en México aprovechando de la mejor manera posible la información disponible. No pretende, como ha sido el caso con otros trabajos relacionados (por ejemplo, Martínez, 1970 y Félix, 1979), producir una base de datos con valores ajustados del ingreso. Más aún, dado que "... el supuesto referido a la regla de asignación es en realidad sólo un buen deseo ya que, sin excepción, todas las reglas de asignación disponibles hasta el momento resultan arbitrarias..." (Leyva, 2004), nosotros hemos decidido iniciar estableciendo con toda claridad el criterio de optimalidad que debe guiar nuestra búsqueda de soluciones en la estimación de la distribución del ingreso. En este sentido, el contraste es claro, las propuestas revisadas por Leyva (2004) han resultado en datos ajustados que satisfacen que su suma expandida coincida con el valor aportado por el SCNM. Sin embargo, no es posible concluir cuáles entre ellas son mejores en algún sentido razonable, pues han sido obtenidas recurriendo a supuestos *ad hoc*, y sin hacer explícita la medida de eficiencia usada. Se puede suponer que esta es otra de las razones por las cuales la medición oficial de la pobreza en México al inicio del siglo XXI no hace uso de ninguna de tales propuestas de ajuste.¹²

En resumen, la metodología usada en el presente trabajo para obtener los resultados que se discutirán es la propuesta en Bustos (2015.a), que tiene el siguiente objetivo: hacer un uso óptimo,¹³ según el criterio MPVR, de la información disponible para estimar una distribución del ingreso corriente de los hogares en México que sea más cercana a la realidad y que resulte menos arbitraria, sin hacer imputaciones ni correcciones a los datos de la ENIGH.

12 En CONEVAL (2019), pp. 100-101, se establece que uno de los criterios usados es "... Emplear la información de ingresos captada bialmente por el INEGI en la ENIGH...", sin hacer referencia a algún ajuste.

13 El término *óptimo* hace referencia a un criterio explícito (por ejemplo, MPVR). Intenta comparar con propuestas para las cuales no es posible determinar si un cambio cualquiera da lugar o no a un mejor resultado numérico, por carecer de dicho criterio explícito. Es posible pensar que, con base en otros criterios, nuestros resultados dejarían de ser óptimos.

El criterio utilizado busca maximizar la pseudo-verosimilitud para la muestra, tomando en cuenta la información proveniente de otras fuentes, incorporada a manera de restricciones que deben ser satisfechas por los valores del vector de parámetros. Dicho planteamiento apareció por primera vez en Bustos (2015.a), y queda expresado como se muestra en (2):

$$Max_{\underline{\theta}, \underline{\lambda}} \left\{ \sum_{i=1}^n \frac{1}{\pi_{(i)}} \ell(\underline{\theta}; Y_{(i)}) - \lambda' (\underline{h}(\underline{\theta}) - \underline{c}) \right\} \quad (2)$$

donde:

$\ell(\underline{\theta}; Y_{(i)})$ representa al logaritmo natural de la función de densidad, evaluada en el i -ésimo valor muestral del ingreso $Y_{(i)}$, como función del vector de parámetros $\underline{\theta}$ cuyo valor óptimo se pretende determinar.

$\pi_{(i)}$, a la probabilidad de inclusión correspondiente a dicha unidad muestral.

$\underline{h}(\underline{\theta})$, a la(s) función(es) de los parámetros cuyos valores están restringidos a tomar valores específicos dados en el vector \underline{c} (ej., SCN) y cuya forma está vinculada a la particular distribución de que se trate.

$\underline{\lambda}$, a un vector de multiplicadores de Lagrange de dimensiones apropiadas.

Con base en los ingresos trimestrales de los hogares de la ENIGH 2012, se ajustaron dos distribuciones de probabilidad:

- Gamma generalizada (GG).
- Beta generalizada tipo II (GB2).

Las formas funcionales de ℓ y de \underline{h} varían con cada selección de distribución. En todos los casos se impuso una restricción: $\underline{h}(\underline{\theta}) = E[Y|\underline{\theta}] = \text{Ingreso trimestral promedio por hogar}$, según el SCNM. Para considerar la información aportada por los registros fiscales anonimizados, se incluyó una segunda. El cuadro 1 resume la forma que han tomado las restricciones usadas a lo largo del desarrollo del proyecto. La primera de ellas obliga a los parámetros del modelo, incluidos los del ajuste óptimo, a tomar valores tales que el promedio de la distribución ajustada coincida con el que se determina a partir de la información del Sistema de Cuentas Nacionales de México.¹⁴ La segunda obliga a coincidir a los promedios de los ingresos de los hogares que ganan más que el umbral φ_{α} , tanto según el modelo como el Sistema de Administración Tributaria (SAT).¹⁵

14 En México, en el 2012, dicho promedio era de 86 410.57 pesos; por lo anterior, todas las funciones ajustadas tienen una media de casi 86 410. El equivalente reportado por la ENIGH para el 2012 se encuentra en casi 38 mil pesos.

15 Se usó el supuesto de que los ingresos de personas físicas mayores a un umbral muy alto son equivalentes a los de un hogar, ya que la base de datos anonimizada de declaraciones anuales de las personas físicas en el SAT no permite agregar los ingresos por hogar.

Cuadro 1

Restricciones usadas en los ejemplos numéricos

Concepto	Restricción:	Interpretación
Ingreso promedio (SCNM)	$h_1(\underline{\theta}) = E[Y \underline{\theta}] = c_1$	El valor del ingreso promedio por hogar en el modelo ajustado coincide con el reportado por el SCNM.
Integral de ingresos (SAT)	$h_2(\underline{\theta}) = E(Y Y > \varphi_{\alpha} \underline{\theta}) = \frac{1}{\alpha} \int_{\varphi_{\alpha}}^{\infty} y f_Y(y \underline{\theta}) dy = c_2$ donde $\alpha = \int_{\varphi_{\alpha}}^{\infty} f_Y(y \underline{\theta}) dy$	El ingreso promedio c_2 del grupo de hogares de tamaño 100% con ingresos mayores al umbral φ_{α} , según el SAT, coincide con el del grupo con ingresos superiores al mismo umbral φ_{α} según el modelo ajustado. ^a

^a La aplicación de este procedimiento no requiere acceso a registros fiscales. Basta con los valores para uno o más umbrales de ingreso, así como los de las correspondientes proporciones.

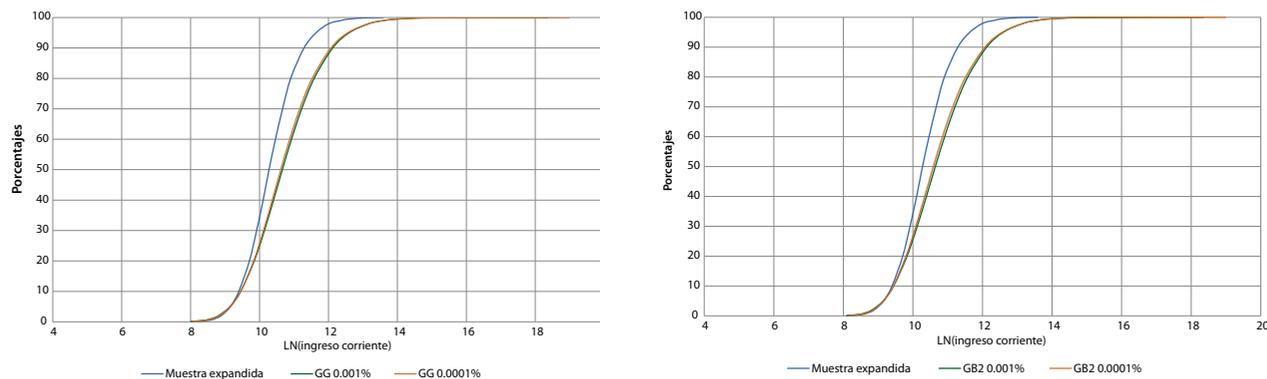
A manera de ejemplificación sobre las ventajas de incorporar información adicional, referimos al lector a los coeficientes de Gini obtenidos en Bustos (2015.a) y en Bustos *et al.* (2017). En el primer caso, al no contar con acceso a registros fiscales, solo se hizo uso de la primera restricción resultando en valores del Gini cercanos a 0.8, para el modelo óptimo. Por su parte, una vez incorporada la información del SAT, los valores reportados para el coeficiente de Gini se ubicaron entre 0.61 y 0.64, para los cuatro ajustes reportados. En ambos casos, los ajustes se realizaron usando información de la ENIGH 2012, la cual reportó un valor del coeficiente de Gini igual a 0.44, subestimando la desigualdad según los resultados anteriores. Análisis posteriores permitieron establecer que, por ejemplo, el umbral correspondiente a un millonésimo de los hogares para el ajuste con una sola restricción resultaba tener un valor 10 veces más grande que lo reportado por la base de datos de declaraciones fiscales para ese año. En otras palabras, dicho pri-

mer ajuste tomaba ingresos correspondientes a los deciles inferiores para transferirlos a los superiores resultando en una aparente sobreestimación del ingreso recibido por hogares en los deciles de ingresos más altos; consecuentemente, de acuerdo con estos resultados, la desigualdad de ingresos también sería sobreestimada. Es claro, por otro lado, que las políticas de recaudación fiscal que es posible instrumentar en uno u otro caso mostrarían también diferencias y montos importantes.

A partir del segundo conjunto de resultados, queda claro que las restricciones impuestas tienen un peso importante en su determinación. En efecto, se realizaron cuatro ajustes considerando dos umbrales diferentes para cada una de las distribuciones. El cuadro que acompaña a las gráficas 2 muestra un resumen de resultados relevantes. Se observará que los valores óptimos del criterio utilizado exhiben una variabilidad muy pequeña. Se apreciará, asimismo, que algo semejante ocurre

Gráficas 2

Comparación ENIGH vs. cuatro modelos ajustados por MPVR (dos distribuciones, dos umbrales c/u)



	ENIGH	MVPR			
		GG		GB2	
		0.0010%	0.0001%	0.0010%	0.0001%
Valor óptimo de MPLVR	N/A	-365 349 780	-364 628 437	-365 541 743	-364 913 147
Restricciones					
$E(X \theta)$	N/A	√	√	√	√
$E(X X > \varphi_{-\alpha}, \theta)$	N/A	√	√	√	√
Coefficiente de Gini	0.45	0.62	0.64	0.61	0.63
Ratio X/I	21.62	52.09	54.16	51.66	53.20

Fuente: cálculos propios a partir de bases de datos ENIGH, SAT y SCNM.

tanto para los coeficientes de Gini como para las relaciones entre el ingreso total del primero y el décimo deciles, denotado por Ratio X/I. A nuestro juicio, los umbrales utilizados hacen poca diferencia; los modelos empleados hacen todavía menos diferencia. A causa de lo anterior, debe quedar claro que en este contexto pierde relevancia el argumento frecuentemente utilizado en el sentido de que el uso de modelos, aunque en contextos diferentes, es arbitrario.

Como puede observarse en las gráficas 3, la consideración de agregados gruesos, como por ejemplo las participaciones en el ingreso de cada uno de los deciles de la población de hogares, parecen reflejar una importante semejanza entre los modelos alternativos. Es claro que las acumuladas en el ingreso de los primeros cinco deciles exhiben una diferencia de apenas 0.3 por ciento. Por su parte, la participación del último decil se encuentra entre

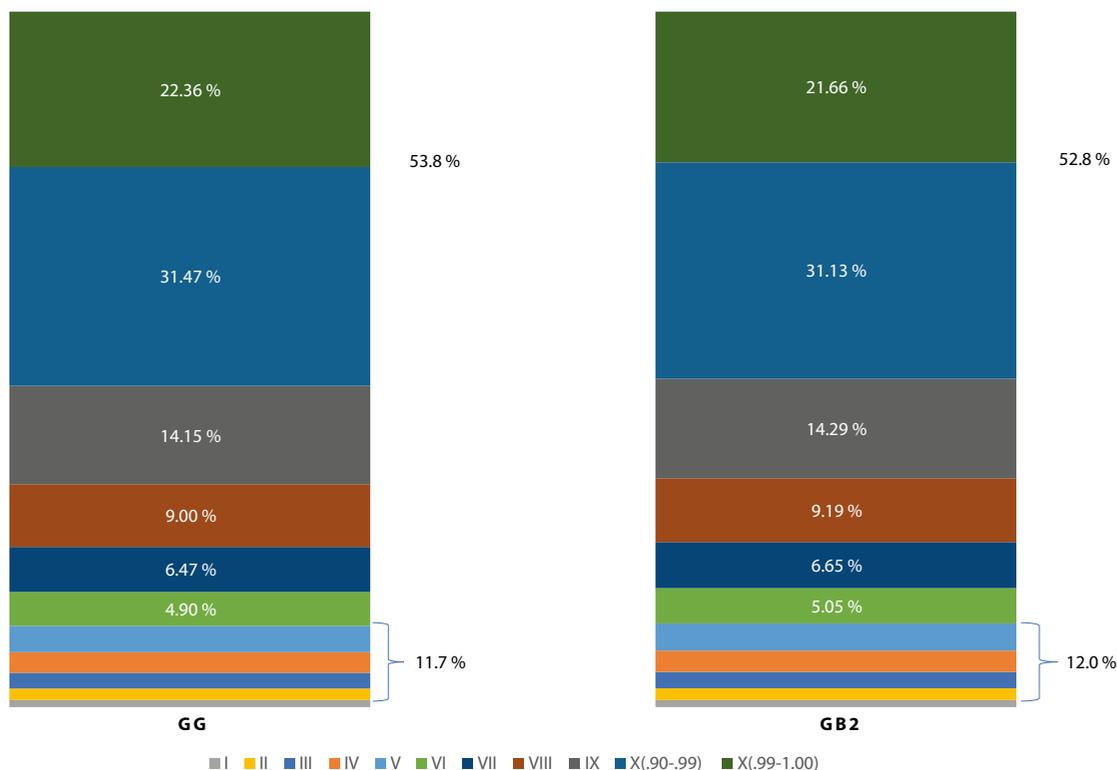
52.8 y 53.8 % del total. Las de 1 % con mayores ingresos difieren en menos de 1 por ciento.

Los anteriores resultados numéricos, en particular la indiferencia entre los modelos, nos llevan a pensar que se ve reducida la arbitrariedad en su elección, así como el peso de los supuestos que la misma conlleva, al considerar el uso simultáneo de fuentes, contra lo que ocurre en otras aplicaciones y contextos.

La gráfica 4 presenta una importante coincidencia entre los promedios de ingresos según la base de datos del SAT y los que se calculan a partir de los modelos ajustados, cuando los hogares son clasificados en rangos según su ingreso. La gráfica 5, en cambio, que muestra el número de hogares en cada rango de ingreso, exhibe diferencias importantes entre ambos conjuntos de cálculos, en particular en los rangos bajos de ingresos. Al menos,

Gráficas 3

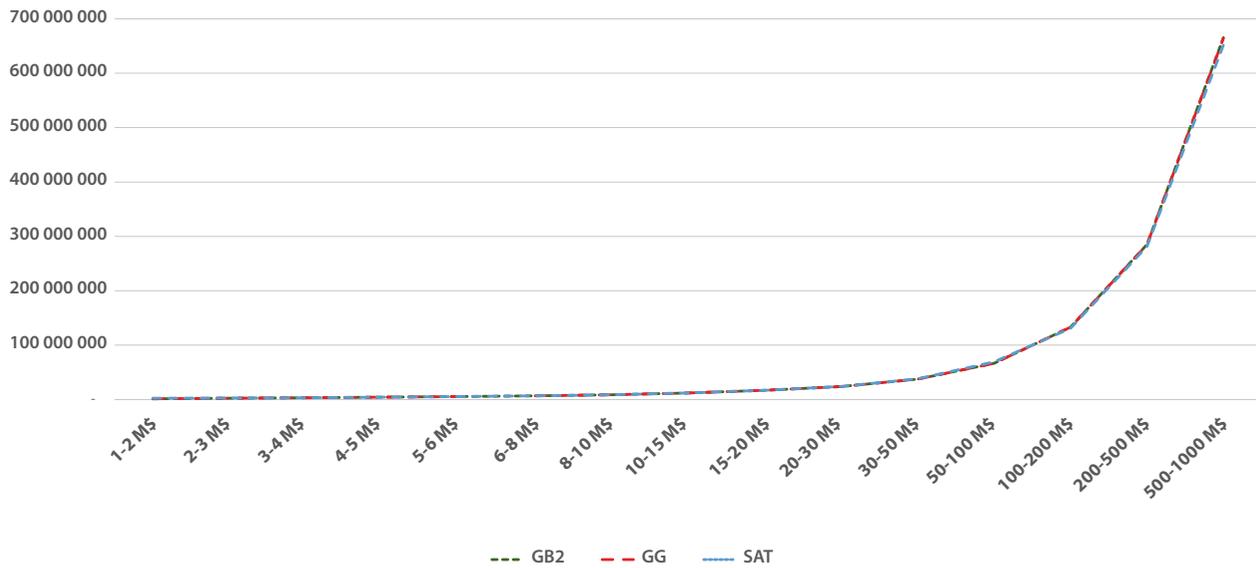
Distribución decilar del ingreso según dos modelos, México, 2012



Fuente: cálculos propios a partir de bases de datos ENIGH, SAT y SCNM.

Gráfica 4

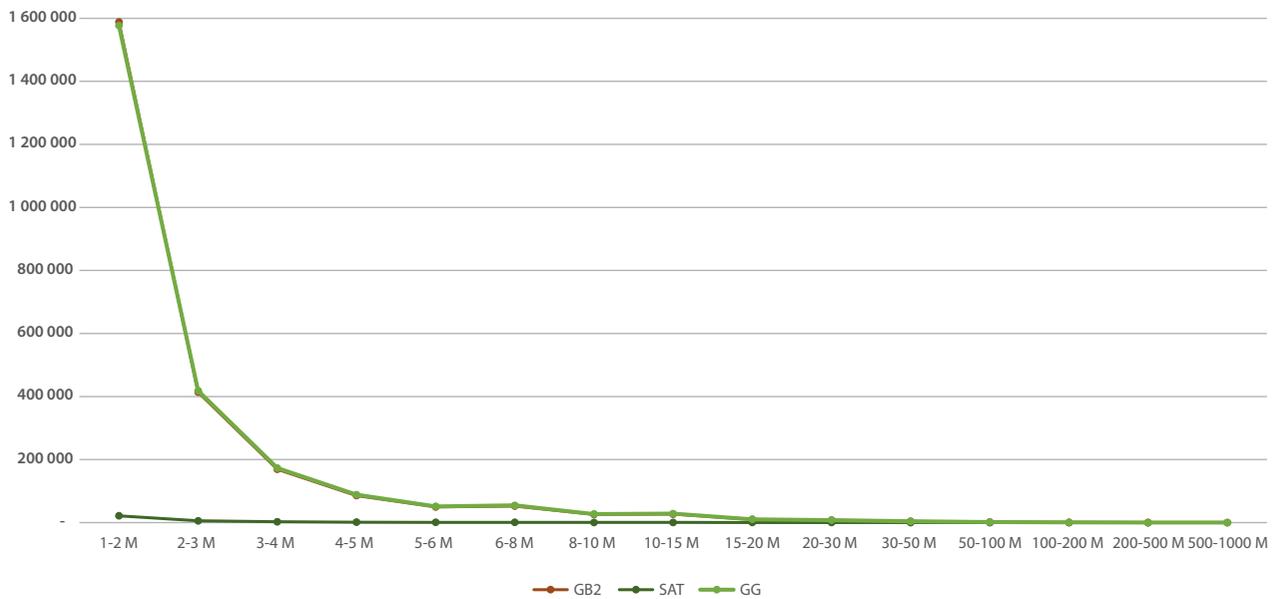
Ingreso promedio anual por rango, según fuente, México, 2014



Fuente: cálculos propios a partir de bases de datos ENIGH, SAT y SCNM.

Gráfica 5

Número de hogares según rango de ingresos y fuente, México, 2014



Fuente: cálculos propios a partir de bases de datos ENIGH, SAT y SCNM.

en parte, la mencionada diferencia parece ser consecuencia del supuesto introducido al hacer uso de la información fiscal; es decir, de utilizar una base de datos cuya unidad de observación es el individuo y no el hogar. En tanto dicho supuesto no parece tener grandes consecuencias para los ingresos altos, el

número de individuos cuyo ingreso se encuentra en uno de los rangos inferiores resultará menor que el de hogares cuyo ingreso, que es la suma de lo percibido por cada uno de sus integrantes, cae en el mismo rango. Es claro que estos hogares no son perceptibles a partir de la base de datos de las declaraciones

de las personas físicas. Así es, a medida que nos alejamos de los ingresos altos, el efecto de las diferentes unidades de observación se tornará más importante. Desde luego, no se descarta que otras causas, tales como nuestro otro supuesto sobre la confiabilidad del dato del Sistema de Cuentas Nacionales de México o la subdeclaración fiscal, estén presentes en las diferencias mostradas. Desafortunadamente, carecemos de información que nos permita determinar la importancia relativa de cada una de ellas.

El ingreso en México a lo largo del tiempo y de la geografía

En la presente sección, a través de un conjunto de resúmenes estadísticos, trataremos de ilustrar la forma en la que nuestros resultados exhiben, primeramente, la evolución de la distribución del ingreso a lo largo del tiempo; acto seguido, se añade la consideración de su distribución geográfica. Para estos fines, se hará uso de las bases de datos de la

ENIGH, o de su Módulo de Condiciones Socioeconómicas (MCS¹⁶), del INEGI; de las anonimizadas de declaraciones fiscales del SAT; así como de los resultados publicados para el sector institucional de hogares del SCNM. Todos ellos para los años 2010, 2012, 2014 y 2016. Asimismo, en todos los casos, el modelo ajustado a los anteriores datos será la gama generalizada con restricciones (GGc).

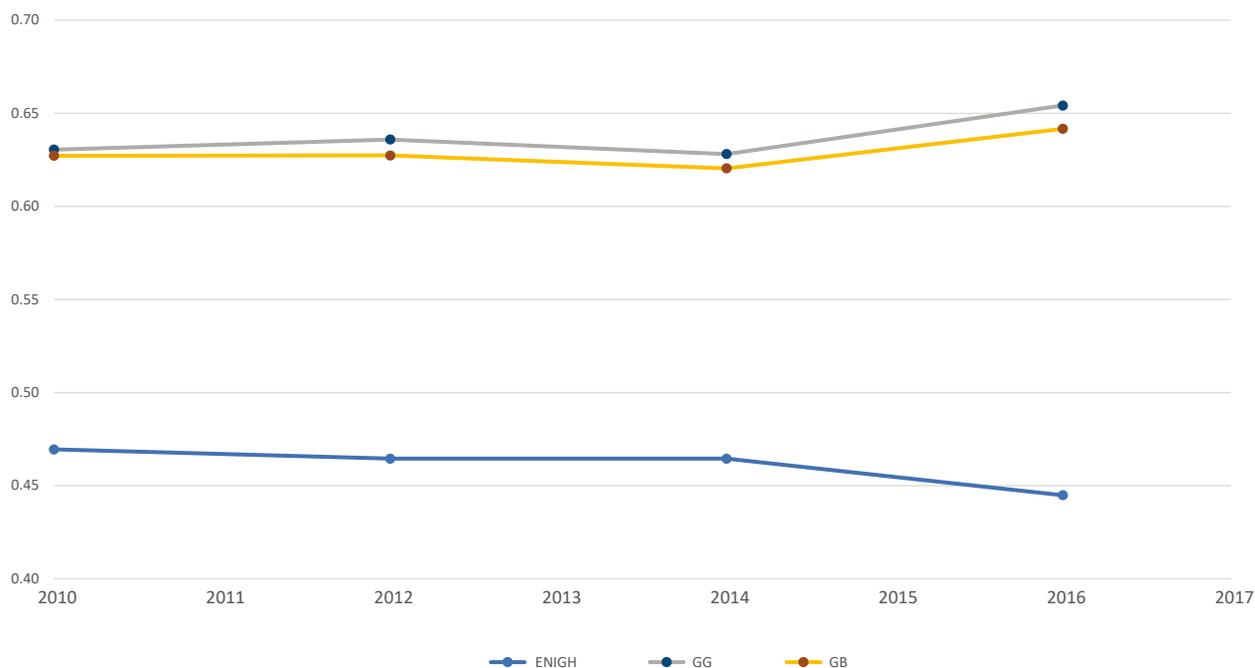
Evolución de la desigualdad en México, 2010-2016

En una primera instancia, concentraremos nuestra atención en el estudio de la desigualdad a lo largo de diversas dimensiones. La gráfica 6 muestra la evolución del coeficiente de Gini a partir tanto de las respectivas ENIGH como del ajuste de la distribución gama generalizada, imponiendo la restricción asociada al umbral un millonésimo, a lo largo del periodo de estudio. Como podrá observarse,

¹⁶ Por ejemplo, <https://www.inegi.org.mx/programas/mcs/2014/>

Gráfica 6

Coefficientes de Gini, México, 2010-2016



Nota: la gráfica muestra también resultados para el ajuste GB.

Fuente: cálculos propios a partir de bases de datos ENIGH-MCS, 2010-2016.

la brecha entre ambas líneas tiene una magnitud casi constante excepto, tal vez, para el último año. En efecto, en tanto, para la encuesta no hay una modificación importante del valor del coeficiente de Gini; para el modelo ajustado, la desigualdad exhibe un incremento que le lleva por vez primera durante el periodo por encima de 0.65.

La razón del total de los ingresos del decil X a los del primero contrasta los ingresos de los dos conjuntos de hogares, cada uno compuesto por 10 % del total, en ambos extremos de la distribución; es decir que, a diferencia del coeficiente de Gini, ignora los resultados para los deciles intermedios. La gráfica 7 parece indicar que, para la encuesta, las brechas entre los extremos han venido reduciéndose al quedar este valor por debajo de las 25 veces. Por su parte, los ajustes por MPVR, que inician con un decremento del primero al segundo año de los considerados, muestran que esta discrepancia ha venido creciendo hasta alcanzar el valor del mayor las 75 veces el valor del menor.

En otras palabras, los valores de las dos estadísticas consideradas, los cuales no se ven afectados

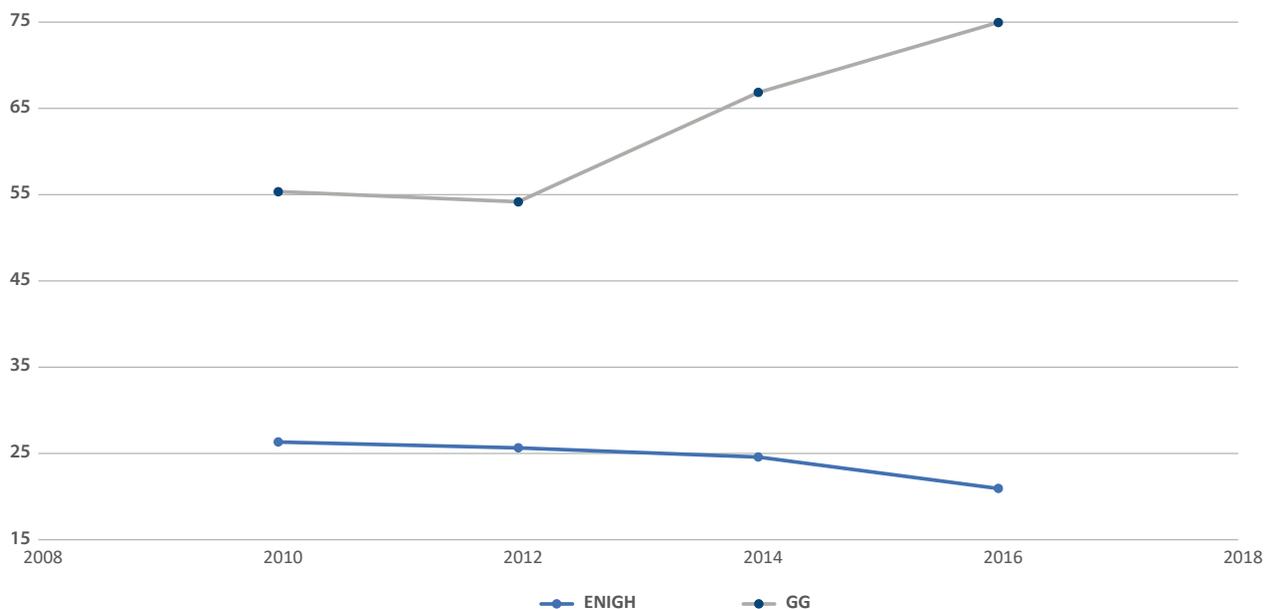
por la inflación, exhiben las aparentes ventajas de hacer un uso más eficiente de la información disponible. En efecto, las imágenes que pintan ambos conjuntos de resultados en nuestras mentes, así como las conclusiones que con fines de política pública es posible extraer de ellas, son contrastantes. Para el caso de los coeficientes de Gini, podría pensarse que los cambios introducidos en la metodología de la encuesta a partir del 2015 podrían haber dado como consecuencia un comportamiento anómalo para el 2016. Sin embargo, cuando consideramos el segundo conjunto de resultados, se puede concluir que la brecha entre ambos ejercicios ya había empezado a incrementarse desde el 2014; es decir, por lo menos un año antes de la introducción de modificaciones a la encuesta. Con base en lo anterior, resulta difícil ignorar la inequidad que se observa en ambos casos para el 2016.

La desigualdad entre entidades federativas, 2010-2016

Es también importante desagregar los resultados anteriores por lo menos hasta el nivel de entidad federativa. Con este fin, se recurrió a los MCS de

Gráfica 7

Razón del total de los ingresos del decil X a los del decil I, México, 2010-2016



Fuente: cálculos propios a partir de bases de datos ENIGH-MCS, 2010-2016.

la ENIGH para los mismos años atendiendo a su mayor tamaño de muestra estatal. A partir de esta información se ajustó un modelo para cada uno de los estados siguiendo el procedimiento anteriormente descrito. Para ello, en vista de la ausencia de información para el sector institucional de los hogares a nivel de entidad, se tornó necesario aproximar la participación estatal porcentual en el ingreso nacional para dicho sector. Aunque se dispone de la que resulta de la propia encuesta, en vista de la subestimación de los totales del ingreso, se decidió intentar el uso de otras fuentes.

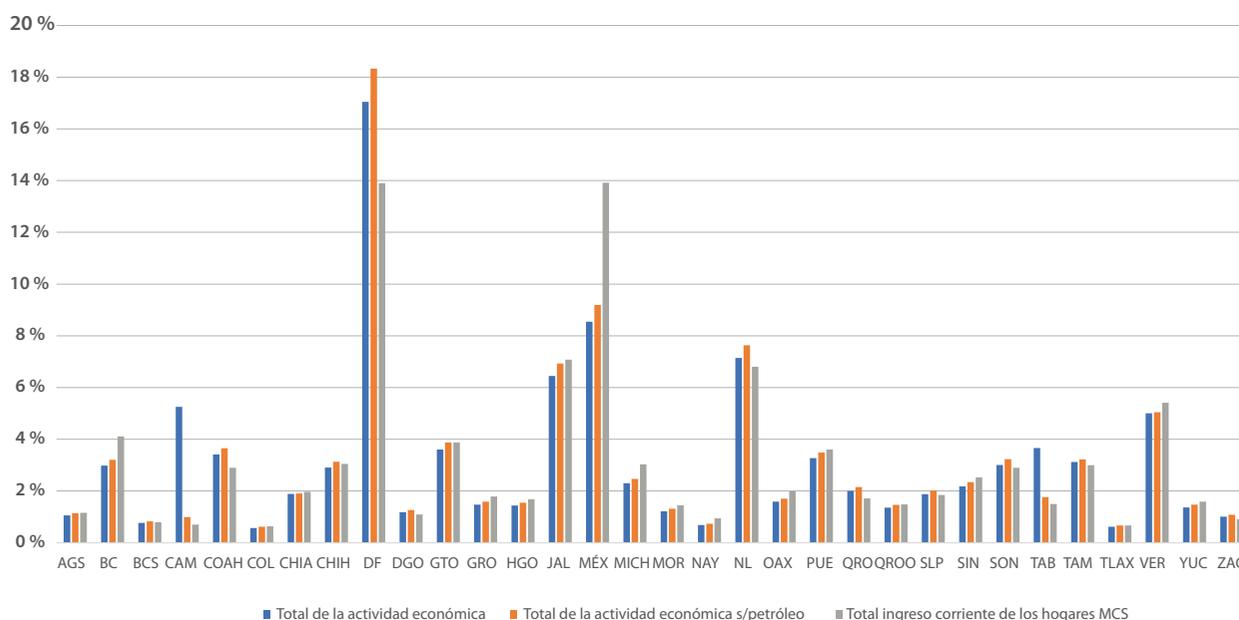
La primera opción fue hacerlo a través de la participación del Producto Interno Bruto Estatal (PIB-E) en el nacional para cada uno de los años, con base en el supuesto de que este y el ingreso total de los hogares por entidad estarían altamente correlacionados. Sin embargo, el PIB de algunas entidades muestra una fuerte influencia de la actividad de extracción de petróleo. Ya que el ingreso por dicha actividad se distribuye a lo largo de todo el territorio, la participación de su PIB-E sobreestima la del ingreso de los hogares en esos estados. Lo anterior nos condujo de manera natural a nuestra segunda opción, dada por el valor

del propio PIB-E, pero del cual se ha restado el valor de la minería petrolera. Como muestra la gráfica 8, la distribución porcentual obtenida de esta forma reduce la participación de las entidades petroleras que exhiben mayor concordancia con la que se obtiene para el ingreso a partir de la encuesta.

Sin embargo, son aún perceptibles dos discrepancias importantes, ahora para entidades no petroleras, según puede apreciarse de la gráfica 8, que ejemplifica para el 2010 la participación porcentual de los estados, tanto en el PIB nacional como en el que no contempla la minería petrolera, y en el ingreso nacional de los hogares. Se trata tanto de la Distrito Federal (hoy Ciudad de México o CDMX) como del estado de México, las que presentan una importante interacción mutua. En efecto, una muy importante parte de la Zona Metropolitana de la Ciudad de México se encuentra localizada en el estado de México. Los residentes en los municipios conurbados trabajan, estudian, compran y se divierten en gran medida en la CDMX, contribuyendo a aumentar el valor de su PIB. Sin embargo, su ingreso es declarado a la encuesta en el lugar donde tienen su residencia.

Gráfica 8

Participaciones estatales en PIB, PIB s/minería petrolera y en ingreso de los hogares, 2010



Fuente: cálculos propios a partir de SCNM y de bases de datos ENIGH-MCS, 2010-2016.

Por lo anterior, recurrir a la participación aportada por el PIB sin la minería petrolera sobreestima la contribución de la CDMX en el ingreso nacional de los hogares, mientras que la del estado de México sería subestimada. Con un PIB agregado que alcanza casi 25 % del PIB nacional, el error de aproximación cometido no podría ser ignorado. La gráfica 9 ilustra como la Ciudad de México y el estado de México distorsionan la participación del ingreso por su alta interdependencia.

El cuadro 2 muestra las sumas de cuadrados de las discrepancias entre la participación porcentual de las entidades en el ingreso nacional de los hogares según la encuesta y diversas cantidades obtenidas a partir de las cuentas estatales. La evidencia que ella aporta, así como la discusión precedente,

nos permite concluir que de entre las tres opciones exploradas, la que aproxima de mejor manera la participación del ingreso estatal de los hogares es la que no introduce sesgos obvios. De esta manera, y en vista de que no se cuenta con otras fuentes igualmente confiables para aproximar la deseada participación, regresamos a nuestra primera intuición distribuyendo de forma proporcional el total nacional de ingresos, según la participación relativa aportada por la propia ENIGH/MCS.

De esta manera se obtuvo la versión estatal de una de las dos restricciones que hemos venido utilizando. En contraste, la base de datos del SAT aporta la entidad en la que se localiza el domicilio fiscal del declarante, por lo que pudo ser usada sin cambio para la segunda de las restricciones

Gráfica 9

Diferencias entre participaciones estatales según MCS y PIB s/minería petrolera



Fuente: cálculos propios a partir de SCNM y de bases de datos ENIGH-MCS, 2010-2016.

Cuadro 2

Sumas de cuadrados de discrepancias entre participaciones porcentuales estatales en el ingreso nacional de los hogares, según ENIGH/MCS, y alternativas provenientes de las cuentas estatales

	2010	2012	2014	2016
PIB-E	0.68	0.77	0.56	0.53
PIB-E, sin minería petrolera	0.45	0.67	0.47	0.49
PIB-E, sin CDMX ni estado de México	0.29	0.23	0.20	0.16
PIB-E, sin minería petrolera, ni CDMX ni estado de México	0.03	0.04	0.04	0.04

mencionadas. Con base en las anteriores consideraciones, es posible obtener un conjunto de resultados a nivel de entidad federativa. En vista de los menores tamaños poblacionales involucrados, los resultados pueden ser fuertemente influenciados por la ocurrencia fortuita de eventos extraordinarios como, por ejemplo, ganar un premio en algún sorteo importante.

A partir de la gráfica 10 queda claro que, de acuerdo con los modelos ajustados, la desigualdad no se distribuye de manera uniforme a lo largo del territorio nacional. A pesar de la variabilidad que exhiben los valores ahí presentados, puede pensarse en la presencia de tres conjuntos de entidades. Al primero de ellos pertenecen aquellas cuyos valores se encuentran por arriba de 0.6 para todos los años considerados. Enseguida, tendríamos al conjunto para las cuales uno o más de los valores están por debajo de 0.6, en tanto que uno o más de ellos se encuentra por encima. Finalmente, se tienen aquellos estados cuyos coeficientes de Gini permanecen por debajo del mencionado valor.

Por otro lado, en tanto algunas de las entidades muestran comportamientos muy semejantes a lo largo del tiempo, otras más presentan una circunstancia cambiante. Vale la pena destacar que, en ninguno de los casos, el 2016 muestra las condiciones más extremas de desigualdad, contra lo que ocurre a nivel nacional. Lo anterior nos lleva a preguntar si es posible que al interior de los estados sea posible que tenga lugar una reducción de la desigualdad cuando, entre ellas, se da un incremento.

La gráfica 11 exhibe la tasa media de crecimiento del coeficiente de Gini, por entidad, para el periodo 2010-2016. Casi la mitad de las entidades muestran un crecimiento para los valores de dicho coeficiente, por lo que en ellas la desigualdad parece haber crecido. Es necesario precisar que un ritmo alto de crecimiento en la desigualdad puede deberse a un valor pequeño al inicio del periodo, como puede ser el caso para Colima y Morelos. Por otro lado, se debe destacar que dos de las entidades con más alta desigualdad en el 2010, Nuevo León y Distrito Federal (hoy Ciudad de México), muestran en

Gráfica 10

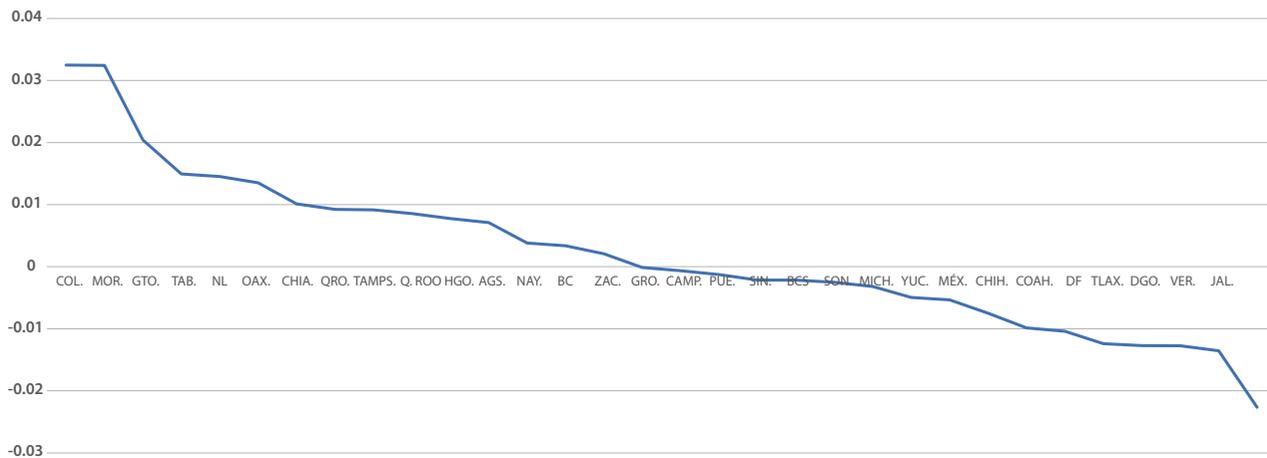
Coeficientes de Gini por entidad, 2010-2016



Fuente: cálculos propios a partir de bases de datos ENIGH-MCS, 2010-2016.

Gráfica 11

Media geométrica del crecimiento anual del coeficiente de Gini, por entidad, 2010-2016



Fuente: cálculos propios a partir de SCNM, de bases de datos ENIGH-MCS y SAT, 2010-2016.

cambio trayectorias divergentes; el valor del coeficiente creció para la primera, en tanto que para la segunda ocurrió lo opuesto. Son estos movimientos en sentidos opuestos para los estados los que nos llevan a pensar que la creciente desigualdad nacional pueda ser explicada más en términos del comportamiento de la desigualdad entre ellas que del que se presenta a su interior.

La anterior pregunta es también aplicable cuando se consideran las comparaciones entre los totales de ingresos para los dos deciles extremos. A partir de la gráfica 12 observamos que, ni para el 2014 ni para el 2016, tales cocientes son uniformemente mayores para todas las entidades. De hecho, es posible apreciar que, para este último año y para una alta proporción de los estados, los valo-

Gráfica 12

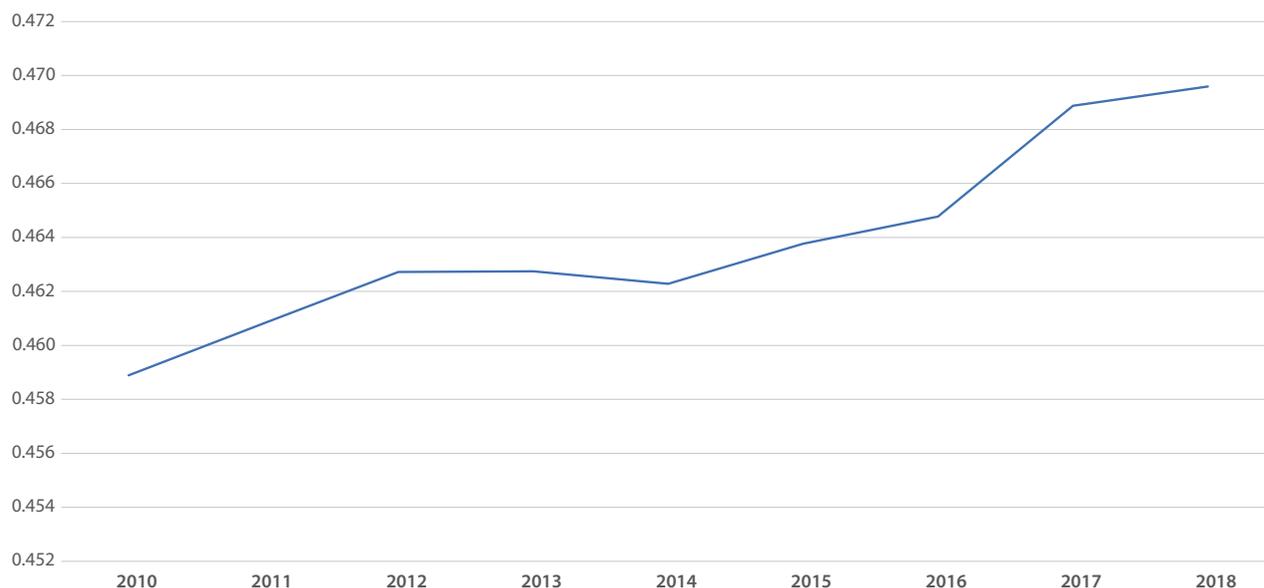
Cocientes entre promedios de ingreso, deciles X a I, por entidad, 2010-2016



Fuente: cálculos propios a partir de bases de datos ENIGH-MCS, 2010-2016.

Gráfica 13

Coeficientes de Gini calculados a partir de los PIB-E s/minería petrolera, 2010-2018



Fuente: cálculos propios a partir deSCNM, 2010-2018.

res de los cocientes resultan ser mínimos al compararse con los de todo el periodo. Lo anterior nos ha hecho sospechar que podríamos estar frente a una situación como la descrita en Lakner (2013), donde se afirma que "... la mayor parte de la desigualdad global es explicada por diferencias entre los países, aunque esta contribución ha declinado a lo largo del tiempo, sugiriendo que los países se han hecho más parecidos. La componente intra-país de la desigualdad global, sin embargo, ha crecido continuamente...".

Evidencia de trayectorias divergentes de la marginación municipal en México, entre 1990 y el 2010, ha sido documentada en Vargas y Cortés (2016), quienes afirman que sus resultados "... proporcionan evidencia clara de que la marginación en México ha sido *decreciente, pero con aumento en la heterogeneidad entre los municipios en los últimos 20 años*. Dicho de otro modo, si bien la marginación ha disminuido en las dos últimas décadas, el país ha experimentado, a nivel municipal, un proceso de divergencia y no de convergencia territorial...". Aplicado al caso de la desigualdad de ingresos, lo anterior permitiría responder positivamente las

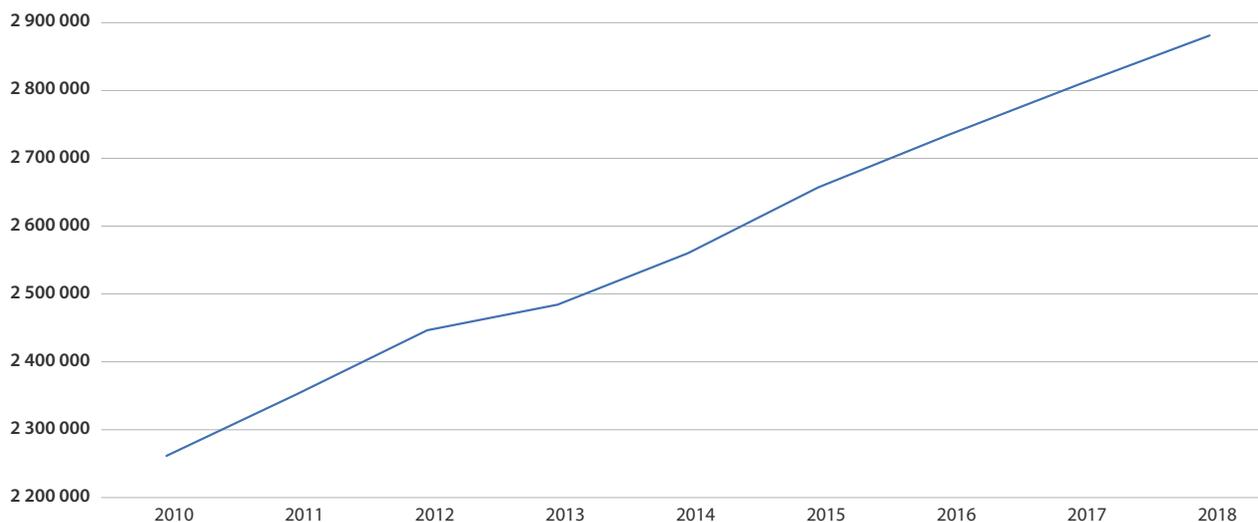
preguntas planteadas arriba afirmando que es posible que al interior de las entidades la desigualdad disminuya, cuando entre ellas crece.

De hecho, con la información publicada por el INEGI para el periodo, es posible calcular medidas de desigualdad a partir de los PIB estatales, sin la minería petrolera. Las gráficas 13 y 14 muestran la evolución del coeficiente de Gini y de la desviación estándar, calculados a partir de dichos PIB. En ambos casos es aparente una tendencia creciente entre el 2010 y 2018. En otras palabras, la desigualdad entre las entidades creció en el periodo, según el SCNM. Este incremento puede ayudar a resolver, así sea parcialmente, la aparente paradoja entre el aumento en la desigualdad de ingresos a nivel nacional, pero no al estatal.

Comentarios finales

Hemos hecho un breve recuento de la metodología MPVR y de algunos de sus resultados. En cuanto a esta, se destacan tres posibilidades con base en los resultados:

Gráfica 14

Desviaciones estándar calculadas a partir de los PIB-E s/minería petrolera, 2010-2018

Fuente: cálculos propios a partir de SCNM, 2010-2016

1. Los datos de la encuesta tienen un papel que jugar en cada ajuste, pero las restricciones impuestas reducen su importancia de modo que sus limitaciones también pierden peso.
2. Los modelos usados en el contexto de dicha metodología juegan también un papel menor.
3. La comparación entre distribuciones ajustadas de ingreso y estadísticas derivadas se lleva a cabo entre las mejores opciones en cada caso, dada la información disponible.

Por lo que toca a los resultados numéricos, aunque a nivel nacional la desigualdad empeoró entre el 2010 y 2016, salvo por algunas excepciones, al interior de las entidades federativas no se aprecia un deterioro importante. Ello nos ha llevado a postular que la desigualdad entre los estados es un factor importante que debe ser tomado en cuenta al intentar explicar su comportamiento a nivel nacional, a lo largo del tiempo. En este momento no disponemos de evidencia suficiente para rechazar dicha conjetura. Sin embargo, se presenta evidencia derivada del SCNM que apoyaría esta afirmación.

Para los resultados anteriores, se cuenta en el caso mexicano con muchos elementos comunes que permiten realizar comparaciones: diseño de muestra, tratamiento para valores faltantes del ingreso, percentiles similares de ingreso según el SAT, asignación del promedio estatal de ingresos según el SCNM y la ENIGH, por mencionar algunos. Los organismos internacionales tienen una tarea más complicada al intentar llevar a cabo comparaciones entre naciones a partir de resultados muestrales. Procedimientos como el sobremuestreo de hogares con ingresos altos o la calibración de resultados con totales exógenos han sido realizados por algunos países, pero no por todos. En consecuencia, se torna más importante cuando se desea llevar a cabo comparaciones entre naciones a lo largo del tiempo, contar con procedimientos estándar que reduzcan el peso de las diferencias, pero evitando el uso de procedimientos *ad hoc*.

Por lo que se refiere a trabajo futuro relacionado con la misma metodología, en este momento estamos explorando:

1. El uso de mezclas de distribuciones con lo que se aumenta el número de parámetros a

estimar y, en consecuencia, el de fuentes alternativas y de restricciones que es posible considerar en el ajuste.

2. Identificar con precisión a la población representada por la muestra mediante el uso de funciones truncadas.
3. Aproximar la proporción que la subdeclaración del ingreso representa de la discrepancia total, así como la que corresponde al truncamiento.
4. Imputar valores del ingreso para unidades muestrales en la base de datos de la encuesta para atender las necesidades de algunos usuarios.

Fuentes

- Altimir, O. "Income distribution statistics in Latin America and their reliability", en: *Review of Income and Wealth*. 33(2), 1987.
- Boghosian, B. M. "Is inequality inevitable?, originally published with the title 'The Inescapable Casino'", en: *Scientific American*. 321, 5, pp. 70-77, November 2019 (DE) <https://www.scientificamerican.com/article/is-inequality-inevitable/>
- Bustos, A. *Estimation of the distribution of income from survey data, adjusting for compatibility with other sources*. Workshop on Measuring Inequalities of Income and Wealth, High-Level Expert Group on the Measurement of Economic Performance and Social Progress. Berlin, Sep. 15-16, 2015.a (DE) <https://fr.slideshare.net/StatsCommunications/hleg-thematic-workshop-on-measuring-inequalities-of-income-and-wealth-alfredo-bustos>
- _____. "Estimation of the distribution of income from survey data, adjusting for compatibility with other sources", en: *Statistical Journal of the IAOS*. 31, pp. 565-577, 2015.b (DE) <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji923>
- Bustos, A. y G. Leyva. "Towards a More Realistic Estimate of the Income Distribution in Mexico", en: *Latin American Policy*. Volume 8, Issue 1, pp. 114-126, June 2017 (DE) <https://onlinelibrary.wiley.com/doi/abs/10.1111/lamp.12114>
- Campos Vázquez, R., E. Chávez Jiménez y G. Esquivel Hernández. *Los ingresos altos, la tributación óptima y la recaudación posible*. Primer lugar, Premio Nacional de Finanzas Públicas, 2014.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). *Metodología para la medición multidimensional de la pobreza en México*. Tercera edición. Ciudad de México, CONEVAL, 2019 (DE) <https://www.coneval.org.mx/Medicion/MP/Paginas/Metodologia.aspx>
- Cortés-Cáceres, F. "El cálculo de la pobreza y la desigualdad a partir de la Encuesta Nacional de Ingresos y Gastos de los Hogares", en: *Comercio Exterior*. Vol. 51, Núm.10, octubre. México, 2001.
- DESA. *Principles and Recommendations for Population and Housing Censuses, Revision 3*. New York, Statistics Division, Department of Economic and Social Affairs, United Nations, 2017 (DE) https://ec.europa.eu/eurostat/ramon/statmanuals/files/PR_on_pop_and_hous_cens_Rev3.pdf, consultado el 18 de junio de 2020.
- Economic Commission for Latin America and the Caribbean (ECLAC). "Income poverty measurement: updated methodology and results", en: *ECLAC Methodologies*. No. 2 (LC/PUB.2018/22-P), Santiago, 2019 (DE) https://repositorio.cepal.org/bitstream/handle/11362/44920/1/S1900554_en.pdf, consultado el 22 de junio de 2020.
- Félix, David. "Income distribution Trends in Mexico and the Kuznets Curves", en: Weinart, Richard y Cynthia Hewlett. *The Political Economy of Brasil and Mexico*. Philadelphia, ISMI Press, 1979.
- Fesseau, M. and P. van de Ven. "Measuring inequality in income and consumption in a national accounts framework", en: *OECD Statistics Brief*. No. 19, pp. 1-10, Nov. 2014.
- Instituto Nacional de Estadística y Geografía (INEGI). *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*. México, INEGI, 2014 (DE) <http://www.inegi.org.mx/est/contenidos/Proyectos/encuestas/hogares/regulares/enigh/presentacion.aspx>, consultado el 15 de junio de 2020.
- _____. *Nueva construcción de la ENIGH 2012. Tabulados*. México, INEGI, 2014 (DE) <http://www3.inegi.org.mx/Sistemas/TabuladosBasicos/tabdirecto.aspx?s=est&c=33501> consultado el 15 de junio de 2020.
- _____. *Sistema de Cuentas Nacionales de México*. Fuentes y metodologías. Año base 2013. Cuentas por Sectores Institucionales. México, INEGI, 2017 (DE) https://www.inegi.org.mx/contenidos/programas/si/2013/metodologias/SCNM_Metodo_CSI_B2013.pdf
- Lakner, Ch. y B. Milanovic. "Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession", en: *World Bank policy research paper*. 6719, 2013.
- Leyva-Parra, G. *El ajuste del ingreso de la ENIGH con la Contabilidad Nacional y la medición de la pobreza en México*. Serie Documentos de investigación. Núm. 19. México, SEDESOL, 2004.
- Martínez de Navarrete, Ifigenia. *La distribución del ingreso en México: tendencias y proyecciones a 1980*. Vol. 1. México, D.F., Siglo Veintiuno Editores, 1970.
- Piketty, T. *El capital en el siglo XXI*. México, D.F., Fondo de Cultura Económica, 2014.
- UNECE, Canberra Group. *Handbook on Household Income Statistics*. Second Edition. Geneva, United Nations, United Nations Economic Commission for Europe, 2011.
- Vargas, D. y Cortés, F. "Análisis de las trayectorias de la marginación municipal en México de 1990 a 2010", en: Cordera et al. (coords.). *Por un México social: contra la desigualdad*. México, Universidad Nacional Autónoma de México, 2016.

Toma de decisiones en ambientes de salud: **modelos de jerarquización analítica de alternativas para la vacunación en México**

Decision-making in Health Care Settings:
**Analytical Ranking Models for
Vaccination Alternatives in Mexico**

Mario Gutiérrez-Lagunes*

* Universidad Autónoma de San Luis Potosí, mario.gutierrez.lagunes@uaslp.mx

Development and manufacture process of a new vaccine. Final production of filled vials of Covid-19 vaccine. Bio science
3D illustration/ 1300036732 / iStock



El Proceso Analítico Jerárquico (AHP, por sus siglas en inglés) es un procedimiento de toma de decisiones utilizado en la asignación de prioridades en problemas de decisión multicriterio. Ante la urgencia de vacunación que se presenta en algunos países, la aplicación de la metodología del AHP aporta confianza en nuestras decisiones en términos de la teoría moderna de la medición. El objetivo de este trabajo es, aplicando el esquema del AHP, evaluar el uso y la pertinencia de vacunación masiva en un ambiente de urgencia por combatir el COVID-19. El análisis y la jerarquización de los criterios para vacunar y los estados de vulnerabilidad con respecto a la salud con estas variables se utilizaron para crear matrices de comparación pareadas. La razón de consistencia entre las decisiones pareadas fue de 0.0114; se concluye que la construcción de estas matrices para analizar y evaluar objetivamente la jerarquización presentada fue consistente.

Palabras clave: proceso analítico jerárquico; teoría de decisiones; matriz de comparación; COVID-19.

Clasificación JEL: C44, C44, C6, I1.

Recibido: 30 de abril de 2021.

Aceptado: 18 de febrero de 2022.

Introducción

A raíz de la crisis provocada por la pandemia a inicios del 2020, y ante la limitada y retardada llegada de la vacuna contra el COVID-19 para la mayoría de los países con el fin de procurar la salud de sus habitantes lo antes posible, casi todos los gobiernos se han encontrado en apremio de mitigar esta contingencia que no cesa. De acuerdo con el informe publicado por Oxfam Novib (OXFAM, 2021), se tiene que la desigualdad entre las naciones es más notoria a nivel global, donde se indica que para el selecto mundo de los mil hombres más ricos del mundo les lleva solo nueve meses recuperar su riqueza previa a la pandemia, en cambio, a las personas más marginadas les llevará, probablemente, más de 10 años superar las consecuencias económicas propias de esta emergencia sanitaria, por lo que la aparición de nuevos pobres es un efecto de esta crisis, siendo las mujeres y los grupos minoritarios los más perjudicados.

The Analytic Hierarchical Process (AHP) is a decision-making procedure used in assigning priorities in multi-criteria decision problems. Given the urgency of vaccination in some countries, the application of the AHP methodology provides confidence in our decisions in terms of modern measurement theory. The objective of this work is, by applying the AHP scheme, to evaluate the use and relevance of mass vaccination in an environment of urgency to combat COVID-19. The analysis and ranking of criteria for vaccination and health vulnerability states with these variables were used to create paired comparison matrices. The consistency ratio between the paired decisions was 0.0114; it is concluded that the construction of these matrices to objectively analyze and evaluate the ranking presented was consistent.

Key words: hierarchical analytical process; decision theory; comparison matrix; COVID-19.

Por otro lado, el esfuerzo del gobierno mexicano para garantizar el acceso a las vacunas con las principales farmacéuticas mundiales —como Pfizer (Estados Unidos de América-Alemania), Astra Zeneca (Gran Bretaña) y CanSino (China), entre otras— se ha visto limitado por la escasez de estas debido, principalmente, por el acaparamiento por parte de los países más ricos, por lo que ha ampliado su estrategia de la obtención de las vacunas a la brevedad posible.

Para lograr existencias que cubran de forma gratuita a toda la población mexicana, el gobierno tuvo que firmar acuerdos de confidencialidad con las farmacéuticas. Desde el inicio de la vacunación (el 24 de diciembre de 2020), hasta el 31 de octubre de 2021, se han aplicado poco más de 126 millones de vacunas, contando con el esquema completo de vacunación contra el COVID-19 81 % de la población (Gobierno de México, 2021). El total de dosis recibidas ha sido de 153 280 045,

iniciando por la población que está en el sector salud y, después, por la más vulnerable, enfocándose en los adultos de 60 años y más de edad. Todavía falta mucho por atender para que se logre la inmunidad de este sector, pero la mayor disponibilidad de personal sanitario, ya vacunados, hace pensar que los enfermos están mejor atendidos hoy en comparación con hace algunos meses.

El objetivo de este estudio empírico es evaluar, bajo el Proceso Analítico Jerárquico (AHP, por sus siglas en inglés), el uso y la pertinencia de vacunación masiva en un ambiente de urgencia para combatir el COVID-19, calculando un vector de jerarquización de las decisiones a realizar sujeto a un índice de consistencia.

La estructura del trabajo es como sigue: primero se presenta una introducción del panorama de la crisis por la escasez de vacunas para combatir el COVID-19 en el mundo, así como la situación en México; enseguida se exponen los criterios, los principios y la vulnerabilidad que se tiene frente al virus SARS-CoV-2 de acuerdo con los Centros para el Control y la Prevención de Enfermedades (CDC, por sus siglas en inglés, 2021) y el grupo técnico asesor para la vacuna en México, así como la distribución de la población de adultos mayores de 60 años y más en el país; después, se explica la metodología del AHP para su aplicación en este trabajo; posteriormente, se analizan los resultados obtenidos de la metodología propuesta en la aplicación de la vacuna; y, por último, se muestran las conclusiones de los hechos más importantes encontrados en esta investigación enfocada a la toma de decisiones resultante.

Criterios

El proceso de desarrollo de la vacuna contra COVID-19 es arduo y complejo: pasar por fases de ensayos clínicos y ser aprobada por instancias nacionales e internacionales para garantizar su eficacia, seguridad y efectividad. Por otra parte, la estrategia de vacunación enfrenta otros retos, como el manejo, la distribución y la administración de las vacunas, ante la escasez de estas.

Para ello, la Organización Mundial de la Salud (OMS) ha establecido un marco de valores del Grupo de Expertos en Asesoramiento Estratégico (SAGE, por sus siglas en inglés) sobre Inmunización (WHO, 2020). Los seis principios que guían a que las vacunas contra el COVID-19 deben ser un bien público mundial son: bienestar humano, respeto igualitario, equidad mundial y nacional, reciprocidad y legitimidad. Asimismo, tiene como objetivo ofrecer orientación a nivel mundial acerca de la asignación de las vacunas contra el COVID-19 entre las naciones, así como orientar a nivel nacional acerca de la determinación de los grupos prioritarios para recibirlas dentro de los países, en especial cuando el suministro es limitado. Cabe hacer mención que este marco solo está diseñado para abordar los aspectos éticos relacionados con la asignación y priorización de estas vacunas, dejando fuera del alcance el diseño de los ensayos y el proceso regulatorio.

De acuerdo con la Política Rectora de Vacunación (PRV) contra el COVID-19 (Cortés *et al.*, 2021), el Grupo Técnico Asesor para la Vacuna (GTAV) COVID-19 en México (formado por expertos en materia de inmunología, vacunación, infectología, sociología, sistemas y economía de la salud) ha recomendado llevar a cabo una estrategia enfocada en reducir el número de muertes asociadas con la enfermedad, diseñada con base en la mortalidad observada en México y apoyada por el Instituto Nacional de Salud Pública, el cual realizó un análisis de la información nacional e internacional, así como la revisión de diversos estudios relacionados con los riesgos de salud, pobreza e inequidad con respecto a la infección y muerte por COVID-19.

Así, acorde con las recomendaciones del GTAV, los cuatro ejes de priorización para la vacunación contra el virus SARS-CoV-2 en México se pueden apreciar en el cuadro 1.

Una vez establecidas estas prioridades de vacunación, anunciadas en la PRV por Cortés *et al.* (2021), se espera reducir el número de personas fallecidas por COVID-19 en México. El eje de priorización 1, *Edad de las personas*, es el principal, con énfasis en vacunar prioritariamente dentro de cada

Cuadro 1

Criterios

a)	Edad de las personas.
b)	Comorbilidades personales.
c)	Grupos de atención prioritaria.
d)	Comportamiento de la epidemia.

Fuente: Política Rectora de Vacunación contra el COVID-19.

grupo etario a quienes viven con comorbilidades. En particular, algunas de las que padecen aquellos que vayan a recibir la vacuna tienen un componente personal de control, ya que muchas de las personas que viven con diabetes, hipertensión o VIH tienen sus condiciones controladas, las cuales ayudan a reducir el riesgo específico. En cuanto a los grupos de atención prioritaria, como individuos con discapacidad y la niñez, entre otros, todavía no cuentan con una autorización de uso en emergencia a nivel mundial. Con respecto al eje de priorización de la vacunación, el comportamiento de la epidemia es dinámico, donde cada entidad federativa tiene diversas epidemias hacia el interior de su territorio, además de que estas afecciones locales dependen también de diferentes factores, el más importante epidemiológicamente hablando es la densidad poblacional.

Vulnerabilidad

De acuerdo con los CDC (2021), los adultos mayores y los grupos de cualquier edad con afecciones subyacentes graves son las personas más vulnerables y con un mayor riesgo de enfermarse gravemente a causa de COVID-19, sin embargo, existen más grupos de individuos que también lo son a este virus.

Los expertos en salud han hecho una lista de las personas más vulnerables ante COVID-19 y que, por lo tanto, deben tener mayores cuidados para reducir el riesgo de presentar complicaciones graves. Las presentadas en la PRV se observan en el cuadro 2.

Con las características asociadas con un mayor riesgo de morir por COVID-19, se establece la es-

Cuadro 2

Grupos priorizados en la orden de prelación

1.	Personal sanitario que enfrenta COVID-19.
2.	Personas adultas mayores.
3.	Personas con comorbilidad o comorbilidades.
4.	Personal docente de las entidades federativas en semáforo epidemiológico verde.
5.	Resto de la población (mayor a 16 años de edad).

Fuente: Política Rectora de Vacunación contra el COVID-19

trategia de contención de defunciones donde se propone una priorización de grupos poblacionales a vacunar. Con base en los datos del Sistema Nacional de Vigilancia Epidemiológica, el GTAV hizo un análisis de mortalidad por grupo de edad en México, donde se encontró que los de 65 a 69 y 60 a 64 años eran los que alcanzaban los mayores porcentajes de mortalidad y los de 0 a 19 y 20 a 24 eran los que presentaban los menores valores. Con los resultados de este análisis, los grupos priorizados en la orden de prelación, establecidos por el GTAV en la PRV, más detallado son:

1. Personal sanitario que enfrenta COVID-19.
2. Personas adultas mayores:
 - a. Mayores de 80 años.
 - b. de 70 a 79.
 - c. de 60 a 69.
3. Personas con comorbilidad o comorbilidades:
 - a. Obesidad mórbida.
 - b. Diabetes mellitus.
 - c. Hipertensión arterial sistémica.
 - d. Enfermedad pulmonar obstructiva crónica (EPOC).
 - e. Asma.
 - f. Enfermedades cerebrovasculares.
 - g. Infección por VIH.
 - h. Enfermedad renal crónica.
 - i. Estados patológicos que requieren de inmunosupresión.
 - j. Cáncer en tratamiento.
4. Personal docente de las entidades federativas en semáforo epidemiológico verde.
5. Resto de la población (mayor a 16 años de edad).

Aunque también existen otras poblaciones vulnerables a contraer COVID-19, como las mujeres embarazadas, personas sin hogar y con discapacidades, entre otras. Cabe mencionar que, por el momento, ninguna vacuna será aplicada a individuos menores de 16 años, hasta que se cuente con la suficiente evidencia de seguridad en esta población; actualmente, ningún ensayo clínico ha incluido a menores de edad y mujeres embarazadas.

Bajo este contexto, y la escasez de vacunas, las pocas que llegan, ¿a quiénes se les aplican?

Para contestar esta pregunta, se estableció la estrategia sugerida por el GTAV en la Política Rectora de Vacunación contra el COVID-19 a la población mexicana, en sus diferentes entidades, ubicando a la población de 60 años y más de edad.

Población de México

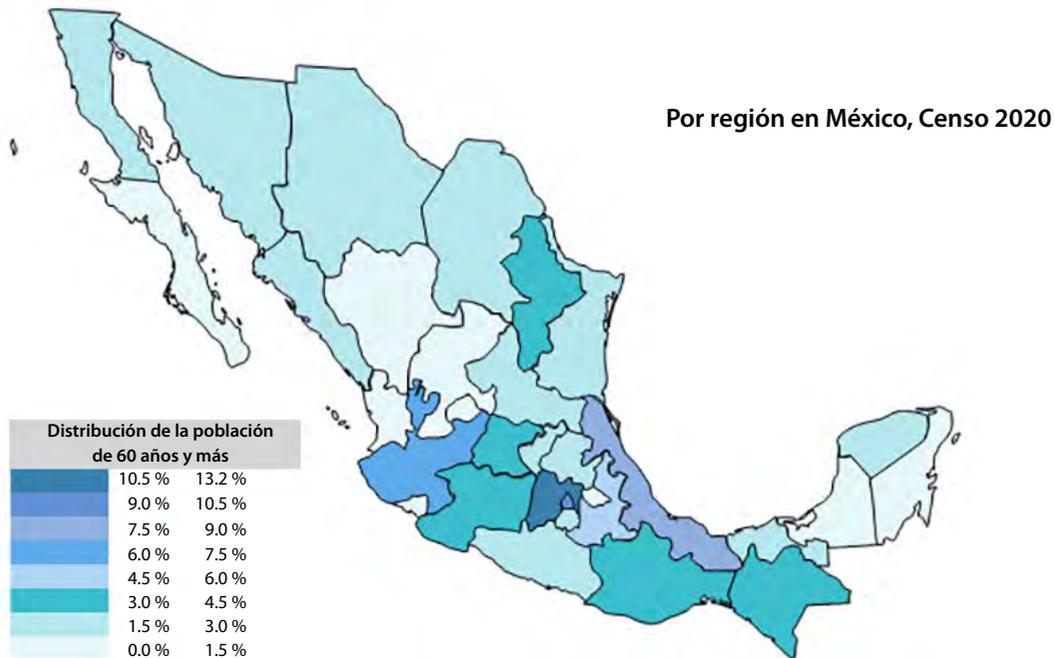
De acuerdo con el Censo de Población y Vivienda 2020 (INEGI, 2020), el total de habitantes de la República Mexicana es de 126 013 024, de los cuales 15 142 976 son personas de 60 años y más, es decir, 12 % de la población pertenece a este rango etario. Las cinco entidades donde se concentra cerca de 42 % de este grupo de edad son: México (12.7 %), la Ciudad de México (9.9 %), Veracruz de Ignacio de la Llave (7.6 %), Jalisco (6.6 %) y Puebla (4.9 %). En contraste, las cinco con menos adultos mayores son: Aguascalientes (0.96 %), Quintana Roo (0.87 %), Campeche (0.68 %), Colima (0.60 %) y Baja California Sur (0.50 %), donde entre todas suman 3.6 % de dicha población (ver mapa).

Teniendo este panorama de la distribución de los adultos de 60 años y más en la República Mexicana, y ante la urgencia de vacunar a lo más importante que posee un país, que es su gente, aunado a la escasez de vacunas, la pregunta es: ¿a quién vacunamos primero de este segmento de la población?

Por otra parte, cabe señalar que 75.7 % de la población mexicana cuenta con asistencia de un seguro médico, donde Chihuahua alcanza una cobertura de 84.4 % y Michoacán de Ocampo es el

Mapa

Distribución de la población de 60 años y más de edad en México



Fuente: elaboración propia con datos del INEGI.

más rezagado en atención de servicios médicos, tiene 62.2 % (INEGI, 2020), por lo que la situación se vuelve más crítica por no contar con la infraestructura suficiente para una emergencia de salud como la que se ha presentado en el país y poder dar solución a la vacunación en un tiempo mínimo.

Marco teórico

El uso de la toma de decisiones se ha vuelto más frecuente debido al empleo de sistemas computacionales que permiten analizar los diferentes recursos tanto organizacionales como humanos, los cuales buscan alcanzar esquemas de optimización. Lo anterior puede ser modelado a través de representaciones formales que clarifican las relaciones entre necesidades, recursos disponibles y objetivos establecidos para alcanzar una meta. Las matemáticas, dentro de la teoría de decisiones, agrupa los criterios y decisiones que, mediante su análisis y procesamiento —cumpliendo ciertas reglas de comportamiento establecido de antemano—, alcancen el óptimo, esto es, las resoluciones más adecuadas, dando consistencia, transparencia y robustez a todo el proceso desde su fase inicial hasta la final del problema.

La aplicación de sistemas analíticos para la toma de decisiones es ampliamente utilizada por instituciones de diferente índole, desde organismos gubernamentales hasta empresas y consorcios, pues provee señales para anticipar soluciones a problemas que amenazan su supervivencia en el mercado. Puede decirse que, de manera general, los modelos de mayor aplicación comprenden una combinación de aspectos tanto de naturaleza normativa como prospectiva. Las representaciones lógicas y simbólicas entre los elementos de un sistema de decisiones emplean recursos semánticos que describen las relaciones entre ellos de forma integral. En consecuencia, la modelación matemática como recurso de comprensión y comunicación ha merecido una creciente atención debido a la diversidad de situaciones que se presentan en un mundo dinámico y, a la vez, complejo, como el actual.

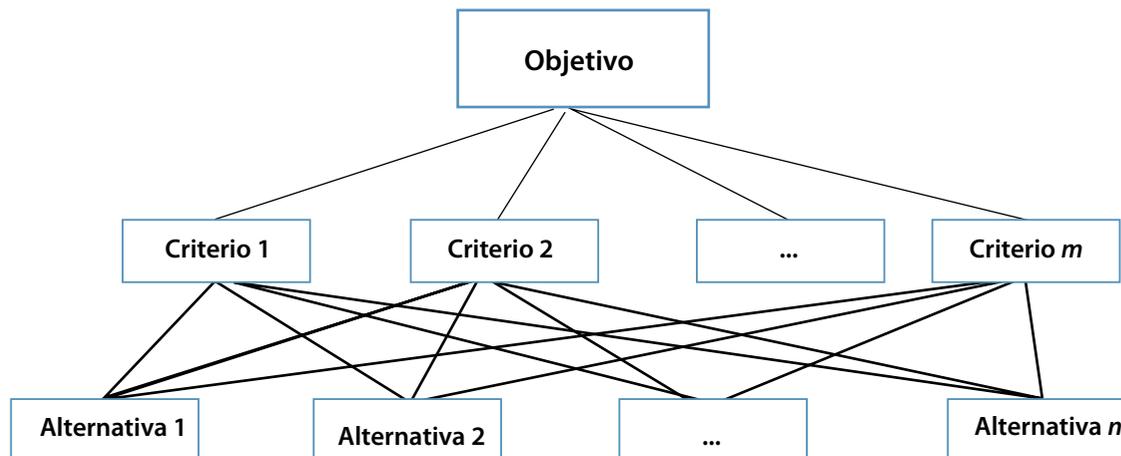
Por lo que respecta de la aplicación del AHP enfocado a la salud, Niño *et al.* (2020) implementaron un modelo de evaluación multicriterio en un municipio de Colombia, identificando las áreas de mayor probabilidad de transmisión del SARS-CoV-2. Por otra parte, Schmidt *et al.* (2015) realizaron una revisión sistemática de la literatura mediante búsquedas en las bases de datos *Pubmed* y *Web of Science* para artículos con las palabras clave en sus títulos o resúmenes: *Proceso Analítico Jerárquico*, *Análisis de decisiones de criterios múltiples* o similares, además de analizar si los autores informaron aspectos importantes y en sus estudios resultantes, afirmando que las investigaciones que aplican AHP en la salud han aumentado.

AHP

La jerarquización analítica surge a principios de la década de los 80 (T. Saaty, 1980) como un recurso metodológico que permite combinar información de carácter cualitativo, con criterios de evaluación cuantitativa. Para construir una decisión, el AHP se basa en la definición de criterios, así como en la identificación de alternativas y en la evaluación de estas con respecto a cada criterio. La teoría de procesos de jerarquización analítica, debido a su capacidad para conjuntar objetivos, criterios y diversas alternativas, es ampliamente recomendable aplicarla en la toma de decisiones bajo condiciones de incertidumbre (ver diagrama).

En el AHP, la experiencia del tomador de decisiones es esencial, ya que sus juicios subjetivos se combinan con la información cualitativa y cuantitativa disponible, dando lugar a la determinación de una priorización fundamentada y consistente para las alternativas propuestas. Así, esta información disponible referente a los criterios, ya sea cualitativa o cuantitativa, se adiciona con los juicios valorativos basados en el conocimiento y experiencia del tomador de decisiones, dando como resultado la relevancia que impacta a cada uno de los criterios establecidos en los términos relativos del conjunto.

Diagrama

Representación de un objetivo con criterios y alternativas múltiples

Fuente: elaboración propia.

Bajo este contexto, cabe mencionar que no existe un acuerdo unánime para elegir el mejor método a implementar en alguna situación en particular, sin embargo, está la aprobación de que esta decisión de juicio debe ser tomada por especialistas en los temas de discusión, y en conjunto con las áreas involucradas. Con respecto al grupo de expertos, existe basta literatura acerca de la importancia en la toma de decisiones, entre ellos se tiene a De Arquer (1996), Petropoulos *et al.* (2018) y Werner *et al.* (2017). Sobre el método Delphi, Linstone & Turoff (1975) y Bolger & Wright (2007) hacen notar la importancia de la experticia de los tomadores de decisiones en situaciones críticas. Un ejemplo del empleo de este método en salud pública es el realizado por el Grupo Síntevi, el cual es un consorcio dedicado a la metodología de la investigación, modelado y análisis de la evidencia científica en México. En su investigación, se describe cómo implementar el método Delphi convencional en tiempos de confinamiento, y se analizan su utilidad y limitaciones para su uso por expertos en salud para la resolución de problemas de tratamiento, diagnóstico o administrativos (Síntevi, 2020).

Cabe mencionar que la teoría del AHP ha tomado mucha atención debido al avance de la tecnología, la cual ha permitido el desarrollo

de aplicaciones computacionales en los procesos operativos en el campo del álgebra matricial, llevando a cabo la simulación de escenarios, el análisis combinatorio, así como la aplicación de métodos numéricos sofisticados para el análisis de sensibilidad.

Metodología de aplicación del AHP

Los principios teóricos que sustentan la metodología del AHP fueron desarrollados hacia las últimas décadas del siglo XX. Fundamentalmente, se parte de las ramas de las matemáticas que tienen mayor relación con el análisis vectorial y álgebra matricial. Thomas Saaty lo creó y desarrolló en diferentes etapas (1977 y 1980), y su fundamentación matemática, en 1986, la cual se basa en estos cuatro axiomas:

Axioma 1. Referente a la condición de los juicios recíprocos. Si A es una matriz de comparaciones pareadas, se cumple que $a_{ij} = 1/a_{ji}$.

Axioma 2. Condición de homogeneidad de los elementos. Los que se comparan son del mismo orden de magnitud o jerarquía.

Axioma 3. Condición de estructura jerárquica o estructura dependiente. Existe dependencia jerárquica en los elementos de dos niveles consecutivos.

Axioma 4. Condición de expectativas de orden de rango. Las expectativas deben estar representadas en la estructura en términos de criterios y alternativas.

Formalmente, se tiene que si $A = \{A_i\}$ donde A_i son las alternativas, $i = 1, \dots, n$.

Una escala para el criterio c es un mapeo P_c , tal que:

$$P_c : A_i \times A_j \rightarrow a_{ij}$$

donde A_i y A_j son las alternativas comparadas bajo el criterio c .

Además, se tiene:

$$a_{ij} > 1 \text{ si y solo si } A_i \text{ domina a } A_j$$

$$a_{ij} = 1 \text{ si y solo si } A_i \text{ es equivalente a } A_j$$

Operacionalmente, los a_{ij} son obtenidos de las asignaciones de los tomadores de decisiones para los datos de la matriz $A = [a_{ij}]$, donde $a_{ij} > 0 \forall i, j$.

Este proceso de jerarquización evalúa (pesa) los diferentes criterios, subcriterios y alternativas en función de su importancia correspondiente en cada nivel, por lo que la técnica AHP está basada en la suposición de que el analista (decisor) puede, de manera más sencilla, elegir un valor de comparación que uno absoluto. Los juicios verbales son trasladados a una escala de puntuación (ver cuadro 3).

Cuadro 3 Continúa
Escala comparativa de T. Saaty (1980)

Juicios	Puntuación
Igual	1
	2
Moderado	3
	4

Cuadro 3 Concluye
Escala comparativa de T. Saaty (1980)

Juicios	Puntuación
Fuerte	5
	6
Muy fuerte	7
	8
Extremadamente fuerte	9

Donde los valores intermedios 2, 4, 6 y 8 son usados como valores de consenso entre dos juicios.

De acuerdo con los axiomas mencionados anteriormente, se construyen el conjunto de matrices de comparación por pares, tanto la de matriz de juicios como las matrices de alternativas, las cuales presentan la siguiente forma:

$$A = \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{bmatrix}$$

Ya obtenidas las matrices de comparación por pares, se procede a calcular las prioridades de cada criterio y/o alternativa comparada.

El proceso matemático de calcular las prioridades se llama síntesis, y en él se calculan los valores y vectores característicos de las matrices de comparación, y después se normalizan.

Sea P la matriz de prioridades de cada criterio (juicio) en función de la meta global:

	Meta global
Criterio 1	P'_1
Criterio 2	P'_2
...	...
Criterio m	P'_m

donde:

m es el número de criterios.

P'_i es la prioridad del criterio i con respecto a la meta global.

Cuadro 4

Matriz de prioridades P'

	Criterio 1		Criterio 2	...	Criterio m
Alternativa 1	P_{11}		P_{12}	...	P_{1m}
Alternativa 2	P_{21}		P_{22}	...	P_{2m}
...	
Alternativa n	P_{n1}		P_{n2}	...	P_{nm}

Donde P_{ij} es la prioridad de la alternativa i con respecto al criterio j .

La matriz de prioridades, P' , que también expresa las prioridades de cada alternativa, en función de cada criterio, se presenta en el cuadro 4.

Finalmente, se tiene que la matriz de prioridad global se obtiene como el producto de P' y P (ver cuadro 5).

Cuadro 5

Matriz de prioridad global

$$\begin{bmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nm} \end{bmatrix} \begin{bmatrix} P'_{11} \\ P'_{21} \\ \dots \\ P'_{m1} \end{bmatrix} = \begin{bmatrix} P_{g1} \\ P_{g2} \\ \dots \\ P_{gn} \end{bmatrix}$$

donde P_{ij} es la prioridad global con respecto a la meta global de la alternativa $i, i = 1, \dots, n$.

Resulta una matriz ($n \times 1$) que indica la ponderación total de cada activo en función de todas las variables explicativas, hayan sido estas cualitativas o cuantitativas.

El resultado de esta multiplicación es el vector propio asociado a la matriz de prioridades global, es decir:

$$\sum_{i=1}^n P_{gi} = 1$$

donde:

$$\vec{V}_p = [P_{g1}, P_{g2}, \dots, P_{gn}]^T$$

El cálculo del vector propio de la matriz de comparación refleja el peso o la importancia relativa que tiene cada componente (criterio o alternativa) cuando son evaluados en su conjunto bajo cierto supuesto. Su cálculo parte de la matriz original A , donde un vector v , distinto de cero, es un vector propio de la matriz A , si para cierto escalar λ se cumple que:

$$Av = \lambda v$$

donde λ es el valor propio de la matriz A asociado al vector propio v . Con este cálculo del vector propio se obtiene la ponderación de cada una de las componentes participativas (juicios o alternativas) que son utilizadas para determinar el impacto (peso) que tiene cada una de ellas.

En cada matriz de comparación es necesario diseñar un método de priorización que permita cuantificar de forma consistente el peso de cada criterio a ser analizado (Wind & Saaty, 1980). Esta técnica AHP permite al tomador de decisiones evaluar la congruencia de los juicios con la *ratio* de inconsistencia (IR). Para ello, se estima el Índice de Consistencia (CI) de una matriz $n \times n$ de juicios, donde CI viene definido por:

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

donde λ_{max} es el máximo valor propio de la matriz de criterios.

Asimismo, se calcula el I_R como:

$$I_R = \frac{CI}{RI}$$

donde RI es el valor aleatorio promedio de CI para una matriz de $n \times n$. Los valores de RI obtenidos por R. Saaty (1987) se aprecian en el cuadro 6.

Cuadro 6
Valores de RI para matrices de diferentes órdenes (R. Saaty, 1987)

N	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Dicho I_R deberá ser cotejado con los que aparecen a continuación, y se considerará que existe consistencia cuando no se superen los valores que se muestran en el cuadro 7.

Cuadro 7
Ratio de consistencia máximo

Tamaño de la matriz	Ratio de consistencia
3	5 %
4	9 %
5 o mayor	10 %

Resultados

Para iniciar el AHP, se presenta la comparación asignada entre pares de la matriz de criterios, de acuerdo con los juicios establecidos por el GTAV acerca de los cuatro ejes de priorización para la vacunación contra el virus SARS-CoV-2 en México (ver cuadro 8).

Cuadro 8
Matriz de criterios

	Criterios			
	a	b	c	d
a	1	2	3	4
b	1/2	1	2	3
c	1/3	1/2	1	2
d	1/4	1/3	1/2	1

Fuente: elaboración propia con base a los criterios del GTAV.

El cuadro se interpreta así, por ejemplo, entre los criterios: el a (*Edad de las personas*) se valora dos veces más que el b (*Comorbilidades personales*), el

a (*Edad de las personas*) se valora tres veces más que el c (*Grupos de atención prioritaria*), el a (*Edad de las personas*) se valora cuatro veces más que el d (*Comportamiento de la epidemia*), y así sucesivamente, hasta comparar todos los criterios entre pares y obtener la matriz de comparación entre pares final, cuyo vector propio correspondiente es:

$$\vec{V}_p = [0.467, 0.277, 0.160, 0.095]$$

donde, de acuerdo con el orden del cuadro 1, destaca la preferencia correspondiente a *Edad de las personas* (0.467 máximo valor de las coordenadas del vector) y la menos preferente es la referente a *Comportamiento de la epidemia* (0.095), lo cual es congruente con los pesos establecidos en la matriz de criterios.

Asimismo, para la construcción de las matrices correspondientes a los grupos priorizados en el orden de prelación (cuadro 2), que van a ser identificadas como las alternativas que se tienen, las cuales se comparan entre pares entre ellas dado cada criterio. Así, para el primero, en el cuadro 9 se muestra una matriz que refleja la asignación de preferencias para cada par de alternativas.

Cuadro 9
Matriz de comparación con respecto al criterio 1

	Criterio 1: <i>Edad de las personas</i>				
	1	2	3	4	5
1	1	2	3	4	5
2	1/2	1	2	3	4
3	1/3	1/2	1	2	3
4	1/4	1/3	1/2	1	2
5	1/5	1/4	1/3	1/2	1

Fuente: elaboración propia con datos del GTAV.

El cuadro se interpreta así: bajo el supuesto del criterio 1 (*Edad de las personas*), la alternativa 1 (*Personal sanitario que enfrenta COVID-19*) se valora dos veces más que la 2 (*Personas adultas mayores*); la 3 (*Personas con comorbilidad o comorbilidades*) se valora dos veces que la alternativa 4 (*Personal docente de las entidades federativas*), y así

sucesivamente, hasta comparar todos los criterios entre pares y obtener la matriz final de comparación mostrada entre pares.

Dada esta matriz del criterio 1, su vector propio correspondiente es:

$$\vec{VC}_1 = [0.419, 0.263, 0.160, 0.097, 0.062]$$

Bajo el criterio 1 (*Edad de las personas*), y de acuerdo con el cuadro 2 de los grupos priorizados, la preferencia de las alternativas se va hacia *Personal sanitario que enfrenta COVID-19* (0.419 máximo valor de las coordenadas del vector) y la última preferencia es hacia *Resto de la población (mayor a 16 años de edad)* (0.062 mínimo valor de las coordenadas del vector), lo cual también es congruente con lo asignado en la priorización.

De manera similar, se construyen las otras matrices bajo el supuesto de los otros criterios.

La matriz P'_g de prioridades global resultante es la mejor aproximación de evaluación de los criterios analizados (T. Saaty, 1980). Este vector de prioridad es:

$$\vec{P}_g = [0.419, 0.263, 0.160, 0.097, 0.062]$$

Nótese que el vector resultante fue igual al resultante del vector de criterios, ya que no variaron en ninguna de las matrices los pesos asignados, por lo que, de acuerdo con el cuadro 2 (grupos priorizados en la orden de prelación), su correspondencia respectiva queda ordenada igual a la establecida, esto es, la de mayor preferencia corresponde a *Personal sanitario que enfrenta COVID-19*, y la de menor preferencia corresponde a *Resto de la población (mayor a 16 años de edad)*.

Es importante señalar que, con otra ponderación de los pesos asignados tanto a los criterios como a las alternativas, el orden de vacunación sería diferente.

Para validar la consistencia, a partir de λ_{max} que es el máximo valor propio, que resultó ser 4.031 de la

matriz de criterios, se calculó el Índice de Consistencia, obteniendo un valor de $CI = 0.0103$, tomando a RI como 0.9, resulta ser $I_R = 0.0114$, el cual, cotejado con el *ratio* de consistencia correspondiente, implica que existe consistencia en la construcción de la matriz. Por lo tanto, se puede considerar que los criterios establecidos fueron congruentes, lo cual implica juicio racional por parte del tomador de decisiones.

Aplicación de la estrategia de vacunación

Para conocer sus resultados, se tomó la base de datos de la Secretaría de Salud (2021). De 1 048 756 registros de pacientes hospitalizados, del 1 de enero de 2020 al 31 de octubre de 2021, 412 310 dieron como resultado de laboratorio positivos a SARS-CoV-2. En la gráfica 1 se muestra el comportamiento del virus conforme a rango de edad y sexo.

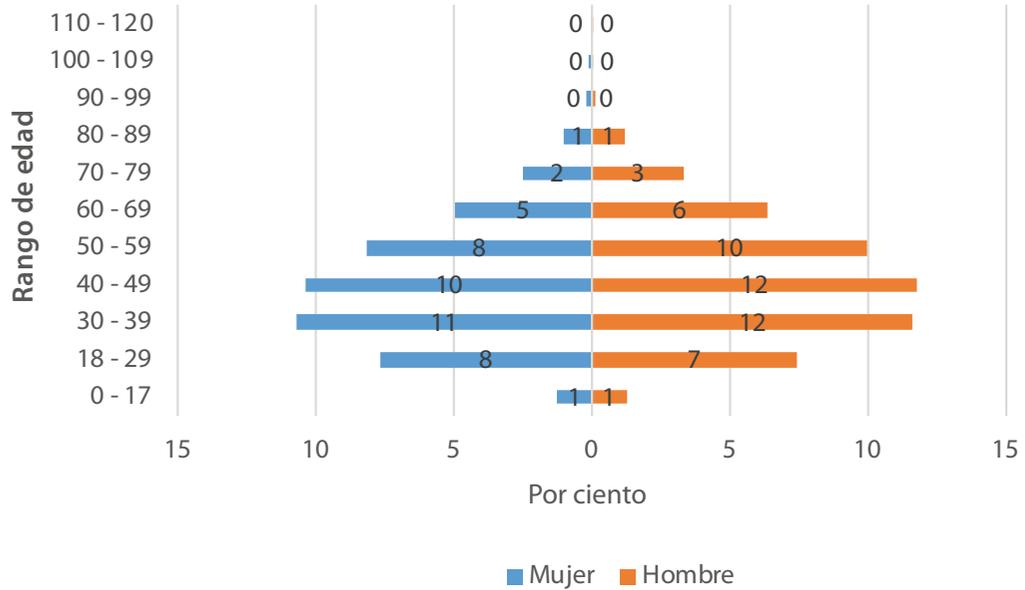
El rango de 30 a 39 años de edad fue donde hubo la mayor concentración de casos, seguido del de 40 a 49; entre los dos captaron 44.5 % de confirmados del SARS-CoV-2.

Asimismo, para conocer la distribución de las defunciones en México por entidad federativa (ver gráfica 2), se tomó la base de datos del gobierno de México con respecto al COVID-19 (2021). Los resultados obtenidos hasta el 31 de octubre de 2021 fueron que 10 estados concentraron 63.6 % de las defunciones a causa de esta enfermedad: México (15.2 %), Ciudad de México (13.8 %), Jalisco (6 %), Puebla (5.3 %), Veracruz de Ignacio de la Llave (5.1 %), Nuevo León (4.5 %), Guanajuato (4.4 %), Baja California (3.4 %), Sinaloa (3 %) y Sonora (2.9 %). El total de muertes al 31 de octubre de 2021 en México fue de 290 278.

Es importante señalar que más de 99 % de los que dieron positivo a SARS-CoV 2 presentaron síntomas de diabetes, EPOC, asma, afecciones inmunosupresoras e hipertensión, así como otras complicaciones, como cardiovascular, obesidad, renal crónica, tabaquismo.

Gráfica 1

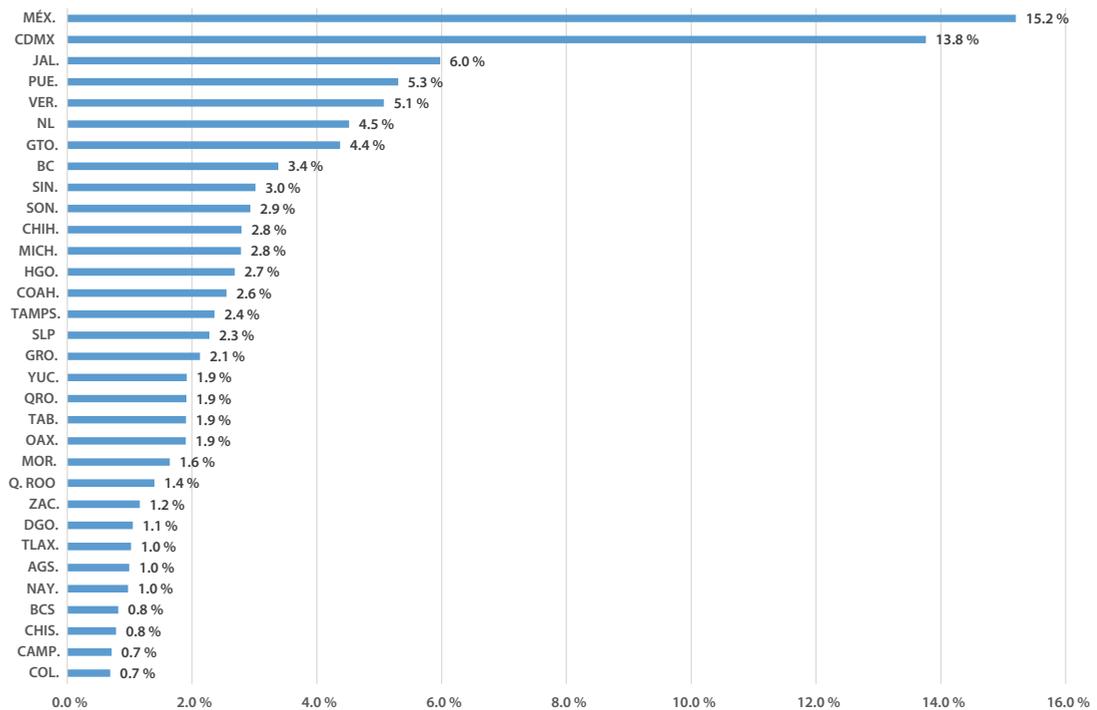
Comportamiento del SARS-CoV-2 confirmado por rango de edad y sexo



Fuente: elaboración propia con datos de la Secretaría de Salud.

Gráfica 2

Distribución de las defunciones por COVID-19 por entidad federativa



Fuente: elaboración propia con datos del gobierno de México.

Conclusiones

En la actualidad, ante situaciones complejas y apremiantes, como es este caso de la escasez de vacunas contra el COVID-19, es importante valerse de metodologías basadas en la experiencia y el conocimiento para una mejor toma de decisiones. En el trabajo desarrollado, y ante la situación de salud que se vive, bajo el supuesto de una matriz de criterios y un conjunto de matrices de prioridades de alternativas, se llegó a un vector de decisiones de alternativas.

El tratamiento de datos cuyo procesamiento está fundamentado por la robustez matemática en su construcción permite obtener resultados altamente satisfactorios y confiables. En este caso, la jerarquización del vector de alternativas resulta satisfactorio, ya que las alternativas por orden de prioridad son representativas de la realidad. Así, la que representa tanto a la jerarquía principal *Personal sanitario que enfrenta COVID-19* como a la del último nivel de jerarquización *Resto de la población (mayor a 16 años de edad)* ayudan a la toma de decisiones ante la escasez de vacunas. Asimismo, se hace énfasis en la consistencia de la construcción de las matrices de comparación, que fortalece el resultado alcanzado.

Además, se debe señalar la importancia que tiene el grupo de expertos en el marco de valores para guiar en la asignación de vacunas contra el COVID-19 entre los países y a nivel nacional en la determinación de los grupos prioritarios para recibirlas, así como en la construcción de las matrices —de criterios y de alternativas—, las cuales son los insumos en este modelo de toma de decisiones, que en ambientes corporativos es un proceso complejo y extenuante, el cual, por lo general, impacta en el entorno de la viabilidad de las decisiones. Por ello, la construcción de escenarios que contemplen la combinación de objetivos, criterios y alternativas múltiples es fundamental para que los planes y programas de desarrollo estratégico tengan un fundamento sólido y ético. El resultado de la estrategia de vacunación establecida por el GTAV en la PRV exhibe la congruencia

esperada en su implementación de acuerdo con nuestras necesidades apremiantes.

Fuentes

- Bolger, F., & G. Wright. "Expertise and Decision Support", en: *Springer Science & Business Media*. 2007.
- Centros para el Control y la Prevención de Enfermedades (CDC). *COVID-19*. Estados Unidos de América, 6 de abril de 2021.
- Cortés, R., R. Gómez, & X. Alba. *Política nacional de vacunación contra el virus SARS-Cov-2, para la prevención de la COVID-19 en México. Documento rector*. Ciudad de México, Política Nacional, Secretaría de Salud, 2021 (DE) de https://coronavirus.gob.mx/wp-content/uploads/2021/01/PolVx_COVID_-11Ene2021.pdf, recuperado en el 2021.
- De Arquer, M. *Fiabilidad humana: métodos de cuantificación, juicio de expertos*. NTP Notas Técnicas de Prevención. Madrid, España, 1996.
- Gobierno de México. *COVID-19 México*. 22 de noviembre de 2021 (DE) <https://datos.covid-19.conacyt.mx/#COMNac>
- Instituto Nacional de Estadística y Geografía (INEGI). *Censo de Población y Vivienda 2020*. México, INEGI, 2020 (DE) <https://www.inegi.org.mx/>
- Linstone, H., & M. Turoff. "The Delphi method: Techniques and Applications", en: Addison-Wesley Pub. Co. *Advanced Book Program*. 1975, pp. 3-12.
- Niño, L., O. Gutiérrez-Lesmes, C. Díaz-Celis, & F. Manrique-Abril. "Riesgo de transmisión de SARS-CoV-2: evaluación espacial multicriterio en un municipio de Colombia, 2020", en: *Rev. Salud Pública*. 22(2), pp. 1-9, 2020.
- OXFAM. *Mega-rich recoup COVID.losses in record.time yet billions will live in poverty for at least a decade*. OXFAM International, 25 de enero de 2021 (DE) <https://www.oxfam.org/en/press-releases/mega-rich-recoup-covid-losses-record-time-yet-billions-will-live-poverty-least>
- Petropoulos, F., N. Kourntzes, K. Nikolopoulos, & E. Siemen. "Judgmental selection of forecasting models", en: *Journal of Operations Management*. 60, pp. 34-46, 2018.
- Saaty, R. "The Analytic Hierarchy Process - What it is and How it is used", en: *Mathl Modelling*. 9(3), pp. 161-176, 1987.
- Saaty, T. "A scaling method for priorities in a hierarchichal structure", en: *Journal Math Psych*. 15(3), pp. 234-281, 1977.
- _____. *The Analytic Hierarchy Process, Planing, Priority Setting and Resource Allocation*. New York, McGraw-Hill, 1980.
- _____. "Axiomatic foundation of the analytic hierarchy process", en: *Management Science*. 32(7), pp. 841-855, 1986.
- Schmidt, K., I. Aumann, I. Hollander, K. Damm, & J. Graf von der Schulenburg. "Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting", en: *Medical Informatics and Decision Making*. Springer Nature, 2015, doi:10.1186/s12911-015-0234-7.

Secretaría de Salud (DE) https://coronavirus.gob.mx/wp-content/uploads/2021/11/2021.10.31_CP_Salud_CTD_COVID-19.pdf, recuperado el 31 de octubre de 2021a.

_____ (DE) http://datosabiertos.salud.gob.mx/gobmx/salud/datos_abiertos/datos_abiertos_covid19.zip, recuperado en octubre de 2021b.

Síntevi. "Consenso experto en tiempos de COVID-19: aplicaciones del método Delphi en materia de salud", en: *Cirugía y Cirujanos*. 89(1), pp. 120-129, 2020 (DE) <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Werner, C., T. Bedford, R. Cooke, A. Hanea, & O. Morales. "Expert judgement for dependence in probabilistic modelling: a systematic literature review

and future research directions", en: *European Journal of Operational Research*. 258(3), pp. 801-819, 2017.

WHO. *WHO SAGE values framework for the allocation and prioritization of COVID-19 vaccination*. World Health Organization. 14 de september de 2020 (DE) https://apps.who.int/iris/bitstream/handle/10665/334299/WHO-2019-nCoV-SAGE_Framework-Allocation_and_prioritization-2020.1-eng.pdf

Wind, Y., T. & Saaty. "Marketing Applications of the Analytic Hierarchy Process", en: *Management Science*. 26, pp. 641-658, 1980, doi:<http://dx.doi.org/10.1287/mnsc.26.7.641>

Uso de los Censos Económicos para la detección de factores de éxito en las empresas de la manufactura automotriz mediante modelos de clúster

Use of Economic Censuses for the Detection of Success Factors in Automotive Manufacturing Companies through Cluster Models

Edgar Manuel Matus Carballo* y Carlos Alberto Jiménez-Bandala**

Las empresas de la industria automotriz (IA) tienden a imitar entre sí prácticas organizacionales que no siempre son exitosas. El administrador o cualquier otro agente debe decidir tan rápido que no puede discernir cuál práctica imitar para llegar al éxito. Esta investigación tiene por objetivo identificar las prácticas exitosas entre empresas de la IA a partir del uso de la información de los Censos Económicos del Instituto Nacional de Estadística y Geografía. Se realiza a partir de un análisis de agrupamiento de datos o *clusterización* que incluyó a 2 934 unidades económicas evaluadas en 162 indicadores que pertenecen a seis dimensiones: *Finanzas, Recursos humanos, Investigación y desarrollo, Innovación tecnológica, Innovación de gestión y Producción*. El éxito fue medido mediante seis indicadores relacionados con el crecimiento, rentabilidad y supervivencia. Los resultados identificaron 52 factores comunes entre las unida-

Companies in the automotive industry (AI) tend to imitate organizational practices that are not always successful. The manager or any other agent must decide so fast that he/she cannot discern which practice to imitate in order to be successful. This research aims to identify successful practices among AI firms using information from the Economic Census of the National Institute of Statistics and Geography. It is carried out from a data grouping or clustering analysis that included 2,934 economic units evaluated on 162 indicators belonging to six dimensions: Finance, Human Resources, Research and Development, Technological Innovation, Management Innovation and Production. Success was measured by six indicators related to growth, profitability and survival. The results identified 52 common factors among the most successful economic units, most of them related to finance. The work demonstrates the

* Universidad La Salle México, macaed00@gmail.com

** Universidad Autónoma del Estado de Quintana Roo, carlos.jimenez@uqroo.edu.mx

Nota: esta investigación se realizó en el marco del proyecto *Patrones de éxito y fracaso en la evolución económica de los negocios identificados a partir de la minería de datos y las redes neuronales artificiales A3-S-129311* del Fondo Sectorial CONACYT-INEGI.

des económicas más exitosas, la mayoría relacionados con finanzas. El trabajo demuestra la importancia de los datos censales y la inteligencia artificial para tomar mejores decisiones.

Palabras clave: isomorfismo organizacional; Método de *Clusterización*; Censos Económicos.

Recibido: 19 de octubre de 2021.

Aceptado: 7 de marzo de 2022.

importance of census data and artificial intelligence in making better decisions.

Key words: organizational isomorphism; Clustering Method; Economic Censuses.



3d rendering cyborg control robot assembly line in car factory/ PhonlamaiPhoto/ iStock

Introducción

Si partimos del supuesto racional de que todas las empresas buscan optimizar sus recursos para maximizar sus beneficios (Coase, 1937), podemos estar de acuerdo en que es tarea central del empresario, administrador o cualquier otro agente que tome decisiones lograr que estos objetivos se alcancen. En algunos campos organizacionales altamente institucionalizados es común que los negocios se

imiten entre sí algunas prácticas; este fenómeno ha sido estudiado con amplitud bajo el nombre de isomorfismo institucional (DiMaggio y Powell, 1991). Desde esta perspectiva, hay un consenso en el que las prácticas isomórficas no responden solo a procesos racionales, es decir, los agentes decisores las imitan sin discernir entre aquellas que resultarán en éxito y las que no; algunas de las imitadas solo responden a la legitimación frente al mercado (Dacin *et al.*, 2002; Hitt *et al.*, 2004).

La industria automotriz (IA) es un ejemplo claro de un campo organizacional altamente institucionalizado, donde se comparten actividades, procesos, indicadores y paquetes tecnológicos y de gestión (McFarland *et al.*, 2008). Algunas prácticas han demostrado de forma general su eficiencia, como el Justo a Tiempo (JIT, por sus siglas en inglés), la Calidad Total o el *Kan-Ban*; sin embargo, muchas otras se replican sin tener certeza en los resultados (Min *et al.*, 2022; Aksom, 2022); en otros casos, aun imitando la misma práctica, los resultados fueron diferenciados según el tamaño de la empresa (Dasanayaka *et al.*, 2022).

Esto complejiza el trabajo del decisor, sus limitaciones se deben, en gran medida, al tiempo que le tomaría hacer el discernimiento entre tipos de prácticas, la cantidad de información que debe reunir para hacerlo y la capacidad de procesamiento que debería tener para tal fin (Simon, 1960). No obstante, en la era de la Industria 4.0, esta *racionalidad limitada* está a punto de superarse, los métodos de análisis masivos de datos (*Big Data*), asociados a capacidades computacionales que imitan las funciones neuronales (*inteligencia artificial*) pueden proveernos información bastante precisa en un tiempo significativamente reducido.

Las aristas de la complejidad aumentarán en la era post-COVID-19; la pandemia obligó a las empresas a adoptar rápidamente herramientas digitales para adaptarse a los ambientes de alta incertidumbre. La IA tuvo severas afectaciones, primero al detener sus actividades en el confinamiento y después por la escasez de insumos durante la etapa de reactivación. Ha sido en esta crisis que ha quedado develado que la imitación de prácticas no necesariamente es racional y, por lo tanto, sus resultados no siempre inciden en el éxito, por ello se hace necesario proponer herramientas para que los agentes decisores puedan discernir entre los tipos de prácticas que imitarán y adoptarán (Pelle y Tabajdi, 2021).

En ese sentido, este trabajo tiene por objetivo identificar las prácticas isomórficas que contribuyen al éxito de las empresas de la IA mexicana

mediante un análisis de clúster de los microdatos de los Censos Económicos 2019¹ del Instituto Nacional de Estadística y Geografía (INEGI, 2019) que representan una gran fuente de información confiable y que hoy, más que nunca, es indispensable. La hipótesis principal es que existen prácticas que son compartidas (imitadas) entre los negocios de este sector de actividad económica, pero que no inciden en los resultados de éxito.

Esta investigación se sustenta en el enfoque teórico del nuevo institucionalismo que sostiene que las prácticas isomórficas en las organizaciones no se llevan a cabo por buscar resultados positivos, sino que hay un entramado de acuerdos implícitos en los que, a partir de compartir ciertas prácticas, se obtiene legitimidad en el campo, aunque explícitamente se declaren como *buenas prácticas* y se promuevan su aplicación (Hitt *et al.*, 2004; Tsai *et al.*, 2021), como sucede en la industria automotriz.

Identificar los factores de éxito permitirá a los agentes económicos tomar mejores decisiones en materia de gestión para definir estrategias de dirección, movilizar recursos hacia una acción determinada o replicar prácticas que tengan efectos positivos en los resultados de las empresas (O'Regan *et al.*, 2005); pero dentro de los ambientes institucionalizados llega a ser muy complicado para los administradores y estudiosos de las organizaciones poder distinguir entre las que se imitan por legitimación y aquellas que inciden en el éxito; esta intuición podría tomar mucho tiempo y años de experiencia; no obstante, la inteligencia artificial y el procesamiento masivo de datos hoy nos permiten acortar el tiempo de forma significativa minimizando, al mismo tiempo, el tamaño del error. Esa es la finalidad última de este trabajo y el problema central que se pretende atender.

Este artículo se divide en cuatro secciones, además de la introducción. En la segunda parte se exponen los fundamentos teóricos que soportan el

¹ Los Censos Económicos se realizan cada cinco años con el objetivo de obtener información estadística básica de todos los establecimientos productores de bienes, comercializadores de mercancías y prestadores de servicios para generar indicadores económicos de México a un gran nivel de detalle geográfico, sectorial y temático.

enfoque institucionalista en la industria automotriz; en la tercera se explica la metodología basada en minería de datos e inteligencia artificial y que, específicamente, conocemos como *clusterización*; en una cuarta se discuten los resultados; y en la quinta se presentan conclusiones.

Isomorfismo organizacional en la industria automotriz mexicana

La manufactura se dedica a la transformación de las materias primas y, junto con las actividades extractivas y de construcción, pertenece al sector secundario. Por su naturaleza es muy heterogénea; por un lado, están los subsectores de alimentos, textiles, madera, papel y cartón, que son mayormente precarios en sus procesos productivos y de trabajo; y por el otro, los relacionados con la petroquímica, la química, los minerales, la metalmecánica y el cómputo, identificados como los más dinámicos e innovadores (Jiménez-Bandala y Sánchez Daza, 2014). En este contexto, la investigación tiene como objeto de estudio la industria automotriz.

La manufactura automotriz es considerada un sector de arrastre, por los encadenamientos productivos que tiene con otras industrias, tanto hacia atrás (vidrio, electrónica, acero, etc.) como hacia adelante (financieras, seguros, etc.) y por las tasas de crecimiento que ha mostrado en los últimos 25 años (Jiménez-Bandala y Pérez, 2019). La IA representa 19 % del valor de producción de toda la industria manufacturera y absorbe entre 35 y 42 % de la Inversión Extranjera Directa de la manufactura (SE, 2019). La IA se concentra, principalmente, en los corredores Centro (Puebla, Hidalgo, Morelos, estado de México y Ciudad de México) y Bajío (Guanajuato, Querétaro y Aguascalientes), así como en la Frontera Norte (Chihuahua, Sonora y Baja California) (Jiménez-Bandala y Pérez, 2019).

La IA utiliza mecanismos altamente institucionalizados, así como rutinas y ceremonias que tienen como finalidad homogeneizar el comportamiento del trabajador y el desempeño de las tareas que requieren un nivel de estandarización dentro de

la cadena global; es decir, el proceso productivo necesita un grado de homogenización interorganizacional, de tal forma que el comportamiento de una organización está determinado por el conjunto de relaciones que tiene con otras organizaciones (Sacomano *et al.*, 2013). Este nivel de interacción las presiona para imitar prácticas o asimilar y adaptar lo que están haciendo las demás (Min *et al.*, 2022). Este proceso puede ser identificado como isomorfismo y es más común en campos organizacionales muy estructurados y profesionalizados (Powell y DiMaggio, 1983; Ansmann y Seyfried, 2022), características comunes de industrias donde predominan grandes empresas transnacionales.

El isomorfismo puede entenderse como un mecanismo de coordinación que produce una convergencia de prácticas y procesos a través de tres diferentes tipos: a) mimético, b) normativo y c) coercitivo (DiMaggio y Powell, 1991). Estas prácticas similares no necesariamente son las más exitosas, por esto, es importante identificarlas y discernir entre ellas.

El tipo coercitivo está relacionado por presiones formales e informales; las segundas provienen de las expectativas culturales en la sociedad sobre la función de las organizaciones, mientras que las primeras vienen de regulaciones gubernamentales, como impuestos, controles de contaminación y regulaciones contables.

El mimético proviene de la incertidumbre, por ejemplo, ambientes de cambios rápidos en el mercado o la tecnología. Las empresas gestionan la incertidumbre imitando a otras, modelan su comportamiento a través de mecanismos como: la contratación de trabajadores de otros negocios, la participación en asociaciones del mismo sector o recurriendo a los mismos consultores, incluso, la base de clientes compartida también puede ser un factor de influencia.

Puede haber tres tipos de imitaciones basadas en: a) la frecuencia, cuando un gran número de empresas reproduce prácticas similares; b) las características, se imitan las prácticas por tamaño o

actividad; y c) resultados, se imitan las prácticas que en apariencia se relacionan con el éxito.

El isomorfismo normativo proviene de la profesionalización, que puede entenderse como la lucha colectiva de los miembros de una ocupación para definir las condiciones y los métodos de su trabajo que establece una base cognitiva y de legitimación. Se pueden identificar dos mecanismos principales: las redes de profesionistas y la educación formal.

En el piso de producción, la industria automotriz se organiza a partir de enclaves comúnmente llamados clústeres, donde hay un gran número de empresas proveedoras que trabajan en torno a una empresa central. La cadena de estas puede ser tan larga como el nivel de especialización de cada uno de los procesos y se clasifica en función de la distancia que hay con respecto a la empresa central. De esta manera, los proveedores directos se conocen como Tier 1, los del primer nivel son el Tier 2 y así sucesivamente. Estas formas de encadenamientos entre organizaciones facilitan el isomorfismo como si se tratara de una *cadena de contagios* (McFarland *et al.*, 2008).

Son las empresas centrales las que detonan, en mayor medida, las transformaciones replicándose a sí mismas en sus subsidiarias, como si se clonaran, y —aunque raro— podría suceder lo contrario de las empresas pequeñas a las grandes como una forma de estrategia subversiva (Morgan y Kristensen, 2006).

Algunas investigaciones sobre la industria automotriz y el isomorfismo dan cuenta que la imitación se propicia en las cadenas de suministros, por ejemplo, Ting Wu *et al.*, (2013) mencionan como ejemplo el JIT y dejan en evidencia que su adopción tuvo más causales poco racionales, como el no perder al cliente, que un verdadero sentido económico para la empresa.

Al principio, los administradores podrían adoptar nuevas prácticas porque ofrecen mejoras en el desempeño o porque los proveedores los han

influido y persuadido; incluso, estas podrían ser adoptadas por legitimación, más que por mejorar el desempeño, por ejemplo, Abrahamson y Rosenkopf (1993) señalan que algunas empresas del mismo campo hacen suyas las mismas prácticas, solo por el hecho de que otras unidades económicas las han adquirido, sin necesidad de discernir sobre sus resultados (de eficiencia o rentabilidad), en una especie de *presión de ferrocarril*. De hecho, esta imitación no siempre es benéfica, por ello, resulta altamente necesario que los agentes que toman decisiones puedan tener los mayores elementos que les permitan distinguir entre tipos de prácticas y proyectar sus consecuencias.

Materiales y métodos

Para este trabajo, se consideró como objeto de estudio a las empresas manufactureras de la industria automotriz que dentro del Sistema de Clasificación Industrial de América del Norte (SCIAN) estuvieran en el subsector 336 (INEGI, 2018) y cuyas ramas se muestran en el cuadro 1.

Se tomó el universo total de las empresas que respondieron en el 2019 el cuestionario de los Censos Económicos 2019: en total 2 934 unidades económicas.²

Con el fin de identificar las prácticas isomórficas, se propuso un modelo matemático de agrupamiento de datos (*clusterización*) que tuviera como criterio de agrupación el grado de éxito de una empresa medido a partir de criterios previamente definidos que consideraron: la rentabilidad, el crecimiento y la permanencia. A partir de los grupos formados se ordenaron de menor a mayor grado de éxito; solo se consideró el que mostró los indicadores más altos y se compararon las características que estadísticamente fueron significativas para considerarse isomórficas. Los datos se procesaron en el Laboratorio de Microdatos de INEGI y por confidencialidad solo se obtuvie-

² Las unidades económicas son la unidad mínima de información de la cual se puede disponer de información.

Cuadro 1

Ramas de las unidades económicas consideradas en la investigación

Nivel	Código	Categoría
Sector	31-33	Industrias manufactureras.
Subsector	336	Fabricación de equipo de transporte.
Rama	3361	Fabricación de automóviles y camiones.
Rama	3362	Fabricación de carrocerías y remolques.
Rama	3363	Fabricación de partes para vehículos automotores.

Fuente: elaboración propia.

ron los resultados agregados. El proceso metodológico se detalla a continuación.

Definición de las variables

Variables de salida (dependientes)

Estas, que en nuestro diseño de investigación podemos llamar dependientes, tienen como finalidad la medición del éxito de las empresas y representan tres categorías: crecimiento, rentabilidad y permanencia, como se muestra en el cuadro 2.

Variables de entrada (independientes)

Estas, que en nuestro diseño de investigación podemos llamar independientes, tienen como fi-

nalidad medir los factores de éxito. Con los reactivos que contiene el cuestionario censal 2019 se construyeron 162 indicadores (*Anexo*) que reflejan prácticas organizacionales y que responden a seis dimensiones: *Finanzas, Investigación y desarrollo (I+D), Innovación en la gestión, Innovación tecnológica, Producción y Recursos humanos.*

Modelo de agrupamiento de datos (clusterización)

La minería de datos es un campo de la estadística para el procesamiento de grandes cantidades de estos; para ello, se apoya en técnicas de inteligencia artificial, cuyo rasgo más significativo es imitar el proceso de pensamiento del cerebro humano a través de códigos y algoritmos computacionales.

Cuadro 2

Indicadores construidos como variables de salida

Indicador	Definición operativa
Crecimiento. Desempeño de inversiones.	Inversión en activos fijos adquiridos menos inversión de activos fijos vendidos entre el total de activos fijos.
Crecimiento. Desempeño de inversiones con respecto al sector.	Resultado de la empresa entre el resultado medio del sector.
Rentabilidad. Nivel de retorno de inversión.	Ingresos totales por ventas y servicios menos los gastos totales por ventas y servicios.
Rentabilidad. Tamaño de la utilidad.	Porcentaje que resulta de 1 menos los totales entre los ingresos totales.
Rentabilidad. Retorno de inversión.	Resultado de la empresa entre el resultado medio del sector.
Supervivencia. Antigüedad de operación.	Resta entre la fecha en la que se realizó el censo y cuando el establecimiento inició operaciones.

Fuente: elaboración propia.

El análisis de clúster es una técnica de la minería de datos. Su objetivo principal es hacer una clasificación de acuerdo con las características que las observaciones presentan donde se plantea encontrar un conjunto de grupos a los que serán asignados los distintos individuos por algún criterio de homogeneidad. Por lo tanto, se hace imprescindible definir una medida de similitud o divergencia para clasificar a los individuos en unos u otros grupos.

Dentro de la teoría del análisis de clúster existen dos clasificaciones principales: a) utilizando un algoritmo de partición: hace referencia a la división de las observaciones con un número k de clústeres ya definidos con anterioridad por el analista; b) empleando un algoritmo jerárquico, método que asigna una jerarquía a las observaciones de acuerdo con sus características y, a su vez, se divide en dos: i) jerárquico disociativo, el cual divide los datos de un gran conglomerado en varios semejantes y ii) jerárquico aglomerativo, que supone a cada observación como un clúster y los va asociando según sus características.

Para poder clasificar a las empresas automotrices, se siguió la metodología de un clúster con algoritmo de partición, basándose en un k con valor de 3, considerando para la industria automotriz los tres primeros Tier. El 1, que es el de las grandes empresas que administran la plantas; el 2, que corresponde a las plantas de producción; y finalmente el 3, que se atribuye a los distribuidores y manejadores de refacciones.

Dentro del análisis, las variables de salida (que están enfocadas en la rentabilidad, supervivencia y crecimiento) se convirtieron en dicotómicas, asignando el valor de 1 cuando el resultante de la empresa estuvo por encima de la media del sector y 0, por debajo, esto con el fin de poder construir la matriz de distancias que se pondrá en el modelo de clúster. En el caso de la supervivencia, la dicotómica detectaba unidades económicas que iniciaron sus operaciones en el mismo año del levantamiento censal (0) y las que tenían mayor antigüedad (1). Posteriormente, se definieron tres posibles grupos

a partir de una clasificación basada en un semáforo y basándose en si se cumplían las condiciones, como se indica a continuación:

$$\text{semáforo} \begin{cases} \text{verde} \rightarrow \text{crecimiento} = 2, \text{rentabilidad} \geq 2, \text{supervivencia} = 1 \\ \text{rojo} \rightarrow \text{supervivencia} = 0 \\ \text{amarillo} \rightarrow \text{otro caso} \end{cases}$$

Esta categorización de las observaciones es un análisis multivariado; para este trabajo se propusieron clústeres basados en particiones con el objetivo de buscar que la función de costo fuera la mínima. De acuerdo con Schubert y Rousseeuw (2019), el modelo PAM (por sus siglas en inglés *Partitioning Around Medoids*), es el más adecuado para formar los grupos porque no es sensible a los datos atípicos, bien conocidos como *outliers*, y que funciona para el estudio de los datos de la industria automotriz dado su origen en la encuesta del SCIAN. El PAM busca los *medoides* de la información a partir de subconjuntos que se eligen de manera aleatoria de las observaciones proporcionadas en la matriz A de distancias, donde tenemos n observaciones con m variables, como en (1); posteriormente, se estandariza sobre las variables dependientes para su uso en la clasificación y así poder reasignarlas con la finalidad de reducir la función de costo (f_{costo}) que está definida como en (2):

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix} \quad (1)$$

$$f_{\text{costo}} = \sum_{i=1}^k \sum_{x_j \in C_i} d(x_j, m_i) \quad (2)$$

Esta última representa la suma de las disimilitudes en cada uno de los puntos $x_j \in C_i$ en el *medoide* m_i elegido del clúster. Esta función de distancia utilizada para $d(x_j, m_i)$ hace uso de la función de distancia de Manhattan que, de acuerdo con Kasambara (2017), se utiliza para poder hacer las mediciones con las medianas en lugar de las medias dado que la media es sensible a los datos atípicos y se define como en (3):

$$d_{\text{man}}(x, y) = \sum_{i=1}^n |(x_i - y_i)| \quad (3)$$

Definida esta función de costos, se deberá calcular la matriz de semejanza para lograr que la información capturada se pueda comparar; en este sentido, se estandariza la matriz de las observaciones para poder generar la función de costo sobre dicha matriz y los *medoides* seleccionados. Así, el objetivo es obtener reasignaciones de las observaciones en grupos (*clusterings*) a través de ese muestreo en el modelo utilizado con la matriz de semejanza y la minimización de la función de costo.

Para realizar la validación del modelo utilizado en el presente trabajo, se utilizaron tres indicadores principales que, de acuerdo con Arbelaitz *et al.* (2013), representan una fuerte validación para el análisis de clúster; estos son:

- Índice de Dunn. Es de razón, donde se estima la distancia mínima de separación entre el diámetro máximo, por lo que se busca que este se vea maximizado.
- Estadístico Pearson-Gamma. Es una adaptación del rango de correlación de la prueba de

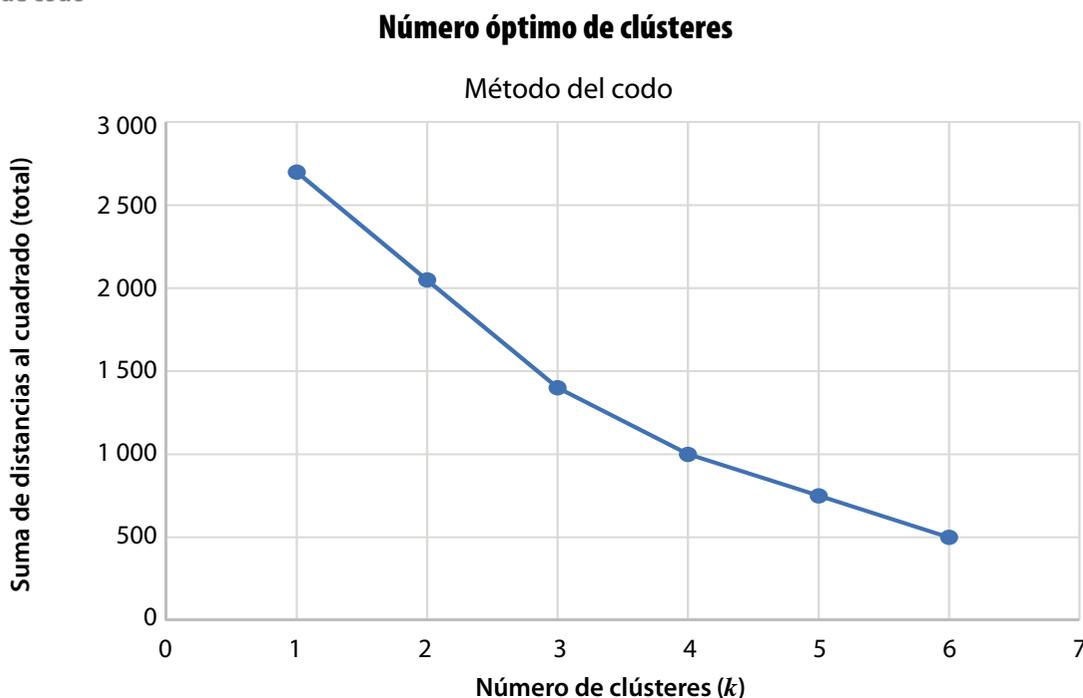
Goodman y Kruskal Gamma para clasificar cantidades, y se busca que también sea alto.

- Índice de Silueta. Es normalizado que busca la distancia entre todos los puntos de un mismo clúster y la divide entre la separación de este mismo. Gracias a esto se toma la silueta promedio, que se pretende sea alta y sin valores negativos.

Resultados

Estos mostraron que el número ideal de clústeres para agrupar las empresas automotrices es de tres, como se observa en la gráfica de codo. Con respecto a los diversos índices construidos para elegir el clúster que mejor se adapta, tenemos que el salto entre las distancias internas del clúster 2 al 3 es la mayor para ese rubro y de igual manera sucede con las distancias internas, situación que nos ayuda a elegir que tres es la mejor opción como número ideal de clústeres.

Gráfica de codo



Fuente: elaboración propia.

Las pruebas estadísticas también comprobaron lo anterior, como se muestra en el cuadro 3.

El clúster 1 agrupó a 1 041 unidades económicas, de las cuales 32 % estuvo en color verde; el 2, a 722 empresas con 15 % en verde; mientras que el 3, a 1 171 negocios, ninguno en verde; por ello, podemos decir con certeza que el clúster 1 es más exitoso que el 2 y, a su vez, el 2 es más que el 3.

En lo que respecta al análisis de los valores de las variables de entrada, se tiene que en el clúster 1 (más exitoso) predominan empresas más intensivas en capital y con indicadores más altos con respecto a los otros grupos en el ámbito de innovación y finanzas. Por su parte, el 2 fue más intensivo en trabajo, pero más dinámico en la adquisición de activo fijo que los otros dos grupos. Por último, el 3 (menos exitoso) fue poco intensivo en capital, orientado en mayor medida a la comercialización y menos a la producción, lo cual puede develarnos que se trataría de las unidades económicas proveedoras y de *outsourcing* de la manufactura automotriz. Lo anterior es congruente tanto con la subdivisión por Tier que se hace de esta industria como con trabajos previos que analizan los contagios isomórficos en la cadena de proveedores, siendo los más cercanos a la empresa central los que logran un mayor nivel de similitud (McFarland *et al.*, 2008; Ting Wu *et al.*, 2013).

Para identificar las prácticas isomórficas que incidieron en mayor medida en el éxito (variables de entrada), se analizaron los valores de todos los

indicadores del clúster 1 y se eligieron los que fueron estadísticamente significativos para diferenciar entre las empresas en color verde y el resto. Este análisis se desarrolló a través del estudio minucioso de las variables donde se tomaron las medias de los indicadores por cada uno de los clústeres y se escogieron aquellas variables donde los valores de las medias tuvieran mayor variación entre dichos clústeres para hacer el estudio de estas variables que son más significativas en el clúster 1. En total, fueron 104 indicadores que reflejaban prácticas diferentes entre empresas más y menos exitosas.

Posteriormente, se analizaron los valores de esos 104 indicadores al interior de las empresas en color verde para identificar las prácticas compartidas; para ello, se calculó el coeficiente de variación (CV) y la mediana de cada indicador para elegir aquellos cuyo CV fuera menor a la mediana, de tal suerte que el resultado muestra los 52 indicadores que reflejan las prácticas más homogéneas de las empresas más exitosas de la industria automotriz. Con este resultado, se demuestra que no todas las prácticas compartidas inciden en el éxito empresarial y podemos suponer que se imitan por efectos isomórficos o por modas teóricas, como sostiene Aksom (2022).

De las seis dimensiones propuestas, solo cuatro fueron importantes para destacar prácticas isomórficas, como se muestra en el cuadro 4.

Por lo que destaca que las prácticas en *Finanzas* son las que más se imitan entre las empresas auto-

Cuadro 3

Pruebas estadísticas para la *clusterización*

Número de clústeres	Distancia interna	Distancia externa	Silueta promedio	Pearson-Gamma
2	1.0120	1.4837	0.3107	0.4468
3	0.7884	1.4605	0.4142	0.5999
4	0.6393	1.4373	0.4775	0.6546
5	0.5189	1.4209	0.5706	0.7058
6	0.3908	1.4037	0.6467	0.7230

Fuente: elaboración propia.

Proporción de los indicadores isomórficos por dimensión

Dimensión	Indicadores propuestos	Indicadores resultantes	Peso ponderado
Finanzas	57	21	40 %
Innovación tecnológica	28	12	23 %
Innovación en la gestión	13	0	0 %
Producción	13	6	12 %
Recursos humanos	18	13	25 %
I+D	25	0	0 %

Fuente: elaboración propia.

motrices y que, al mismo tiempo, tienen incidencia en el éxito. Cabe destacar el acceso al crédito y el tipo de uso. Las unidades económicas exitosas utilizan el crédito en mayor medida para el apalancamiento de otras deudas y menos para la compra de activos (locales, vehículos e insumos). Esto nos lleva a deducir una relación en el tamaño de los negocios más grandes con los más exitosos, ya que según Dasanayaka *et al.* (2022), la tendencia en las grandes empresas es recurrir al *leasing*, que adelgaza los activos (*lean*). De forma congruente, se encontró que las empresas más exitosas muestran tasas más altas en gastos de arrendamiento y más bajas en mantenimiento y publicidad como proporción de los gastos totales. También, tuvieron las tasas más bajas de ingreso por comisiones, arrendamiento y prestación de servicios, lo cual significa que están enfocadas en la producción del bien principal, que es congruente con resultados como los obtenidos por Pearce y Robbins (1993) y Quinn y Hilmer (1994).

En segundo lugar, destaca la gestión de *Recursos humanos*, y entre las prácticas con mayor incidencia sobresalen: a) el pago de un mayor volumen de prestaciones; pero b) una menor proporción en el reparto de utilidades (PTU); c) incorporación del trabajo femenino; d) menor proporción del trabajo sin salario; pero e) mayor proporción del que es suministrado por una filial de la empresa (terciarización del empleo).

Lo anterior resulta lógico, tanto las variables de *Finanzas* como las de *Recursos humanos* están re-

flejando un alto nivel de flexibilidad y estructuras esbeltas que permiten cambios más rápidos; en contraste, las empresas menos exitosas, cuyos indicadores fueron opuestos, son entonces más rígidas porque su proporción de activos tangibles es más alto y el peso de los costos salariales, también, lo que provoca cambios más lentos, tal como se halló en trabajos como los de Boudreau (1996), Min *et al.* (2022) y Tsai *et al.* (2021).

La tercera dimensión con mayor peso fue la *Innovación tecnológica*. Las prácticas isomórficas son el mayor uso de internet (tanto para los procesos de la empresa como para la gestión del personal) y haber recibido apoyos públicos, evidencia que se ha mostrado en trabajos como el de Jiménez-Bandala y Pérez (2019).

En cuarto lugar estuvieron las variables asociadas a la *Producción*, en la que se destacan prácticas como reducción de costos de materias primas, mayor número de días y horas laboradas al año (hubo menos paros técnicos) y una menor proporción de los gastos en fuerza de trabajo (intensivas en capital), muy de la mano de lo que señala Blumberg (1998) y Tsai *et al.* (2021) sobre el rediseño del tamaño organizacional.

Las dimensiones *Innovación en la gestión* e *I+D* no resultaron representadas porque suponemos no son un elemento diferenciador entre empresas más exitosas; al contrario: para el sector resulta inherente asumirlas, como ocurre en aquellos contextos que exigen mayores niveles de calidad y formación

educativa (Ansmann y Seyfried, 2022); recordemos que el primer filtro de descarte de variable fue la diferencia entre más y menos exitosas.

Conclusiones

El análisis organizacional desde la perspectiva institucional ha sido utilizado en mayor medida para la comprensión, pero poco en la aplicación en la gestión y el direccionamiento de los cambios en la organización; la aportación de este trabajo pretende abonar en esta área poco explorada. También busca aportar la inclusión del *Big Data* y la inteligencia artificial en los estudios organizacionales, de forma más específica en la corriente institucionalista, a partir del procesamiento de grandes cantidades de datos, en este caso los provenientes de los Censos Económicos por medio de metodologías novedosas, como el agrupamiento o *clusterización*.

Los resultados del análisis de los datos censales para la industria automotriz comprueban que se trata de un campo altamente institucionalizado, que se explica a partir de las múltiples relaciones que se tienen entre empresa central y su cadena de proveedores, por lo que es alto el contagio de las prácticas y, por lo tanto, el isomorfismo, como se planteó en la hipótesis. De los 162 indicadores analizados que pueden considerarse prácticas organizacionales, al menos 52 resultaron ser estadísticamente homogéneos entre las empresas más exitosas y diferentes con respecto a las que menos lo son. Podemos decir que estos indicadores tienen incidencia en el éxito (rentabilidad, supervivencia o crecimiento) de la organización, no obstante, de acuerdo con la teoría, su imitación podría tener otras finalidades no racionales, como la legitimación o el seguimiento de modas administrativas.

Se identificó que la mayor parte de las empresas exitosas comparten características comunes en sus procesos, como ser intensivas en capital, y por ello es congruente que la distinción más grande entre las más y las menos exitosas tenga relación con las dimensiones de *Finanzas y Recur-*

sos humanos, y se orienten hacia un nivel menor de precarización del trabajo.

La continuación de este tipo de trabajos nos permitirá comprender de mejor manera los mecanismos de isomorfismo, así como una distinción de las causas y sus resultados para poder incidir de mejor manera en la transformación de nuestras organizaciones.

Fuentes

- Abrahamson, E. y L. Rosenkopf. "Institutional and competitive bandwagons: using mathematical modelling as a tool to explore innovation diffusion", en: *Academy of Management Review*. 18(3), 1993, pp. 487-517.
- Aksom, H. "Reconciling conflicting predictions about transience and persistence of management concepts in management fashion theory and new institutionalism" en: *International Journal of Organizational Analysis*. 30(2), 2022, pp. 430-453. <https://doi:10.1108/IJOA-10-2020-2445>.
- Ansmann, M. y M. Seyfried. "Isomorphism and organizational performance: Evidence from quality management in higher education", en: *Quality Assurance in Education*, 30(1), 2022, pp. 135-149. <https://doi:10.1108/QAE-07-2021-0114>.
- Arbelaitz, O., I. Gurrutxaga, J. Muguerza, J. M. Pérez y I. "Perona. An extensive comparative study of cluster validity indices", en: *Pattern Recognition*. 46(1), 2013, pp. 243-256.
- Blumberg, D. "Strategic assessment of outsourcing and downsizing in the service market", en: *Managing Service Quality*. 8 (1), 1998, pp. 5-18.
- Boudreau, J. *Human Resources and Organization Success*. Ithaca, NY, Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, 1996.
- Coase, R. "The Nature of the firm", en: *Economica*. 4(16), 1937, pp. 386-405 (DE) <https://doi.org/10.1111/j.1468-0335.1937.tb00002.x>.
- Dacin, M. T., J. Goodstein y W. R. Scott. "Institutional theory and institutional change: Introduction to the special research forum", en: *Academic Management Journal*. 45, 2002, pp. 45-56.
- Dasanayaka, C. H., N. Gunarathne, D. F. Murphy y P. Nagirikandalage. "Triggers for and barriers to the adoption of environmental management practices by small and medium-sized enterprises: A critical review", en: *Corporate Social Responsibility and Environmental Management*. 1(16), 2022 (DE) <https://doi:10.1002/csr.2244>.
- DiMaggio, W. y P. Powell. *The new institutionalism in organizational analysis*. Chicago, University of Chicago Press, 1991.
- Hitt, M. A., D. Ahlstrom, M. T. Dacin, E. Levitas, y L. Svobodina. "The institutional effects on strategic alliance partner selection in transition economies: China vs. Russia", en: *Organization Science*. 15(2), 2004, pp. 173-85.

- INEGI. *Censos Económicos 2019. Resultados definitivos*. México, INEGI, 2019.
- _____. *Sistema de Clasificación Industrial de América del Norte*. México, INEGI, 2018.
- Jiménez-Bandala, C. y N. Pérez. "Estado y sector automotriz: la configuración de la estructura empresarial. Los enjuegos macro y microeconómicos del desarrollo", en: *Revista Latinoamericana De Investigación Social*. 2(1), 2019, pp. 1-17.
- Jiménez-Bandala, C. y G. Sánchez Daza. "La flexibilidad laboral en México. Los efectos de las relaciones neoliberales de trabajo", en: Ventura, Teresa (coord.). *Las expresiones del mundo laboral*. México, BUAP, 2014, pp. 15-34.
- Kassambara, A. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. Volume 1. STHDA, 2017.
- McFarland, R. G., J. M. Bloodgood y J. M. Payan. "Supply chain contagion", en: *Journal of Marketing*. 72(2), 2008, pp. 63-79.
- Min, Y., Y. Liao y Z. Chen. "The side effect of business group membership: How do business group isomorphic pressures affect organizational innovation in affiliated firms?", en: *Journal of Business Research*. 141, 2022, pp. 380-392 (DE) <https://doi.org/10.1016/j.jbusres.2021.11.036>.
- Morgan, G. y P. Kristensen. "The contested space of multinationals: Varieties of institutionalism, varieties of capitalism", en: *Human Relations*. 59(11), 2006, pp. 1467-1490 (DE) <https://doi.org/10.1177/0018726706072866>.
- O'Reagan, N., A. Ghobadian y M. Sims. "The link between leadership, strategy, and performance in manufacturing SMEs, J. Small Bus", en: *Strategy*. 15(2), 2005, pp. 45-57.
- Pelle, A. y G. Tabajdi. "Covid-19 and transformational megatrends in the european automotive industry: Evidence from business decisions with a central and eastern european focus", en: *Entrepreneurial Business and Economics Review*. 9(4), 2021, pp. 19-33 (DE) <https://doi.org/10.15678/EBER.2021.090402>.
- Pearce, J y K. Robbins. Toward improved theory and research on business turnaround, *Journal of Management*, 19(3), 1993 pp. 613-636 (DE) [https://doi.org/10.1016/0149-2063\(93\)90007-A](https://doi.org/10.1016/0149-2063(93)90007-A).
- Powell, P. y W. DiMaggio. "The new institutionalism: avenue of collaboration", en: *Journal of Institutional and Theoretical Economics*. 154(4), 1983, pp. 696-705.
- Quinn, J. y F. Hilmer. "Strategic Outsourcing", en: *Sloan Management Review*. 35(4), 1994 (summer), pp.43-55.
- Sacomano, M., O. Serra y C. Kirschbaum. "Isomorphism and Institutional Control in a Modular Plant of the Automotive Industry", en: *Revista Brasileira de Gestão de Negócios*. 15(49), 2013, pp. 529-544.
- Secretaría de Economía (SE). *Boletín de información de la industria automotriz*. México, PROMEXICO-SE, 2019.
- Simon, H. A. "Some further notes on a class of skew distribution functions", en: *Information and Control*. 3(1), 1960, pp. 80-88.
- Schubert, E. y P. Rousseeuw. "Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms", en: *Similarity Search and Applications*. Springer International Publishing, 2019, pp. 171-187 (DE) https://doi.org/10.1007%2F978-3-030-32047-8_16
- Ting Wu, E. M., D. Hinton y P. Quintas. "Isomorphic mechanisms in manufacturing supply chains: a comparison of indigenous Chinese firms and foreign-owned MNCs", en: *Supply Chain Management: An International Journal*. 18(2), 2013, pp. 161-177 (DE) <http://dx.doi.org/10.1108/13598541311318809>.
- Tsai, M., J. Wang y Y. Chen. "Effect of social identity on supply chain technology adoption of small businesses", en: *Asia Pacific Management Review*. 26(3), 2021, pp. 129-136 (DE) <https://doi.org/10.1016/j.apmr.2020.12.001>.

Indicadores construidos como variables de entrada

Indicador	Definición operativa
Tasa de gastos administrativos.	Porcentaje de gastos atribuido al pago de servicios legales, contables, administrativos, ingeniería, diseño e investigación sobre gastos totales.
Tasa de gastos por publicidad.	Porcentaje de gastos atribuidos al pago de publicidad sobre gastos totales.
Tasa de gastos de oficina.	Porcentaje de gastos atribuido al pago de servicios y objetos de papelería sobre gastos totales.
Tasa de gastos por servicios de agua.	Porcentaje de gastos atribuidos al consumo de agua sobre gastos totales.
Tasa de gastos por servicios de energía eléctrica.	Porcentaje de gastos atribuidos al consumo energía eléctrica sobre gastos totales.
Tasa de gastos por servicios de combustibles.	Porcentaje de gastos atribuidos al consumo combustibles sobre gastos totales.
Tasa de gastos por comunicación.	Porcentaje de gastos atribuidos a servicios de comunicación sobre gastos totales.
Tasa de gastos por otros conceptos.	Porcentaje de gastos atribuidos a otros conceptos no ligados a producción (vigilancia, vestimenta, capacitación) sobre gastos totales.
Tasa de gastos por mantenimiento.	Porcentaje de gastos atribuidos al pago de mantenimiento de edificios, maquinaria y otros bienes sobre gastos totales.
Tasa de gastos por rentas.	Porcentaje de gastos atribuidos al pago de rentas y alquileres sobre gastos totales.
Tasa de publicidad nacional.	Porcentaje de gasto de publicidad nacional con respecto al gasto total de publicidad.
Tasa de publicidad extranjera.	Porcentaje de gasto de publicidad extranjera con respecto al gasto total de publicidad.
Tasa de gastos de viáticos.	Porcentaje de gastos de viáticos con respecto a gastos totales.
Tasa de gastos por servicios.	Porcentaje de gastos atribuido al pago de pago a tercero por servicios sobre gastos totales.
Cociente de los gastos no operacionales entre el ingreso.	Porcentaje de gastos atribuidos a gastos fiscales financieros y donaciones, sobre ingresos totales.
Cociente de los gastos no operacionales entre el gasto.	Porcentaje de gastos atribuidos a gastos fiscales, financieros y donaciones sobre gastos totales.
Tasa de gastos de seguros.	Porcentaje de gastos de seguros con respecto a gastos totales.
Tasa de ingreso por comisiones.	Porcentaje de ingreso atribuido a las comisiones sobre ventas sobre ingresos totales.
Tasa de ingreso por arrendamiento.	Porcentaje de ingreso atribuido a la renta y alquileres ingresos sobre ingresos totales.
Tasa de ingreso por prestación de servicios.	Porcentaje de ingreso atribuido a prestación de servicios sobre ingresos totales.
Tasa de ingreso por otras actividades.	Porcentaje de ingreso atribuido a otros ingresos por sus actividades sobre ingresos totales.
Cociente de los ingresos no operacionales entre el ingreso total.	Porcentaje de ingreso atribuido a trámites fiscales, financieros y donaciones sobre los ingresos totales de la empresa.
Cociente de los ingresos no operacionales entre los gastos no operacionales.	Razón del total de ingresos fiscales sobre los gastos fiscales.
Tasa de ingreso por venta de productos.	Porcentaje de ingreso atribuido a las ventas de productos elaborados, manufacturados o fabricados sobre ingresos totales.

Indicadores construidos como variables de entrada

Indicador	Definición operativa
Tasa de ingreso por reventa de mercancías.	Porcentaje de ingreso atribuido a la reventa de mercancías sobre ingresos totales.
Tasa de ingreso por servicios de maquila.	Porcentaje de ingreso atribuido al cobro por maquila y transformación de materias primas propiedad de terceros sobre ingresos totales.
Rotación de activos.	Ingresos por ventas sobre total de activos.
Rotación de inventarios.	Costo de ventas/(Inventario final más inventario inicial entre dos).
Costo de ventas.	Inventario inicial menos inventario final más compras sobre total de ventas.
Ausencia de cuentas bancarias por causas endógenas.	No tiene cuenta bancaria debido a decisión propia de la empresa excluyendo las cuestiones del mercado (opciones: 0561.1, 0561.2).
Ausencia de cuentas bancarias por causas exógenas.	No tiene cuenta bancaria debido cuestiones del mercado excluyendo decisiones propias de la empresa (opciones: 0561.3).
Ausencia de cuentas bancarias por causas endógenas y exógenas.	No tiene cuenta bancaria debido a decisión propia de la empresa y por cuestiones de la empresa.
Uso de crédito para creación o apertura de negocio.	Variable categórica para uso de crédito para creación o apertura de negocio.
Uso de crédito para equipamiento o ampliación del negocio.	Variable categórica para uso de crédito para equipamiento o ampliación del negocio.
Uso de crédito para compra de local o vehículo.	Variable categórica para uso de crédito para compra de local o vehículo.
Uso de crédito para pago de deudas.	Variable categórica para uso de crédito para pago de deudas.
Uso de crédito para adquisición de insumos en mercado nacional.	Variable categórica para uso de crédito para adquisición de insumos en mercado nacional.
Uso de crédito para adquisición de insumos en mercado internacional.	Variable categórica para uso de crédito para adquisición de insumos en mercado internacional.
Uso de crédito para pago de salarios.	Variable categórica para uso de crédito para pago de salarios.
Procedencia de los créditos, préstamos o financiamiento recibido de bancos.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de bancos.
Procedencia de los créditos, préstamos o financiamiento recibido de cajas de ahorro popular.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de cajas de ahorro popular.
Procedencia de los créditos, préstamos o financiamiento recibido de proveedores.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de proveedores.
Procedencia de los créditos, préstamos o financiamiento recibido de familiares o amigos.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de familiares o amigos.
Procedencia de los créditos, préstamos o financiamiento recibido de gobierno.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de gobierno.
Procedencia de los créditos, préstamos o financiamiento recibido de prestamistas privados o agiotistas.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de prestamistas privados o agiotistas.
Procedencia de los créditos, préstamos o financiamiento recibido de propietarios o socios.	Variable categórica para procedencia de los créditos, préstamos o financiamiento recibido de propietarios o socios.
Ausencia de crédito por causas endógenas.	Carece de crédito debido a decisión propia de la empresa excluyendo las cuestiones del mercado.
Ausencia de crédito por causas exógenas.	Carece de crédito debido a cuestiones del mercado excluyendo decisiones propias de la empresa.
Ausencia de crédito por causas endógenas y exógenas.	Carece de crédito debido a decisión propia de la empresa y por cuestiones de la empresa.
Tasa de salarios con respecto al gasto (sin liquidaciones).	Porcentaje de pago atribuido a los salarios sin contar liquidaciones sobre los gastos totales.

Indicadores construidos como variables de entrada

Indicador	Definición operativa
Tasa de salarios con respecto al gasto (con liquidaciones).	Porcentaje de pago atribuido a los salarios contando liquidaciones sobre los gastos totales.
Tasa de gastos por comisiones.	Porcentaje de gastos atribuido al pago de honorarios o comisiones sin sueldo sobre gastos totales.
Tasa de crecimiento de tecnología.	Porcentaje de crecimiento de inversión destinado a tecnología con respecto a los años 2011 al 2013.
Tasa de crecimiento de tecnología con respecto a total de inversiones.	Porcentaje de inversión destinado a <i>software</i> de cómputo y bases de datos en el 2013 sobre total invertido entre el 2011, 2012 y 2013.
Tasa de crecimiento de patentes.	Porcentaje de crecimiento de inversión destinado a patentes con respecto a los años 2011 al 2013.
Tasa de crecimiento de patentes con respecto a total de inversiones.	Porcentaje de inversión destinada a patentes, marcas y otros productos en el 2013 sobre total invertido entre el 2011, 2012 y 2013.
Tasa de cambio de activo fijo para uso propio (2009-2014).	Variación porcentual en el activo fijo que la empresa produjo para su uso propio entre el 2009 y 2014 con respecto a los activos fijos.
Tasa de cambio de activo fijo para uso propio (2014-2019).	Variación porcentual en el activo fijo que la empresa produjo para su uso propio entre el 2014 y 2019 con respecto a los activos fijos.
Patentes registradas (2011).	Presencia de patentes registradas en el 2011.
Patentes registradas (2012).	Presencia de patentes registradas en el 2012.
Patentes registradas (2013).	Presencia de patentes registradas en el 2013.
Patentes compradas (2011).	Presencia de patentes compradas en el 2011.
Patentes compradas (2012).	Presencia de patentes compradas en el 2012.
Patentes compradas (2013).	Presencia de patentes compradas en el 2013.
Tasa de crecimiento de investigación.	Porcentaje de crecimiento de inversión destinado a investigación con respecto a los años 2011 al 2013.
Tasa de crecimiento de investigación con respecto a total de inversiones.	Porcentaje de inversión destinado a I+D en el 2013 sobre total invertido entre el 2011, 2012 y 2013.
Innovación universitaria (2011).	Presencia de innovación con universidades en el 2011.
Innovación universitaria (2012).	Presencia de innovación con universidades en el 2012.
Innovación universitaria (2013).	Presencia de innovación con universidades en el 2013.
Innovación empresarial (2011).	Presencia de innovación con empresas no asociadas en el 2011.
Innovación empresarial (2012).	Presencia de innovación con empresas no asociadas en el 2012.
Innovación empresarial (2013).	Presencia de innovación con empresas no asociadas en el 2013.
Innovación con clientes (2011).	Presencia de innovación con clientes en el 2011.
Innovación con clientes (2012).	Presencia de innovación con clientes en el 2012.
Innovación con clientes (2013).	Presencia de innovación con clientes en el 2013.
Innovación propia (2011).	Presencia de innovación propia en el 2011.
Innovación propia (2012).	Presencia de innovación propia en el 2012.
Innovación propia (2013).	Presencia de innovación propia en el 2013.
Tasa de subcontratación.	Porcentaje de personal contratado de manera externa.
Tasa de personal que realizó innovación de procesos.	Porcentaje de empleados que hicieron innovación de procesos.
Tasa de personal que realizó innovación de <i>marketing</i> .	Porcentaje de empleados que hicieron innovación de <i>marketing</i> .
Tasa de personal que realizó innovación de gestión de la empresa.	Porcentaje de empleados que hicieron innovación de gestión de la empresa.

Indicadores contruidos como variables de entrada

Indicador	Definición operativa
Tasa de cambio de gasto de papelería (2009-2014).	Variación porcentual en el consumo de papelería entre el 2009 y 2014 con respecto a gastos totales.
Tasa de cambio de gasto de papelería (2014-2019).	Variación porcentual en el consumo de papelería el 2014 y 2019 con respecto a gastos totales.
Tasa de cambio de gasto de asesoría (2009-2014).	Variación porcentual en gastos de asesoría entre el 2009 y 2014 con respecto a gastos totales.
Tasa de cambio de gasto de asesoría (2014-2019).	Variación porcentual en gastos de asesoría entre el 2014 y 2019 con respecto a gastos totales.
Tasa de gastos de asesoría.	Porcentaje de gastos de asesoría con respecto a gastos totales.
Tasa de cambio de consumo de bienes y servicios (2009-2014).	Variación porcentual en el consumo de bienes y servicios entre el 2009 y 2014 con respecto a gastos totales.
Tasa de cambio de consumo de bienes y servicios (2014-2019).	Variación porcentual en el consumo de bienes y servicios entre el 2014 y 2019 con respecto a gastos totales.
Tasa de cambio de consumo de combustibles (2009-2014).	Variación porcentual en el consumo de combustibles entre el 2009 y 2014 con respecto a gastos totales.
Tasa de cambio de consumo de combustibles (2014-2019).	Variación porcentual en el consumo de combustibles entre el 2014 y 2019 con respecto a gastos totales.
Tasa de cambio de consumo de electricidad (2009-2014).	Variación porcentual en el consumo de electricidad entre el 2009 y 2014 con respecto a gastos totales.
Tasa de cambio de consumo de electricidad.	Variación porcentual en el consumo de electricidad entre el 2014 y 2019 con respecto a gastos totales.
Tasa de gasto de ingeniería.	Porcentaje de gastos de ingeniería con respecto a gastos totales.
Tasa de familiares, becarios, voluntarios sin sueldo base.	Porcentaje de familiares, becarios o voluntarios sin sueldo base con respecto al total de empleados.
Tasa de empleo no dependiente.	Porcentaje de empleados que no dependen de esta empresa con respecto al total de empleados.
Tasa de subcontratación.	Porcentaje de personal contratado por otra razón social con respecto al total de empleados.
Tasa de subcontratación femenina.	Porcentaje de mujeres contratadas por otra razón social con respecto a todas las mujeres contratadas.
Tasa de trabajo bajo comisiones.	Porcentaje de empleados que trabajan bajo comisiones con respecto al total de empleados.
Tasa de pago de comisiones.	Porcentaje de los pagos que se realizaron hacia los empleados con comisiones con respecto a todas las remuneraciones.
Promedio de pago de comisiones.	Cuánto se pagó, en promedio, a los empleados por comisiones.
Tasa de ventas sobre bienes nuevos.	Porcentaje de ventas de bienes o servicios nuevos.
Tasa de ventas sobre bienes mejorados.	Porcentaje de ventas de bienes o servicios mejorados.
Tasa de ventas de sobre bienes sin cambios.	Porcentaje de ventas de bienes o servicios sin cambios.
Financiamiento para innovación con recursos propios (2011).	Financiamiento mediante recursos propios en el 2011.
Financiamiento para innovación con créditos (2011)	Financiamiento mediante créditos en el 2011.
Financiamiento para innovación con apoyos gubernamentales (2011).	Financiamiento mediante apoyos gubernamentales en el 2011.
Otros métodos de financiamiento para innovación (2011).	Financiamiento mediante otros medios en 2011.
Financiamiento para innovación con recursos propios (2012)	Financiamiento mediante recursos propios en el 2012.

Indicadores construidos como variables de entrada

Indicador	Definición operativa
Financiamiento para innovación con créditos (2012).	Financiamiento mediante créditos en el 2012.
Financiamiento para innovación con apoyos gubernamentales (2012).	Financiamiento mediante apoyos gubernamentales en el 2012.
Otros métodos de financiamiento para innovación (2012).	Financiamiento mediante otros medios en el 2012.
Financiamiento para innovación con recursos propios (2013).	Financiamiento mediante recursos propios en el 2013.
Financiamiento para innovación con créditos (2013).	Financiamiento mediante créditos en el 2013.
Financiamiento para innovación con apoyos gubernamentales (2013).	Financiamiento mediante apoyos gubernamentales en el 2013.
Otros métodos de financiamiento para innovación (2013).	Financiamiento mediante otros medios en el 2013.
Tasa de personal que realizó innovación de tecnologías.	Porcentaje de empleados que hicieron innovación de tecnologías.
Servicio de internet.	Presencia de servicio de internet en la institución.
Servicio de extranet.	Presencia de servicio de extranet en la institución.
Servicio de intranet.	Presencia de servicio de intranet en la institución.
Servicio de internet.	Variable categórica que muestra si se tuvo servicio de internet.
Operaciones bancarias en internet.	Uso de internet para realizar operaciones bancarias.
Trámites gubernamentales en internet	Uso de internet para realizar trámites gubernamentales.
Compra-venta de servicios o productos en internet.	Uso de internet para realizar compra-venta de servicios o productos.
Búsqueda de información en internet.	Uso de internet para realizar búsqueda de información.
Gestión del negocio en internet.	Uso de internet para realizar la gestión del negocio.
Equipos de cómputo.	Presencia de equipos de cómputo en la empresa.
Tasa de renovación de activos fijos de cómputo.	Porcentaje de adquisición de activos fijos de cómputo con respecto a la venta de activos fijos de cómputo.
Tasa de personal que usa internet.	Porcentaje de personal que usó internet.
Tasa de gastos mercancías en reventa.	Porcentaje de gastos atribuido al pago de mercancías compradas para reventa sobre gastos totales.
Tasa de gastos por materias primas.	Porcentaje de gastos atribuido al pago por compra de materias primas y auxiliares sobre gastos totales.
Tasa de gastos por materiales.	Porcentaje de gastos atribuido al pago por compra de materiales para la prestación de servicios sobre gastos totales.
Tasa de gastos para envases.	Porcentaje de gastos atribuido al pago por consumo de envases y empaques sobre gastos totales.
Tasa de gastos por maquilar.	Porcentaje de gastos atribuido al pago por servicios de maquila sobre gastos totales.
Tasa de gastos de fleteo.	Porcentaje de gastos atribuido al pago de transporte y logística por productos vendidos sobre gastos totales.
Costo unitario de la mano de obra.	Razón de remuneraciones totales sobre ingresos por ventas.
Tasa de gastos de regalías.	Porcentaje de gastos de regalías con respecto a gastos totales.
Tasa de productividad laboral.	Razón de ingresos por ventas sobre horas diarias promedio anualizadas.
Remuneraciones medias.	Razón de remuneraciones totales sobre horas diarias promedio anualizadas
Tasa de pago por liquidación con respecto a la masa de remuneraciones.	Porcentaje de pagos atribuidas a las liquidaciones con respecto a la masa salarial total. Resulta de dividir los gastos que se pagaron por concepto de liquidaciones entre el total de remuneraciones.
Empresa filial.	Muestra si la empresa que proporcionó el personal es filial de la empresa.

Indicadores contruidos como variables de entrada

Indicador	Definición operativa
Proporción salarial del personal de producción.	Porcentaje del pago de salarios al personal de producción, ventas o servicios sobre el total de la masa salarial sin contar liquidaciones.
Proporción salarial del personal administrativo.	Porcentaje de pagos a personal administrativo, contable y de dirección sobre el total de la masa salarial sin contar liquidaciones.
Tasa de personal dependiente.	Porcentaje de empleados contratados directamente por la empresa con respecto al total de empleados.
Tasa de personal dedicado a la producción, venta o prestación de servicios.	Porcentaje de trabajadores dedicados a producir, vender o prestar servicios con respecto al total de empleados.
Tasa de personal dedicado a tareas administrativas.	Porcentaje de empleados que realizan trabajo administrativo con respecto al total de empleados.
Tasa de personal no remunerado.	Porcentaje de empleados a los que no se remunera con respecto al total de empleados.
Tasa de trabajo sin sueldo base.	Porcentaje de trabajadores que no tuvieron un sueldo base con respecto al total de empleados.
Tasa de trabajo femenino.	Porcentaje de mujeres contratadas en la empresa sobre el total de la plantilla laboral.
Tasa de personal femenino no remunerado.	Porcentaje de mujeres propietarias, familiares y otras trabajadoras no remuneradas con respecto al total de propietarios, familiares y otros trabajadores no remunerados. Resulta de dividir la cantidad de mujeres no remuneradas entre el total de trabajadores no remunerados.
Tasa de mujeres sin sueldo base.	Porcentaje de mujeres sin sueldo base con respecto a todos los empleados sin sueldo base.
Tasa de mujeres familiares, becarias o voluntarias sin sueldo base.	Porcentaje mujeres familiares, becarias o voluntarias sin sueldo base con respecto a todos los familiares, becarios o voluntarios sin sueldo base.
Tasa de pago de prestaciones.	Porcentaje de gastos que toma la empresa para prestaciones sociales con respecto a gastos totales.
Tasa de pago de indemnizaciones.	Porcentaje de gastos de la empresa en indemnización con respecto a gastos totales.
Tasa de seguridad social efectivamente pagada.	Porcentaje de pagos atribuidos a la seguridad social sobre el total sin contar liquidaciones.
Tasa de otras prestaciones sociales efectivamente pagadas.	Porcentaje de pagos atribuidos a otras prestaciones sociales sobre el total sin contar liquidaciones.
Tasa de reparto de utilidades con respecto a la masa de remuneraciones.	Porcentaje de pagos atribuidos al reparto de utilidades sobre el total sin contar liquidaciones.

Fuente: elaboración propia.

Índices de precios de renta de la vivienda en México basados en la ENIGH

Rental Housing Price Indexes in Mexico Based on ENIGH

Jesús López-Pérez y Francisco de Jesús Corona Villavicencio*

* Instituto Nacional de Estadística y Geografía (INEGI), jesus.lopezp@inegi.org.mx y franciscoj.corona@inegi.org.mx, respectivamente.

Nota: nuestro agradecimiento a Gerardo Leyva Parra, director general adjunto de Investigación, por sus valiosas aportaciones realizadas; también a Xavier Herrera, quien aportó puntuales comentarios sobre el borrador de este artículo.



Casas de Guanajuato / KIKILOMBO / iStock

Este documento aborda el problema de la construcción de índices de precios de vivienda en alquiler para el mercado de vivienda mexicano, tanto para las rentadas como aquellas ocupadas por sus dueños. Utilizando datos de cinco ediciones de la Encuesta Nacional de Ingresos y Gastos de los Hogares, primero se construye un Indicador de Calidad de las Viviendas (ICV) para reducir la variabilidad de 26 variables categóricas relacionadas con materiales de vivienda, acceso a servicios públicos y equipamiento de esta; luego, presentamos tres metodologías de índices de precios que son construidos a partir del ICV y dos variables geográficas. Este puede reducir, al menos, 77 % de la variabilidad de las variables categóricas, con el cual se puede construir en un modelo parsimonioso para estimar precios de alquiler ajustados por la calidad de la vivienda. El Índice de Precios Hedónicos estima un incremento anual promedio de 2.6 % del 2010 al 2018 a nivel nacional; además, el Índice Estratificado de Precios muestra que las localidades rurales han experimentado mayor volatilidad que las urbanas, mientras que el Índice Espacial de Precios permite afirmar que Campeche, Ciudad de México y Quintana Roo, usualmente, presentan precios más elevados que la entidad de referencia (estado de México), mientras que Durango, Tlaxcala y Coahuila de Zaragoza suelen tenerlos más bajos que esta.

Palabras clave: precios de vivienda; Índice de Precios Hedónicos; *Propensity Score Matching*; Análisis de Correspondencia Múltiple.

Recibido: 6 de julio de 2021.
Aceptado: 31 de marzo de 2022.

Introducción

El mercado de vivienda residencial enfrenta el reto de contar con precios que reflejen la calidad de los activos que, a diferencia de otros mercados, se caracterizan por ser muy heterogéneos; cada vivienda es diferente. Mientras que, por ejemplo, en la construcción de índices de precios de alimentos es posible registrar y dar seguimiento a un mismo producto (pan de caja de cierto tamaño o el kilo de pollo, por ejemplo) y este no varía de ciudad en ciudad, esto no es posible de medir de manera

This paper addresses the problem of constructing rental housing price indexes for the Mexican housing market, both for rented and owner-occupied housing. Using data from five editions of the National Survey of Household Income and Expenditures (ENIGH), we first construct a Housing Quality Indicator (HQI) to reduce the variability of 26 categorical variables related to housing materials, access to public services and housing equipment; then, we present three price index methodologies that are constructed from the HQI and two geographic variables. This can reduce at least 77% of the variability of the categorical variables, which can be used to build a parsimonious model to estimate rental prices adjusted for housing quality. The Hedonic Price Index estimates an average annual increase of 2.6 % from 2010 to 2018 at the national level; in addition, the Stratified Price Index shows that rural localities have experienced greater volatility than urban ones, while the Spatial Price Index allows us to affirm that Campeche, Mexico City and Quintana Roo usually have higher prices than the reference entity (state of Mexico), while Durango, Tlaxcala and Coahuila de Zaragoza usually have lower prices than the reference entity (state of Mexico).

Key words: housing prices; Hedonic Price Index; Propensity Score Matching; Multiple Correspondence Analysis.

directa para los precios de las viviendas, los cuales se ven influenciados, principalmente, por su ubicación geográfica y el conjunto de atributos propios, que también se pueden considerar únicos. Otros aspectos que inciden en la medición son la depreciación (que disminuye su valor), las renovaciones (que lo incrementa) y el hecho de que las transacciones son esporádicas. Así, en la construcción de índices de precios, un aspecto crítico es lograr la comparabilidad en periodos sucesivos de precios transaccionales de bienes y servicios semejantes (Silver, 2016).

En México se han desarrollado índices de precios de vivienda para compra de vivienda nueva y usada (SHF, 2020); sin embargo, en la actualidad no se conoce de estudios en el país relacionados con la construcción de índices de precios de renta que utilicen la información de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH).¹ Si bien existe, como parte del Índice Nacional de Precios al Consumidor, información del genérico *Renta de vivienda* para tres niveles (alta, media y baja), este se realiza con base en un muestreo no probabilístico (INEGI, 2018).

En este trabajo se estiman índices de precios de renta de vivienda en México ajustados por la calidad de las viviendas, con base en tres metodologías rescatadas de Zamudio *et al.* (2020) que pretenden capturar la heterogeneidad de estas: Índice de Precios Hedónicos (IPH), Estratificado de Precios (IEstrP) y Espacial de Precios (IEspP). Los datos provienen de la ENIGH, la cual es bienal (ediciones del 2010 al 2018). Se utiliza la información tanto de viviendas rentadas, de las que se conoce el monto de renta pagado, así como de las utilizadas por sus dueños o prestadas, de las cuales se sabe la cantidad que se estaría pagando si se rentaran (renta imputada) junto con los datos de las características físicas de las viviendas. La aportación de esta investigación consiste en desarrollar un Indicador de Calidad de las Viviendas (ICV) mediante la técnica de Análisis de Correspondencia Múltiple (MCA, por sus siglas en inglés), el cual resume en una sola variable toda la información de las características de las viviendas disponible en la ENIGH. La inclusión de este en los índices de precios permite controlar estadísticamente los efectos de las características únicas de cada vivienda.

El presente estudio es relevante para apoyar el análisis de políticas públicas de vivienda orientadas al uso de inmuebles rentados. En México no se cuenta con programas sociales de apoyo en este sentido, por ejemplo, como los sugeridos en Bergman *et al.* (2019) para apoyar la movilidad social al permitir que familias puedan restablecerse en

localidades con mejor acceso a servicios educativos y de salud. La información que aquí se genera puede ser de utilidad para diseñar programas de apoyo a la renta de vivienda y también para apoyar en la estimación de rentas imputadas para las cifras de contabilidad nacional.

Enseguida de esta introducción, este documento se divide en cinco apartados: en primera instancia se presenta una revisión de literatura; después se describen los datos de la ENIGH y las variables empleadas; en tercer lugar se muestra el Indicador de Calidad de las Viviendas; luego se dan los resultados de los índices de precios de renta estimados; y en la última parte se emiten conclusiones con base en los resultados de los tres métodos.

Mercado de renta de vivienda en México

Nos interesa, en particular, el mercado de vivienda rentada, que en el país asciende a 13 % del total de viviendas (INEGI, ENIGH 2018). En naciones de América Latina, la proporción de hogares que rentan una vivienda alcanza 21 %, mientras para las que pertenecen a la Organización para la Cooperación y el Desarrollo Económicos (OCDE) es de 24 por ciento.²

La política de vivienda en México de recientes administraciones se ha caracterizado por incentivar la demanda mediante el financiamiento a la adquisición de inmuebles a través de las entidades financieras del gobierno: INFONAVIT para trabajadores de la economía formal privada y FOVISSSTE para los del servicio público. Muestra de ello son los fuertes niveles de colocación de créditos hipotecarios: mientras que en el 2000 se otorgaron, aproximadamente, 477 mil financiamientos, en el 2008 llegaron a un máximo de 2 millones 10 mil y, a raíz de la gran recesión de ese año, decrecieron hasta 945 mil en el 2019, antes de la pandemia (SNIIV³).

1 Programa estadístico del Instituto Nacional de Estadística y Geografía (INEGI).

2 OECD. *Affordable Housing Database* – <http://oe.cd/ahd>

3 Sistema Nacional de Información e Indicadores de Vivienda (SNIIV), <https://sniiv.sedata.gob.mx/Dashboard/Inicio>

Por su parte, el mercado de vivienda en renta se ha desarrollado en menor medida. Entre el 2010 y 2020, de acuerdo con los censos de población y vivienda, la proporción de unidades alquiladas pasó de 14 a 17.3 %, un aumento de 3.3 % en 10 años. Además, la misma fuente señala que hay 15.8 % del total de viviendas que son prestadas, lo que supondría la existencia de un mercado informal de rentas. Una dimensión importante de los mercados de renta es que estos se encuentran más desarrollados en localidades urbanas. Según la Encuesta Nacional de Vivienda (ENVI),⁴ el porcentaje de vivienda alquilada en ciudades de más de 100 mil habitantes es de 21.8; en las de 15 mil a menos de 100 mil, de 19.7; en las de 2 500 a menos de 15 mil, de 11.5; y en las más pequeñas, de menos de 2 500 habitantes, se ubica en 4.1.

El arrendamiento de una vivienda se genera cuando este representa una mejor opción para ciertos hogares en los que comprar una propiedad no se ajusta a sus posibilidades. De acuerdo con la ENVI 2020, la distribución porcentual de viviendas rentadas por motivo principal de alquiler es: no tiene acceso a crédito o recursos (51.4), por facilidad para mudarse (22), la mensualidad es menor que la hipoteca (9.7), no le interesa comprar (5.4), prefiere invertir en su persona (3.3), otros (8). De hecho, en ONU-Hábitat e INFONAVIT (2018) se indica que, tomando como referencia el precio promedio del inmueble financiado por los organismos de vivienda, los hogares más pobres tardarían 120 años para pagarlo y más de 47 años para liquidar uno de tipo económico.

La utilización de vivienda desocupada, que en México asciende a 650 mil viviendas deshabitadas (SEDATU, 2019), a través de mecanismos que fomenten la renta tiene beneficios para las ciudades al evitar la edificación de nuevas unidades que demandan servicios públicos e infraestructura y aprovechar el inventario de inmuebles existentes.

4 Programa estadístico del INEGI.

Metodologías

En la literatura de índices de precios se han desarrollado, principalmente, cuatro que permiten hacer compatibles viviendas en diferentes locaciones y con distintas características: regresión hedónica, estratificación, ventas repetidas y basado en avalúos (Eurostat, 2013). Con base en la información disponible proveniente de la ENIGH, en este trabajo se desarrollan los primeros dos: IEstrP e IPH. Además, siguiendo a Zamudio *et al.* (2020), se realiza un ejercicio de IEspP con la metodología de Galvis y Carrillo (2013).

Metodología del IPH

Se emplean regresiones hedónicas, cuyos fundamentos fueron propuestos inicialmente por Lancaster (1966) y Rosen (1974). En el modelo hedónico se supone que los individuos escogen una vivienda con base en el bienestar que esperan recibir de los atributos que la componen, por lo que es posible estimar precios ajustados por calidad, más allá de buscar la estimación directa de cada una de las características.

En el caso particular de viviendas, tenemos que el precio $P_{i,t}$ de una vivienda i en el periodo t es una función de un número fijo, K , de características medidas por las cantidades $Z_{iK,t}$ y un error aleatorio $\varepsilon_{i,t}$ (ruido blanco). Considerando $T+1$ periodos, desde el periodo base 0 hasta el T , tenemos:

$$P_{i,t} = f(z_{i1,t}, \dots, z_{iK,t}, \varepsilon_{i,t}), \quad (1)$$

desde $t = 0, \dots, T$.

La representación anterior es una generalización del modelo *del desarrollador* que descompone el precio de una vivienda solo en dos características: el costo de la tierra (L , usualmente metros cuadrados) y el de la estructura (S , el total de los materiales utilizados para la construcción), de manera que el precio lo podemos definir por la expresión:

$$p_{i,t} = \beta_t L_{i,t} + \gamma_t S_{i,t} + \varepsilon_{i,t}, \quad (2)$$

para $t = 1, \dots, T$, el número de periodos; $i = 1, \dots, N$, el total de viviendas; y $\varepsilon_{i,t}$ es un término de error aleatorio con media 0. En este modelo, los factores L y S pueden ampliarse para considerar otras H características que afectan la calidad de la tierra $X_{i1,t}, \dots, X_{iH,t}$ y otras J características que impactan la calidad de la estructura $Y_{i1,t}, \dots, Y_{iJ,t}$, con lo que llegamos a la expresión:

$$p_{i,t} = \beta_t \left[1 + \sum_{k=1}^K X_{ik,t} \eta_k \right] L_{i,t} + \gamma_t \left[1 + \sum_{j=1}^J Y_{ij,t} \lambda_j \right] S_{i,t} + \varepsilon_{i,t} \quad (3)$$

Como parte de las características de la calidad de la tierra se pueden considerar: si la vivienda se encuentra ubicada en una intersección de calles o en un callejón, cerca de un centro comercial o no y diversas variables *dummy* de ubicación. Las siguientes variables se pueden considerar que afectan la calidad de la estructura: número de recámaras, baños y garajes; antigüedad de la vivienda; tipos de material de piso y techo; presencia de patio; aire acondicionado; etcétera.

Para poder estimar las contribuciones marginales de las características, se utilizan métodos de regresión, por lo que esta última ecuación se reexpresa como un modelo paramétrico. Las dos mejores especificaciones hedónicas son el modelo lineal:

$$p_{i,t} = \beta_{0,t} + \sum_{k=1}^K \beta_{k,t} z_{ik,t} + \varepsilon_{i,t} \quad (4)$$

y el log-lineal:

$$\ln p_{i,t} = \beta_{0,t} + \sum_{k=1}^K \beta_{k,t} z_{ik,t} + \varepsilon_{i,t}, \quad (5)$$

donde $\beta_{0,t}$ y $\beta_{k,t}$ son el intercepto y los parámetros de las características a ser estimadas. Es preferible utilizar el modelo log-lineal cuando se está en presencia de heteroscedasticidad en los errores (varianza no constante en los errores) dado que los precios tienden a distribuirse log-normal. En la

práctica, muchas de las variables explicatorias pueden ser categóricas en vez de continuas, y están representadas por un conjunto de variables *dummy* que toman el valor 1 si la vivienda tiene la categoría en comento o 0 en caso contrario.

Habiendo estimado los parámetros $\beta_0, \beta_1, \dots, \beta_K$, se realizan las predicciones de los precios de una vivienda característica, N^* , para lo cual se hace uso de las características promedio de esta, \bar{x} , y se procede enseguida a la estimación del índice de precios del inmueble como se describe a continuación. En primer lugar, se construye un índice de precios tipo Laspeyres entre el periodo 0 y el t :

$$I_{0t}^L = \frac{\exp(\hat{\beta}_{0t}) \exp\left(\sum_{k=1}^K \hat{\beta}_{kt} \bar{x}_{k0}\right)}{\exp(\hat{\beta}_{00}) \exp\left(\sum_{k=1}^K \hat{\beta}_{k0} \bar{x}_{k0}\right)} \quad (6)$$

Enseguida, se calcula también uno de tipo Paasche:

$$I_{0t}^P = \frac{\exp(\hat{\beta}_{0t}) \exp\left(\sum_{k=1}^K \hat{\beta}_{kt} \bar{x}_{kt}\right)}{\exp(\hat{\beta}_{00}) \exp\left(\sum_{k=1}^K \hat{\beta}_{k0} \bar{x}_{kt}\right)} \quad (7)$$

donde $\hat{\beta}_{00}$ es el coeficiente estimado del intercepto en el periodo base; $\hat{\beta}_{0t}$, el del intercepto en el periodo t ; $\hat{\beta}_{k0}$, el de la i -ésima característica (precio hedónico de la característica i), en el periodo base; $\hat{\beta}_{kt}$, el de la i -ésima característica (precio hedónico de la característica i), en el periodo t ; \bar{x}_{k0} , el valor promedio de la i -ésima característica de la vivienda en el periodo base; y \bar{x}_{kt} , el valor promedio de la i -ésima característica de la vivienda en el periodo t .

Finalmente, con base en ambos índices, se obtiene su media geométrica para obtener el índice tipo Fisher:

$$I_{0t}^F = \sqrt{\left(I_{0t}^L I_{0t}^P \right)}. \quad (8)$$

Metodología del IEstrP

El Índice Estratificado de Precios está basado en la separación del total de viviendas en la muestra en un número m de submuestras o estratos, y permite establecer la importancia o peso que cada una de estas submuestras tiene respecto a todas las demás, el cual podemos representar a través de la siguiente expresión:

$$P_t = \sum_{m=1}^M w_{m,0} P_{m,t} \quad (9)$$

donde $P_{m,t}$ es el precio del estrato m que compara el precio promedio en el periodo t con el de un periodo anterior o periodo base 0, y donde $w_{m,0}$ denota el peso del estrato m en el periodo base. Los pesos representan la ponderación del estrato al total, basada usualmente en el número de viviendas, y que se mantienen fijos respecto al periodo base, que suele ser de un año aun cuando la periodicidad del índice sea mensual o trimestral.

El siguiente paso es agregar los precios y las cantidades con base en los ponderadores en un solo índice de precios; para ello, se emplean el Índice de Laspeyres, P_t^L , en el periodo t , que está dado por:

$$P_t^L(P_0, P_t, Q_0) = \frac{\sum_{m=1}^M P_{m,t} Q_{m,0}}{\sum_{m=1}^M P_{m,0} Q_{m,0}} \quad (10)$$

De manera similar se puede definir la fórmula para el Índice de Paasche, P_t^P , de la siguiente manera:

$$P_t^P(P_0, P_t, Q_t) = \frac{\sum_{m=1}^M P_{m,t} Q_{m,t}}{\sum_{m=1}^M P_{m,0} Q_{m,t}} \quad (11)$$

Con estos dos índices se obtiene el Índice de Fisher, definido como el promedio geométrico de P_t^L y P_t^P :

$$P_t^F(P_0, P_t, Q_0, Q_t) = \left[P_t^L(P_0, P_t, Q_0) \times P_t^P(P_0, P_t, Q_t) \right]^{1/2} \quad (12)$$

Metodología del IEspP

Esta se basa en el ejercicio de Galvis y Carrillo (2013), quienes hacen una aplicación para la vivienda urbana en Colombia. Para calcular índices de precio espaciales, se utilizan también regresiones hedónicas, pero en primer lugar se realiza un procedimiento estadístico, conocido como *matching*, para hacer comparables las muestras entre dos regiones espaciales. Posteriormente, se comparan los precios de una vivienda en la entidad control vis a vis los de una similar en la otra entidad; para ello, se estiman coeficientes para los atributos de la vivienda, y al final se calculan precios promedio que se utilizan para construir el índice.

El procedimiento de emparejamiento empleado es el algoritmo PSM (por sus siglas en inglés, *Propensity Score Matching*) para homologar muestras uno a uno, el cual busca determinar el efecto que ocasiona un suceso sobre un grupo de individuos u observaciones. Fue propuesto por Rosenbaum & Rubin (1983) y Rubin (1976). A través del PSM se pueden identificar viviendas con características similares entre una entidad control y otra, de esta manera se busca que los índices reflejen los diferenciales de precios y no las variaciones en los atributos de las viviendas.

Sea x el vector de covariables para una vivienda en particular y sea z la variable que indica si esta se encuentra en el estado de control ($z = 1$) o en la entidad de emparejamiento ($z = 0$). La medida de propensión (*PS propensity score*), $e(x)$, es la probabilidad condicional de estar en el estado de control dadas las covariables, $e(x) = P(z=1|x)$. De esta manera, con respecto a las viviendas de la entidad control son seleccionadas aquellas en la otra entidad que tienen la mínima diferencia de $e(x)$, es decir, se busca que tengan la misma distribución de x . Para determinar el valor de $e(x)$, se utilizan mode-

los *probit*, cuyos resultados son utilizados a manera de función distancia, es decir, se elige el par de viviendas, una en la entidad control y la otra en la de interés, cuya distancia sea menor. En este sentido, para una vivienda j del estado de México (entidad control) será similar a la vivienda i del estado r si se satisface la siguiente relación, comparando una vivienda a la vez:

$$D(i) = \{j \in EdoMex :: \min(|p_i - p_j|)\},$$

donde $D(\cdot)$ es la contrafactual de la vivienda i en la entidad r , p es el *PS*, y el término *EdoMex* denota la entidad control estado de México.⁵

Una vez que se ha encontrado el conjunto de parejas de viviendas homogéneas se realizan regresiones hedónicas, con la metodología descrita en la sección de índices de precios hedónicos, con lo cual se permite asumir que las diferencias en precio representan el diferencial en el costo de vida de cada entidad o debido a factores de localización. Para comparar los precios entre las entidades, se construyen índices de precios Paasche, Laspeyres y de Fisher, que combina a los dos primeros; la expresión para obtener este resultado es la siguiente:

$$I_{r, EdoMex} = \sqrt{\frac{\exp\left(\hat{\alpha}_{it} + \sum_{i=1}^m \hat{\beta}_{it} \bar{x}_{it}\right) \exp\left(\hat{\alpha}_{it} + \sum_{i=1}^m \hat{\beta}_{it} \bar{x}_{i0}\right)}{\exp\left(\hat{\alpha}_{i0} + \sum_{i=1}^m \hat{\beta}_{i0} \bar{x}_{it}\right) \exp\left(\hat{\alpha}_{i0} + \sum_{i=1}^m \hat{\beta}_{it} \bar{x}_{i0}\right)}} \times 100 \quad (13)$$

donde $\hat{\alpha}_{it}$ es el coeficiente estimado del intercepto en el periodo t ; $\hat{\alpha}_{i0}$, el del intercepto en el periodo base; \bar{x}_{it} , el valor promedio de la i -ésima característica de la vivienda en el periodo t ; \bar{x}_{i0} , el valor promedio de la i -ésima característica de la vivienda en el periodo base; $\hat{\beta}_{it}$, el coeficiente estimado de la i -ésima característica (precio hedónico de la ca-

racterística i), en el periodo t ; y $\hat{\beta}_{i0}$, el de la i -ésima característica (precio hedónico de la característica i), en el periodo base.

Esta expresión permite comparar los precios entre la entidad federativa r y la control, donde entre más bajo sea el valor del índice, más barata es la entidad federativa en comento respecto al estado de México. La comparación relativa de los precios de vivienda que genera el Índice Espacial de Precios de vivienda ayuda a reconocer las disparidades existentes en los precios de vivienda entre las entidades federativas, lo cual puede motivar el diseño de políticas públicas que reconozcan estas diferencias, por ejemplo, para la modificación de políticas de otorgamiento de créditos.

Datos

En este apartado se presenta un resumen de las bases de datos utilizadas tanto para la construcción del ICV como en la elaboración de los índices de precios. Se utiliza la información de la ENIGH para los años 2010, 2012, 2014, 2016 y 2018.

La variable de mayor interés para este trabajo es el importe pagado por el uso de la vivienda. Para la de alquiler corresponde al monto del pago mensual reportado por el encuestado (renta), mientras que para las viviendas ocupadas por sus dueños o prestadas corresponde a la estimación del pago (renta estimada o imputada), que es el precio que el encuestado considera que estaría pagando por un inmueble similar en el mercado, o bien, el importe que cobraría si este no lo estuviese habitando y lo rentara, o en el caso de las viviendas financiadas puede corresponder al pago por concepto de hipoteca.

En las gráficas 1 se muestra una comparación de la distribución de las rentas pagadas y las imputadas.

Se observa que estas son similares, solo se ven diferencias en la parte inferior de la distribución, en la que el importe de la imputada está por debajo del de renta, pero aproximadamente después

⁵ La implementación se realiza en el programa R utilizando la librería *matchit* (Ho, Imai, King, & Stuart, 2011).

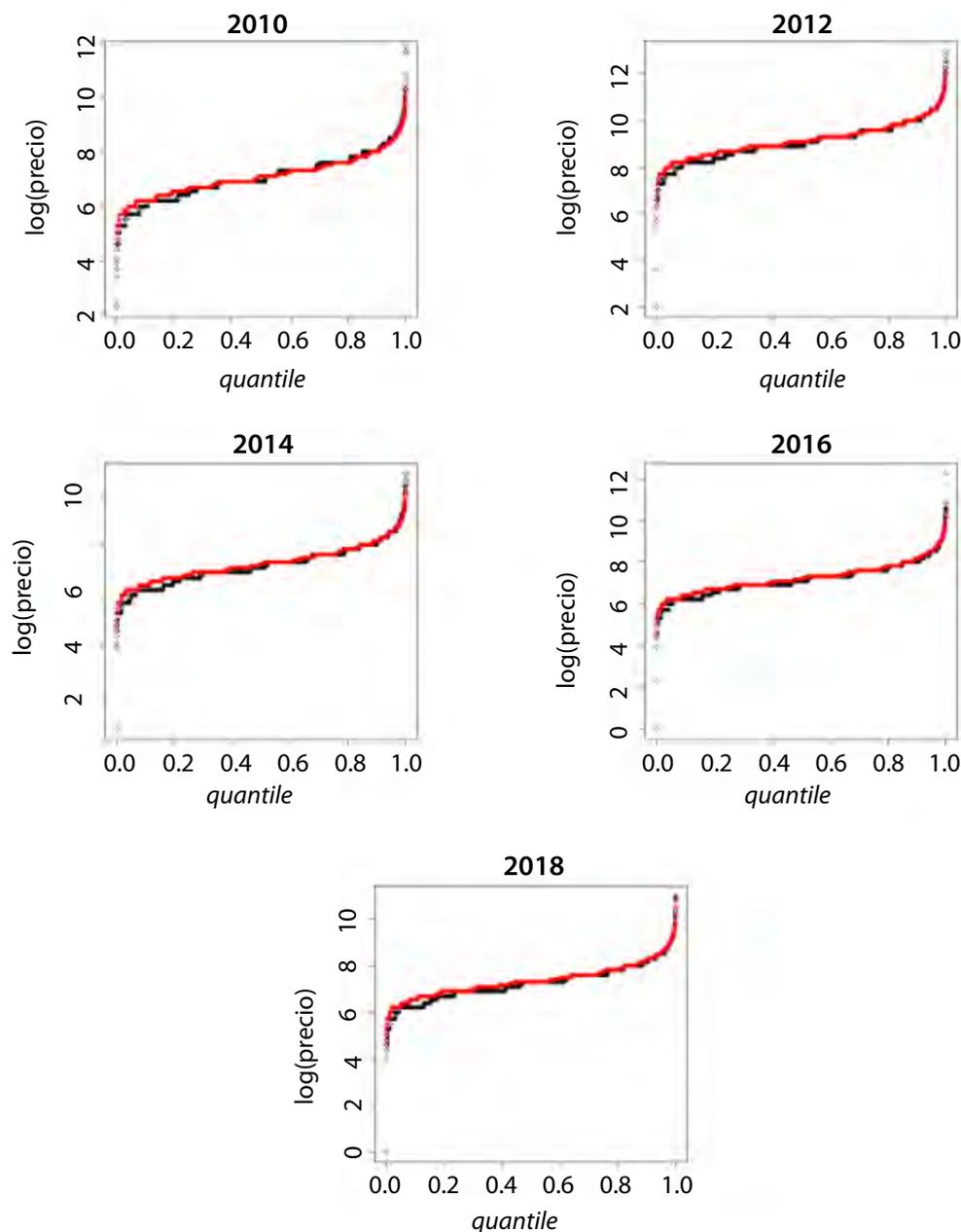
del cuantil 0.25 son muy semejantes; es decir, se puede considerar que, para los niveles inferiores de la distribución, los dueños de los inmuebles subestiman de manera ligera⁶ el valor en el que se podría rentar la propiedad con esas características en el mercado, lo cual no ocurre en la parte alta de la distribución.

Más aún, las rentas imputadas enfrentan el sesgo de reflejar el valor de uso de quien la habita, que está determinado por la provisión del servicio de alojamiento, del acceso a condiciones para higiene personal o para la preparación de alimentos, razón por la que el valor asociado a la producción doméstica de estos servicios podría no necesariamente reflejar el valor de mercado del activo en renta.

6 Entre 3.6 y 6% para el percentil 5 y entre 2.8 y 5.1% para el 10.

Gráficas 1

Distribución de las variables renta y renta imputada en la ENIGH



Nota: los puntos en color rojo muestran los cuantiles de las viviendas rentadas y los que están en negro son los cuantiles de las viviendas ocupadas por sus dueños.
Fuente: INEGI. ENIGH. Ediciones 2010 a 2018.

Los datos de estas cinco ediciones de la ENIGH permiten también considerar diversas variables sobre las características de las viviendas, las cuales podemos agrupar en cuatro grupos: i) edificación de la vivienda (siete: material de techos, paredes y pisos; tipología de la vivienda; uso compartido; número de cuartos; antigüedad), ii) acceso a servicios (seis: disponibilidad de agua, agua en sanitario, drenaje, disponibilidad de electricidad, combustible, eliminación de basura), iii) equipamiento (13: cocina, fregadero, lavadero, regadera, tinaco, cisterna, piletta, calentador, medidor de luz, bomba de agua, tanque de gas, aire acondicionado, calefacción) y iv) ubicación (dos: entidad federativa y tamaño de localidad).

Debido a los cambios metodológicos a través de los años en cada edición de la Encuesta y a la presencia de datos faltantes, se tomaron las siguientes consideraciones:

- Es necesario homologar las categorías contempladas en cada variable con el fin de hacerlas comparables entre sí. En el *Anexo 1* se muestra un resumen de las variables de los grupos i, ii y iii arriba descritos; ahí se observan además las frecuencias de las categorías definidas para cada variable con base en los documentos metodológicos de cada edición de la ENIGH (2010, 2012, 2014, 2016 y 2018).
- Con respecto a los datos faltantes, se identifica esta situación en particular para la variable de antigüedad, donde se tiene que hasta 15 % de las viviendas no cuentan con esta información. Dado que la ausencia de este dato no refleja

cero años, y la antigüedad usualmente está asociada al mejoramiento del inmueble, se decide imputar los datos faltantes.⁷

- Se imputan dichos registros de forma aleatoria a partir de un conjunto de posibles donadores con base en las variables relacionadas con la ubicación geográfica y tipo de vivienda⁸. En el *Anexo 2* se muestran gráficas con el resultado de la imputación de la variable antigüedad.

Con respecto a las variables de ubicación geográfica, se tiene cobertura en las 32 entidades federativas y, además, se cuenta con información por tamaño de localidad.

En el cuadro se muestra un resumen de la distribución de las viviendas por tamaño de localidad.

Indicador de Calidad de las Viviendas

Esta sección está basada en el artículo de Arévalo y Ruiz-Castillo (2006), en el que se realiza un ejercicio de estimación de precios hedónicos para el mercado español de vivienda con base en dos encuestas en hogares de los años 1980/1981 y 1990/1991, y se encuentra que el mercado de rentas puede ser bien representado por un índice de calidad de las viviendas, dos variables sobre la ubicación geográfica de las viviendas y el año de ocupación.

7 La alternativa es dejar fuera esos registros, sin embargo, esto podría generar desbalances en los estratos de la Encuesta, por lo que los resultados no serían consistentes a nivel entidad federativa.

8 Se utiliza la función *match impute* de la librería VIM del paquete R.

Cuadro

Distribución de viviendas por tamaño de localidad

Tamaño de localidad por número de habitantes	Frecuencia (%)				
	2010	2012	2014	2016	2018
>= 100 000	48.6	42.4	46.4	38.3	36.7
>= 15 000 y < 100 000	15.3	16.3	15.1	13.0	12.9
>= 2 500 y < 15 000	14.1	17.7	15.6	13.4	13.2
< 2 500	22.0	23.7	22.9	35.3	37.3

Fuente: INEGI. ENIGH. Ediciones 2010-2018.

De esta manera, con la finalidad de estimar índices de precios de vivienda ajustados por la heterogeneidad de las viviendas se elaboró el ICV para sintetizar la mayor parte de la variabilidad asociada a las características de estas en un reducido número de variables. Dado que las características en las que se desea reducir dimensionalidad se observan en categorías (y no en un continuo numérico), se emplea la técnica MCA, la cual permite reducir múltiples variables categóricas en un reducido número al tiempo que se extrae la mayor variabilidad —ver Husson y Pagès (2017) o Tenenhaus y Young (1985) para una amplia explicación—; esta técnica es análoga a la de componentes principales para variables continuas.

De esta forma, al incorporar el ICV en sustitución de las 26 variables que lo componen (por ejemplo, en la regresión de precios hedónicos), se tiene un modelo parsimonioso que aprovecha al máximo la información de la ENIGH. Además, el Indicador permite capturar la interacción entre estas variables, con lo que, de otro modo, incrementaría considerablemente el número de variables dicotómicas adicionales, generando saturación del modelo, reduciendo el número de grados de libertad.

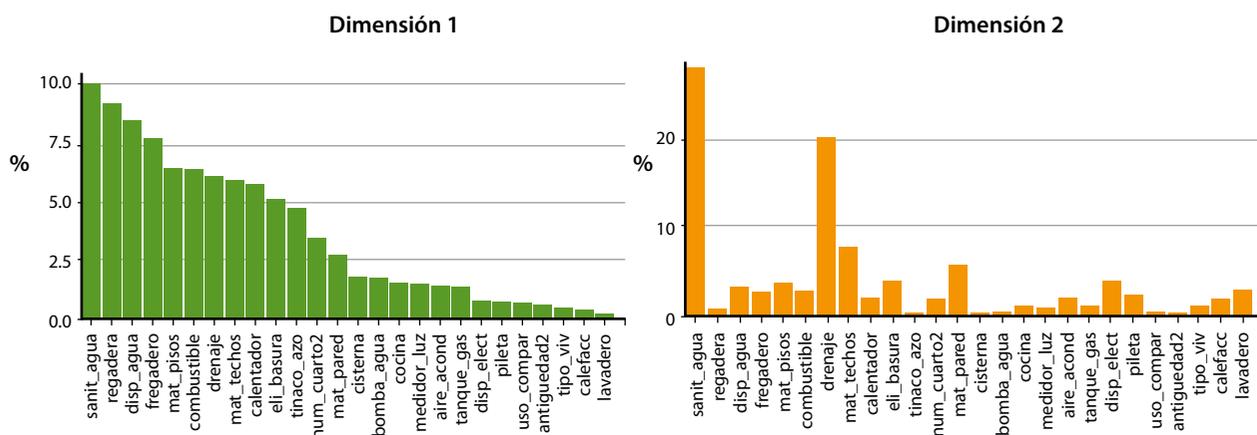
Para la construcción del ICV₉ se consideran variables que están relacionadas con las características de edificación, acceso a servicios y equipamiento, y que se describen en el *Anexo 1*. Las relacionadas con la ubicación de la vivienda se consideran factores determinantes del precio de renta, por lo que estas, entidad federativa y tamaño de localidad no se incluyen en la construcción del Indicador, sino que se incorporan directamente en las diferentes metodologías de precios.

El porcentaje de inercia (varianza explicada) calculado con el criterio de Benzécri (1979) para la primera dimensión se ubica en 79.31, 79.04, 79.08, 76.79 y 76.86 para los cinco años de la Encuesta, mientras que para la segunda asciende a 3.48, 3.07, 2.85, 3.41 y 3.27, respectivamente. Estos porcentajes son razonablemente comparables con el trabajo de Arévalo y Ruiz-Castillo (2006), donde el porcentaje de inercia de los primeros dos factores es de 86.2 y 12.1, en ese orden, para la encuesta de 1980/1981, y 73.2 y 10.5 para la de 1990/1991, para los dos años de encuesta que ahí se estudian. En las gráficas 2 se muestra la relación entre cada una

9 La estimación del MCA se realiza a través de la librería *FactoMineR* del paquete R.

Gráficas 2

Contribución de la variable a las dimensiones 1 y 2 en el MCA para la ENIGH 2018



Nota: el porcentaje representa la contribución de cada variable a las dimensiones 1 y 2.

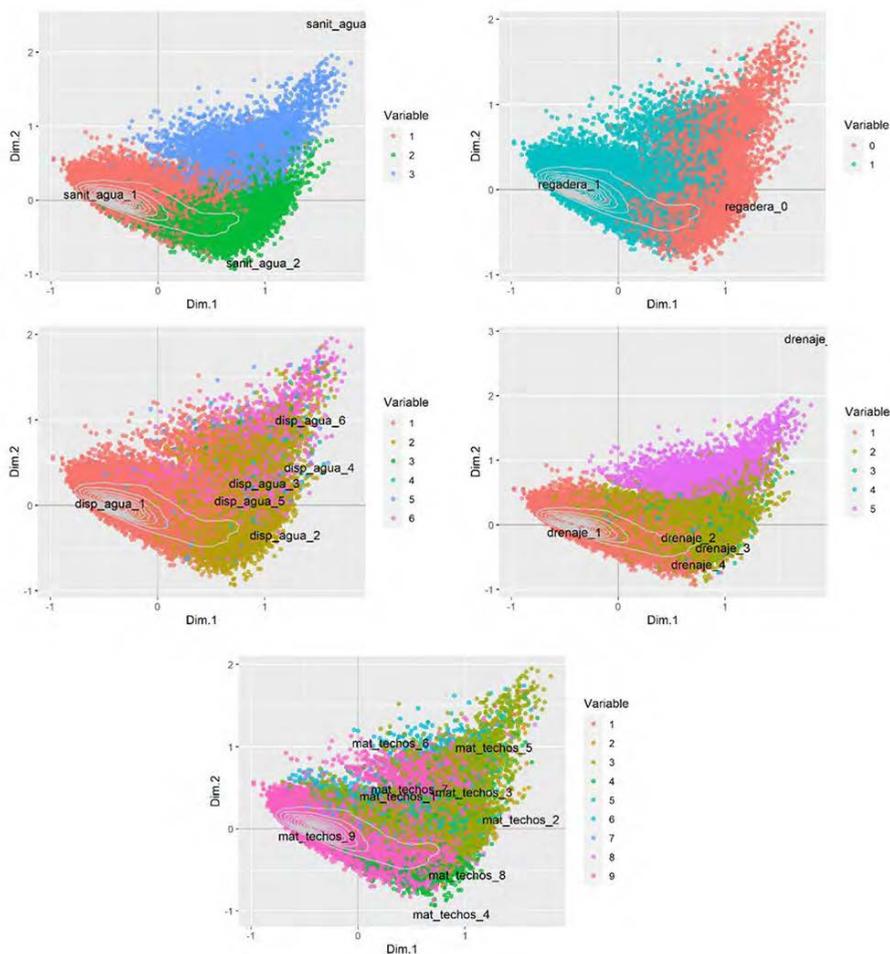
de las variables con las primeras dos dimensiones para la ENIGH 2018. Una interpretación geométrica de estos resultados es la siguiente: los n -datos de las viviendas en el p -espacio se pueden expresar en un nuevo conjunto de coordenadas, que coinciden con los ejes de la hiperelipsoide centrada en la vivienda con característica promedio, y cuyos ejes están dados, son los *eigen*vectores de la matriz de varianzas y covarianzas.

Se puede apreciar que las variables agua en sanitario, regadera y disponibilidad de agua están más asociadas con la primera dimensión, mientras que las de agua en sanitario, acceso a drenaje y material en techos son las que más contribuyen a la segunda.

Además, al observar la distribución de las viviendas en el plano construido a partir de las dimensiones 1 y 2 para las variables más representativas mencionadas en el párrafo anterior, se puede identificar una clara separación entre las categorías asociadas dentro de cada variable. Más aún, se ve que se distribuyen siguiendo una estructura parabólica, que es conocida como efecto Guttman (Greenacre, 1991), el cual revela que, mientras el primer factor resume la estructura de todas las categorías, el segundo muestra una oposición entre categorías extremas (de baja frecuencia) y categorías promedio (alta frecuencia) de una variable. En las gráficas 3 se muestran las variables más representativas mencionadas en el párrafo anterior.

Gráficas 3

Distribución de las viviendas en el plano de las dimensiones 1 y 2 para las variables agua en sanitario, regadera, disponibilidad de agua, drenaje y material en techos



Nota: el orden de las categorías dentro de cada variable corresponde al orden en que se muestran en el Anexo 1.

A partir de las gráficas 2, podemos identificar las características típicas de las viviendas: aquellas más alejadas del centro de gravedad (0,0) son las menos parecidas, mientras que las más cercanas a este son las que ocurren con mayor frecuencia. Con base en el análisis de los resultados, se elige la primera componente como reflejo de la calidad de la vivienda, ya que permite identificar qué características de los inmuebles producen mayor variabilidad y, con ello, diferenciar aquellos que se separan más respecto a una vivienda promedio.

Por último, en la gráfica 4 se muestra la densidad de los indicadores de calidad de las viviendas estimados para las cinco ediciones de la ENIGH (2010-2018).

Se puede observar que la estimación del Indicador es consistente a través de las cinco ediciones de la Encuesta Nacional de Ingresos y Gastos de los Hogares, y genera un continuo que se puede asociar a la calidad de las viviendas, el cual permi-

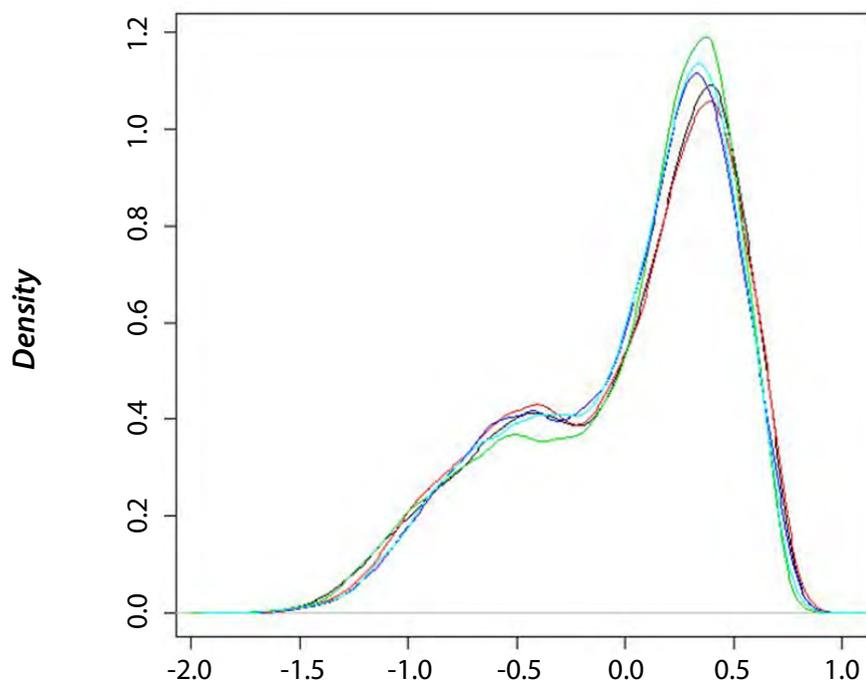
te capturar la información de las 26 variables categóricas con un total de 93 niveles en sus variables.

Resultados de índices de precios

Aquí se muestran gráficamente los índices elaborados bajo las metodologías descritas en la segunda sección. Se dan las variaciones porcentuales bienales por entidad y nacional, mientras que en los anexos 3, 4 y 5 se muestran cuadros con los índices de precios en niveles. Note que, al tratarse de índices, estos son útiles, principalmente, para hacer comparaciones temporales dentro de cada estado, y debe tenerse cuidado en hacerlas sobre la evolución de los índices entre las entidades, pues al estar las cifras normalizadas con respecto a su valor inicial no es válido afirmar que un estado ha crecido/decrecido más o menos que otro, sino que se puede decir que esa entidad ha crecido/decrecido más o menos que otra respecto al valor inicial de cada una de ellas, y no en términos absolutos.

Gráfica 4

Densidades del Indicador de Calidad de las Viviendas para las diferentes ediciones de la ENIGH



Nota: línea negra, 2010; roja, 2012; verde, 2014; azul marino, 2016; y cian, 2018.

Índice de Precios Hedónico

El modelo de precios hedónico incorpora las variables del ICV, entidad federativa y tamaño de localidad. Al incluir el ICV como variable en la regresión hedónica, el porcentaje de varianza explicada se ubica en 44.2, 43.9, 47.3, 46.1 y 48.0 para las cinco ediciones de la ENIGH, respectivamente. Esto compara de manera favorable con la alternativa de introducir en el modelo todas las variables, donde la varianza explicada se ubica alrededor de 41.4 % en promedio para las ediciones de la ENIGH consideradas. Además, las tres variables incluidas resultan significativas. Una vez que se obtienen los coeficientes de la regresión hedónica, se utiliza la mediana de las características promedio $mediana\left(\exp\left(\sum_{i=1}^m \beta_i x_{it}\right)\right)$ y se multiplica por los coeficientes de la regresión multiplicados, así se aprovecha toda la distribución disponible en la muestra que se pudiera perder al utilizar promedios $\exp\left(\sum_{i=1}^m \hat{\beta}_i \bar{x}_{it}\right)$. A través del uso de la mediana se mitiga la problemática que surge ante la presencia de datos atípicos, que no se corrige al utilizar la media de los datos.

En las gráficas 5 se muestran las variaciones porcentuales de los índices de precios hedónicos calculados para las 32 entidades federativas.

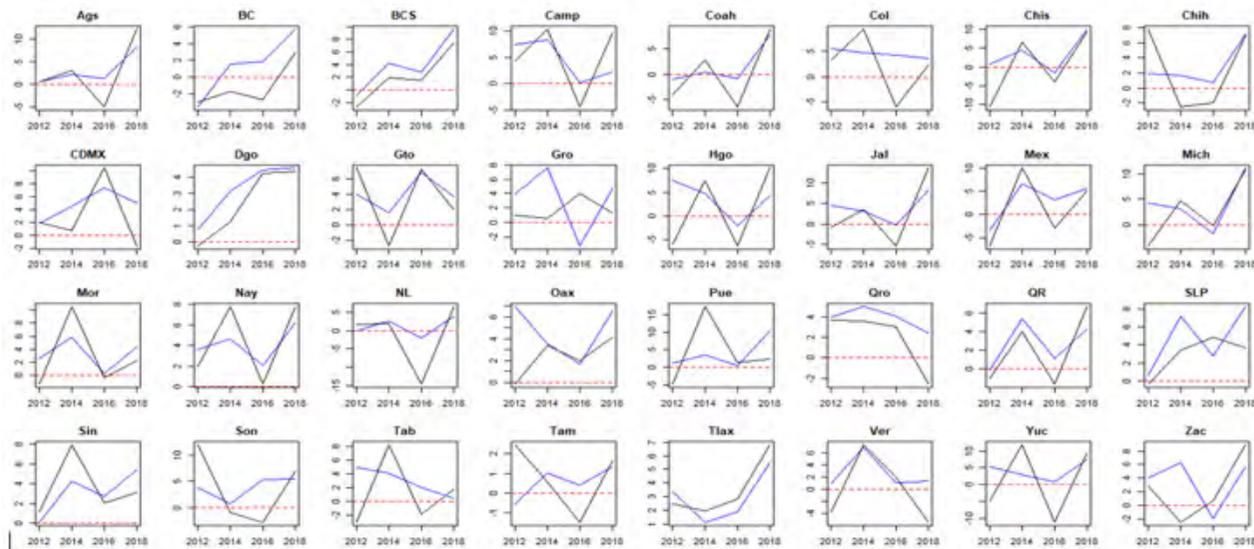
Se observa de estas que los índices de precios (línea azul) muestran una dinámica más suavizada que los precios promedio (negra), es decir, los índices de precios hedónicos, al incorporar la información de las características de la vivienda reflejan de manera más fidedigna las variaciones en los precios de la vivienda en renta. Así, durante el periodo 2010-2018, el aumento promedio de estos para las 32 entidades federativas fue de 29.1 %, lo que implica un incremento medio anual de 2.6 por ciento.

Índice Estratificado de Precios

Para la elaboración de los índices estratificados de precios de renta se definieron cuatro estratos en función de las variables disponibles en la base de datos: i) año de la ENIGH, ii) entidad federativa, iii) cuantil del ICV (0-0.25, 0.25-0.5, 0.5-0.75, 0.75-1) y

Gráficas 5

Índices de precios hedónicos de renta de vivienda, 2010-2018. Variaciones porcentuales bienales. Base 2010 = 100



Nota: la línea azul representa el índice de precios y la negra es la evolución de los precios promedio.

iv) tamaño de localidad (1 = menos de 2 500 hab., 2 = de 2 500 a 14 999 hab., 3 = de 15 000 a 99 999 hab., 4 = 100 000 y más hab.). Además, se utiliza la variable de tamaño de localidad para elaborar dos subíndices para localidades urbanas (con más de 2 500 hab.) y rurales (con menos de 2 500 hab.); y se calculan índices agregados a nivel nacional. Se utiliza el 2010 como año base para el cálculo de los ponderadores, los cuales se mantienen fijos en el cálculo de los índices de los siguientes años; dichos ponderadores se calculan utilizando los factores de expansión de la ENIGH.

En las gráficas 6 se muestra la evolución de las variaciones porcentuales a lo largo del tiempo de los diferentes índices de precios calculados para cada entidad federativa.

De las gráficas 6 podemos ver la variación de precios respecto a los del 2010. Esta metodología permite separar los resultados en localidades urbanas (línea roja) y rurales (azul), y el agregado de

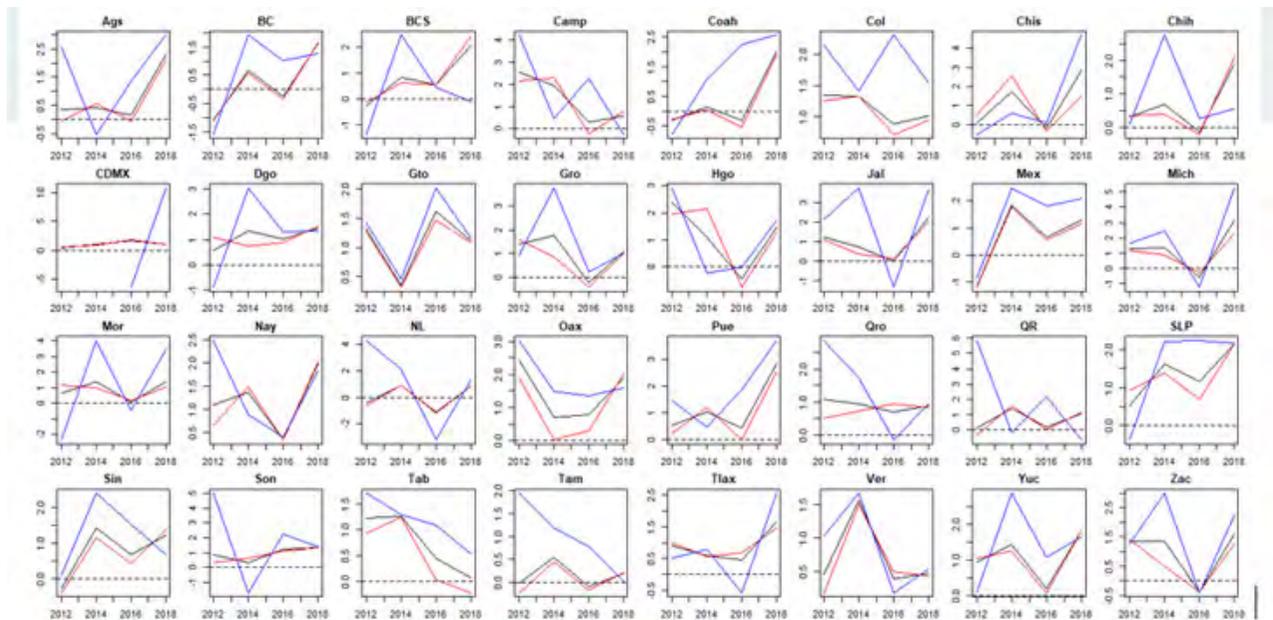
ambas (negra); así, podemos observar que la evolución de los precios total refleja primordialmente la dinámica de las localidades urbanas, ya que estas dos líneas se encuentran usualmente juntas. La variación en localidades rurales es más volátil que en las urbanas. Además, la variación promedio del periodo 2010-2018 para las 32 entidades federativas fue de 3.7 por ciento.

Índice Espacial de Precios

Utilizando la metodología PSM seorean las viviendas en cada entidad con respecto a un inmueble similar en la entidad control, en este caso el estado de México, y se hacen comparaciones de los precios a través del cálculo de índices. Se elige este ya que es una entidad con una amplia variedad de localidades rurales y urbanas. De esta forma, al calcular un índice por encima del 100 indica que, en promedio, los precios en la entidad de análisis están por encima del estado control y, viceversa, cuando están

Gráficas 6

Índices estatificados de precios de renta de vivienda, 2010-2018. Variaciones porcentuales bienales. Base 2010 = 100



Nota: la línea negra representa el índice total de la entidad; la roja, el índice para localidades urbanas; y la azul es para localidades rurales.

por debajo son, en promedio, inferiores a los de la entidad control.

En las gráficas 7 se muestran los índices para los cinco años en que se tiene información de la ENIGH (2010 a 2018).

De las gráficas 7 podemos ver que se marca en color azul la entidad federativa de referencia (México) y el resto de los estados se muestra en orden descendiente respecto a dicha entidad. Destaca que los que en promedio observan precios por encima de la entidad control son Campeche (+18.9 %), Ciudad de México (+17.5 %), Quintana Roo (+15 %), Jalisco (+14.8 %) y Querétaro (+13.3 %); en ese mismo sentido, las entidades cuyos precios son más parecidos a la de control son Colima (+0.5 %), Veracruz de Ignacio de la Llave (+0.5 %), Nuevo León (-0.4 %) y Guerrero (-0.6 %); por su parte, Durango (-21.2 %), Tlaxcala (-19.2 %), Coahuila de Zaragoza (-19.2 %), Chihuahua (-18 %) y Zacatecas (-15.7 %) regularmente registran precios por debajo del estado de México. El resto de las entidades presenta

precios que no muestran una consistencia regular superior o inferior a los de México.

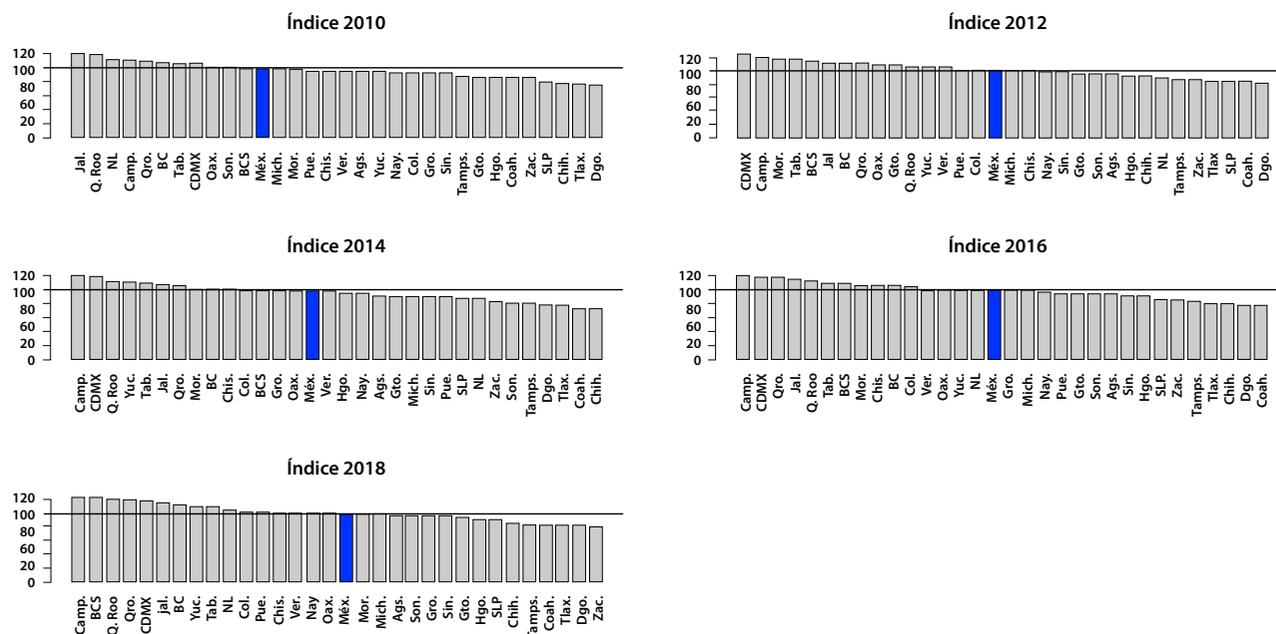
Conclusiones

En esta investigación se cuantificaron las variaciones en el alquiler de vivienda en México por entidad federativa para el periodo 2010-2018; para ello, se utilizaron tres metodologías de índices de precios: Hedónico, Estratificado y Espacial. Se explotaron los datos de la ENIGH, de la cual se obtiene un amplio conjunto de variables sobre las características físicas, la ubicación y el acceso a servicios de las viviendas. La principal aportación del presente estudio es la elaboración del ICV, el cual resume en una sola variable entre 77 y 80 % de la variabilidad asociada a las características de las viviendas, y que se incorpora posteriormente para el cálculo de los índices de precios.

Es clara la ventaja de reducir la dimensionalidad de variables ya que al introducir esta en los mode-

Gráficas 7

Índices espaciales de precios de renta de vivienda, 2010-2018



Nota: las barras indican la relación del precio de cada entidad respecto al estado de México (entidad control).

los para construir los índices de precios se incorpora de manera eficiente la vasta información de las viviendas; es decir, es posible emplear variables subjetivas que reflejen una valoración del bien. Se podría utilizar, por ejemplo, la información de calificaciones asignadas por los usuarios para valorar viviendas o algún otro indicador que refleje la calidad de estas.

Las tres metodologías de índices de precios exploradas en este trabajo permiten abordar la problemática sobre la evolución de los precios de renta de vivienda en México desde diferentes perspectivas, ya sea ajustando por la calidad de los inmuebles o al separar por tamaños de localidad o emparejando viviendas en entidades federativas que pueden estar muy distantes. Los resultados de la investigación sugieren que, con base en el Índice de Precios Hedónico, el alquiler se ha incrementado alrededor de 2.6 % en promedio cada año entre el 2010 y 2018 a nivel nacional. Por otro lado, el Índice Estratificado de Precios permite separar las variaciones de precios con respecto al tamaño de localidad (urbana y rural); así, se puede ver que la evolución de los precios en localidades rurales ha sido más volátil que en las urbanas. Por último, a través del Índice Espacial de Precios, podemos comparar precios de viviendas iguales en dos entidades federativas para un año en particular, y así identificar en qué estados recurrentemente hay inmuebles con precios más elevados con respecto al estado de México (como Campeche, Ciudad de México y Quintana Roo) y en cuáles se tienen con precios inferiores a la entidad de referencia (como Durango, Tlaxcala y Coahuila de Zaragoza).

Los resultados de este trabajo pueden ser de utilidad para los encargados de políticas públicas, por ejemplo, para la asignación presupuestal de apoyos a la renta de vivienda considerando los diferenciales de precios ajustados por la calidad de las viviendas. Por otro lado, pueden servir de referencia para futuras investigaciones sobre alquiler que cuenten con bases de datos de mayor frecuencia, para lo cual se podrían emplear herramientas de web *scrapping* de sitios de bienes raíces en internet. Las limitaciones del presente estudio

son, principalmente, la disponibilidad temporal de información, ya que la ENIGH se encuentra disponible en forma bienal entre 2010 y 2018, por lo que no hay datos para los años intermedios; se podría explorar la incorporación de otras fuentes, como la Encuesta Nacional de los Hogares (aunque solo se tiene para los años 2014, 2015, 2016 y 2017).

Fuentes

- Arévalo, R. y J. Ruiz-Castillo. "On the imputation of rental prices to owner-occupied housing", en: *Journal of the European Economic Association*. 4(4), 2006, pp. 830-861.
- Benzécri, J. P. "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum", en: *Cahiers de l'Analyse des Données*. 4, 1979, pp. 377-378.
- Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz y C. Palmer. *Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice*. Working Paper No. 26164. National Bureau of Economic Research, 2019.
- Eurostat. *Handbook on Residential Property Prices Indices (RPPIs)*. Bélgica, Comisión Europea, 2013.
- Galvis, L. y B. Carrillo. "Índice de precios espacial para la vivienda urbana en Colombia: una aplicación con métodos de emparejamiento", en: *Revista de Economía del Rosario*. 16(1), 2013, pp. 25-59.
- Greenacre, M. "Interpreting Multiple Correspondence Analysis", en: *Applied Stochastic Models and Data Analysis*. 7, 1991, pp. 195-210.
- Ho, D., K. Imai, G. King, & E. Stuart. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference", en: *Journal of Statistical Software*. Vol. 42, Núm. 8, 2011, pp. 1-28.
- Husson, F., S. Lê y J. Pagès. *Exploratory multivariate analysis by example using R*. CRC press, 2017.
- INEGI. *Índice Nacional de Precios al Consumidor. Documento metodológico: base segunda quincena de julio del 2018*. Aguascalientes, México, INEGI, 2018.
- _____. *Nueva construcción de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2010: Descripción de la base de datos*. Aguascalientes, México, INEGI, 2011.
- _____. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2012: Descripción de la base de datos de la nueva construcción de variables*. Aguascalientes, México, INEGI, 2013.
- _____. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2014: Descripción de la base de datos. Nueva construcción de variables*. Aguascalientes, México, INEGI, 2015.
- _____. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2016: Nueva serie, descripción de la base de datos*. Aguascalientes, México, INEGI, 2017.

_____. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2018: Nueva serie, descripción de la base de datos*. Aguascalientes, México, INEGI, 2019.

Lancaster, K. "A new approach to consumer theory", en: *Journal of Political Economy*. 74(2), 1966, pp. 132-157.

ONU-Habitat e INFONAVIT. *Vivienda y ODS en México*. 2018.

Rosen, S. "Hedonic prices and implicit markets: product differentiation in pure competition", en: *Journal of Political Economy*. 82(1), 1974, pp. 34-55.

Rosenbaum, P. R. y D. R. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects", en: *Biometrika*. 70(1), 1983, pp. 41-55.

Rubin, D. "Multivariate Matching Methods that are Equal Percent Bias Reducing. I: Some Examples Biometrics", en: *Biometrics*. 32(1), 1976, pp. 109-120.

Secretaría de Desarrollo Agrario, Territorial y Urbano (SEDATU). *Programa Nacional de Vivienda 2019-2024*. México, SEDATU, 2019.

Sociedad Hipotecaria Federal (SHF). *Índice SHF de Precios de Vivienda*. México, SHF, 2020.

Silver, M. "Real-Estate Price Indexes. Availability, Importance, and New Developments", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. 7(1), 2016, pp. 4-25.

Tenenhuis, M. y F. W. Young. "An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data", en: *Psychometrika*. 50(1), 1985, pp. 91-119.

Zamudio, F. J., M. A. Andrade y G. Becerril. "Un sistema integral de indicadores de vivienda para México", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. 11(2), 2020, pp. 36-51.

Anexos

Anexo 1

Continúa

Descripción de variables de la ENIGH ediciones 2010 a 2018

Núm.	Variable	Categorías	Frecuencia (%)				
			2010 N* = 27 990 507	2012 N* = 29 809 830	2014 N* = 30 700 388	2016 N* = 32 109 725	2018 N* = 33 439 834
Edificación de la vivienda							
1	Material de techos						
		Material de desecho	0.1	0.1	0.3	0.2	0.2
		Lámina de cartón	1.6	1.5	1.2	0.9	0.8
		Lámina de metal	12.6	14.5	13.9	14.7	16.2
		Lámina de asbesto	5.7	5.6	4.5	5.3	5.6
		Palma o paja	0.6	0.7	0.7	0.9	0.8
		Madera o tejamanil	3.1	3.8	3.4	5.0	4.8
		Terrado con viguería	1.2	1.1	0.9	0.9	0.8
		Teja	2.0	2.6	2.1	1.8	1.7
		Losa de concreto o viguetas con bovedilla	73.0	70.2	73.3	70.4	69.1
2	Material de pared						
		Material de desecho	0.1	0.1	0.2	0.2	0.2
		Lámina de cartón	0.3	0.2	0.2	0.2	0.2
		Lámina de metal	0.6	0.6	0.6	0.8	0.7
		Carrizo o bambú	0.4	0.4	0.3	0.2	0.2

Descripción de variables de la ENIGH ediciones 2010 a 2018

Núm.	Variable	Categorías	Frecuencia (%)				
			2010 N* = 27 990 507	2012 N* = 29 809 830	2014 N* = 30 700 388	2016 N* = 32 109 725	2018 N* = 33 439 834
		Embarro	0.4	0.3	0.3	0.4	0.4
		Madera	4.7	5.0	4.4	4.9	5.2
		Adobe	8.0	9.0	7.1	6.9	7.0
		Ladrillo o tabique	85.5	84.5	86.8	86.4	86.2
3	Material de pisos						
		Tierra	3.7	3.0	2.9	3.0	2.9
		Cemento o firme	52.8	55.9	53.5	55.6	56.4
		Madera, mosaico u otro recubrimiento	43.5	41.2	43.6	41.4	40.7
4	Tipología de vivienda						
		Casa única o casa sola	91.9	93.7	95.0	95.0	95.2
		Vertical	5.1	3.5	2.6	2.4	2.3
		Comparte terreno o comparte terreno y construcción o dúplex - hilera	2.8	2.6	1.7	1.9	1.8
		Azotea	0.1	0.1	0.1	0.1	0.1
		No construido para habitación	0.1	0.1	0.7	0.7	0.6
5	Uso compartido						
		Uso solo	93.1	93.1	93.4	92.6	92.9
		Uso compartido	6.9	6.9	6.6	7.4	7.1
6	Antigüedad						
		Menor o igual a 10 años	29.4	27.7	35.0	33.0	29.1
		Mayor de 10 y menor o igual a 20 años	31.3	31.7	29.3	30.6	31.1
		Mayor a 20 y menor o igual a 30 años	21.4	21.2	18.7	18.5	19.2
		Mayor a 30 y menor o igual a 50 años	14.1	15.9	13.9	14.7	17.0
		Mayor a 50 años	3.8	3.5	3.2	3.3	3.7
7	Número de cuartos						
		Menor o igual a 3 cuartos	42.7	44.9	44.0	47.5	48.0
		4 cuartos	26.1	26.6	27.8	27.2	27.4
		5 cuartos	17.3	16.5	16.6	15.2	14.7
		Mayor a 5 cuartos	13.9	12.0	11.7	10.2	9.8

Descripción de variables de la ENIGH ediciones 2010 a 2018

Núm.	Variable	Categorías	Frecuencia (%)				
			2010 N* = 27 990 507	2012 N* = 29 809 830	2014 N* = 30 700 388	2016 N* = 32 109 725	2018 N* = 33 439 834
Acceso a servicios							
8	Disponibilidad de agua						
		Agua entubada dentro de la vivienda	73.2	71.0	73.0	71.7	71.8
		Agua entubada fuera de la vivienda, pero dentro del terreno	20.2	22.0	19.1	20.1	19.9
		Agua entubada de llave pública (o hidrante)	0.3	0.3	0.3	0.3	0.3
		Agua entubada que acarrear de otra vivienda	0.8	0.8	0.7	0.8	0.7
		Agua de pipa	1.1	0.9	1.6	2.3	2.2
		Agua de un pozo, río, lago, arroyo y otro	4.4	5.0	5.3	4.9	5.2
9	Agua en sanitario						
		Tiene descarga directa de agua	68.3	67.0	68.0	66.8	66.5
		Le echan agua con cubeta	24.0	25.5	25.2	25.6	26.2
		No se le puede echar agua	7.7	7.5	6.9	7.6	7.4
10	Acceso a drenaje						
		Red pública	74.4	72.9	73.7	68.1	66.2
		Fosa séptica	18.6	20.6	20.7	25.2	27.3
		Tubería que da a barranca	0.8	0.9	0.9	1.1	1.1
		Tubería que da a río, lago o mar	0.5	0.4	0.5	0.5	0.4
		No tiene	5.7	5.2	4.3	5.2	5.1
11	Disponibilidad de electricidad						
		Servicio público	98.9	98.8	98.8	98.7	98.6
		Planta particular	0.1	0.1	0.1	0.2	0.1
		Panel solar	0.1	0.2	0.2	0.2	0.2
		Otra	0.4	0.4	0.4	0.5	0.6
		No hay luz	0.6	0.6	0.5	0.5	0.5
12	Tipo de combustible						
		Leña	13.7	15.9	16.1	16.3	18.3

Descripción de variables de la ENIGH ediciones 2010 a 2018

Núm.	Variable	Categorías	Frecuencia (%)				
			2010 N* = 27 990 507	2012 N* = 29 809 830	2014 N* = 30 700 388	2016 N* = 32 109 725	2018 N* = 33 439 834
		Carbón	0.4	0.3	0.4	0.3	0.4
		Gas de tanque	77.9	75.1	75.6	75.6	73.8
		Gas natural o de tubería	6.5	6.9	6.7	6.0	5.7
		Electricidad	0.7	0.8	0.7	1.1	1.3
		Otro combustible	0.8	1.0	0.5	0.6	0.6
13	Eliminación de basura						
		La recoge un camión o carrito de basura o contenedor o depósito	86.2	84.6	85.2	83.8	81.7
		La tiran en el basurero público	11.3	12.9	12.2	13.1	14.9
		La queman	1.3	1.4	1.4	1.9	2.1
		La entierran	0.4	0.4	0.4	0.5	0.5
		La tiran en un terreno baldío o calle	0.4	0.5	0.4	0.4	0.4
		La tiran en barranca o grieta o río, lago o mar	0.4	0.4	0.3	0.3	0.4
14	Cocina						
		Disponible	6.7	7.9	6.5	6.8	7.5
		No disponible	93.3	92.1	93.5	93.2	92.5
15	Fregadero						
		Disponible	39.8	41.4	39.7	40.4	41.4
		No disponible	60.2	58.6	60.3	59.7	58.6
16	Lavadero						
		Disponible	13.5	15.8	15.7	17.8	17.9
		No disponible	86.5	84.2	84.3	82.2	82.1
17	Regadera						
		Disponible	30.4	31.2	27.1	27.1	27.7
		No disponible	69.6	68.8	72.9	72.9	72.3
18	Tinaco en azotea						
		Disponible	46.4	48.7	48.2	49.9	48.3
		No disponible	53.6	51.3	51.8	50.1	51.7
19	Cisterna						
		Disponible	79.1	82.6	82.2	83.6	83.7
		No disponible	20.9	17.4	17.8	16.4	16.3

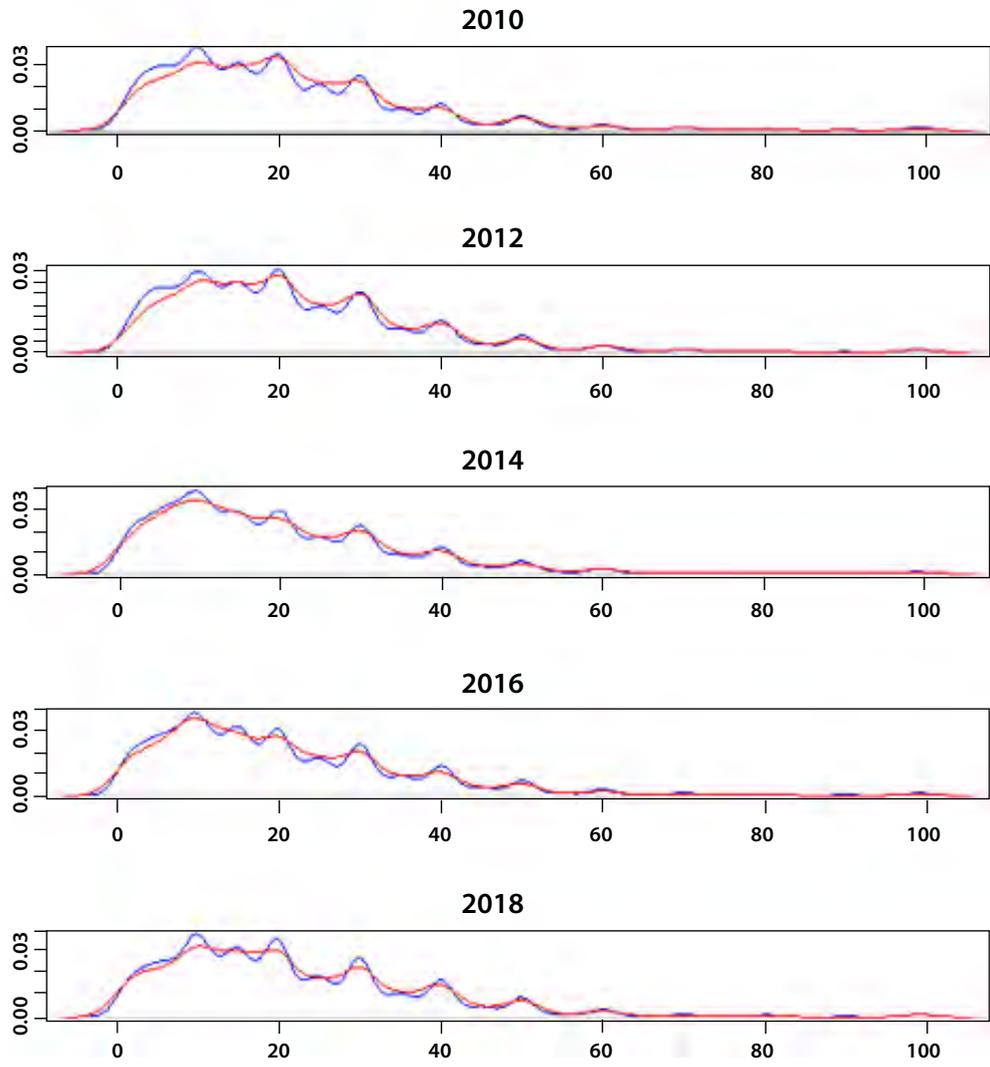
Descripción de variables de la ENIGH ediciones 2010 a 2018

Núm.	Variable	Categorías	Frecuencia (%)				
			2010 N* = 27 990 507	2012 N* = 29 809 830	2014 N* = 30 700 388	2016 N* = 32 109 725	2018 N* = 33 439 834
20	Pileta						
		Disponible	60.4	61.1	64.6	63.7	62.1
		No disponible	39.6	38.9	35.4	36.3	37.9
21	Calentador						
		Disponible	52.9	56.2	57.1	61.4	61.1
		No disponible	47.1	43.8	42.9	38.6	38.9
22	Medidor de luz						
		Disponible	5.5	5.4	6.0	6.9	7.2
		No disponible	94.5	94.6	94.0	93.1	92.8
23	Bomba de agua						
		Disponible	76.4	79.8	78.8	79.1	78.5
		No disponible	23.6	20.2	21.2	20.9	21.5
24	Tanque de gas						
		Disponible	89.4	90.3	90.7	91.6	91.3
		No disponible	10.7	9.7	9.3	8.4	8.7
25	Aire acondicionado						
		Disponible	84.4	83.1	81.0	80.7	80.1
		No disponible	15.6	16.9	19.0	19.3	19.9
26	Calefacción						
		Disponible	98.0	97.3	96.8	97.1	96.9
		No disponible	2.0	2.7	3.2	2.9	3.1

* Cálculos a partir de los factores de expansión asociado a cada hogar en la muestra para calcular el total de hogares a nivel nacional.

Fuente: INEGI. ENIGH. Ediciones 2010, 2012, 2014, 2016, 2018.

Densidades de la variable antigüedad



Nota: la línea roja muestra la densidad de la variable antigüedad imputada y la azul, la densidad de la variable antigüedad de las viviendas no imputadas.

Índice de Precios Hedónico de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Aguascalientes	100.0	100.9	105.4	108.0	125.9
Baja California	100.0	92.8	95.7	99.3	110.6
Baja California Sur	100.0	98.2	106.4	112.4	133.9
Campeche	100.0	115.0	134.4	134.7	140.8
Coahuila de Zaragoza	100.0	97.8	98.6	96.8	111.8
Colima	100.0	110.9	121.2	131.3	140.9
Chiapas	100.0	101.6	110.4	106.6	127.9
Chihuahua	100.0	103.7	107.2	108.6	124.4
Ciudad de México	100.0	103.7	113.0	129.5	142.4
Durango	100.0	101.5	107.9	117.5	128.3
Guanajuato	100.0	107.9	111.3	126.6	135.8
Guerrero	100.0	107.9	124.3	116.3	127.4
Hidalgo	100.0	114.9	125.8	120.3	130.4
Jalisco	100.0	109.0	115.6	115.2	134.1
México	100.0	93.1	105.5	112.1	124.6
Michoacán de Ocampo	100.0	108.6	115.7	111.6	136.8
Morelos	100.0	105.3	117.4	118.2	128.5
Nayarit	100.0	107.3	117.2	122.0	137.3
Nuevo León	100.0	99.9	105.1	100.7	108.3
Oaxaca	100.0	113.8	122.0	126.1	142.4
Puebla	100.0	102.5	109.5	110.6	133.5
Querétaro	100.0	108.1	119.0	128.9	135.2
Quintana Roo	100.0	99.7	110.4	112.9	122.7
San Luis Potosí	100.0	101.1	115.7	122.1	142.4
Sinaloa	100.0	100.2	108.8	114.7	127.1
Sonora	100.0	107.7	109.5	121.2	134.5
Tabasco	100.0	109.9	118.8	123.5	124.6
Tamaulipas	100.0	98.9	100.9	101.7	104.4
Tlaxcala	100.0	106.7	109.0	113.1	125.5
Veracruz de Ignacio de la Llave	100.0	101.9	116.2	118.8	122.1

Índice de Precios Hedónico de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Yucatán	100.0	110.5	117.1	119.1	136.8
Zacatecas	100.0	108.0	121.7	117.0	130.4
Nacional	100.0	103.0	111.5	115.3	128.3

Índice Estratificado de Precios de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Aguascalientes	100.0	100.3	100.8	100.9	103.3
Baja California	100.0	98.9	99.6	99.3	100.9
Baja California Sur	100.0	99.7	100.6	101.1	103.2
Campeche	100.0	102.6	104.6	104.9	105.5
Coahuila de Zaragoza	100.0	99.7	99.8	99.5	101.5
Colima	100.0	101.4	102.7	103.6	104.7
Chiapas	100.0	100.1	101.8	101.6	104.5
Chihuahua	100.0	100.3	101.0	100.8	102.7
Ciudad de México	100.0	100.4	101.5	103.2	104.3
Durango	100.0	100.5	101.9	102.9	104.4
Guanajuato	100.0	101.3	101.7	103.3	104.5
Guerrero	100.0	101.4	103.2	102.9	104.0
Hidalgo	100.0	102.4	103.5	103.0	104.6
Jalisco	100.0	101.2	102.0	101.9	104.2
México	100.0	98.8	100.7	101.3	102.6
Michoacán de Ocampo	100.0	101.3	102.7	102.1	105.3
Morelos	100.0	100.6	102.1	102.1	103.6
Nayarit	100.0	101.1	102.5	102.8	104.9
Nuevo León	100.0	99.6	100.5	99.3	100.2
Oaxaca	100.0	102.4	103.1	104.0	105.9
Puebla	100.0	100.5	101.6	102.0	104.9
Querétaro	100.0	101.1	102.0	102.8	103.7

Anexo 4

Concluye

Índice Estratificado de Precios de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Quintana Roo	100.0	100.2	101.6	101.8	102.9
San Luis Potosí	100.0	100.5	102.2	103.4	105.6
Sinaloa	100.0	99.7	101.2	101.8	103.1
Sonora	100.0	100.8	101.2	102.4	103.8
Tabasco	100.0	101.2	102.5	103.0	103.0
Tamaulipas	100.0	100.0	100.6	100.5	100.7
Tlaxcala	100.0	100.9	101.5	102.0	103.7
Veracruz de Ignacio de la Llave	100.0	100.5	102.0	102.4	102.9
Yucatán	100.0	100.9	102.4	102.6	104.4
Zacatecas	100.0	101.4	102.7	102.3	104.0
Nacional	100.0	100.4	101.5	102.0	103.5

Anexo 5

Contiúna

Índice Espacial de Precios de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Aguascalientes	95.3	93.7	94.0	92.8	96.4
Baja California	108.2	113.6	102.5	105.4	112.0
Baja California Sur	100.4	114.6	101.2	109.7	122.0
Campeche	110.5	118.3	121.9	120.2	123.9
Coahuila de Zaragoza	86.2	83.6	74.3	77.1	82.9
Colima	93.0	101.6	101.4	103.5	103.2
Chiapas	96.2	98.2	101.8	105.5	101.4
Chihuahua	78.9	92.0	73.1	79.3	86.6
Ciudad de México	107.1	125.5	118.3	117.7	118.6
Durango	74.2	82.1	78.0	78.4	81.7
Guanajuato	87.3	94.8	93.3	93.9	93.8
Guerrero	92.8	108.4	100.7	99.9	95.2
Hidalgo	86.5	92.9	96.1	91.4	91.3

Índice Espacial de Precios de Renta de Vivienda

Entidad	2010	2012	2014	2016	2018
Jalisco	120.1	113.7	109.0	115.5	115.7
México	100.0	100.0	100.0	100.0	100.0
Michoacán de Ocampo	99.4	98.3	92.4	97.1	98.6
Morelos	97.7	117.4	102.8	105.5	99.9
Nayarit	94.9	97.7	95.4	96.3	100.4
Nuevo León	114.4	90.2	88.8	100.9	103.5
Oaxaca	103.3	109.0	100.1	101.9	100.3
Puebla	96.2	101.6	91.1	94.0	102.0
Querétaro	109.6	110.4	108.3	117.0	121.2
Quintana Roo	119.7	106.5	114.5	112.5	121.8
San Luis Potosí	79.7	85.5	89.0	87.0	91.2
Sinaloa	92.3	96.5	92.0	91.8	94.9
Sonora	101.3	94.5	82.4	93.0	96.2
Tabasco	107.8	115.4	112.3	111.9	109.8
Tamaulipas	88.4	87.2	82.0	83.4	83.3
Tlaxcala	77.0	86.0	77.9	80.5	82.5
Veracruz de Ignacio de la Llave	96.0	103.7	99.2	102.2	101.3
Yucatán	95.1	104.7	113.5	101.7	110.1
Zacatecas	85.2	86.7	83.3	85.0	81.5

Nota: el Índice representa el precio promedio de la entidad respecto a los precios promedio del estado de México.

Indicadores de agua limpia y saneamiento *en la Ciudad de México y el reto ante un desarrollo sostenible de los recursos hídricos*

Clean Water and Sanitation Indicators *in Mexico City and the Challenge of Sustainable Development of Water Resources*

Berenice García Cabrera*

* Comisión Nacional del Agua (CONAGUA), berenice.garcia@conagua.gob.mx



sweet and clean drinking water/ zms/ iStock

Desde hace muchos años se ha cobrado conciencia de que el recurso hídrico es escaso y está en grave peligro, es decir, en muchas regiones se encuentra en lo que se denomina *estrés hídrico*. Para garantizar su disponibilidad, es importante tener un enfoque integrado; comprender que el problema tiene una naturaleza compleja; y considerar los aspectos sociales, económicos, jurídicos, ambientales y culturales que existen alrededor de esta problemática del agua en la Ciudad de México. Se discute la importancia de la responsabilidad social para un desarrollo sostenible y el rol del administrador ante este desarrollo y se analizan los datos e indicadores actuales en cuanto al acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente en la entidad.

Palabras clave: índices; sostenibilidad; responsabilidad social; administración sustentable.

Recibido: 23 de marzo de 2022.
Aceptado: 3 de junio de 2022.

Introducción

Uno de los aspectos que se consideró al elegir el tema de este artículo fue que el término sostenibilidad es relativamente nuevo en México. El presente trabajo puede aportar elementos que contribuyan a la toma de decisiones para cumplir los Objetivos de Desarrollo Sostenible (ODS) en la Ciudad de México, específicamente el 6: *Agua limpia y saneamiento*.

En este documento se expone la problemática existente en la entidad en torno al agua y cómo el diseño de indicadores apegados a prácticas sostenibles, que han sido aplicadas internacionalmente, puede contribuir a la sostenibilidad del recurso hídrico. También, se presenta un panorama de los conflictos sociales acerca de la disponibilidad de agua en zonas, colonias y demarcaciones territoriales que registran escasez, lo cual confirma la necesidad de contar con indicadores que puedan dar seguimiento a ese ODS.

For many years there has been a growing awareness that the water resource is scarce and in serious danger, i.e., in many regions it is under what is called water stress. To ensure its availability, it is important to have an integrated approach; to understand that the problem has a complex nature; and to consider the social, economic, legal, environmental and cultural aspects that exist around this water issue in Mexico City. The importance of social responsibility for sustainable development and the role of the administrator in this development is discussed and current data and indicators are analyzed in terms of access to drinking water, sanitation facilities, quality and efficient use in the entity.

Key words: indices; sustainability; social responsibility; sustainable management.

Por otro lado, se discute la importancia de la responsabilidad social para un desarrollo sostenible y el rol del administrador ante este; también, se aborda la relación con el concepto de sostenibilidad y su conexión con la administración.

Finalmente, se presentan las conclusiones, donde se destaca la necesidad de tener indicadores y datos que muestren el grado de acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente del agua en la entidad.

Crisis del agua en la Ciudad de México

De acuerdo con datos del Instituto Nacional de Estadística y Geografía (INEGI), esta se consolida como uno de los centros urbanos más importantes del continente americano. Cuenta con una extensión territorial de 1 485 km² (ver figura 1) y, de acuerdo con los datos del INEGI (2017), la urbe re-

presenta 0.1 % del territorio nacional, lo que revela que la entidad es la de menor extensión territorial en comparación con otras. Al norte, este y oeste colinda con el estado de México y al sur, con Morelos (INEGI, 2017).

Figura 1

Ciudad de México



Fuente: INEGI. *Conociendo a la Ciudad de México*. México, INEGI, 2017 (DE) <http://t.ly/F0w8>

La Cuenca del Valle de México, en la cual se asienta la entidad, se localiza sobre la región más elevada de la mesa central del territorio nacional; cuenta con una superficie aproximada de 9 600 km², de los cuales la Ciudad de México ocupa 1 485 km²; esta extensión territorial solo considera la mancha urbana. La cuenca hidrográfica a la que pertenece se considera como cerrada o endorreica, es decir, sus escurrimientos pluviales no tienen salida al mar. Por su topografía, se encuentra rodeada de sierras, montañas y volcanes, de entre los cuales el Ajusco, Popocatepetl e Iztaccíhuatl son los más importantes (Legorreta, 2006).

Abastecer de agua a la megalópolis representa un reto de grandes dimensiones, tanto desde el punto de vista geográfico como del económico y social. La ciudad presenta condiciones geográficas muy particulares, ya que se localiza a más de 2 200 metros sobre el nivel del mar, y ocupa el segundo lugar en el número de habitantes con 9 209 944 personas. El estado de México se sitúa en el primer lugar, con una población de 16 992 418 habitantes

(INEGI, 2020). Suministrar el vital líquido a la población requiere el bombeo constante y permanente, así como recorrer grandes distancias, potabilizarlo y entregarlo a las viviendas. Por esta razón, el agua representa un factor limitante en el desarrollo sostenible de la Ciudad de México. El aumento exponencial de la población, la sobreexplotación del acuífero, lo obsoleto de la infraestructura hidráulica y la carencia de una clara conciencia sobre el valor real del agua han ocasionado que su escasez se registre en la mayoría de las demarcaciones territoriales de la urbe (Torres, 2017).

Así, el acuífero de la entidad no resiste más la sobreexplotación debido a que, para satisfacer las necesidades de los diversos usos, se extrae más agua de la que se recarga. Tal situación ha ocasionado que la ciudad presente hundimientos diferenciados en diversas zonas, lo que, a su vez, ha propiciado rupturas en las tuberías y fugas, así como el desabasto de agua. La infraestructura hidráulica ha rebasado su vida útil (Torres, 2017).

Con base en los registros históricos disponibles, se estima que el hundimiento de la Ciudad de México inició en la década de los 40 del siglo pasado. Actualmente, los datos nos muestran que los hundimientos van de los 6 cm/año a más de 30 en algunas zonas. En los últimos 60 años, el centro de la urbe se ha hundido cerca de 10 metros; estas depresiones han dañado edificaciones, el drenaje profundo, los sistemas de bombeo y la infraestructura hidráulica de abastecimiento (SEDEMA, 2016).

Para revertir la sobreexplotación del acuífero, se necesita que la recarga de este sea mayor a la extracción. Sin embargo, entre otros factores, las razones por las cuales el acuífero de la Ciudad de México tiene un balance hídrico negativo son el incremento desmedido de la población, el desperdicio y la ineficiencia en su aprovechamiento (80 % del agua de lluvia llega al drenaje profundo y se mezcla con las aguas negras, las cuales solo equivalen a 20 %). Por otra parte, la infraestructura hidráulica ha rebasado su vida útil; esto ocasiona que entre 35 y 40 % del agua potable suministrada se pierda en fugas (SEDEMA, 2016).

Situación del recurso hídrico en la Ciudad de México

De acuerdo con el Censo de Población y Vivienda 2020, la proporción de viviendas particulares habitadas que disponen de agua entubada es de 98.9 %; por lo que respecta al drenaje, la cobertura es de 99.7 % (INEGI, 2020). Se debe aclarar que el Censo no especifica si el porcentaje de acceso al agua potable y drenaje son gestionados de manera segura, es decir, líquido disponible en todo momento y exento de contaminación; por otro lado, si las viviendas cuentan con instalaciones de saneamiento que no se comparten con otras familias, en las que los residuos se eliminan de manera segura, estos porcentajes no miden el acceso real, la calidad y la continuidad del agua. Si nos basamos en estos datos, estaríamos evadiendo nuestra realidad y nos alejaríamos de la sostenibilidad del recurso; la verdad es que no todas las personas que viven en la ciudad cuentan con agua todos los días o con instalaciones de saneamiento dignas y seguras.

Por otra parte, de acuerdo con la Encuesta Nacional de los Hogares (INEGI, 2016), del total de hogares que reciben agua entubada dentro de la vivienda o fuera de esta, pero dentro del terreno, se tienen los siguientes datos (ver gráfica 1): 79.7 % recibe agua diariamente; 9.3 %, cada tercer día; 4.3 %, dos veces por semana; 4 %, solo una vez

por semana; y 2.8 %, de vez en cuando o no recibe, es decir, 525 046 hogares no tienen continuidad en el suministro de agua, lo cual equivale a 1 810 486 habitantes. Esto demuestra la falta de accesibilidad y de sostenibilidad del recurso.

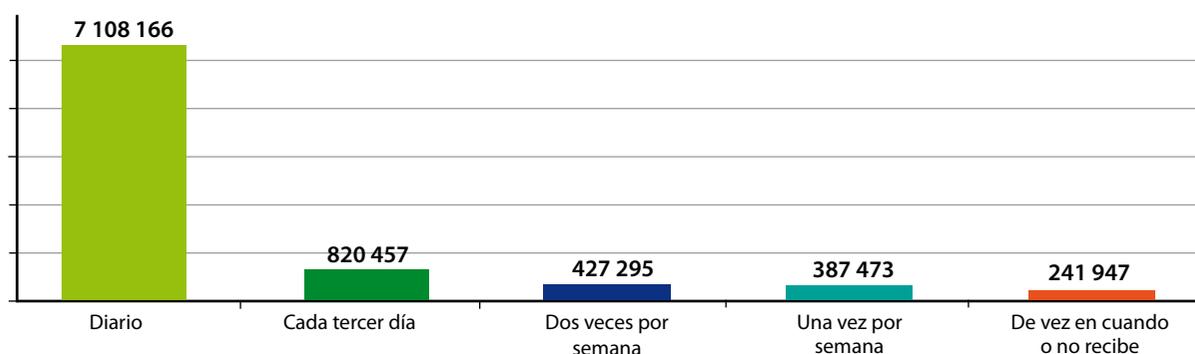
En relación con las descargas de aguas residuales, en la ciudad se generan 22 510 098 litros por segundo; sin embargo, solo 15 % son tratadas por el Sistema de Aguas de la Ciudad de México (SACMEX): 62 % de estas se utilizan para el riego agrícola en la zona sur de la entidad y el restante 38 % se destina a los procesos industriales, riego de áreas verdes y para llenar los canales y lagos, como los de Xochimilco, Bosque de Chapultepec y el de San Juan de Aragón. No obstante que la Ciudad de México cuenta con 25 plantas de tratamiento, se necesitaría una gran inversión por parte del gobierno federal para tratar 100 % de las aguas residuales que se generan (SEDEMA, 2016).

Conflictos sociales relacionados con el recurso hídrico

Para equilibrar el problema de sobreexplotación y escasez del agua en la Ciudad de México, se ha optado por regular la demanda; en otras palabras, la distribución al acceso del recurso y el volumen dotado a la urbe siempre será menor en relación

Gráfica 1

Frecuencia en la dotación del agua



Fuente: elaboración propia con base en INEGI. Encuesta Nacional de los Hogares 2016.

con la demanda, lo cual ha ocasionado que se presenten conflictos relacionados con el agua (Torres, 2017). La distribución de agua en la entidad no es equitativa ni accesible para toda la población: miles de habitantes se ven obligados a buscar otro tipo de fuentes para dotarse del vital líquido, lo cual, sin embargo, no garantiza que el agua tenga una calidad adecuada y, en muchos casos, suele costar más cara (Legorreta, 2006).

La gestión del agua en la ciudad no es ni asequible ni sostenible ni equitativa. Un ejemplo de esto es la batalla que libran algunas mujeres de Iztapalapa: pueden pasar toda la noche esperando pipas; incluso, algunas se suben al camión con los choferes, lo que las expone a situaciones de riesgo. En un reportaje elaborado por *The New York Times* en el 2017 se recolecta la historia de una de esas mujeres: ella comenta que se tiene que formar desde las tres de la mañana para pedir una pipa y que, lamentablemente y por lo general, no siempre alcanza para todos. La misma persona refiere que algunas venden el lugar en la fila para ayudar con el ingreso familiar; agrega que los barrios ricos del otro lado no tienen que molestarse en pensar en el agua (Kimmelman, 2017).

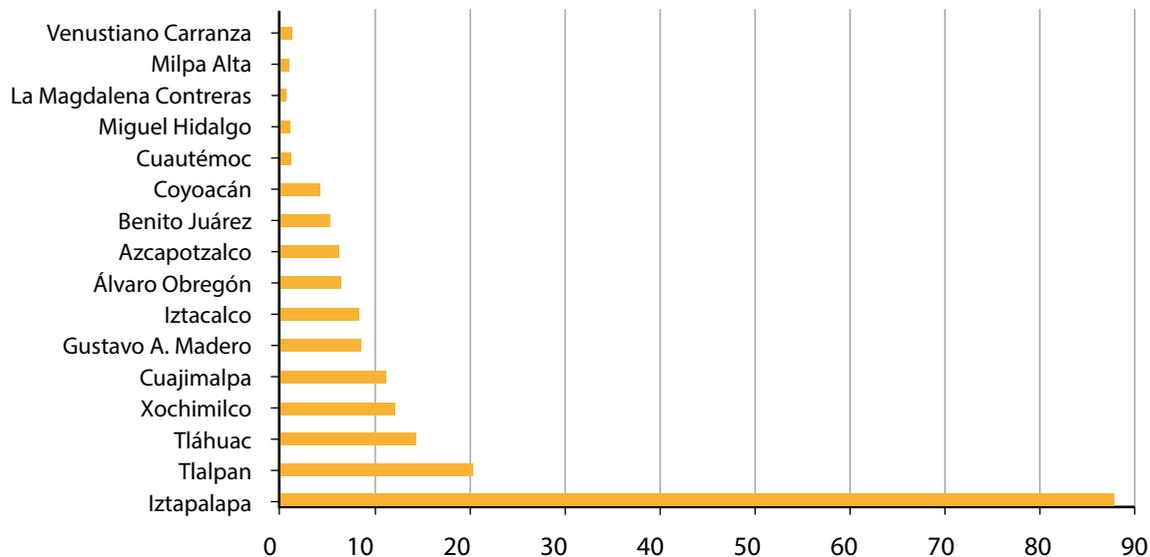
Debido a las situaciones como la anterior, el Instituto Mexicano de Tecnología del Agua (IMTA) ha puesto en marcha el Observatorio de Conflictos por el Agua en México (OCAM). Este sitio web es una herramienta informativa sobre pugnas de relevancia nacional vinculadas con el vital líquido. El objetivo principal del OCAM es generar un banco de datos en relación con las desavenencias sociales sobre el agua del 2007 al 2018.

La base del OCAM cuenta con 7 311 noticias del 2007 al 2018 a nivel nacional. Del total de estas, se identificaron 235 notas relacionadas con la problemática del agua en la Ciudad de México, de las cuales 78 % corresponde a eventos vinculados a conflictos sociales registrados en la urbe (OCAM, 2018), como muestra la gráfica 2.

De acuerdo con la gráfica 2, las demarcaciones territoriales con más conflictos sociales relacionados con los recursos hídricos son Iztapalapa, Tlalpan y Tláhuac; es importante mencionar que estas tres son las que presentan mayor índice de tandeo y la menor frecuencia con la que llega el agua a las tomas domiciliarias; también, son las que más índice de asentamientos irregulares presentan.

Gráfica 2

Conflictos sociales por el agua en la Ciudad de México, 2007-2018



Fuente: elaboración propia con base en datos del Observatorio de Conflictos por el Agua en México.

Responsabilidad social y el reto en el sector público

En opinión de Robbins y Coulter (2010), para entender la noción de responsabilidad social, primero debemos tener en cuenta dos conceptos similares:

- **Obligación social:** es el compromiso que tienen las empresas para realizar acciones sociales, las cuales derivan de su deber de satisfacer las responsabilidades tanto económicas como legales; este concepto refleja la visión clásica de que la responsabilidad social de la administración es la maximización de las utilidades.
- **Sensibilidad social:** atendiendo a las necesidades de la sociedad, la empresa se compromete a realizar acciones dirigidas a estas necesidades, lo cual significa que los negocios van más allá de solo maximizar utilidades. Esta visión socioeconómica incluye la protección y el bienestar de la sociedad.

Los autores definen la responsabilidad social como "... la intención de un negocio, más allá de sus obligaciones legales y económicas, para hacer las cosas correctas y actuar de modo que beneficie a la sociedad..." (Robbins y Coulter, 2010, p. 93). Por su parte, Chiavenato (2014) sostiene que se habla de ese concepto cuando una empresa decide realizar acciones que sean en favor del bienestar de la sociedad y de la organización.

Si la responsabilidad social se concreta en la forma en la que Chiavenato expresa, para que la administración pública pueda adoptar este concepto, se necesita que los servicios que presta el Estado se coordinen tanto con otras instancias estatales como con los sectores privado y social en razón de que muchos de los servicios públicos son monopolísticos, rasgo que dificulta la interacción con los diferentes sectores (Gaete, 2009). Por consiguiente, la responsabilidad social aportará beneficios y avances a las instancias de gobierno debido a que proporcionarán a los servicios las herramientas necesarias para relacionarse con los otros actores.

En este mismo sentido, Corredor (2005, p. 7) sostiene que "... las organizaciones públicas, aunque parezca redundante, en el cambio de paradigma que se perfila, tienen que asumir en su gestión la responsabilidad social, con base en los cuatro factores que comprende el concepto: servicio con calidad, responsabilidad social propiamente dicha, vinculación cooperativa y balance social...". Estos cuatro factores son los siguientes, en palabras más amplias:

1. **Servicio con calidad:** si el Estado presta servicios confiables y de calidad, crecerá su credibilidad y se logrará un crecimiento.
2. **Responsabilidad social propiamente dicha:** las organizaciones públicas fueron creadas para desarrollar acciones encaminadas a eso; sin embargo, en muchos países actualmente existe un déficit en los servicios públicos.
3. **Vinculación cooperativa:** son las relaciones que hacen los gobiernos para que exista una participación social con la finalidad de buscar soluciones a los problemas que enfrentan y generar acciones que garanticen el bienestar de la sociedad.
4. **Balance social:** es la evaluación a la administración pública que indica su cumplimiento ante la sociedad, lo que fortalecerá la relación sociedad-gobierno.

Es importante mencionar que una de las funciones de la administración pública es la relación del Estado con la ciudadanía; por lo anterior, los gobiernos necesitan rediseñar sus procesos con la finalidad de mejorar los servicios que prestan, así como aumentar su transparencia y responsabilidad; de esta forma, se podrá incorporar a sus funciones un comportamiento socialmente responsable.

En suma, la administración pública se enfrenta a un gran reto, pues actualmente la sociedad demanda mejores servicios. Por tal razón, los gobiernos deben desarrollar capacidades más eficientes que les permitan fincar relaciones con los distintos sectores (privado, académico y social) para hallar soluciones a los problemas que enfrentan. Adaptar el concepto de responsabilidad social permitirá a

los Estados atender las necesidades sociales, económicas, culturales o ambientales, lo que traerá un beneficio a la sociedad.

Responsabilidad social y desarrollo sostenible

Hemos señalado que la responsabilidad social son las acciones de las empresas para generar un beneficio en la población y de la propia organización. Tal definición se aplica tanto al sector privado como al público; en consecuencia, si un desarrollo sostenible es tener buenas prácticas que lleven a un equilibrio entre un crecimiento económico, social y ambiental, es factible determinar que ambos conceptos coinciden en dirigir sus esfuerzos hacia un beneficio de la organización y del medio social y ambiental que los rodea.

Así, si esta mancuerna es buscar dicho beneficio, "... las crisis no pueden afrontarse con los mismos principios de simplificación con que fueron creadas [por lo cual] se requiere un planteamiento más profundo para explicar, comprender y atender la emergencia socioambiental..." (Azamar y Parada, 2019, p. 33). En consonancia, podemos afirmar que es necesario rediseñar los procesos y objetivos que contribuyan a frenar o resolver las dificultades que enfrentamos actualmente; por tal razón, es preciso que las organizaciones públicas y privadas sean conscientes de la importancia de incorporar el concepto de responsabilidad social porque con ello podrán generarse políticas que ayuden a revertir o frenar los problemas socioambientales.

Cabe señalar que estos dos conceptos ya se ven en muchos informes empresariales, hecho que revela la disposición de las empresas para mejorar su ejercicio económico y, por lo tanto, para aumentar su competitividad. Además, las organizaciones han optado por diseñar indicadores que muestren su desempeño ambiental y social (Núñez, 2003).

Entonces, hoy en día existe una nueva percepción de lo que es la responsabilidad social y se ha

dejado de ver como acciones filantrópicas. Las organizaciones cada día se preocupan más por los problemas económicos, sociales y ambientales, por lo cual han implementado diversas estrategias que ayudan a resolverlos, y estas, hay que decirlo, están encaminadas a un desarrollo sostenible, es decir, cada día se busca más el equilibrio entre el crecimiento económico, el cuidado al medioambiente y el beneficio de la sociedad.

Para finalizar, es posible afirmar que no existe un desarrollo sostenible sin responsabilidad social porque las estrategias, los objetivos, los programas o las políticas elaboradas bajo el segundo concepto son los que llevarán a las organizaciones al primero; por ende, habrá un crecimiento económico, se tomarán en cuenta las necesidades de la sociedad (presente y futura) y se cuidará al ambiente. En síntesis, es fundamental que tanto el sector privado como el público y social acepten que todas sus acciones repercuten en el medio que los rodea, lo cual significa que debemos actuar de una forma socialmente responsable.

El rol del administrador ante un desarrollo sostenible

En algún tiempo se veía al desarrollo como un crecimiento financiero. No fue sino con la aportación de Elton Mayo cuando comenzó a reconocerse al hombre como un ser social dentro de la organización (citado en Martínez Crespo, 2005). Sin embargo, aún no se consideraba el cuidado del ambiente como parte del desarrollo. Cuando se adoptó el concepto de sostenibilidad, empezó a tomarse en cuenta que el desarrollo de un país no solo era un crecimiento económico, sino que también se busca este en el bienestar de la sociedad y el cuidado de los recursos naturales.

Si bien aún existen problemas ambientales y desigualdad social en el mundo, los países han creado políticas o estrategias socialmente responsables para llegar a un desarrollo sostenible. En este punto, surgen nuevas preguntas: ¿cuál es el rol del administrador ante estos problemas?, ¿cuál

es su aportación ante el desarrollo sostenible? En la siguiente cita, Saavedra (1996, p. 157) da respuesta a estas interrogantes: “Le confiere a la administración la función de apoyo operativo y como acción de gobierno organizacional, un rol reivindicativo en el alcance del desarrollo sostenible, entendiéndolo no sólo en una perspectiva microeconómica de manejo de empresas, corporaciones e instituciones públicas y privadas, sino en una aceptación globalizante de la sociedad como un todo, donde los vertiginosos cambios le dan relevancia y oportunidad...”.

Entonces, el rol del administrador ante un desarrollo sostenible se inserta en su responsabilidad social para buscar y alcanzar el bienestar presente y futuro, así como, la implementación de acciones que contribuyan al cuidado del medioambiente (Saavedra, 1996).

Ahora bien, las teorías clásicas de la administración veían la responsabilidad social del administrador solo como la maximización de utilidades, es decir, como obligación social. Por su parte, Milton Friedman (citado en Saavedra, 1996, p. 157) afirma que, cuando “... los administradores deciden invertir recursos en el bienestar social, esos costos son transferidos a los diferentes agentes del circuito organización-mercado...”; en otras palabras, cuando una organización invierte en crear beneficio social o cuidado en el medioambiente, estos costos se verán reflejados en los costos de sus bienes o servicios, lo que traerá como consecuencia pérdida en las ventas, disminución en las utilidades y, por ende, los trabajadores podrían tener repercusiones, como la baja de salarios o despidos.

Cabe señalar que las teorías administrativas han evolucionado y los administradores han trabajado para adaptar los sistemas administrativos y traer un beneficio tanto a la sociedad como al ambiente. Por ello, en las organizaciones se ha incrementado el concepto de responsabilidad social bajo la visión de generar objetivos, programas o procesos que cumplan con mejorar y proteger el bienestar de la sociedad, el cuidado del ambiente y los intereses de estas (Saavedra, 1996).

De esta manera, los administradores deben ver a la responsabilidad social con una visión integradora y holística, donde puedan integrarse los requerimientos de un desarrollo sostenible. Si en la búsqueda del crecimiento económico tomamos en cuenta las necesidades de la sociedad, así como los efectos que están generándose en el medioambiente, y si logramos un equilibrio entre estos tres aspectos, entonces se podrá llegar a un desarrollo sostenible, lo cual será un gran paso, porque, como mencionamos anteriormente, este no existe sin acciones socialmente responsables.

Para Robbins y Coulter (2010, p. 187), “... la administración es universalmente necesaria en todas las organizaciones, por lo que necesitamos encontrar nuevas vías para mejorar la forma de administrarlas...”. Por consiguiente, los administradores tienen la obligación de encontrar nuevas estrategias —bajo el concepto de responsabilidad social— que ayuden a llegar a un desarrollo sostenible.

Indicadores de sostenibilidad

Un indicador es un valor que resulta de un conjunto de parámetros; este valor simplifica y/o comunica la realidad de una situación compleja (Rodríguez-Ortega y Flores-Martínez, 2008). De acuerdo con la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 1998), se define a los indicadores como: “... un valor con un significado más amplio que el directamente asociado al valor del parámetro mismo...”; ahora bien, la Real Academia Española (RAE) define el concepto como: “Dato o información que sirve para conocer o valorar las características y la intensidad de un hecho o para determinar su evolución futura”.

Ahora bien, el diseño de indicadores relacionados con la sostenibilidad tuvo mayor impulso a partir de 1972 con la Conferencia de las Naciones Unidas y la Cumbre de la Tierra de Río en 1992; también, cobraron más importancia al aprobarse la *Agenda 21*. En estos eventos, se emitieron recomendaciones para su formulación y uso con el propósito de medir en los países los

avances en el cumplimiento de los objetivos planteados (CEPAL, 2018).

De acuerdo con Rodríguez-Ortega y Flores-Martínez (2008, p. 19), "... un conjunto de indicadores provee información clave para dar una imagen clara y completa acerca del estado actual del sistema o fenómeno determinado y, además, proporciona información suficiente para tomar decisiones que permitan dirigir el sistema hacia los objetivos seleccionados y determinar el nivel de éxito de las acciones...". Por lo anterior, podemos decir que un conjunto de indicadores permite hacer perceptible el problema o fenómeno de interés; por otra parte, comunican, miden y cuantifican el objeto de estudio.

A nivel mundial, los indicadores se han convertido en los métodos estándar para medir el progreso de los países; en el caso de este artículo, son para cuantificar los avances de sostenibilidad relacionados con los recursos hídricos. Por sí solo, el valor que resulta de un conjunto de indicadores no arroja mucha información; este resultado debe ser analizado e interpretado por los tomadores de decisiones para establecer si se está o no haciendo un cambio (Rodríguez-Ortega y Flores-Martí-

nez, 2008). También ayudan a comparar las buenas prácticas entre las ciudades o naciones. La OCDE (1998) establece tres criterios que deben tener los indicadores: a) relevancia tanto para quienes toman decisiones como para la población en general, b) solidez analítica y c) que faciliten la medición del fenómeno o problemática observada. En el cuadro 1 se muestran más con detalle estos criterios de acuerdo con la OCDE (1998).

Por lo anterior, podemos decir que los indicadores se pueden utilizar de tres maneras: a) herramientas para explicar un fenómeno o problema observado; b) herramientas piloto, es decir, para dar un primer acercamiento; y c) herramientas de evolución, en otras palabras, para darle seguimiento a través del tiempo al problema o fenómeno.

De acuerdo con Rodríguez-Ortega y Flores-Martínez (2008), los indicadores apegados a los criterios de sostenibilidad pueden interpretarse como las herramientas que ayudan a evaluar el grado de avance en los países con miras a un desarrollo sostenible. Si bien desde los últimos 30 años se ha avanzado en la elaboración de aquellos relacionados con la sostenibilidad, algunas naciones

Cuadro 1

Criterios de los indicadores

Meta	Los indicadores deben:
Relevancia	<ul style="list-style-type: none"> • Proporcionar una visión sobre la presión, estado y respuesta del problema observado. • Fácil interpretación y mostrar tendencias en el tiempo. • Responder a los cambios. • Comparación internacional. • Ser aplicables a nivel nacional o regional, según sea el caso.
Solidez analítica	<ul style="list-style-type: none"> • Estar bien fundados y motivados. • Basarse en consensos internacionales. • Relacionarse con el desarrollo económico y con los sistemas de información.
Medición	<ul style="list-style-type: none"> • Razonable relación costo/beneficio. • Estar bien documentados. • Actualizarse periódicamente.

Fuente: elaboración propia con base en OCDE (1998).

han desarrollado herramientas más eficaces que otras, un ejemplo de ellos es Canadá, que desde finales de la década de los 80 empezó con la iniciativa de este tipo de indicadores; otro es Nueva Zelanda. El trabajo de estos, aunado al apoyo que tienen tanto político como financiero, han generado indicadores de sostenibilidad más desarrollados que otros países. En América Latina se observa que la elaboración de aquellos apegados a los criterios de sostenibilidad está más relacionada con los organismos gubernamentales, sin embargo, el apoyo político y financiero es menor que en otras naciones, lo cual resulta en que la experiencia y la velocidad en su desarrollo sea menor. En esta región, quienes lideran su desarrollo son México, Chile, Colombia, Costa Rica y Brasil (Rodríguez-Ortega y Flores-Martínez, 2008).

Importancia del diseño de indicadores con criterios de sostenibilidad para México

Se mostró la significancia de estos para monitorear y dar seguimiento a los ODS; sin embargo, ¿cuál es la relevancia para México de diseñarlos con este carácter? De acuerdo con Pochat (2008), para un desarrollo sostenible, se debe tener un diagnóstico de la situación actual. Por otro lado, Núñez (2003) comenta que muchas organizaciones han optado por diseñar indicadores que muestren su desempeño ambiental y social.

Ahora bien, en el 2016, Jeffrey Sachs y colaboradores presentaron el informe del índice de los ODS a nivel mundial. El objetivo de este fue evaluar los datos e indicadores con los que contaban los países para darles seguimiento.

De acuerdo con este informe, es de suma importancia tener indicadores y datos confiables para que los ODS se conviertan en instrumentos prácticos para la resolución de problemas; por lo anterior, se basó en la información más reciente que había en 149 de los 193 Estados que se comprometieron a cumplir los ODS (Sachs *et al.*, 2016).

La metodología que utilizaron Jeffrey Sachs y sus colaboradores fue la siguiente:

1. La relevancia de los indicadores para el seguimiento de los ODS y comparables.
2. Confiabilidad de los datos.
3. Si los datos son de acceso público y reciente.
4. Si existen datos por tema específicos (variables de los ODS).

Derivado de lo anterior, el resultado que obtuvo México fue de 63.4; los países que tuvieron mayor puntuación son Suecia, Dinamarca y Noruega, con más de 80 puntos. Por otro lado, los que registraron valores más bajos son Congo, Liberia y República Centroafricana, con menos de 30 puntos (Sachs *et al.*, 2016).

Con base en los datos anteriores, México tiene 63.4 % de posibilidad de cumplir con los Objetivos de Desarrollo Sostenible. Si bien estamos por arriba de muchos países, aún nos falta mucho para cumplir los ODS. Por consiguiente, es de suma importancia que se diseñen indicadores apegados a los criterios de sostenibilidad que nos ayuden a su cumplimiento. En nuestro caso, existe una ausencia de datos a nivel regional. Por otro lado, otra de las aportaciones del informe de Jeffrey Sachs *et al.* (2016) fue la detección de la falta de indicadores y datos específicos para el ODS 6, *Agua limpia y saneamiento*, el cual es objeto de este artículo; por lo anterior, resulta relevante que el país cuente con indicadores para este objetivo.

La importancia de los indicadores o de los datos actualizados no solo es para mostrar un cumplimiento o avance de los ODS; más bien, todos los sectores (público, privado, social y académico) debe cobrar conciencia de que enfrentan problemas socioambientales muy graves y solo con un rediseño de ellos podrán hacer frente a esta problemática y tomar las mejores decisiones en favor de la sociedad y del ambiente.

Por otro lado, México enfrenta un gran reto en cuanto a la implementación de los conceptos de responsabilidad social y desarrollo sostenible. Re-

cordemos que estos van de la mano; si la administración pública los adapta en su día a día, podrán establecerse políticas adecuadas, lo cual dará como resultado diseñar procedimientos adecuados y tener indicadores que puedan medir los logros establecidos.

Perspectiva sobre los datos e indicadores en la Ciudad de México

La información de los indicadores propuestos se buscó a nivel nacional, local e internacional para tener una perspectiva de aquellos a los que se podría tener acceso, su justificación, la unidad de medición, la temporalidad y las variables que usan cada uno. Se revisaron las bases de datos de dependencias gubernamentales (como el INEGI, la SEMARNAT, la CONAGUA, el SACMEX) y otras (por ejemplo, las del Banco Mundial y la ONU-Agua). Los datos generales que se obtuvieron de esta indagación están relacionados con las variables objeto de esta investigación: acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente. Se debe mencionar que, en la búsqueda, se encontró que a nivel nacional se carece de concordancia entre instituciones y no se cuenta con indicadores o datos que puedan servir para el seguimiento y monitoreo del ODS 6.

Un ejemplo de esto es que los datos más recientes con los que se cuenta son de la Encuesta

Intercensal 2015¹ o algunos otros que provienen del Censo de Población y Vivienda realizado por el INEGI en el 2010. Otro es la página de internet de la *Agenda 2030*, cuyo objetivo puede señalarse del siguiente modo: “El Sistema de Información de los Objetivos de Desarrollo Sostenible (SIODS) es una herramienta desarrollada conjuntamente con la Coordinación de Estrategia Digital Nacional de la Presidencia de la República y el Instituto Nacional de Estadística y Geografía (INEGI), y pone a disposición de los usuarios la información sobre el avance en el seguimiento de la Agenda 2030 para el Desarrollo Sostenible, la cual México adoptó como un compromiso de Estado”.

La información que se muestra en la página mencionada en el párrafo anterior tiene carácter oficial y, por lo tanto, debe servir para el diseño de las políticas públicas, así como para los reportes internacionales que rinda el país. En lo que toca a esta página oficial de la *Agenda 2030*, no se tienen todos los datos a nivel local que estén relacionados con el acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente; si bien se cuenta con algunos relacionados con las variables de este artículo, estos también provienen de la Encuesta Intercensal 2015.

En el cuadro 2 se establece una comparación con las variables que utiliza el INEGI en relación con

¹ Programa estadístico del INEGI.

Cuadro 2

Continúa

Indicadores del INEGI y de la ONU-Agua

Indicador	ONU-Agua	INEGI	Observaciones
6.1 Lograr el acceso universal y equitativo al agua potable, segura y asequible para todos.	Agua corriente en las viviendas, patios o parcelas, disponible en todo momento y exenta de contaminación, incluidos hogares, escuelas, centros de atención sanitaria y lugares de trabajo.	Agua entubada dentro de la vivienda, fuera de la vivienda, pero dentro del terreno, llave comunitaria en otra vivienda, de una pipa, pozo, río, arroyo o lago, de recolección de lluvia; excluye locales no construidos para habitación, viviendas móviles y refugios.	De acuerdo con las disposiciones internacionales, el acceso al agua debe estar disponible en todo momento. Se deben incluir escuelas, centros de salud y de trabajo; las variables mencionadas son excluidas por el INEGI. Si no se toman en cuenta todas las variables, no se podrá tener una visión general de la sostenibilidad respecto de este indicador.

Indicadores del INEGI y de la ONU-Agua

Indicador	ONU-Agua	INEGI	Observaciones
6.2 Lograr el acceso a servicios de saneamiento e higiene adecuados y equitativos para todos, y poner fin a la defecación al aire libre, prestando especial atención a las necesidades de las mujeres y las niñas, así como a las personas en situación de vulnerabilidad.	Acceso a instalaciones de saneamiento en sus hogares que no se comparten con otras familias, inodoros con cisterna, fosas sépticas y letrinas (con losa y ventiladas), incluidos hogares, escuelas, centros de atención sanitaria y lugares de trabajo.	Drenaje conectado a la red pública de alcantarillado o fosa séptica, barranca o grieta, río, lago o mar; excluye locales no construidos para habitación, viviendas móviles y refugios.	Al igual que el indicador anterior, el INEGI excluye a las escuelas, centros de salud y trabajo; asimismo, menciona como variable la defecación al aire libre (barranca, grieta, río, lago o mar). Es importante mencionar que una de las iniciativas a nivel mundial es terminar con esta práctica, por lo cual las variables con las que cuenta el INEGI en este indicador no cubren los aspectos de sostenibilidad planteados por la ONU-Agua.
6.3 Mejorar la calidad del agua reduciendo la contaminación, eliminando el vertimiento y minimizando la emisión de productos químicos y materiales peligrosos, reduciendo a la mitad del porcentaje de aguas residuales sin tratar y aumentando considerablemente el reciclado y la reutilización sin riesgos a nivel mundial.	Aguas residuales generadas por los hogares y actividades económicas que se tratan de forma segura. Parámetros de oxígeno disuelto, conductividad eléctrica, nitrógeno, fósforo, nitrato y pH.	Este indicador no se encuentra especificado en el catálogo de indicadores del INEGI; solo se cuenta con la cobertura de agua tratada.	El INEGI solo cuenta con información del 2013 relacionada con las descargas realizadas por los organismos operadores del sector privado y paraestatal. La importancia de contar con este indicador es medir la calidad del agua y aumentar el tratamiento de aguas residuales y la reutilización de estas.
6.4 Aumentar el uso eficiente de los recursos hídricos en todos los sectores y asegurar la sostenibilidad de la extracción y el abastecimiento de agua para frenar su escasez, así como reducir considerablemente el número de personas que padecen escasez de agua.	Valor económico por volumen de agua extraída en metros cúbicos; incluye actividades económicas, con especial atención en la agricultura, la industria y el sector de servicios. Extracción de agua realizada por todas las actividades económicas, y se centra en la agricultura, la industria manufacturera, la electricidad, el tratamiento y el suministro de agua.	Volumen de agua concesionada para usos agrícola, doméstico, industrial y termoeléctricas.	En el catálogo del INEGI solo se cuenta con el volumen de agua concesionada, con el cual se podrían calcular los niveles de estrés hídrico: extracción de agua en proporción con los recursos disponibles; sin embargo, no se cuenta con el cambio en la eficiencia del uso del agua con el tiempo (valor económico).

Fuente: elaboración propia con base en datos del INEGI (2015) y ONU-Agua (2017).

las que propone la ONU-Agua, lo cual demuestra que no existe concordancia entre ambas. En consecuencia, el resultado es que no se pueda dar seguimiento al cumplimiento del ODS 6.

Considerando lo anterior, diseñar y proponer indicadores respecto al acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente es relevante para evaluar el avance ante un desarrollo sostenible del agua en la Ciudad de México y tener indicadores que puedan mostrar el estado en

el que se encuentra la urbe; todo ello, con la finalidad de que los tomadores de decisiones puedan generar estrategias para frenar o revertir la situación actual de los recursos hídricos.

Conclusiones

Como se ha descrito a lo largo de este artículo, la Ciudad de México presenta serios problemas de abastecimiento de agua, hundimientos diferen-

ciados, sobreexplotación del acuífero, infraestructura obsoleta que origina ruptura de tuberías y fugas. Aunado a ello, aún existe una cultura de ilusión de abundancia del recurso agua, lo que ha dificultado tener un servicio eficiente. A ello se suma que las características geográficas de la urbe elevan la complejidad en cuanto al acceso de agua potable, instalaciones de saneamiento, calidad y el uso eficiente.

En 1975, con la publicación del *Plan Nacional Hidráulico*, se integraron los primeros balances de disponibilidad y usos del agua a nivel nacional, regional y local, así como los retos a enfrentar. Hoy por hoy, la Ciudad de México está frente a los umbrales de resentir una severa crisis de agua; de ahí la importancia de administrar y preservar corresponsablemente un recurso vital y escaso.

Cabe señalar la necesidad de contar con indicadores y datos, lo cual se fundamenta en este artículo en razón de que a nivel local hay carencia de indicadores que posibiliten medir el grado de avance ante un desarrollo sostenible en la urbe. Por otra parte, se encontró que los datos oficiales simplemente no existen o, en algunos casos, tienen una antigüedad de 10 años; por lo anterior, no muestran la realidad actual en la que viven millones de habitantes.

Si bien es cierto que los indicadores son una herramienta que ayuda a la planificación y gestión, esta no proporciona una visión sistémica de un desarrollo sostenible; más bien, auxilia en relevar tendencias en el tiempo sobre los fenómenos o cambios en los que se requiere más análisis y una posible acción.

Uno de los aspectos más importantes que se encontró durante la elaboración de este artículo —más que conclusión, sería una recomendación— es el hecho de la falta de información actualizada (los últimos datos sobre agua y saneamiento en la Ciudad de México provienen de la Encuesta Nacional de los Hogares 2016), la inconsistencia entre los pocos datos respecto al desarrollo sostenible en México y la falta de compromiso de los

gobiernos; lo anterior se fundamenta con la información asentada en el informe *Consulta a OSC sobre la Agenda 2030 ante tercer INV del Gobierno de México*; en este se asienta que fue a partir del 2020 cuando se iniciaron los trabajos para la definición y selección de metas e indicadores de la *Agenda 2030*, la cual entró en vigor desde el 1.º de enero de 2016 (Becerra y Pérez, 2021). Si continuamos con la misma tendencia, el deterioro de los recursos naturales —no únicamente de los recursos hídricos— será cada día mayor, y revertir esa situación será imposible, por lo que no solo el gobierno debería tomar conciencia de ello. Es necesario que el sector privado, la Academia y la sociedad entera se involucren para formular políticas públicas socialmente responsables que puedan contribuir de manera eficiente y eficaz en la problemática del agua en la Ciudad de México.

Finalmente, retomamos la reflexión de Azamar y Parada (2019) en cuanto a las diferentes visiones y problemas que enfrentamos ante un desarrollo sostenible. Debemos adoptar una visión sistémica en materia de sostenibilidad y no dejar los esfuerzos solo en un discurso; es necesario cobrar conciencia de que todas nuestras acciones tienen impacto en los recursos naturales y condición de vida. Por lo anterior, el reto como administradores ante un desarrollo sostenible se inserta en la responsabilidad social de buscar y alcanzar el bienestar social presente y futuro, así como implementar acciones que contribuyan al cuidado del medioambiente.

Fuentes

- Azamar Alonso, A. y M. Parada. *Tendiendo puentes para una sustentabilidad integral*. México: Universidad Autónoma Metropolitana. México, Ciudad de México, Universidad Autónoma Metropolitana, 2019, pp. 84-103 (DE) <https://t.ly/MKEz> consultado en el 2021.
- Becerra, Laura y Roberto Pérez. *Consulta a OSC sobre la Agenda 2030 ante tercer INV del Gobierno de México 2021*. Ciudad de México, 2021 (DE) <https://t.ly/a6QG> consultado en el 2022.
- CEPAL. *Agenda 2030 y los Objetivos de Desarrollo Sostenible. Una oportunidad para América Latina y El Caribe, publicación de las Naciones Unidas*. Santiago, Naciones Unidas, 2018 (DE) <https://t.ly/gKaG> consultado en el 2020.

- Chiavenato, Idalberto. *Introducción a la teoría general de la administración*. México, McGraw Hill, 2014.
- Corredor, J. *El cambio de paradigma en la visión ética de la gestión pública*. Santiago, CLAD, 2005 (DE) <https://t.ly/XX4f> consultado en el 2019.
- Diario Oficial de la Federación (DOF). *Acuerdo por el que se da a conocer el resultado de los estudios técnicos de las aguas nacionales subterráneas del acuífero Zona Metropolitana de la Ciudad de México*. México, 26 de septiembre de 2016 (DE) <https://t.ly/ZRvP> consultado en el 2020.
- Gaete Quezada, R. "Aplicaciones de la responsabilidad social a la nueva gestión pública", en: *DIALNET*. Chile, DAAPGE. ISSN-e 1851-3727. 11, 2009, pp. 1-28 (DE) <https://t.ly/wHZp> consultado en el 2020.
- INEGI. *Encuesta Nacional de los Hogares (ENH) 2016*. México, INEGI (DE) <https://t.ly/0iT1q> consultado en el 2021.
- _____. *Censo de Población y Vivienda 2020*. México, INEGI (DE) <https://t.ly/1nRz> consultado en el 2021.
- _____. *Conociendo a la Ciudad de México*. Séptima edición. México, INEGI, 2017 (DE) <http://t.ly/F0w8> consultado en el 2022.
- Kimmelman, M. "Ciudad de México, al borde de una crisis por el agua", en: *The New York Times*. México, 2017 (DE) <https://t.ly/qU1C> consultado en el 2020.
- Legorreta, J. *El agua y la Ciudad de México, de Tenochtitlán a la megalópolis del siglo XXI*. México, Universidad Autónoma Metropolitana-Azcapotzalco, 2006.
- Martínez Crespo, J. *Administración y organizaciones. Su desarrollo evolutivo y las propuestas para el nuevo siglo*. Colombia, Universidad de Medellín, 2005, pp. 8-16.
- Núñez, G. *La responsabilidad social corporativa en un marco de desarrollo sostenible*. Chile, CEPAL, 2003 (DE) <https://t.ly/C5B3> consultado en el 2020.
- OCAM. *Observatorio de Conflictos por el Agua en México*. 2018 (DE) <https://t.ly/HP-g> consultado en el 2022.
- OCDE. *Toward sustainable development: environmental indicators*. París, OCDE, 1998.
- Pochat, V. *Principios de gestión integrada de los recursos hídricos. Bases para el desarrollo de planes nacionales*. Argentina, Global Water Partnership, Central América, 2008 (DE) <https://t.ly/m0oN> consultado en el 2021.
- Real Academia Española (RAE) (DE) <https://dle.rae.es/> consultado en el 2021.
- Robbins, S. y M. Coulter. *Administración*. México, Prentice Hall, 2010, pp. 99-126.
- Rodríguez-Ortega, C. y A. Flores-Martínez. "El Sistema Nacional de Indicadores Ambientales (SNIA)", en: López, B. J. y G. M. de L. Rodríguez. *Desarrollo de indicadores ambientales y de sustentabilidad en México*. México, Universidad Nacional Autónoma de México, Instituto de Geografía, 2008, pp. 15-26.
- Saavedra, S. "La administración frente al desarrollo sustentable", en: *Revista de Economía*. 1996, pp.139-170 (DE) <https://t.ly/P9WW> consultado en el 2020.
- Sachs, J., G. Schmidt-Traub, Ch. Kroll, G. Lafortune y G. Fuller. *Índices y paneles de los ODS*. Nueva York, Bertelsmann Stiftung and Sustainable Development Solutions, 2016 (DE) <https://t.ly/6Elh> consultado en el 2020.
- Secretaría del Medio Ambiente de la Ciudad de México (SEDEMA). *Cuidar el agua es cosa de todos*. México, Subdirección de Procesos Escolarizados de la Dirección de Educación Ambiental de la SEDEMA, 2016 (DE) <https://t.ly/RBCo> consultado en el 2021.
- Torres, L. *La gestión del agua potable en la Ciudad de México. Los retos hídricos de la CDMX: Gobernanza y sustentabilidad*. México, Instituto de Investigaciones Parlamentarias, Distrito Federal, 2017 (DE) <https://t.ly/0gG1> consultado en el 2021.

Identificación de usuarios ansiosos para detectar condiciones de salud mental en la población tuitera de México

Identification of Anxious Users to Detect Mental Health Conditions in the Twitter Population of Mexico

Ana Miriam Romo Anaya, Víctor Silva Cuevas e Irving Gibran Cabrera Zamora*

Basándonos en múltiples estudios que muestran que las palabras que usa una persona en sus escritos transmiten información sobre su personalidad, motivaciones, estado de ánimo y emocional e incluso su nivel económico, podemos entonces pretender que la información subyacente de los usuarios de *Twitter* permite generar indicadores del estado de ánimo e indicadores asociados con la prevalencia de condiciones de salud mental. Como primera aproximación, este trabajo se enfoca en identificar a quienes más expresan sentimientos negativos en sus textos acompañados de palabras relacionadas con ansiedad y que los hacen diferenciables de los demás. Para esto, usamos toda la muestra de tuits recolectados por el INEGI desde el 2017 al 2020 que se utilizó para la construcción del Índice del Estado de Ánimo de los Tuiteros en México junto con el recurso léxico *LIWC*

Based on multiple studies that show that the words a person uses in their writings convey information about their personality, motivations, moods, emotional states, and even economic level, we can claim that the underlying information of Twitter users allows us to generate mood indicators as well as indicators associated with the prevalence of mental health conditions. As a first approximation, this work focuses on identifying those who express more negative feelings in their texts accompanied by anxiety-related words and who differentiate them from others. For this we use the entire sample of tweets collected by INEGI from 2017 to 2020 that was used for the construction of the Mood Index of Twitter users in Mexico along with the lexical resource *LIWC* (Linguistic Inquiry Word Count) designed to analyze psychological implications of word use. This resource

* Instituto Nacional de Estadística y Geografía (INEGI), miriam.romo@inegi.org.mx, victor.silvac@inegi.org.mx e irving.cabrera@inegi.org.mx, respectivamente.

Nota: colaboradores en el área de investigación del INEGI; los juicios contenidos en este artículo son responsabilidad exclusiva de los autores y no reflejan una posición oficial del Instituto al respecto.

(*Linguistic Inquiry Word Count*) diseñado para analizar implicaciones psicológicas del uso de las palabras muy empleado por psicólogos, sociólogos y comunicólogos.

Los resultados nos informan que, a pesar de que la captación de usuarios y sus textos en estos cuatro años está descendiendo, existe una incidencia creciente de personas con textos ansiosos cargados de valencia negativa a través del tiempo. Como un monitoreo de la prevalencia de ansiedad, se construyó un índice para cada tuitero recolectado en el periodo de referencia en las 32 entidades federativas del país. Este nos reveló el impacto que han tenido eventos acontecidos en estos últimos años, como el sismo del 2017 en ciertos estados y el inicio del confinamiento por la pandemia por COVID-19 en el 2020.

Palabras clave: *Twitter*; Índice de Ansiedad; aprendizaje estadístico; ánimo tuitero; *datificación* de la salud; INEGI.

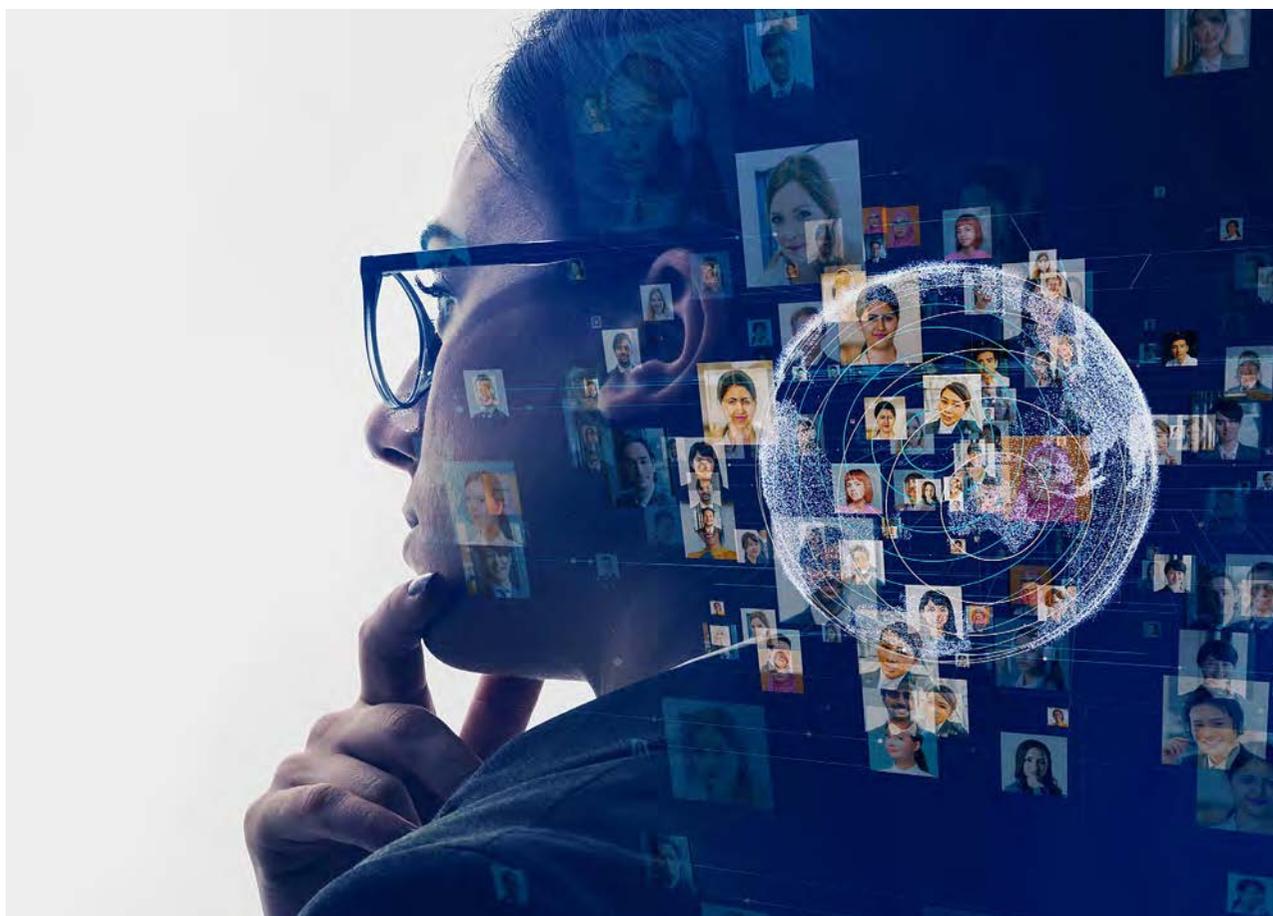
Recibido: 21 de febrero de 2022.

Aceptado: 29 de junio de 2022.

has often been used by psychologists, sociologists and communication scholars.

The results inform us that, although the number of users and texts has decreased in these four years, there is an increasing incidence of people with texts reflecting anxiety, loaded with negative valence over time. As a monitoring of the prevalence of anxiety, an index was constructed for each Twitter user collected in the reference period in the 32 states of the country. This revealed the impact of events that have occurred in recent years, such as the 2017 earthquake in certain states, and the beginning of the COVID-19 pandemic confinement in 2020.

Key words: *Twitter*; Anxiety Index; statistical learning; Twitter mood; datafication of health; INEGI.



Al (Inteligencia Artificial). Red de comunicación.

Introducción

La extracción de conocimiento de manera automática con distintos tipos de datos es un área de la inteligencia artificial que está cobrando cada vez más interés en distintos sectores. Con mucho éxito, se ha realizado una gran variedad de estudios que usan la enorme cantidad de datos disponibles en el mundo de la tecnología y que, junto con las técnicas del aprendizaje estadístico, han logrado diagnosticar y predecir más información de interés con alta precisión.

La tendencia tecnológica que logra convertir o mapear los datos cualitativos obtenidos de las acciones de las personas a través del uso de dispositivos y/u ordenadores a datos cuantificables para ser medidos y descubrir nuevo conocimiento o *insight*¹ se conoce como *datificación*. Gracias a ella, el procesamiento y análisis de datos no estructurados obtenidos de estos medios (texto, imágenes y audio) están ayudando a medir el impacto de eventos relevantes, opiniones políticas, preferencias, comportamientos y sentimientos de los individuos. Es por ello que iniciativas gubernamentales, instituciones y empresas regionales y globales reconocen el valor de estos datos, por lo que están invirtiendo en capital humano e infraestructura tecnológica para analizar y predecir el comportamiento de la población a la hora de comprar un producto o de opinar sobre una decisión política, entre otros aspectos.

En particular, en el mundo de las redes sociales, los datos que se pueden obtener y procesar son inagotables, pues muchos de sus usuarios comparten sus opiniones y emociones día a día. Sus palabras pueden reflejar la personalidad, el estado de ánimo e incluso su nivel económico. Por ello, estas plataformas se consideran una fuente potencial de información para el estudio de una gran variedad de temas.

¹ Es un término utilizado en psicología proveniente del inglés, que se puede traducir al español como *visión interna* o, más genéricamente, *percepción* o *entendimiento*.

El Instituto Nacional de Estadística y Geografía (INEGI), por ejemplo, propuso un indicador que mide el estado de ánimo de los tuiteros en México (aplicación web²) por medio de técnicas de análisis de sentimiento, el cual se obtiene mediante una muestra recolectada diariamente de tuits georreferenciados clasificándolos de manera automática como un mensaje con carga positiva o negativa. Esta información llega a ser útil cuando se relaciona con eventos políticos, sociales, deportivos o económicos que se dan en el país para conocer el sentimiento de cierta parte de la población ante estos sucesos.

Dentro de este mismo enfoque del análisis de sentimiento, se puede determinar no solo la valencia del texto, sino también la carga emocional negativa integrada por miedo, ira, tristeza o asco, entre otras que pueden existir detrás de la información que el individuo expone en sus redes sociales; es un tópico que está contribuyendo de manera importante al campo de la medicina y la psicología, pues su alcance es tal que, si llegan a ser muy intensas, estas pueden afectar la calidad de vida del individuo, mostrando un gran riesgo de padecer un trastorno mental, como depresión y ansiedad, con la posibilidad de llegar al suicidio (Piqueras Rodríguez *et al.*, 2009).

Según Richard Layard (2017), el efecto de un problema de salud mental como la depresión va más allá del bienestar social: afecta la producción en las empresas debido a que este padecimiento prevalece más en las personas en edad laboral. Se hace notable la reducción de la concentración y memoria del trabajador, aumentan sus ausencias y se generan más gastos médicos que, a su vez, merman los presupuestos de las agencias públicas de salud. Toda esta información muestra la importancia no solo de atender los casos ya diagnosticados de padecimientos mentales, sino también detectar las posibles situaciones en riesgo de padecerlos.

Este tipo de estudios se relacionan directamente con la *datificación* de la atención médica, también

² <https://www.inegi.org.mx/app/animotuitero/#/app/multiline>

llamada *datificación de la salud* (Ruckenstein et al., 2017), que se desarrolla con diferentes registros, como la investigación médica basada en datos (de gobierno, biobancos), la infraestructura de la salud pública, la monitorización continua de pacientes, biosensores implantables, el uso de internet para la interacción médico paciente, prácticas de autocuidado, dispositivos de salud, así como *fitness* portátiles y aplicaciones para teléfonos inteligentes. Estos datos recolectados son de gran interés para el gobierno, los especialistas o las instituciones públicas de la salud, pues reconocen su potencial para prevenir y mitigar las cargas físicas y financieras de las *enfermedades del estilo de vida actual*, como obesidad, diabetes y enfermedades cardiovasculares. Además, proporciona evidencia para la generación de programas o políticas de salud pública, y a los profesionales de la salud les ayuda a un temprano diagnóstico para enfocarse en dar pronto seguimiento a quienes necesiten más atención.

El estudio presentado en este artículo pretende generar información relevante acerca de la emoción negativa *ansiedad* en los mensajes de los tuiteros en México, identificando *el grado de ansiedad* y *en qué momentos se incrementa* como indicadores en el tema de salud pública y bienestar social utilizando datos no tradicionales (los cuales no se obtienen por medio de censos o encuestas).

Consideramos que los tuits georreferenciados a los que accede el INEGI desde el 2016 son una base de datos suficiente para lograr estimaciones sobre los patrones de escritura de los tuiteros, que indicarán la vulnerabilidad de expresar emociones negativas que los hace diferenciable de los demás. Además, nos permite realizar comparaciones de tipos de tuiteros por unidad geográfica a través de los años.

Como punto de partida, decidimos empezar a identificar la *ansiedad* que se puede ver reflejada en la escritura de los tuiteros tomando como hipótesis de que esta forma parte de las emociones negativas presentes comúnmente en padecimientos mentales. Algunos psicólogos (Piqueras

Rodríguez et al., 2009) indican que la ansiedad es el resultado de una expresión patológica del *miedo* y mencionan que "... La expresión conductual es la confrontación, la reacción agresiva y lleva al individuo, para mitigar su malestar, a conductas no saludables, como beber, fumar y no alimentarse sanamente...". Además, la Organización Mundial de la Salud (OMS) ha manifestado varias veces que tanto la depresión como la ansiedad son consideradas como enfermedades mentales y que estos padecimientos se están convirtiendo en un problema mundial, pues sucede en países ricos y pobres (OMS, 2021).

Realizar esta medición complementará la información que se obtiene del Índice del Estado de Ánimo de los Tuiteros en México pues, aparte de cuantificar el sentimiento negativo de los textos de cada tuitero, identifica el grado de ansiedad que los hace diferenciables de otros.

La estrategia consistió en asociar un valor a cada tuitero, el cual cuantifica qué tanta ansiedad negativa o positiva expresa en sus textos. Hacemos un comparativo por entidad geográfica y temporalidad al comparar este índice de todos los tuiteros recolectados para el 2017, 2018, 2019 y 2020. Realizar esta comparación ayudó a identificar el impacto que tienen ciertos eventos políticos, económicos y sociales en la población.

Además, perfilamos aquellos tuiteros que suelen expresar más palabras de ansiedad con polaridad negativa, ya que estos pueden ser candidatos idóneos para ser evaluados por psicólogos para asignar de manera cualitativa una propensión a padecer o no alguna condición de riesgo en cuanto a su salud mental. Para descubrir la incidencia de este tipo de usuarios, su tipificación se llevó a cabo por año.

Conscientes estamos de que se deben considerar factores que no siempre son detectados de forma automática para decidir si un usuario tiende a tener un problema real de salud mental; por ejemplo, la detección de ciertas emociones no es altamente precisa debido a que el tono de la expresión

textual puede ser interpretada de formas distintas según la zona geográfica de donde provenga el tuitero. Hay regiones en este país en las que sus habitantes suelen ser más sarcásticos o bromistas que en otras. Es entonces muy importante ratificar de manera cualitativa que el grupo de tuiteros que detectemos con métodos cuantitativos en realidad esté padeciendo ansiedad, por lo que expertos en psicología y/o comunicación deberán validar la detección.

En los dos siguientes apartados se relata cómo se están aplicando las técnicas estadísticas y recursos léxicos con los datos de las redes sociales para pronosticar el riesgo de que un usuario pueda estar padeciendo un problema de salud mental con base en lo que expresan en dichas plataformas.

Sobre los datos que usamos y su procesamiento se detalla en la tercera parte del este documento, continuando con un análisis exploratorio de estos donde se reflejan tendencias a expresar más ansiedad en ciertos meses relacionados con eventos políticos y sociales.

En las dos penúltimas secciones hablamos de la construcción del Índice de Ansiedad, así como la aplicación de métodos de aprendizaje no supervisado y supervisado para la detección de los usuarios con mayor ansiedad; por último, presentamos algunas conclusiones y consideraciones.

Técnicas de aprendizaje estadístico para la obtención de información

Las empleadas en el campo de la inteligencia artificial son algoritmos estadísticos y matemáticos que ayudan a identificar de manera automática patrones en datos y a resolver problemas complejos, por ejemplo, de *clasificación* y *agrupamiento* en distintos sectores de la ciencia.

Uno de los pioneros en aprovechar los modelos estadísticos para descubrir patrones de salud mental de los usuarios a través de redes sociales fue De Choudhury (2013), quien recopiló datos mediante

una base de usuarios diagnosticados clínicamente con depresión y que, a su vez, estuvieran inscritos en *Twitter*. A través de sus publicaciones un año antes de ser diagnosticados, se midieron los estilos lingüísticos, los atributos de sus propios grupos sociales (por medio de gráficos de redes), la carga emocional de sus textos y los tópicos a los que hacen referencia. Con un método de aprendizaje supervisado, alcanzó 72 % de precisión en el diagnóstico de depresión.

Actualmente, el autor David Brooks, columnista de *The New York Times*, señala en su artículo “How Artificial Intelligence Can Save Your Life...” (2019) que estos métodos son: “... [técnicas] aterradoras, difíciles de entender, y bastante asombrosas pues ayudan a salvar la vida de personas, ya que con estas técnicas se puede detectar y diagnosticar a tiempo a personas que intentarán suicidarse...”. Hace referencia a varios estudios que han usado estas técnicas para el diagnóstico temprano de padecimientos mentales.

El artículo escrito por Eric Topol en su libro llamado *Deep Medicine...* (2019) muestra un estudio en el que se introduce la información de los registros médicos de los pacientes a un algoritmo de aprendizaje estadístico para predecir quién es probable de cometer suicidio, y lo hizo correctamente en 93 % de las veces.

Otro trabajo muy innovador es el que llevó a cabo una empresa llamada Mindstrong³ fundada por un equipo de especialistas en neurociencia, medicina e inteligencia artificial. En ella se está desarrollando una aplicación que controla la forma en la que la persona escribe, con qué frecuencia borra caracteres y cuánto se desplaza mientras usa otras aplicaciones, tratando de descubrir cómo cambia su cognición mental a lo largo del tiempo. Los datos que recopila se cifran y analizan en forma remota mediante aprendizaje automático, y se comparten con el paciente y su médico. A diferencia de otras aplicaciones que tienen la misma finalidad, Mindstrong asegura que —considerando las

³ <https://mindstronghealth.com/>

interacciones físicas de los usuarios con sus teléfonos identificando no lo que hacen, sino cómo lo hacen— puede ofrecer pistas más precisas de rastrear estos problemas con el tiempo.

Además, existen otras investigaciones (Reece y Danforth, 2017) donde la unidad de análisis no es solo el texto, ahora se vuelve la atención hacia las imágenes, de manera que analizaron 43 950 fotos de *Instagram* de 166 personas. Sus algoritmos lograron 70 % de precisión al reconocer quién estaba deprimido.

En México también hay estudios con el mismo objetivo: identificar de manera temprana indicios de riesgo de personas propensas a depresiones, tristeza y/o suicidios en función de los textos que escriben en sus plataformas digitales (Rodríguez-Esparza *et al.*, 2016).

Recursos usados en trabajos relacionados

El tener un conjunto de usuarios etiquetados con ansiedad o depresión (conjunto de entrenamiento) diagnosticados por especialistas psicólogos o psiquiatras es la clave para generar un conjunto ideal de entrenamiento para que, junto con un buen método de aprendizaje estadístico supervisado, se obtengan predicciones confiables con relación a que un usuario exhiba un riesgo de ansiedad o depresión. Sin embargo, no siempre se tiene el tiempo, el dinero y la disponibilidad de información para conformar uno con estas características. Es necesario acudir a otros recursos más económicos para identificar a estos usuarios; por ejemplo, según lo demostraron Tsugawa *et al.* (2015), el uso de un *bag of words*⁴ relacionado con depresión o tristeza es útil para identificar emociones negativas.

Otra forma para detectar de manera automática una colección de usuarios en *Twitter* con algún pro-

⁴ Bolsa de palabras relacionadas con un tema que sirven como un filtro para detectar aquellos textos de tuits que mencionen dichas palabras.

blema de salud mental es el de localizar a los usuarios que hayan expresado públicamente tener una afección, ya sea creyendo que pueden tener apoyo de otros que padecen lo mismo o para combatir el tabú de las enfermedades mentales o también como explicación de sus comportamientos (Coppersmith *et al.*, 2014; y Nadeem *et al.*, 2014). Por lo regular usan un filtrado de tuits buscando expresiones como “me diagnosticaron X”. Esta X puede ser algún trastorno, según lo que se necesita detectar, como: depresión, ansiedad, trastorno bipolar o de estrés postraumático, entre otros. En estos estudios se está suponiendo que, debido a que existe un estigma alrededor de los problemas de salud mental, parecería poco probable que se declare padecerlos cuando en realidad no es este el caso. Aun así, se sabe que en este tipo de redes se suelen dar declaraciones a modo de bromas y/o sarcasmos, por lo que es muy importante que se verifique manualmente que el significado del texto sea genuino.

Otro enfoque es el de realizar una identificación de categorías de emociones en un texto usando recursos léxicos, es decir, diccionarios que contienen un conjunto de palabras etiquetadas por emociones. La mayoría de estos recursos están en el idioma inglés; algunos de ellos tienen traducción al español, pero no todos son evaluados manualmente para asegurarse de que la connotación de la palabra sea en verdad etiquetada con la emoción correcta. En el cuadro 1 se describen algunos diccionarios desarrollados en español disponibles para los interesados en estos temas.

Es importante mencionar que *LIWC* se consigue como un programa computacional ejecutable comercial que hace un conteo de las palabras para 74 categorías agrupadas en cinco procesos psicológicos: social, afectivo, cognitivo, biológico y personal. Además, proporciona el número de palabras relacionadas con el lenguaje estándar de escritura, por ejemplo: total de pronombres, total de pronombres personales (1.ª persona, 1.ª persona plural, 2.ª persona, 3.ª persona singular, pronombres, verbos —pasado, presente, futuro—, adverbios, negaciones y otros).

Cuadro 1

Algunos recursos léxicos marcados por emociones en el idioma español

Diccionario	Clases	Creado por	Idioma	Verificado manual	Referencia
<i>SEL Spanish Emotion Lexico</i>	Tristeza, alegría, ira, enojo, sorpresa, desagrado.	Camacho, V., G. Sidorov, & N. Galicia (2016).	Español	Sí	https://www.cic.ipn.mx/~sidorov/
<i>NFC National Research Council Canada</i>	Alegría, sorpresa, confianza, miedo, ira, tristeza, aversión, anticipación.	Saif Mohammad and Peter Turney.	Inglés, traducido a 40 idiomas.	No	https://saifmohammad.com/WebPages/lexicons.html
<i>Linguistic Inquiry Word Count (LIWC)</i>	Ansiedad, tristeza, enfado, emoción negativa, emoción positiva.	Pennebaker, J. W., R. L. Boyd, K. Jordan & K. Blackburn (2015).	Inglés, español, alemán, holandés noruego, italiano y portugués.	Sí	https://liwc.wpengine.com/

Un ejemplo del uso de *LIWC* es una herramienta expuesta en la web *Analyze Words*,⁵ la cual arroja un indicador que mide el aspecto emocional, social y racional del tuitero con base en sus últimas 25 publicaciones teniendo solo el nombre de la cuenta pública del usuario.

Datos y su procesamiento

Este trabajo se realizó usando los tuits públicos y georreferenciados dentro del territorio nacional recolectados por el INEGI desde el 2017 al 2020. Se cuenta con alrededor de 2 millones de tuits por mes. A partir de estos datos, detectamos a todos los usuarios que hayan publicado más de 25 do-

cumentos por año, identificando así a aproximadamente 600 mil. La gráfica 1 muestra la distribución por año. Notamos que la captación de usuarios va disminuyendo a través del tiempo.

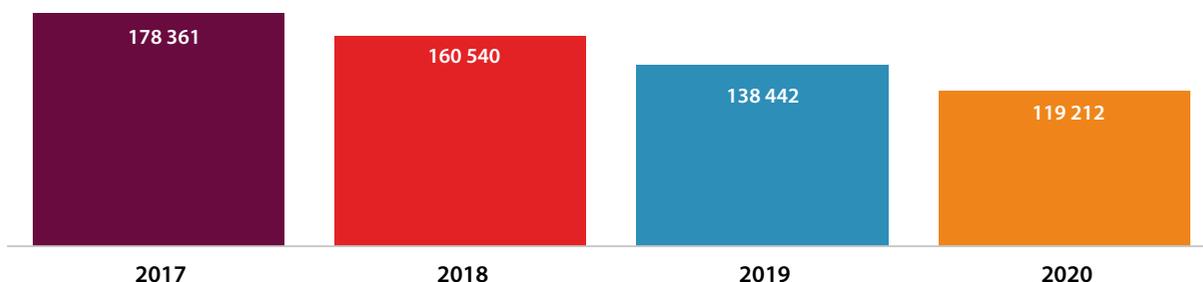
Se adquirió el *LIWC*, programa ejecutable diseñado para hacer los conteos de palabras pertenecientes a cada clase cuando se imputa una serie de documentos, sin embargo, el formato y la dimensión de los tuits a estudiar diferían de las del programa, por lo que fue necesario crear nuestro propio código para la lectura y conteo. Decidimos realizarlo con dos tipos *software*, *Python* y *R*, con la finalidad de llevar a cabo una comparación de resultados y garantizar la calidad de estos.

Además, se implementó un ambiente específico para el almacenamiento y las consultas bajo *Spark*

⁵ <https://www.analyzewords.com/science.php>

Gráfica 1

Número de tuiteros que publicaron más de 25 tuits por año



Fuente: INEGI. *Estado de ánimo de los tuiteros en México* (DE) <https://www.inegi.org.mx/app/animotuitero/#/app/multiline>

y *Apache Zeppelin*, pasando de un formato muy general, como *json*, a archivos *Parquet*, logrando disminuir significativamente la velocidad de consultas.

Como sabemos, la limpieza y el preproceso de datos es una etapa importante en cualquier estudio donde se analiza información no estructurada: quitar caracteres y símbolos especiales ayuda a eliminar ruido cuando se requiere interpretar resultados y reducir el texto a palabras básicas, como *tokens* y lemas, y facilita su conteo. Para nuestro estudio, eliminamos enlaces http, signos de puntuación, espacios innecesarios, emojis, etiquetas (#), menciones (@), y *tokenizamos*.⁶ Ni los números ni *stopwords*⁷ fueron omitidos, pues estos los contabiliza *LIWC* para diagnosticar la forma de escritura del usuario.

Se procesaron 124 627 640 tuits de 600 mil usuarios desde enero del 2017 a noviembre del 2020. Con estos datos se construyó una matriz de infor-

mación donde cada registro es un usuario y las columnas están dadas por características que representan sus textos (ver cuadro 2).

Para la identificación del municipio correspondiente a cada tuit, se utilizó el elemento *bounding_box*, que contiene un mínimo de cuatro pares de puntos geográficos (latitud, longitud) que hace referencia a la zona geográfica donde el texto fue creado; con estos se genera un polígono que será identificado por su valor central (centroide). Se realizó un cruce de estos valores con los polígonos que identifican a cada municipio del país para asociarlo con el tuit de acuerdo con el polígono donde cae el centroide. Con este método logramos 99 % de clasificación correcta de municipios a todos los tuits.

Aunque tenemos el conteo de palabras de todas las clases de *LIWC* por cada tuitero, este trabajo solo se concentra en la clase *ansiedad*; algunas palabras con esta etiqueta se muestran en el cuadro 3. Sin duda, esta matriz de información será de gran utilidad para investigaciones más robustas futuras al presente estudio.

6 Proceso de segmentar el texto en unidades lingüísticas: palabras, signos, números, alfanuméricos, etcétera.

7 Palabras muy comunes y poco informativas desde el punto de vista léxico, por ejemplo, conjunciones (y, o, ni, que), artículos (la, los, el), preposiciones (a, en, por, etcétera).

Cuadro 2

Información obtenida para cada usuario

Concepto	Descripción
id usuario	Identificador único del usuario.
N_word_clase_i	Número de palabras para cada clase de <i>LIWC</i> .
Tweet_positivos	Número de tuits positivos según ánimo tuitero (AT).
Tweet_negativos	Número de tuits negativos según AT.
Tweet_neg_ansi	Número de tuits negativos y que tenga al menos una palabra de cada categoría de <i>LIWC</i> .
Tweet_pos_ansi	Número de tuits positivos y que tenga al menos una palabra de cada clase <i>i</i> .
Clase_i	Conteo de palabras para la clase <i>i</i> de <i>LIWC</i> .
Clase_Ansiedad	Conteo de palabras relacionadas con <i>ansiedad</i> .
N_tweet	Número de tuits.
N_word	Número de palabras totales.
N_word_liwc	Número de palabras captadas por <i>LIWC</i> .
Entidad	Entidad federativa donde el tuitero publica la mayoría de sus tuits.
Municipio	Municipio donde el tuitero publica la mayoría de sus tuits.
Mes	Mes en el que el tuitero realiza sus documentos.

Cuadro 3

Palabras relacionadas a la clase *ansiedad* en LIWC

abruma*	apremia*	chocad*	desconcert*	enoj*	falte*	inadecuad*	miserable	preocupad*	sobresalta	sobresalté	temerari*	trauma
	aprensi*	compulsivo	desdichad*	escandal*	faltó	inciért*	nervios*	preocupé	sobresaltáis	sobresaltó	temid*	turba
abyect*	apuro*	confundi*	deseos*	espanto*	<i>fear</i>	incómod*	obsesión	preocupe*	sobre-saltamos	suda*	temiendo	turbad*
aflicci*	<i>ashamed</i>	confus*	desequilib*	esperpento*	histéric*	incomodo	obsesion*	preocupo	sobresaltan	sudando	temor*	turbar
afligi*	asust*	<i>confused</i>	desespera*	estres	horrend*	inquieta*	pánic*	preocupó	sobresaltando	sudar*	tensa*	<i>upset</i>
agita*	aterrad*	crisis	desgana*	estrés	horrible*	inquieta*	pavor	presión*	sobresaltar	sudó	tensión	urgencia
alarm*	aterroriza*	critic	deshonra*	estresante	horror	insegur*	peligr*	presiona*	sobresaltará*	suscep-tible*	tension*	vacilante*
angusti*	aturdid*	<i>critic</i>	desorde*	estruja*	horroriz*	intimid*	pena*	problema	sobresaltare*	susto*	tenso*	vengue*
ansiedad*	aturdi-miento	depresión	<i>disturb</i> *	estupefact*	humilla*	<i>late</i>	petrific*	reacia*	sobresaltare*	temblar*	terco*	vergonzos*
anticipa*	aturdir	depresiones	embarazos*	excitable	impacien*	mello	prejuicio*	reacio*	sobresaltaron	temblor*	terror*	vergüenza*
aplstar	aversi*	deprimid*	emoción	faltaron	impotencia	miedo*	preocupa	riesgo	sobresaltas	teme	tímid*	vulnerable*
aprehensión	<i>breakdown</i>						preocupaci*	sacud*	sobresaltast*	temer	torcedura*	<i>worried</i>

Nota: las palabras con * son raíces mejor conocidas como *steam* (en inglés), i. e., *angusti* es la raíz de varias palabras: angustiar, angustiado, angustia, etcétera.

Análisis exploratorio de datos

Una vista general del comportamiento de la información a través del tiempo por medio de gráficas puede ayudar a obtener conocimiento a partir de patrones naturales de los propios datos, o bien, buscar elementos que confirmen si las hipótesis o técnicas establecidas al inicio de cualquier estudio son las adecuadas.

La gráfica 2 nos muestra la evolución por año de los tipos de tuits de todos los tuiteros. Como se puede notar, existe una disminución de tuits recolectados en estos cuatro años. Tanto los positivos como los negativos (de acuerdo con la etiqueta que proporciona el clasificador del ánimo tuitero —AT—) decaen, aunque la velocidad de disminución de los últimos es menor que la de positivos. Se aprecia un ligero aumento de *ansioso-negativos*, los cuales son definidos por aquellos que tienen al menos una palabra relacionada con ansiedad y que son clasificados como negativos por el AT. Una disminución de tuits *ansioso-positivos* y el número de palabras relacionadas con ansiedad por año parece constante con un ligero aumento en el 2018. La información que nos da esta gráfica nos lleva a pensar que, aun cuando los tuiteros y los tuits van en disminución, pueden existir subpoblaciones de

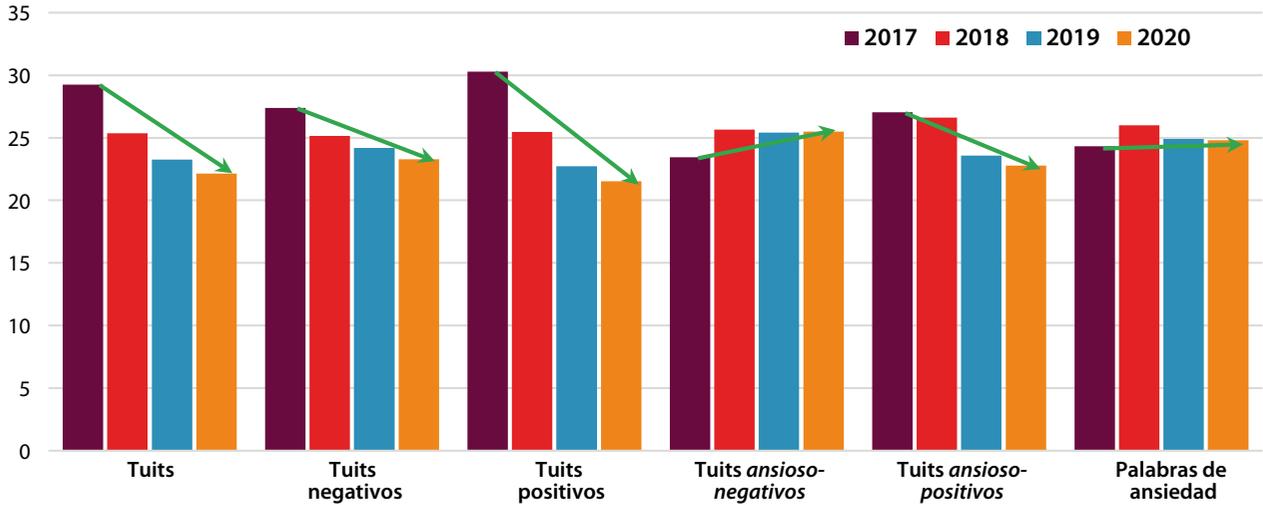
usuarios que expresan cierta temática que no necesariamente sigue la misma tendencia de disminución de la población total, por ejemplo, el caso de los que escriben textos *ansioso-negativos*, hecho que se demuestra más adelante al usar la técnica de aprendizaje supervisado.

Un análisis de los textos, tuiteros y palabras de ansiedad a través de los cuatro años, pero ahora por mes, nos revela ciertas fluctuaciones interesantes. Observamos en la gráfica 3 que la cantidad de tuits positivos siempre rebasa a los negativos y, efectivamente, tienden a disminuir a través del tiempo. Además, de manera general, se aprecia una relación entre estos dos tipos de tuits, si aumentan los positivos, se incrementan los negativos y viceversa. Note que estas fluctuaciones de aumento se dan en determinados meses de cada año.

En la gráfica 4, donde se muestra el conteo de palabras relacionadas con ansiedad por mes y año, se destacan barras más altas en ciertos meses. Resulta muy tentador empezar a relacionar estos aumentos con eventos importantes que sucedieron en cada año, por ejemplo, septiembre del 2017, mes en el que ocurrió un sismo; en mayo-julio del 2018 tuvieron lugar las elecciones presidenciales; enero del 2019, desabasto por gasolina; y marzo

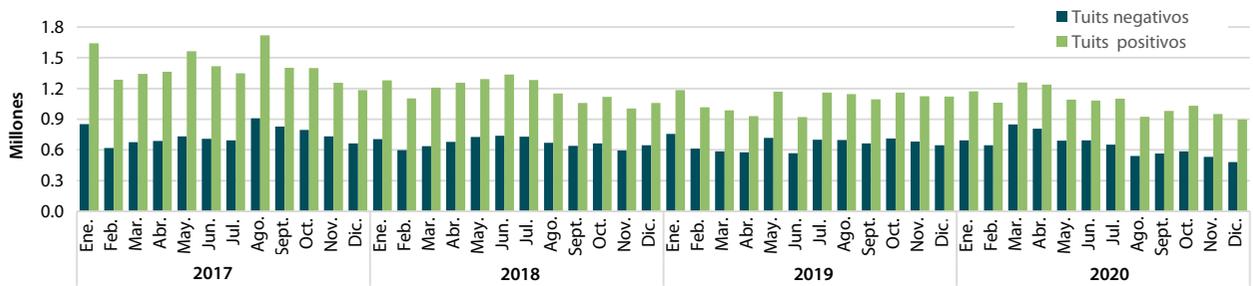
Gráfica 2

Evolución en el tiempo del tipo de tuits (porcentaje)



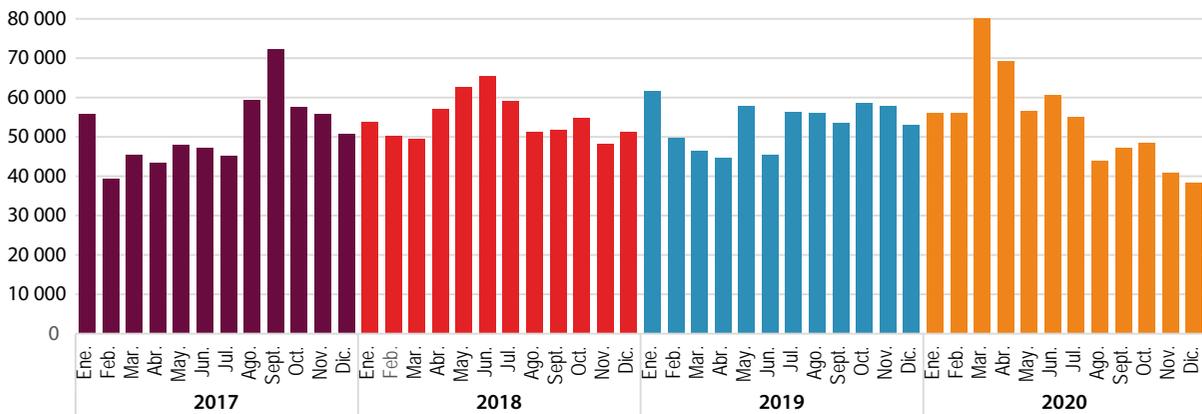
Gráfica 3

Distribución de los tuits negativos y positivos por mes a través de los años



Gráfica 4

Número de palabras relacionadas con ansiedad por mes a través del tiempo



del 2020 empezó el confinamiento en México por la pandemia de COVID-19.

Una pregunta natural que surge después de analizar la gráfica anterior es: ¿hacia qué valencia se inclinan las palabras relacionadas con ansiedad, a la negativa o a la positiva? Note que en la gráfica 5 solo se examinan a los tuiteros que expresaron palabras de ansiedad. Las barras de color verde oscuro representan a quienes publicaron textos clasificados como negativos; las de color verde claro, los positivos. Vemos que existe una incidencia mayor de tuiteros con textos *ansioso-negativos* que *ansioso-positivos*, y la presencia de los primeros se hace más notable en los mismos meses en los que tuvimos eventos que ya mencionamos.

Confirmando lo antes dicho, la correlación estadística asociada de las palabras de ansiedad vs. los textos negativos y positivos de todos los tuiteros por año se presentan en el cuadro 4, aunque existe una mayor correlación hacia lo negativo; la que hay con positivos no es tan baja, pues presenta un valor mayor a 0.6 en todos los años, incluso, en el

2018, estos coeficientes se igualan; por lo tanto, también deben contemplarse en el estudio y es por ello que en la siguiente sección se construye un índice que informa qué tan positivos o negativos son los textos que muestran ansiedad.

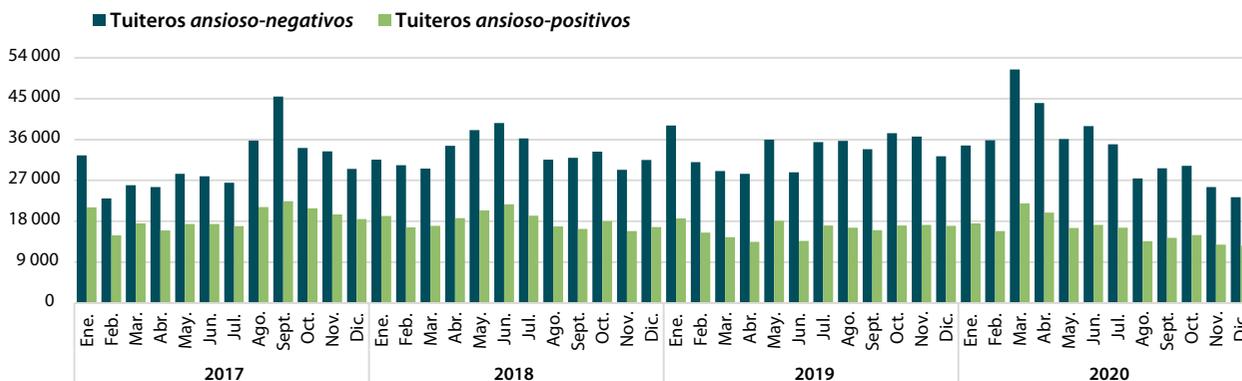
Como resumen, la información de las gráficas anteriores nos muestra que existen meses relevantes donde la población tuitera expresa más negatividad y ansiedad, que existe una tendencia más pronunciada a expresar ansiedad con contexto negativo que positivo y que los tuiteros *ansioso-negativos* van aumentando en cada año a pesar de que la población usuaria es cada vez menor.

Generación del Índice de Ansiedad por tuitero

Para realizar comparaciones entre usuarios y poder diferenciar si sus expresiones sobre ansiedad se encuentran en un contexto positivo o negativo, se pensó en generar un indicador que refleje esas diferencias.

Gráfica 5

Número de tuiteros ansiosos por mes a través del tiempo



Cuadro 4

Correlación de tuits *ansioso-negativos* y *ansioso-positivos* por año

Año	Tuits <i>ansioso-negativos</i>	Tuits <i>ansioso-positivos</i>
2017	0.61	0.56
2018	0.72	0.71
2019	0.76	0.63
2020	0.85	0.63

Asignación de clave en textos para el cálculo del Índice de Ansiedad

Polaridad de textos (AT)	Tuits con al menos una palabra relacionada con ansiedad	Tuits que no tienen palabras relacionadas con ansiedad
Positivos	<i>B</i>	<i>C</i>
Negativos	<i>A</i>	<i>D</i>

Cada usuario identificado en nuestra base de datos tendrá un valor entre -1 y 1, que indica el grado de ansiedad en función de sus textos: si este se acerca a -1, entonces podemos decir que tiende a expresar ansiedad en un contexto negativo; si se acerca a 0, el usuario es neutro en cuanto a su ansiedad; y si se aproxima a 1, entonces la mayoría de sus textos contienen palabras de ansiedad con valencia positiva.

Hicimos una partición excluyente de las clases de AT para cada tuitero, como se muestra en el cuadro 5, de tal manera que *N* representa la suma de todos sus textos.

El Índice de Ansiedad está determinado solo por las cantidades *A* y *B*:

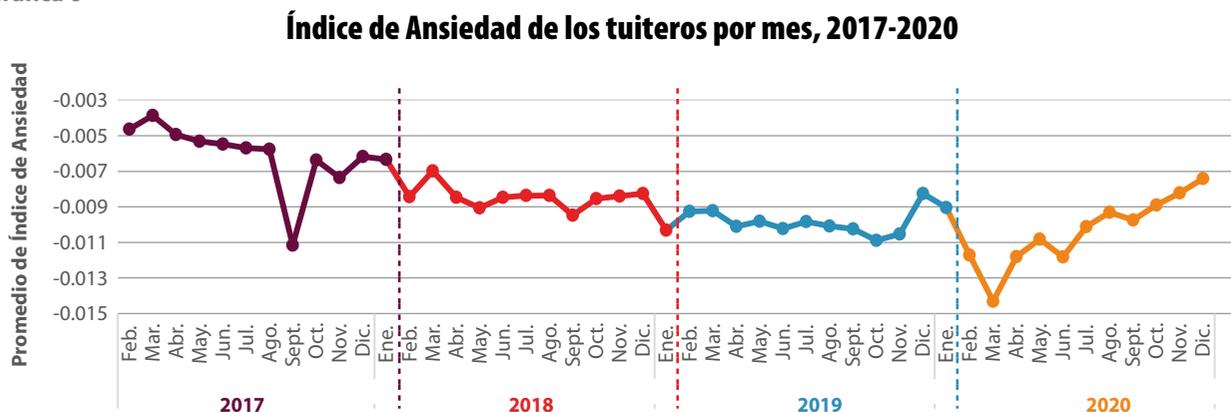
$$I_{ansiedad} = \frac{B}{N} - \frac{A}{N} = \frac{B-A}{N} \quad (1)$$

Note que si $A = 0$ y $B > 0$, entonces todos los textos con palabras de ansiedad del tuitero se clasificaron como positivos; por lo tanto, el Índice tiene un valor de $I_{ans} = 1$. Caso contrario, si $B = 0$ y $A > 0$, entonces no hubo textos *ansioso-positivos* y el Índice es $I_{ans} = -1$.

Un índice igual a 0 indica que el usuario es neutro en ansiedad; sin embargo, esto se puede presentar en dos casos: 1) si *A* es igual a *B*, el usuario expresa ansiedad, pero sus textos positivos igualan a sus negativos; por lo tanto, no tiene un sesgo de ansiedad que muestre una valencia de sus textos y 2) no usa palabras de ansiedad, por lo que *A* y *B* son iguales a 0; este usuario puede expresar gran cantidad de negativos (positivos), pero esa negatividad (positividad) no tiene nada que ver con la ansiedad.

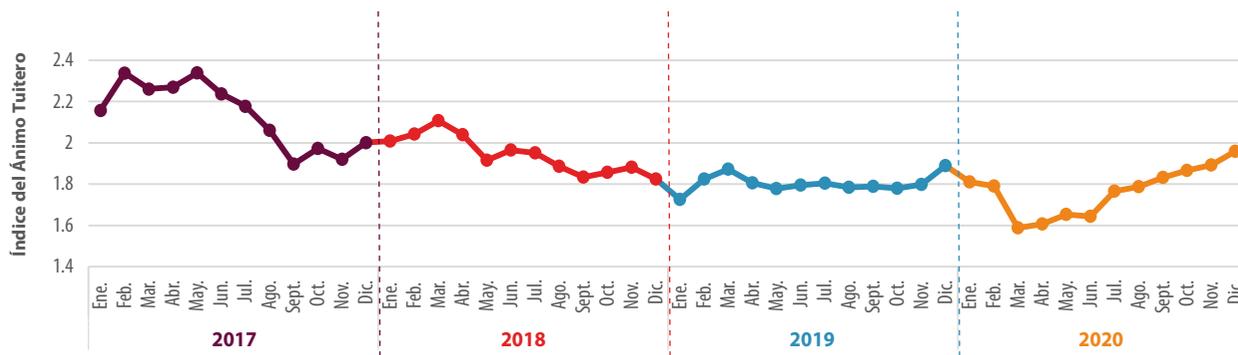
La gráfica 6 representa el promedio del Índice de Ansiedad de los tuiteros por mes desde el 2017 al 2020. Los picos más bajos en toda la serie corresponden a septiembre del 2017 y marzo del 2020. Para fines de comparación, en la gráfica 7 se visualiza el Índice del Ánimo Tuitero para el mismo periodo. Ambas gráficas muestran una tendencia similar, coinciden en ciertos aumentos o disminuciones a pesar de que las unidades de medición son diferentes; por un lado, el Índice del Ánimo Tuitero corresponde a la clasificación del propio texto de los tuits y el de Ansiedad, a la clasificación del tuitero. Ambos oscilan en distintas escalas. Podemos decir que no son excluyentes, pero tampoco

Gráfica 6



Gráfica 7

Índice del Ánimo Tuitero por mes, 2017-2020



complementarios. Los valores más bajos del Índice del Ánimo Tuitero en toda la serie se dan en enero del 2019 y marzo del 2020.

El Índice de Ansiedad ahora se representa por entidad federativa en las gráficas de 8, 10, 11 y 12. Podemos notar diferentes comportamientos para los distintos años; por ejemplo, en el 2017, las entidades con valores más bajos durante septiembre son: Chiapas, Ciudad de México, Guerrero, Morelos, Puebla y Tlaxcala; estas fueron las más afectadas por los sismos sucedidos en ese mes. Para confirmar esta sospecha, nos dimos a la tarea de encontrar las palabras más frecuentes asociadas con ansiedad en dichos estados para septiembre,

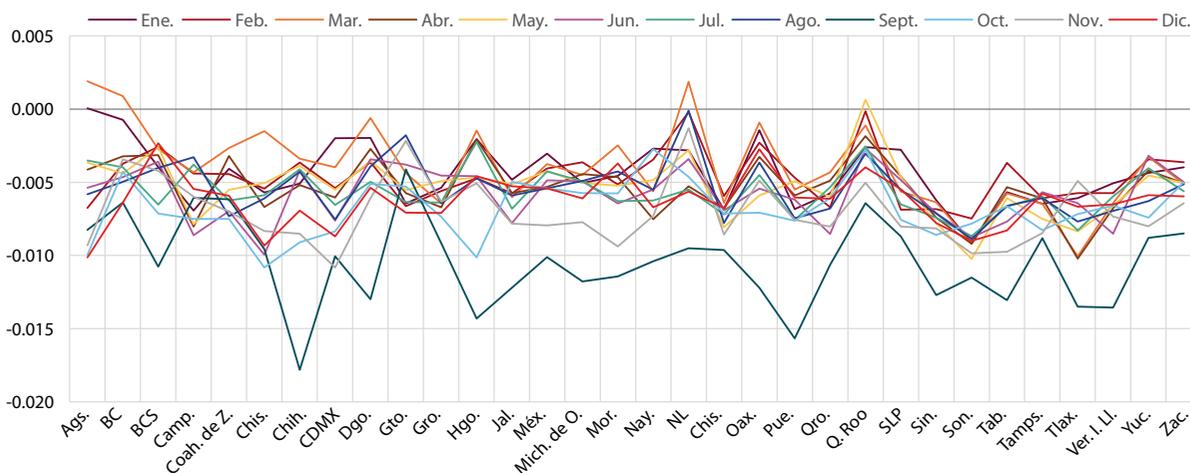
las cuales se muestran en la gráfica 9 a través de nubes de palabras de algunas entidades afectadas.

Se observa que los valores del Índice cambian en los siguientes años; el lector puede identificar los extremos y relacionarlos con algún suceso acontecido en el país. De estas graficas se desprende una interpretación meramente tendencial; para corroborar las asociaciones con eventos, hay que verificarlo profundizando el análisis.

Para el 2020, según la gráfica 11, destaca marzo con los índices más bajos para casi todas las entidades, hecho que claramente podemos relacionar con el inicio de la cuarentena por la pandemia de COVID-19.

Gráfica 8

Índice de Ansiedad por mes según entidad federativa, 2017



Gráfica 9

Top ten de palabras relacionadas con ansiedad en septiembre del 2017 para entidades seleccionadas

Guerrero



Chiapas



Puebla

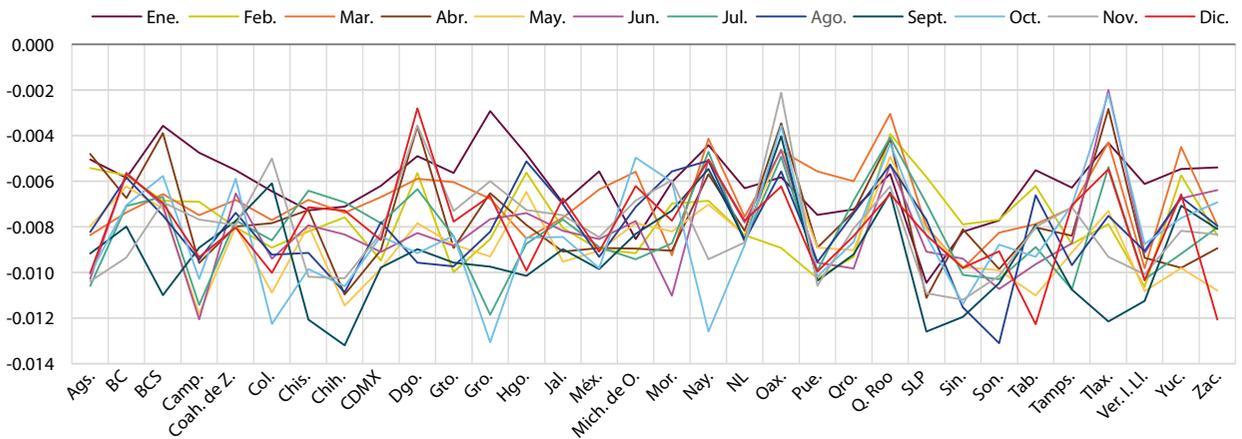


Tlaxcala



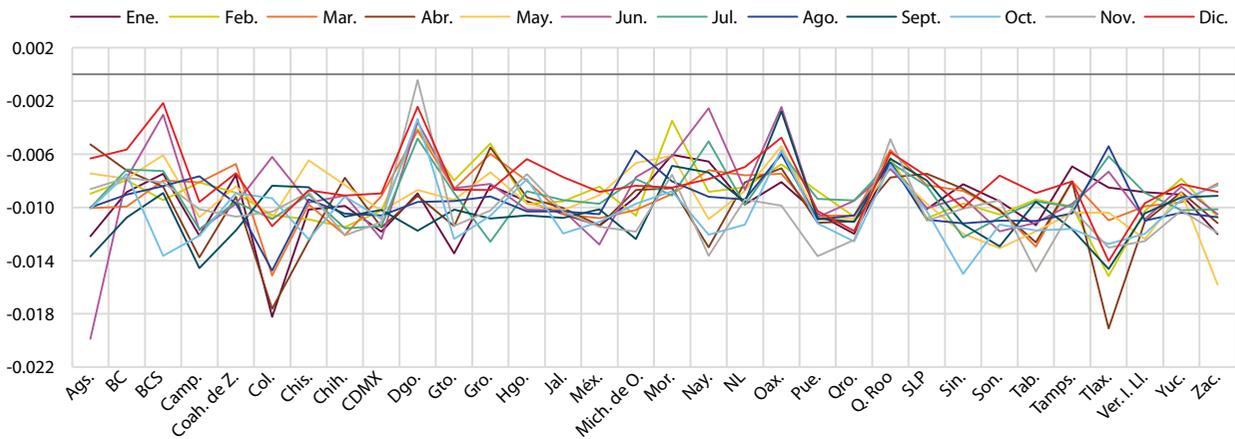
Gráfica 10

Índice de Ansiedad por mes según entidad federativa, 2018



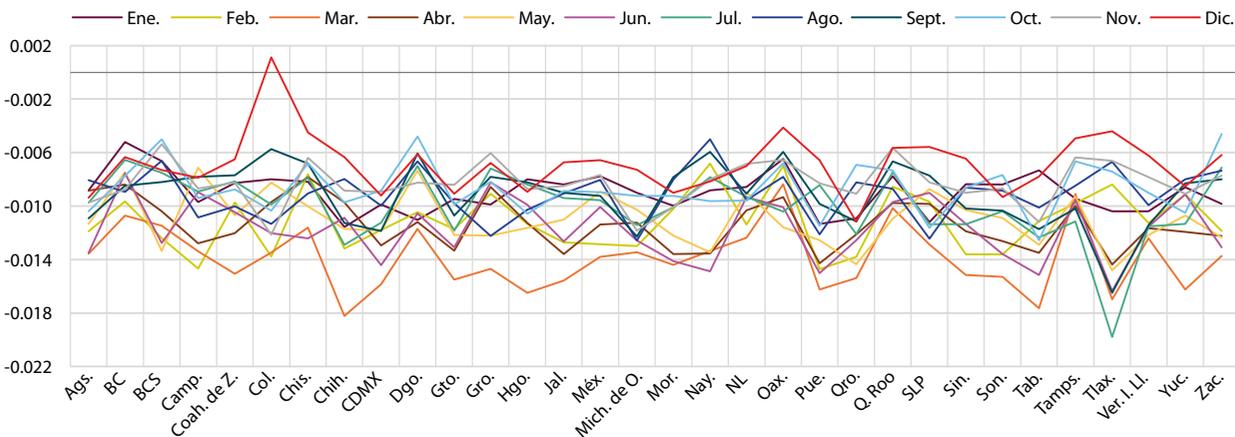
Gráfica 11

Índice de Ansiedad por mes según entidad federativa, 2019



Gráfica 12

Índice de Ansiedad por mes según entidad federativa, 2020



Para diagnosticar la presencia de otras emociones presentes en los tuitos, se tiene la posibilidad de construir índices similares para casi todas las clases de LIWC, de tal manera que puedan ser comparados y correlacionados. En pruebas previas a este análisis, notamos que las clases relacionadas con afecciones positivas más expresadas por los usuarios son *alegría* y *sorpresa*. Sobre las afecciones negativas, se puede decir que la ansiedad está muy relacionada con la emoción *miedo* (tal como lo señalan Piqueras Rodríguez *et al.*, 2009); la *tristeza*, con el *enojo*; y las palabras que expresan *repulsión* suelen ser menos utilizadas por los tuitos.

Si quisiéramos detectar a los usuarios que expresan más ansiedad con más carga negativa en sus textos que los distingue de los demás, usar este índice como referencia no sería lo más adecuado. Imagine a un tuitos que escribió solo un documento *ansioso-negativo*, por lo que su valor es -1; ahora, si existiera otro que escribió 10 documentos *ansioso-negativos* y ninguno *ansioso-positivo*, su índice también es igual a -1. El segundo ha usado más palabras de ansiedad que el primero y, sin embargo, el índice los iguala. La siguiente sección explica la construcción de grupos para diferenciar a este tipo de usuarios.

Construcción de tipos de ansiosos de acuerdo con la formación de grupos

Bajo el supuesto de que cuantas más palabras de ansiedad y más textos negativos escriba un tuitero, este tendrá mayor probabilidad de padecer un riesgo de salud mental, y si esta situación se corrobora con la evaluación cualitativa por parte de los expertos psicólogos, podemos entonces conformar una base de datos etiquetada para futuros estudios. Este es uno de los objetivos que buscamos al perfilar a los usuarios por medio de la técnica estadística *clustering*; el otro es el de medir la prevalencia de este tipo de tuiteros a través de los años, tal como lo sospechamos desde la gráfica 2.

Primeramente, la agrupación fue realizada tomando a todos los usuarios captados en el 2017 con más de 25 tuits. Se usaron tres variables para el agrupamiento: número de tuits *ansioso-negativos*, número de tuits *ansioso-positivos* y número de palabras relacionadas con ansiedad. Probamos diferentes técnicas de agrupamiento como *k-medias* y las basadas en mezcla de funciones gaussianas. Considerando el criterio *Bayesian Information Criterion (BIC)* decidimos quedarnos con seis grupos y usar un modelo con distribución elipsoidal con iguales tamaños, igual forma y diferentes orientaciones (EEV) (Scrucca *et al.*, 2016).

Como resultado de la agrupación, obtenemos las etiquetas que identifican al grupo de pertenencia de cada usuario en ese año; esta categorización se convierte en la base para generar un clasificador que determine las etiquetas de los tuiteros para los años subsecuentes, es decir, el proceso determinará el grupo al que pertenece el tuitero según la semejanza que tiene con los del año base. Por lo tanto, los grupos formados en el 2017 sirven como un conjunto de patrones fijos en los que el modelo se basa para asignar a nuevos usuarios al patrón más similar, de tal manera que sea más sencillo notar el crecimiento o disminución de tuiteros a través de los años bajo un año base. En otras palabras, se usa un método

de aprendizaje supervisado, suponiendo que las etiquetas de los usuarios del 2017 son las que determinan la diferencia de escritura de los usuarios.

Con la información que nos presenta el cuadro 5 es posible caracterizar a cada grupo formado en términos de los promedios de las variables usadas para la agrupación y otras que no fueron usadas. Podemos notar sus similitudes y diferencias.

El grupo 1 está caracterizado por muy pocos tuiteros, tan solo 67 que escriben mucho más texto que los demás; en promedio, cada tuitero escribe más de 3 mil tuits al año categorizados en su mayoría como positivos; sin embargo, tienden a expresar más palabras de ansiedad con una media de 167 por tuitero. Los textos *ansioso-negativos* son casi el doble que sus *ansioso-positivos*.

El grupo 2 está compuesto por tuiteros que, aunque no están en el extremo como en el grupo 1, sí reflejan una negatividad ansiosa en sus textos, con un promedio de 38 palabras de ansiedad por usuario y 24 tuits *ansioso-negativos* vs. 9.4 *ansioso-positivos*.

El grupo 3 es más neutro en las expresiones de ansiedad negativa y positiva.

Por otro lado, el 4 está compuesto por tuiteros que expresan palabras de ansiedad, pero con tendencia positiva, sus tuits *ansioso-positivos* son poco más de seis veces de sus *ansioso-negativos*; podemos llamar a este grupo como los tuiteros *ansioso-positivos*. Aunque pareciera que los grupos 4 y 5 tienen características similares, el algoritmo los diferencia al obtener promedios menores de este último.

Por su parte, en el 6 quedan clasificados la mayoría de los tuiteros, quienes no muestran ansiedad.

Pondremos atención en los grupos 1 y 2, pues creemos que pueden ser más vulnerables a seguir expresando textos negativos con palabras de ansiedad.

El método de clasificación (aprendizaje supervisado) que usamos fue el de análisis discriminante

Cuadro 6

Promedios de las variables por grupo resultado de la agrupación

Grupo	1	2	3	4	5	6
Número de tuiteros	67	2 092	10 830	2 194	894	162 284
Tuits	3 449.51	892.61	376.2	728.8	285.58	155.99
Tuits negativos	1 373.01	412.34	120.16	122.05	55.18	51.73
Tuits positivos	2 076.49	480.27	256.04	606.75	230.39	104.26
Tuits <i>ansioso-negativos</i>	93.19	24.27	5.3	2.75	1.68	2.06
Tuits <i>ansioso-positivos</i>	54.78	9.48	4.43	13.58	8.06	1.1
Palabras de ansiedad por tuitero	167.04	38.34	11.1	16.57	12.01	3.21
Índice de Ansiedad	-0.024	-0.02	-0.0097	0.031	0.013	-0.024

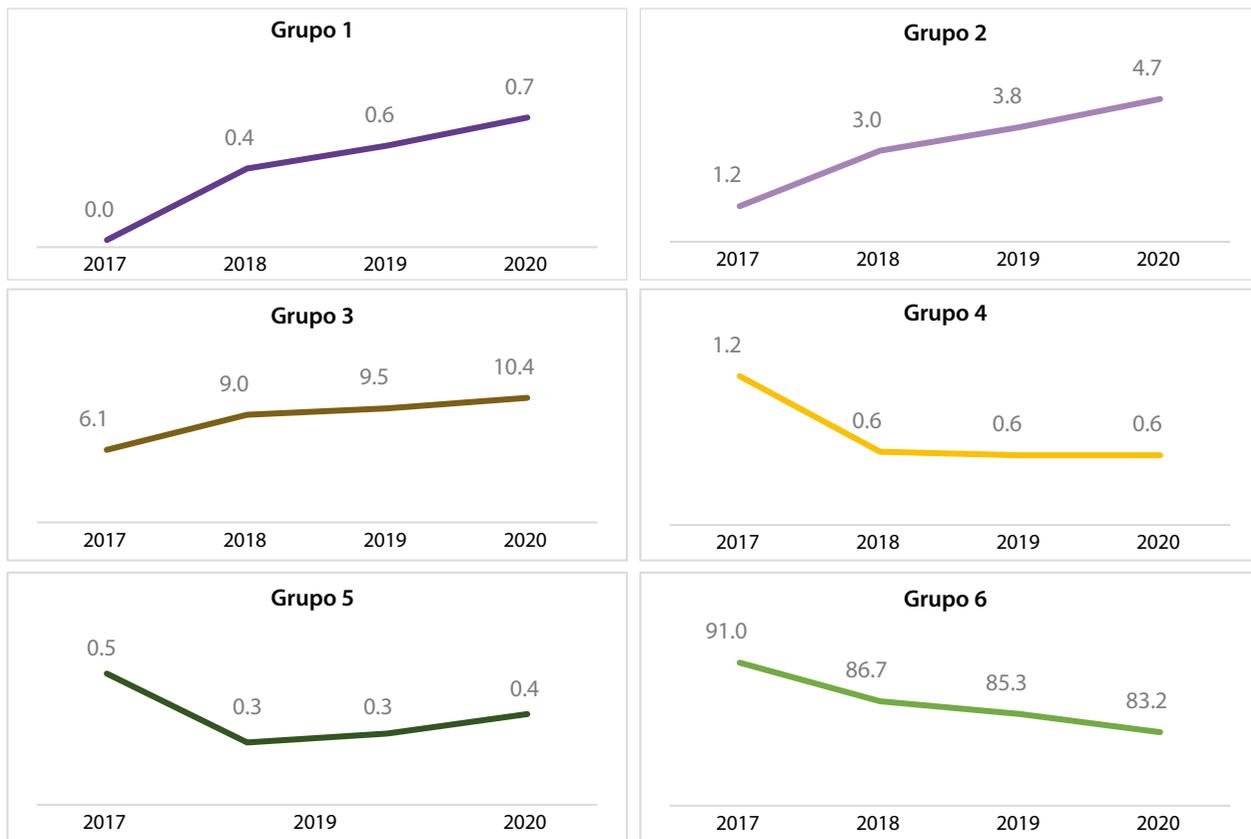
basado en mezclas, donde se asume que la densidad de cada clase sigue una distribución gaussiana. El corpus de usuarios del 2017, que está dado por un total de 178 361, se dividió en 70 % como base de entrenamiento y 30 % de prueba. Para una validación cruzada con $k = 10$ interacciones, ob-

tuvimos un error de predicción del clasificador de 0.030 (0.0012 desviación estándar).

La evolución de los usuarios en cada grupo a través de los años se muestra en las gráficas 13. Para los grupos más ansiosos (1-3), notamos una tendencia

Gráficas 13

Distribución porcentual de los tuiteros en cada grupo a través de los años



creciente de tuiteros a través de los años; mientras que para los menos ansiosos (4-6) ocurre lo contrario.

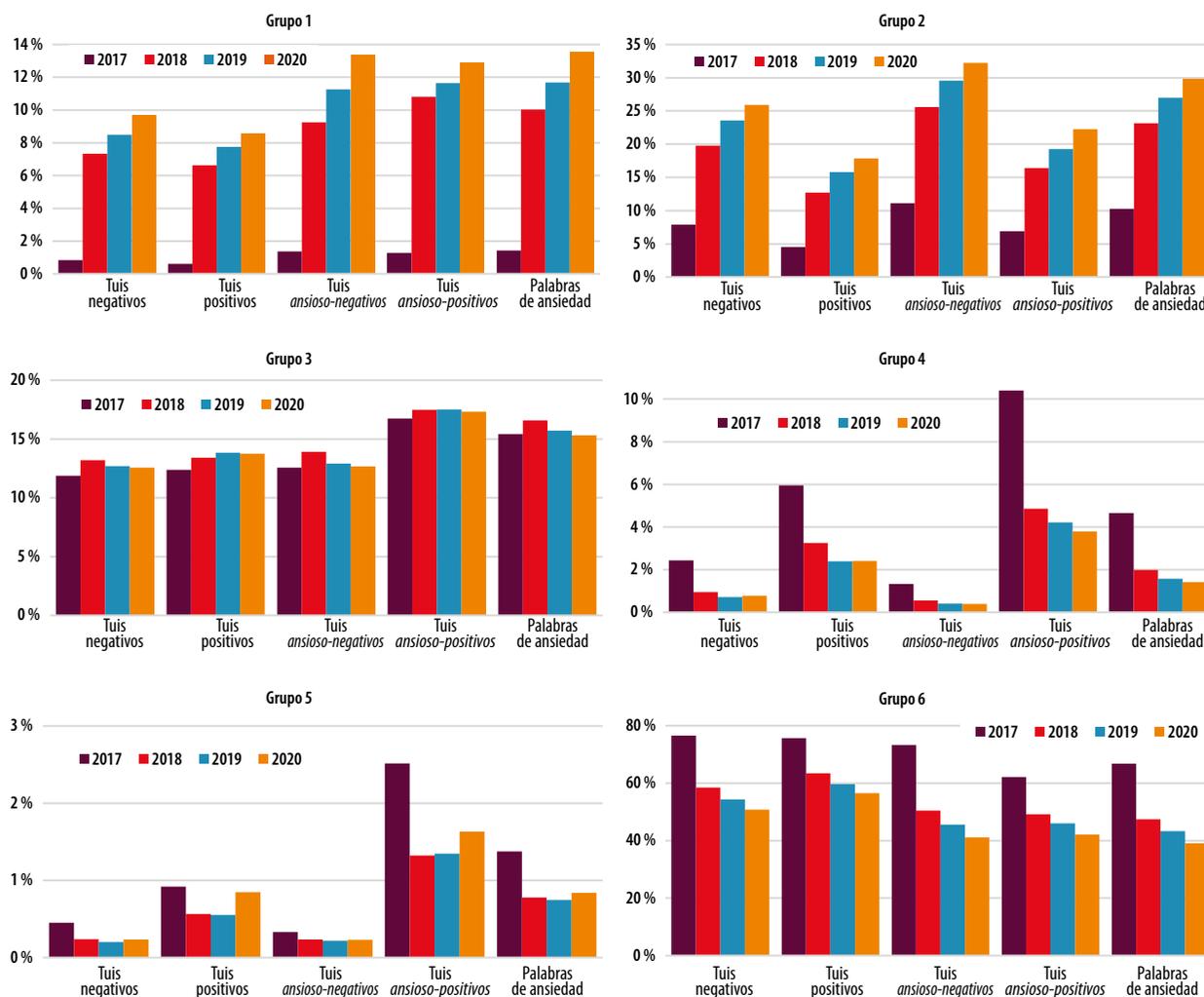
El crecimiento de tuiteros tipo 1 y 2 a través de los cuatro años implica que los tuits *ansioso-negativos* también van en aumento, tal como se refleja en las gráficas 14. Estos dos grupos de usuarios tienen un comportamiento diferente al de todos los tuiteros de nuestra muestra contemplados en la gráfica 2.

Aunque existe un aumento de usuarios clasificados en el grupo 3 en los cuatro años, el por-

centaje de sus tipos de textos se mantiene casi constante; es decir, no hay cambios identificables en la tendencia de escritura de los usuarios de este grupo de un año a otro. Por otra parte, los categorizados en el grupo 4 en el 2017 son más *ansioso-positivos*, sin embargo, quienes se clasificaron en este mismo grupo durante el 2020 se caracterizan por escribir con menos ansiedad positiva. Algo similar sucede para el 5. Es importante resaltar que los tuiteros del 6 son cada vez menos, por lo tanto, hay una disminución de sus tipos de textos.

Gráficas 14

Porcentaje de los tuiteros en cada grupo a través de los años



Conclusiones y consideraciones

Como ya lo comentamos en la introducción, la disposición de datos de redes sociales atrae a los investigadores en distintas áreas precisamente por la gran variedad de temas que abarcan para tener conocimiento sobre gustos, eventos, opiniones y sentimiento de los usuarios prácticamente a un costo cero, a diferencia de los obtenidos de fuentes oficiales, como encuestas y censos. Sin embargo, emplear los de las redes no proporciona un cierto grado de confiabilidad en las estimaciones o predicciones que se realicen, situación que sí se tiene con los datos oficiales; es decir, existen ventajas notables al usar datos abiertos, pero también hay riesgos metodológicos y éticos. Lo ideal es implementar estrategias metodológicas para combinar estos dos tipos de fuentes, como lo están haciendo ya los organismos nacionales e internacionales.

Aun así, a pesar de todo lo que se está haciendo al respecto y que se planea hacer para combinar datos tradicionales con no tradicionales, en este ejercicio, donde examinamos la ansiedad como una emoción intrínseca a un problema de salud mental y cómo es el comportamiento de una muestra relativamente grande de usuarios de *Twitter* referente al uso de palabras relacionadas con ansiedad y la negatividad de sus textos, establecemos una estadística a través de datos no tradicionales que muestra indicadores en el tema de salud pública y bienestar social que puede interesar a las autoridades gubernamentales, a profesionales o planificadores de la salud y organizaciones no gubernamentales implicadas o interesadas en proporcionar servicios de salud mental como evidencia para la generación de programas o políticas de salud pública.

Bajo una perspectiva de estudio basada en el análisis cuantitativo y automático de datos, observamos en nuestros resultados:

- Tendencias de ansiedad a nivel poblacional en ciertos periodos coincidentes con eventos desafortunados en el tiempo, como el sismo del 2017 en algunas entidades del país y la ansie-

dad mostrada en el inicio de la cuarentena por COVID-19.

- El Índice de Ansiedad proporciona un indicador de alerta del bienestar social.
- La prevalencia creciente de usuarios *ansioso-negativos* a través del tiempo.
- Diagnóstico inicial de posibles usuarios afectados que serán etiquetados por expertos para la conjunción de una base de datos que sirva para realizar predicciones del grado de ansiedad de otros usuarios.
- Los resultados complementan aún más la información que muestra el ánimo tuitero sobre la salud mental de la población tuitera en México.

Este estudio abre la puerta para el análisis de lenguaje asociado a otros trastornos mentales considerando las estrategias que han empleado otros investigadores al usar *LIWC* como diagnóstico de personalidades de una población de usuarios o de personas en lo individual.

Contamos con una estructura de información bastante amplia para robustecer esta investigación. Algunos trabajos futuros que podríamos explorar: identificar palabras usadas con frecuencia que acompañen a otra afección emocional negativa y observar la diferencia de uso de esas palabras con las emociones antagónicas; comprobar si existe un patrón similar en nuestro país con respecto a otras naciones donde los estudios indican que las personas deprimidas (Argamon, 2005) tienden a usar más pronombres en primera persona, mientras que las que reflejan miedo en sus textos son propensas a expresar más palabras en tercera persona; identificar las cinco grandes (*big-five*) diferentes personalidades de los usuarios: neuroticismo, extroversión, apertura a la experiencia, conciencia y afabilidad (Costa y McCrae, 1992); así como generar más indicadores emocionales según las clases que detecta *LIWC* y correlacionarlos para una detección más robusta de salud mental.

Fuentes

Brooks, David. "How Artificial Intelligence Can Save Your Life. The machines know you better than you know yourself", en: *The New York Times*. 24 de junio de 2019.

- Coppersmith, Glen, Mark Dredze y Craig Harman. "Quantifying Mental Health Signals in Twitter", en: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Baltimore, Maryland, USA. Association for Computational Linguistics, 2014, pp. 51-60.
- Costa, P.T. Jr., & R. R. McCrae. Four ways five factors are basic", en: *Personality and Individual Differences*. 13, 1992, pp. 653-665 (DE) Google Scholar.
- De Choudhury, Munmun, Michael Gamon, Scott Counts y Eric Horvitz. "Predicting depression via social media", en: *Seventh Int. AAAI Conf. Weblogs Soc. Media. Proceedings of the International AAAI Conference on Web and Social Media*. 7(1), 2013, pp. 128-137 (DE) <https://ojs.aaai.org/index.php/ICWSM/article/view/14432>
- Díaz Rangel, Ismael, Grigori Sidorov y Sergio Suárez-Guerra. "Creation and evaluation of a dictionary tagged with emotions and weighted for Spanish", en: *Onomázein*. Núm. 29, junio. Santiago de Chile, Pontificia Universidad Católica de Chile, 2014, pp. 34-46.
- Fontaine, Johnny R. J., Klaus R. Scherer, Etienne Roesch y Phoebe C. Ellsworth. "The World of Emotions Is Not Two-Dimensional", en: *Psychological Science*. 18, núm. 12, 2017, pp. 1050-1057.
- Layard, Richard. *The economics of mental health" With modern psychological therapy, mentally ill people can become more productive and more satisfied with life*. LSE, UK and IZA, German, IZA World of Labor, 2017.
- Mohammad, Saif M. y Peter Turney. *Emotions Evoked by Common Words*. Ottawa, Ontario, Canada, K1A, Institute for Information Technology, National Research Council Canada, 2010.
- Nadeem, Moin, Mike Horn, Glen Coppersmith, Johns Hopkins University y Sandip Sen. *Identifying depression on Twitter*. PhD, University of Tulsa, 2014.
- OMS. *Depresión*. OMS, 2021 (DE) https://www.who.int/health-topics/depression#tab=tab_1, consultado el 28 de febrero de 2021.
- Piqueras Rodríguez, José Antonio, Victoriano Ramos Linares, Agustín Ernesto Martínez González y Luis Armando Oblitas Guadalupe. "Emociones negativas y su impacto en la salud mental y física", en: *Suma Psicológica (Fundación Universitaria Konrad Lorenz)*. 16, núm. 2, diciembre de 2009, pp. 85-112.
- Reece, Andrew G. y Chris M. Danforth. "Instagram photos reveal predictive markers of depression", en: *EPJ Data Science*. 6, núm. 15, 2017.
- Rodríguez-Esparza, Luz Judith, Diana Barraza-Barraza, Jesús Salazar-Ibarra y Rafael Gerardo Vargas-Pasaye. "Índice de riesgo al suicidio en México utilizando Twitter", en: *Revista de Ciencias de la Salud*. 5, núm. 15, 2016, pp. 1-13.
- Ruckenstein, Minna Susanna, and Natasha Dow Schull. "The datafication of health", en: *Annual Review of Anthropology*. 46, 2017, pp. 261-278.
- Scrucca, Luca, Michael Fop, T. Brendan Murphy y Adrian E. Raftery. "Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models", en: *The R Journal*. 8, núm. 20, 2016, pp. 5-33.
- Stirman, Shannon Wiltsey y James W. Pennebaker. "Word use in the poetry of suicidal and nonsuicidal poets", en: *Psychosomatic Medicine*. 63, núm. 4, 2001, pp. 517-522.
- Topol, Eric. *Deep Medicine, How Artificial Intelligence Can Make Healthcare*. New York, Basic Books, 2019.
- Tsugawa, Sho, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh y Hiroyuki Ohsaki. "Recognizing Depression from Twitter Activity", en: *33rd Annual ACM Conference on Human Factors in Computing Systems*. Editado por ACM Digital Library, 2015, pp. 3187-3196.

Colaboran en este número

Liliana Meza González

De nacionalidad mexicana. Es licenciada en Economía por el Instituto Tecnológico Autónomo de México (ITAM), maestra en Economía por la Universidad de Houston y el ITAM y doctora en Economía Laboral por la Universidad de Houston. Entre el 2017 y 2021 fue investigadora en el Instituto Nacional de Estadística y Geografía (INEGI). Actualmente, es profesora e investigadora en el Departamento de Estudios Internacionales de la Universidad Iberoamericana Ciudad de México. Es miembro del Sistema Nacional de Investigadores (SNI) con nivel 1.

Contacto: liliana.meza@ibero.mx

Jael Pérez Sánchez

De nacionalidad mexicana. Es licenciado en Economía por la Universidad Autónoma de Aguascalientes (UAA) y concluyó la maestría de Estadística Oficial en el Centro de Investigación en Matemáticas, A. C. (CIMAT); además, cuenta con un diplomado en Demografía por El Colegio de México (COLMEX). En la actualidad, ocupa el puesto de subdirector de Estandarización de Clasificaciones y Estrategias de Codificación en el INEGI y es vicepresidente regional de la Federación Nacional de Colegios de Economistas.

Contacto: jael.perez@inegi.org.mx

Víctor Alfredo Bustos y de la Tijera

De nacionalidad mexicana. Es licenciado en Actuaría por la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM), maestro en Estadística e Investigación de Operaciones por el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM y doctor en Estadística por la London School of Economics. Ingresó al INEGI en 1991, donde actualmente es investigador en la Dirección General Adjunta de Investigación. Sus áreas de interés son, entre otras: el ajuste de modelos a partir de encuestas, donde destaca la estimación de la distribución del ingreso con datos de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), compatible con las cuentas nacionales; la estimación para dominios pequeños en temas como la desocupación y la migración; la clasificación multivariada con aplicación a la medición del bienestar y la marginación; y la construcción, mediante criterios explícitos, de índices temporales multivariados para seguir la evolución de una economía. Fue elegido al Consejo del Instituto Internacional de Estadística (ISI) por el periodo 2019-2023.

Contacto: alfredo.bustos@inegi.org.mx

Ana Miriam Romo Anaya

De nacionalidad mexicana. Es licenciada en Matemáticas Aplicadas por la UAA y maestra en Ciencias en Estadística Oficial por el CIMAT. Desde el 2009 ha participado en la Dirección General Adjunta de Investigación del INEGI en el desarrollo y análisis de técnicas estadísticas, como: muestreo, modelos lineales y no lineales, técnicas de agrupamiento y clasificación, optimización de funciones paramétricas y el desarrollo computacional estadístico; ha sido catedrática en diferentes universidades impartiendo cursos de Matemáticas, Estadística y *Machine Learning*.

Contacto: miriam.romo@inegi.org.mx

Mario Gutiérrez-Lagunes

De nacionalidad mexicana. Es matemático con Maestría en Investigación de Operaciones por la UNAM y doctor en Ciencias Financieras por el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). En el ámbito laboral, se desempeña como profesor-investigador de tiempo completo en la Universidad Autónoma de San Luis Potosí (UASLP)-Zona Media. Tiene experiencia profesional en la industria aérea y la banca de desarrollo, principalmente. Es evaluador voluntario del Panel Experto del programa Impulso a Emprendimientos de Alto Impacto del Instituto Nacional del Emprendedor en colaboración con el Banco Mundial. Sus líneas de investigación actuales son: finanzas corporativas, proyectos de inversión, sectorización económica y desarrollo económico.

Contacto: mario.gutierrez.lagunes@uaslp.mx

Edgar Manuel Matus Carballo

De nacionalidad mexicana. Es pasante de la Licenciatura en Actuaría en la Facultad de Negocios de la Universidad La Salle México. Ha participado en proyectos de investigación, entre ellos del Fondo Sectorial CONACYT-INEGI. Ha publicado artículos de investigación en revistas arbitradas y participado en congresos nacionales e internacionales.

Contacto: macaed00@gmail.com

Carlos Alberto Jiménez-Bandala

De nacionalidad mexicana. Es doctor en Estudios Organizacionales por la Universidad Autónoma Metropolitana-Iztapalapa. Se desempeña como profesor-investigador del Departamento de Mercadotecnia y Negocios de la Universidad Autónoma del Estado de Quintana Roo, Campus Cancún. Es miembro del SNI con nivel 1. Ha participado como responsable técnico en proyectos de investigación, entre ellos del

Fondo Sectorial CONACYT-INEGI. Ha publicado artículos de investigación en revistas arbitradas y participado en congresos nacionales e internacionales.

Contacto: carlos.jimenez@uqroo.edu.mx

Jesús López-Pérez

De nacionalidad mexicana. Es licenciado en Economía por el ITESM y maestro en Estadística Aplicada por la misma institución. En la actualidad, es subdirector de Investigación del INEGI en temas relacionados con el análisis econométrico de series de tiempo. Anteriormente, ocupó diversos cargos en áreas de administración y análisis de riesgo crediticio en instituciones financieras del sector público y privado.

Contacto: jesus.lopezp@inegi.org.mx

Francisco de Jesús Corona Villavicencio

De nacionalidad mexicana. Es licenciado en Economía por la Universidad Autónoma de Baja California (UABC), maestro en Estadística Aplicada por el ITESM y doctor en Economía y Métodos Cuantitativos por la Universidad Carlos III de Madrid (UC3M). Actualmente, es Investigador en el INEGI y sus líneas de interés están relacionadas con el análisis econométrico y pronóstico de series de tiempo, así como en *Sport Analytics*. Ha publicado en revistas arbitradas de circulación internacional. Pertenece al SNI con nivel I.

Contacto: franciscoj.corona@inegi.org.mx

Berenice García Cabrera

Mexicana. Es doctora en Administración por la Universidad Intercontinental. Labora desde hace 11 años en la Comisión Nacional del Agua (CONAGUA), donde en la actualidad se desempeña como subdirectora de área en la Coordinación de Recaudación y Fiscalización

Contacto: berenice.garcia@conagua.gob.mx

Víctor Silva Cuevas

De nacionalidad mexicana. Es ingeniero en Sistemas Computacionales por el Instituto Tecnología de Zacatecas y tiene la Maestría en Ciencia de Datos e Información por el Centro de Investigación e Innovación en TIC. Desde el 2017 ha participado en la Dirección General Adjunta de Investigación del INEGI en distintos proyectos basados en estadísticas experimentales a partir de fuentes de datos no convencionales, aplicando técnicas para el procesamiento de grandes cúmulos de información, técnicas para el procesamiento

del lenguaje natural y el uso de modelos de *machine learning* y *deep learning* para clasificación supervisada y no supervisada. Además, cuenta con conocimiento en ingeniería de datos y desarrollo de *software*.

Contacto: victor.silvac@inegi.org.mx

Irving Gibran Cabrera Zamora

De nacionalidad mexicana. Es licenciado en Matemáticas Aplicadas por la UAA y Maestro en Ciencias por la Universidad Tecnológica de Múnich. En el ámbito laboral, participa en la Dirección General Adjunta de Investigación del INEGI desde el 2019 en proyectos que incluyen diseño de cuestionarios, tratamiento y transformación de resultados de encuestas y diseño de indicadores, y en el Laboratorio de Ciencia de Datos está encargado de diseñar e implementar flujos de datos, creación de servicios de infraestructura y modelos de aprendizaje automático.

Contacto: irving.cabrera@inegi.org.mx

Política y lineamientos editoriales

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA es una publicación cuatrimestral que sirve de enlace entre la generación de la información estadística y geográfica oficial y la investigación académica para compartir el conocimiento entre especialistas e instituciones con propósitos similares.

Se publicarán sólo artículos inéditos y originales relacionados con la situación actual del uso y aplicación de la información estadística y geográfica a nivel nacional e internacional.

Es una revista técnico-científica, bilingüe, cuyos trabajos son arbitrados por pares (especialistas), bajo la metodología doble ciego, con los siguientes criterios de evaluación: trabajos inéditos, originalidad, actualidad y oportunidad de la información, claridad en la definición de propósitos e ideas planteadas, cobertura de los objetivos definidos, estructura metodológica adecuada y congruencia entre la información contenida en el trabajo y las conclusiones.

El resultado del proceso de dictaminación se comunica por correo electrónico y contempla tres variantes: recomendado ampliamente (con modificaciones menores), recomendado (pero condicionado a modificaciones sugeridas) y no recomendado (rechazado). Dos dictámenes aprobados, se notifica al autor que se publica y se envía a corrección de estilo; un aprobado y uno rechazado, se le solicita realizar cambios; y dos rechazados, se notifica la no publicación.

Indizaciones y registros

- LATINDEX Catálogo (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal).
- CLASE (Citas Latinoamericanas en Ciencias Sociales y Humanidades).
- REDIB (Red Iberoamericana de Innovación y Conocimiento Científico).

Lineamientos para publicar

Se publicarán trabajos en español e inglés: artículos de investigación, revisión y divulgación; ensayos; metodologías; informes técnicos; comunicaciones cortas; reseñas de libros; revisiones bibliográficas y estadísticas, entre otros.

1. El artículo —o cualquier otro tipo de escrito de los mencionados— deberá entregarse con una carta dirigida al editor responsable de REALIDAD, DATOS Y ESPACIO. REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA en la que se proponga el texto para su publicación, que se declare que es inédito y que no ha sido postulado de manera paralela en otro medio. Asimismo, deben incluirse los datos completos del(os) autor(es), nacionalidad(es), institución(es) de adscripción y cargo(s) que ocupa(n), domicilio(s) completo(s), correo(s) electrónico(s) y teléfono(s). Esto debe dirigirse a la atención de la M. en C. Virginia Abrín Batule, virginia.abrin@inegi.org.mx (tel. 5278 10 00, ext. 1161).
2. El trabajo se debe presentar en versión electrónica (formato *Word* o compatible) con: a) extensión no mayor de 20 cuartillas; b) letra Helvética, Arial o Times de 12 puntos y c) interlineado de 1.5 líneas. El material adicional al texto se requiere por separado: a) las imágenes, con resolución de 300 ppp y un tamaño no menor a 17 centímetros de base (ancho) en formato JPG o TIF —no remuestrear (ampliar) imágenes de menor resolución—; si son líneas o mapas, deben entregarse en formato vectorial (EPS o Ai), en caso de incluirse imágenes en mapa de bits, incrustarlas o enviarlas con el nombre con el cual se creó el vínculo (conservando los requerimientos de resolución y tamaño estipulados); y para fotografías, éstas no deben ser menores a 5 megapíxeles; b) las fórmulas o expresiones matemáticas tienen que elaborarse con el editor de ecuaciones propio de *Microsoft*[™], pero en caso de usar *software* de terceros, incluir en la entrega PDF testigo en el cual figuren exactamente cómo deben representarse; c) las gráficas, que incluyan el archivo en *Excel* con el cual se desarrollaron o, en su defecto, la imagen JPG legible, de origen, en alta resolución; y d) los cuadros, que sean editables, no se deben insertar como imagen.
3. La colaboración debe incluir: título del trabajo (en español e inglés o viceversa); resúmenes del trabajo en español e inglés (que no excedan de un párrafo de 10 renglones); palabras clave en español e inglés (mínimo tres, máximo cinco); bibliografía u otras fuentes; así como breve(s) semblanza(s) del(os) autor(es) que no exceda(n) de un párrafo de cinco renglones y que incluya(n) nacionalidad(es), grado(s) académico(s), principal(es) experiencia(s) profesional(es), adscripción(es) laboral(es) actual(es) y dirección(es) electrónica(s) de contacto.
4. Las referencias bibliográficas u otras fuentes deberán presentarse al final del artículo de la siguiente manera: nombre(s) del(os) autor(es) comenzando por el(los) apellido(s); título de la publicación con cursivas (si se trata de un artículo, debe estar entrecorrido, seguido de coma y la preposición en con dos puntos y, enseguida, el título de la revista o libro donde apareció publicado, con cursivas); país de origen; editorial; lugar y año de edición; página(s) consultada(s). En el caso de las fuentes electrónicas (páginas web) se debe seguir el mismo orden que en las bibliográficas, pero al final se pondrá entre paréntesis DE (dirección electrónica), la fecha de consulta y la liga completa. Se tienen que omitir aquellas que se mencionen como notas a pie de página. Si se aplica la opción de incluir en cuerpo de texto la referencia de nombre de autor y año de la fuente consultada entre paréntesis, si deben aparecer todas las referencias mencionadas.

Página electrónica: <http://rde.inegi.org.mx>

Editorial Guidelines and Policy

REALITY, DATA AND SPACE INTERNATIONAL JOURNAL OF STATISTICS AND GEOGRAPHY is a four-monthly publication that connects statistics and geographic official information with academic research in order to share knowledge among specialists and institutions with similar aims.

We will publish only original and unpublished articles related to the current use and appliance of statistical and geographical information at both national and international levels.

It is a technical-scientific and bilingual magazine, with articles previously peer-reviewed by specialists under a double-blind methodology with the following evaluation criteria: unpublished works, originality, information related to opportunity and current affairs, we expect clarity in the definition of aims and ideas stated, defined objectives coverage, accurate methodological structure and coherence between the information of the paper as well as its conclusions.

The result of the paper-assessment process is delivered by email, and it involves three possibilities: fully recommended (with slight modifications), recommended (on condition of suggested modifications) and not recommended (i.e. rejected). When there are two reports of approval, the author gets notified that his/her paper will be published and it is sent to a style editing process. When one report approves the paper for publication and another one rejects it, the author is requested to make some changes for the text to be published. If the text submitted receives two non-favourable reports, the author is notified that the text will not be published.

Index and Registers

- LATINDEX Catalogue (Online Regional Information System for Scientific Journals from Latin America, the Caribbean, Spain and Portugal).
- CLASE (Latin American Quotations in Humanities and Social Sciences)
- REDIB (Latin American Net of Innovation and Scientific Knowledge)

Publishing Guidelines

Articles will be published in Spanish or English: research, revision and scientific-spreading articles; methodologies; technical reports; short texts; book reviews; and bibliographical and statistical revisions, among others.

1. The article —or any other kind of text from those aforementioned— must be delivered with an attached letter addressed to the chief editor of Reality, Data and Space. International Statistics and Geography Magazine in which the text intended for publication will be submitted. There it must be stated that the text has not been published, and that it has not been submitted for publication in any other media. The names in full of the authors must be included, as well as their nationalities, adscription institutions, position in those institutions, postal address, e-mail address, and telephone numbers. This must be addressed to MSc Virginia Abrín Batule, Virginia.abrin@inegi.org.mx (tel (+52) (55) 52.78.10.00, extension 1161).
2. The article must be submitted in an electronic version (a Microsoft Word file or a compatible one) with the following format: a) the text should not exceed the 20 pages of length; b) typography must be Helvetic, Arial or Times (12 points); and c) there should be a 1.5 line spacing in each paragraph. Additional material to the text will be delivered separately: a) images with a resolution of 300 ppp and no smaller than 17 cm width will be delivered in format JPG or TIF —please do not amplify images with lower resolution—. If the added materials are lines or maps, these must be delivered in vectorial format (EPS or Ai). If there are images in bits map, these must be embedded or attached with the name of the original file with which the link was created (keeping the resolution and size requirements above stated). As regards to photographs, these should not be inferior as 5 megapixels; b) mathematical expressions or formulae have to be created with the equations editor by Microsoft[™], but in case of using third-parties software, please attach a witness PDF in which the exact representation of mathematical formulae or expressions is contained; c) graphics must include the Excel file in which they were created or a legible image in the original JPG format in high resolution; and d) charts must be editable, and must not be inserted as images.
3. The text must include the following: the article's title (both in English and Spanish); the abstract of the article—both in English and Spanish (not longer than a 10-line paragraph); key words—both in English and Spanish (three as minimum and five as maximum); bibliography and other sources; as well as brief biographical sketches of the authors not exceeding a five-line paragraph each including nationalities, academic titles, main professional experiences, current work-related affiliations, and electronic addresses for the authors to be contacted.
4. Bibliographical references and other sources must be included at the end of the article in the following way: author's name (Surname first), and publication's title (in italics). If it is an article, the title must be in quotation marks followed by a comma and the preposition "en" with semicolon (in Spanish), then it should appear the title of the book or magazine in which the article was published (in italics); country of origin; publishing house, edition year, and consulted pages. As regards to electronic sources (web pages) the same order of the bibliographical references must be followed, but at the end the word "EA" (as for Electronic Address) ("DE" in Spanish) must be added within parenthesis followed by consultation date and the complete reference link. Those web links referred previously as footnotes, must be omitted in this section. However, if the name of the author and the year of the consulted source were included in the main body of the text within parenthesis, all these must be included as part of the bibliographical references.

Webpage: <http://rde.inegi.org.mx>

