

APLICACIONES DE LA METODOLOGÍA ESTADÍSTICA EN DIFERENTES DISCIPLINAS

COORDINADORES
JULIA AURORA MONTANO RIVAS
CECILIA CRUZ LÓPEZ
SERGIO FRANCISCO JUÁREZ CERRILLO



ISBN: 978-607-59328-1-1



UNIVERSIDAD VERACRUZANA
FACULTAD DE ESTADÍSTICA E INFORMÁTICA
ESPECIALIZACIÓN EN MÉTODOS ESTADÍSTICOS
UV-CA-107: METODOLOGÍA Y APLICACIONES
DE LAS TÉCNICAS Y MODELOS ESTADÍSTICOS

APLICACIONES DE LA METODOLOGÍA ESTADÍSTICA EN DIFERENTES DISCIPLINAS

COORDINADORES

JULIA AURORA MONTANO RIVAS
CECILIA CRUZ LÓPEZ
SERGIO FRANCISCO JUÁREZ CERRILLO

AUTORES

ANGEL FERNANDO ARGÜELLO ORTIZ, CECILIA CRUZ LÓPEZ, ELISEO GABRIEL ARGÜELLES, DARÍO SEBASTIÁN PAVÓN DÍAZ, JOSÉ JUAN MUÑOZ LEÓN, IRMA JANETH PRADO MÉNDEZ, JULIA AURORA MONTANO RIVAS, DARÍO ALEJANDRO SEGOVIA HERNÁNDEZ, JOSÉ MANUEL MENDOZA RIVERA, ERNESTO PEREDO RIVERA, JOSEFA CAROLINA FORTUNO HERNÁNDEZ, IRVING DENNIER ALATORRE BARILLAS, ÁNGEL JUAN SÁNCHEZ GARCÍA, EMMANUEL MORALES GARCÍA, ERNESTO PEDRO MENÉNDEZ ACUÑA, CARLOS ADRIÁN ALARCÓN ROJAS, TEOBALDO ORTIZ RAMÍREZ, CLAUDIO RAFAEL CASTRO LÓPEZ, ARELI KEILA ACATITLA DÁVILA, LILIAM MARIE PINZÓN, KEILA RAQUEL MEDINA AGUILAR, MARÍA YESENIA ZAVALETA SÁNCHEZ, MARYCRUZ ÁLVAREZ JIMÉNEZ, JESÚS HERNÁNDEZ SUÁREZ, SERGIO FRANCISCO JUÁREZ CERRILLO, ISABELLE BAROIS BOULLARD, CYNDY NOHEMY CARDONA CLAROS, ALEJANDRA ORTIZ GAZCA, ZOYLO MORALES ROMERO, WENDY SANGABRIEL CONDE, KEISY DANIELA PARRA HENAO, JULIANA SANDOVAL HERNÁNDEZ, RAFAEL SÁNCHEZ GREGORIO, JORGE VALENZUELA GONZÁLEZ, GIBRÁN RENOY PÉREZ TOLEDO.

EDITORIAL

©RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C. 2022



ISBN: 978-607-59328-1-1



EDITA: RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C.
DUBLÍN 34, FRACCIONAMIENTO MONTE MAGNO
C.P. 91190. XALAPA, VERACRUZ, MÉXICO.
CEL 2282386072
PONCIANO ARRIAGA 15, DESPACHO 101.
COLONIA TABACALERA
DELEGACIÓN CUAUHTÉMOC
C.P. 06030. MÉXICO, D.F. TEL. (55) 55660965
www.redibai.org
redibai@hotmail.com

Sello editorial: Red Iberoamericana de Academias de Investigación, A.C.
(978-607-59328)
Primera Edición, Xalapa, Veracruz, México.
Presentación en medio electrónico digital
Formato PDF 20 MB
Fecha de aparición 27/10/2022
ISBN 978-607-59328-1-1

Directorio Institucional

Dr. Martín Gerardo Aguilar Sánchez
Rector de la Universidad Veracruzana

Dra. Elena Rustrían Portilla
Secretaria Académica de la Universidad Veracruzana

Dra. Marisol Luna Leal
Abogada General de la Universidad Veracruzana

Dr. José Luis Sánchez Leyva
Director General del Área Económico-Administrativa de la Universidad Veracruzana

Dr. Luis Gerardo Montané Jiménez
Director de la Facultad de Estadística e Informática

Dra. Julia Aurora Montano Rivas
Coordinadora de la Especialización en Métodos Estadísticos

Publicación subvencionada por el Programa de Apoyo para el Fortalecimiento de los Cuerpos Académicos de la Universidad Veracruzana en apoyo al UV-CA-107 Metodología y Aplicaciones de las Técnicas y Modelos Estadísticos perteneciente a la Facultad de Estadística e Informática.

Agradecimientos

Se agradece el apoyo de la Secretaría Académica de la Universidad Veracruzana, a la Dirección General de Desarrollo Académico e Innovación Educativa, a la Dirección General del Área Económico-Administrativa y a la Dirección de la Facultad de Estadística e Informática para la publicación de este libro. Nuestro más sincero agradecimiento a todas las autoridades de dichas dependencias que siempre brindaron su apoyo incondicional, sin ustedes esto no sería posible. Así también a todos los estudiantes de la Especialización en Métodos Estadísticos generación 2021 por el esfuerzo realizado para que cada uno de sus trabajos estuviera plasmado como un capítulo de libro. Asimismo, a todos los dictaminadores que se tomaron el tiempo de evaluar el contenido y enviar sugerencias para la mejora de este. Un agradecimiento especial a los integrantes del UV-CA-107: Metodología y Aplicaciones de las Técnicas y Modelos Estadísticos y miembros del NAB de la Especialización en Métodos Estadísticos por colaborar guiando a los estudiantes para que la metodología estadística fuera aplicada correctamente.

Dictaminadores

Candidato a Dr. Carlos Abraham Carballo Monsivais
Centro de Investigaciones en Matemáticas, CIMAT Zacatecas

Dr. Diego Soto Hernández
Universidad de la Sierra Sur, Oaxaca

Dr. Héctor Moreno Loera
Facultad de Economía, Contaduría y Administración, Universidad Juárez del Estado de Durango

Dr. José Luis Ávila Valdez
Universidad Popular Autónoma del Estado de Puebla

Dr. Juan Andrés Sánchez García
Facultad de Arquitectura, Universidad Veracruzana

Dr. Juan Felipe Martínez Montoya
Colegio de Postgraduados, Campus San Luis Potosí

Dr. Juan Ruiz Ramírez
CA: Manejo Sustentable de Recursos Hortofrutícolas, Universidad Veracruzana

Dra. Katuska Fernández Morales
Instituto de Investigación y Desarrollo Educativo, Universidad Autónoma de Baja California

Dr. Pere Grima Cintas
Universidad Politécnica de Cataluña, España

Dr. Roberto Behar Gutiérrez
Universidad del Valle, Cali, Colombia

Mtra. Tania Guadalupe Ramírez Guerra
Tecnológico de Monterrey, Campus Cuernavaca

Dr. Víctor Leyva Sánchez
Pontificia Universidad Católica de Valparaíso, Chile

Presentación

En la actualidad, la metodología estadística y las habilidades en un software estadístico juegan un papel fundamental en todo proyecto de investigación que requiera del análisis de datos para obtener un nuevo conocimiento.

Desde el punto de vista científico, la aplicación de las técnicas estadísticas se apega a una secuencia o estrategia de análisis, es decir, se inicia con el planteamiento del problema a resolver, seguido de la formulación de hipótesis y objetivos; se identifica el diseño de selección de la unidad de estudio, se procede con la colección de la información (datos), así como, su ordenamiento para su análisis estadístico de modo que puede darse respuesta a los objetivos e hipótesis y por último la interpretación de resultados de los análisis. Con este proceso se puede arribar a la mejor toma de decisión, para confirmar un conocimiento ya existente u obtener uno nuevo.

La estadística es la ciencia del análisis de datos, que estudia por qué los sujetos se agrupan o por qué son diferentes de acuerdo con algunas características en común. Desde el punto de vista teórico, con la estadística se determinan modelos para conocer asociaciones entre características, para clasificar sujetos, hacer comparaciones, etc., principalmente para probar hipótesis de investigación.

Múltiples disciplinas tienen su soporte de la investigación en la estadística, por ello se puede decir que esta ciencia se puede considerar como una técnica transversal, ya que el biólogo, psicólogo, agrónomo, pedagogo, economista, climatólogo, entre otros, requieren de esta herramienta para desarrollar sus investigaciones y comprobar sus hipótesis, resultados que contribuyen en su conocimiento.

En la Especialización en Métodos Estadísticos, los alumnos son usuarios de la estadística ya que provienen de diferentes disciplinas. Cada uno llega con su proyecto de investigación, el cual está relacionado con su profesión; por ello, llegan con miras, de no solo aprender cómo aplicar las técnicas estadísticas para solucionar su problema, sino también acrecentar sus conocimientos y así continuar con mayor confianza sus investigaciones en su profesión. Al obtener resultados y emitir las conclusiones de su proyecto, adquieren la certeza de que la investigación está contribuyendo en beneficio de la sociedad.

En este documento se recompilan los proyectos concluidos de los alumnos de la generación 2021, mediante capítulos que muestran el trabajo conjunto entre alumno y docentes que forman el Núcleo Académico Básico de la Especialización y que además son integrante del Cuerpo Académico: Metodología y Aplicaciones de las Técnicas y Modelos Estadísticos. Es importante mencionar que en algunos proyectos colaboraron investigadores externos, lo que muestra la vinculación con otras instituciones.

El documento presenta una diversidad de temas y la contribución en las diferentes áreas de conocimiento, lo cual permite comunicar a los lectores algunos de los problemas actuales a los que se enfrenta nuestra sociedad y como han sido abordados a través de la investigación realizada mostrando la aplicación de las técnicas estadísticas.

Los proyectos que se presentan están a disposición del lector interesado.

Xalapa, Veracruz, octubre de 2022.

Los coordinadores

Índice

Capítulo 1	Eficiencia presupuestal en la atención de las desigualdades sociales municipales en México 2010-2020. <i>Angel Fernando Argüello Ortiz, Cecilia Cruz López, Eliseo Gabriel Argüelles</i>	7
Capítulo 2	Proyecciones económicas del sector cultural en México: una fuente de crecimiento desaprovechada. <i>Darío Sebastián Pavón Díaz, Cecilia Cruz López, José Juan Muñoz León</i>	22
Capítulo 3	Análisis espacial de la pobreza en el municipio de Xalapa-Enriquez de 2010 – 2020. <i>Irma Janeth Prado Méndez, Julia Aurora Montano Rivas</i>	40
Capítulo 4	Pronóstico del desempleo en México: una aplicación de series de tiempo multivariadas. <i>Darío Alejandro Segovia Hernández, Eliseo Gabriel Argüelles, Cecilia Cruz López</i>	57
Capítulo 5	Desigualdad y crecimiento económico: evidencia de la hipótesis del colesterol para México del 2000 – 2019. <i>José Manuel Mendoza Rivera, Ernesto Peredo Rivera, Josefa Carolina Fortuno Hernández.</i>	70
Capítulo 6	Algoritmos clasificatorios para estudiar la relación entre rendimiento escolar y uso del teléfono inteligente en estudiantes universitarios de algebra lineal. <i>Irving Dennier Alatorre Barillas, José Juan Muñoz León, Cecilia Cruz López, Ángel Juan Sánchez García.</i>	85
Capítulo 7	Comportamiento del intervalo de confianza y de cobertura para la proporción en la multinomial. <i>Emmanuel Morales García, Julia Aurora Montano Rivas, Ernesto Pedro Menéndez Acuña</i>	108
Capítulo 8	Influencia de la actividad física en sobrevivientes de cáncer de mama y colorrectal. <i>Carlos Adrián Alarcón Rojas, Ernesto Pedro Menéndez Acuña, Julia Aurora Montano Rivas</i>	122
Capítulo 9	Desgaste profesional y compromiso laboral en médicos residentes del estado de Veracruz. <i>Teobaldo Ortiz Ramírez, Claudio Rafael Castro López, Julia Aurora Montano Rivas, Carlos Adrián Alarcón Rojas</i>	137
Capítulo 10	Análisis estadístico de los factores de riesgo asociados con la actividad sexual en adolescentes mexicanos entre 2012 – 2018. <i>Areli Keila Acatitla Dávila, Claudio Rafael Castro López, Liliam Marie Pinzón</i>	153
Capítulo 11	Predicción de casos y mortalidad por diabetes, cambios alimentarios para le reducción del impacto. <i>Keila Raquel Medina Aguilar, María Yesenia Zavaleta Sánchez</i>	174
Capítulo 12	Análisis de componentes principales en altas dimensiones aplicado a la metabolómica de suelos del café con <i>Pontoscolex corethrurus</i>. <i>Marycruz Álvarez Jiménez, Jesús Hernández Suárez, Sergio Francisco Juárez Cerrillo, Isabelle Barois Boullard.</i>	190
Capítulo 13	Validación de modelos de calidad de hábitat para el Chigüiro (<i>Hydrochoerus hydrochaeris</i>) en la Orinoquía Colombiana. <i>Cyndy Nohemy Cardona Claros, Sergio Francisco Juárez Cerrillo</i>	200

Capítulo 14	Inoculación con hongos <i>Micorrízicos Arbusculares</i> como alternativa para optimizar el aprovechamiento de maíces criollos sembrados en policultivo. <i>Alejandra Ortiz Gazca, Zoylo Morales Romero, Wendy Sangabriel Conde</i> 215
Capítulo 15	Variables asociadas a la selección de hábitat de tres especies de tortugas Dulceacuicolas de la familia <i>Podocnemididae</i> presentes en la reserva natural Bojonawi (Vichada, Colombia), durante la estación seca. <i>Keisy Daniela Parra Henao, Zoylo Morales Romero</i> ... 226
Capítulo 16	Relación entre algunos caracteres morfológicos y fisiológicos de odonatos y sus patrones de distribución geográfica. <i>Juliana Sandoval Hernández, María Yesenia Zavaleta Sánchez, Cecilia Cruz López</i> 242
Capítulo 17	Comparación de la mirmecofauna en agroecosistemas de café (<i>Coffea arabica L.</i>) con diferente estructura arbórea en el centro de Veracruz. <i>Rafael Sánchez Gregorio, Cecilia Cruz López, Jorge Valenzuela González, Gibrán Renoy Pérez Toledo</i> 254

Eficiencia presupuestal en la atención de las desigualdades sociales municipales en México 2010-2020

Angel Fernando Argüello Ortiz¹, Cecilia Cruz López², Eliseo Gabriel Arguelles³

¹Facultad de Estadística e Informática, Universidad Veracruzana, aarguello@uv.mx

²Facultad de Estadística e Informática, Universidad Veracruzana, ceccruz@uv.mx

³Facultad de Estadística e Informática, Universidad Veracruzana, egabriel@uv.mx

Resumen

La asignación y distribución del presupuesto municipal, ha generado distancias cada vez más grandes entre las necesidades sociales de la población y su atención mediante el gasto público. Esto justifica, que la perspectiva financiera de las instituciones municipales deba apoyarse bajo criterios analíticos multi, inter y transdisciplinarios; con bases científicas y estrechamente relacionadas con la vida económica, política, administrativa y social, que resalten la importancia de las operaciones financieras con instrumentos de decisión de las finanzas públicas modernas que disminuyan las condiciones de rezago. Dichas desigualdades, evaluadas a nivel municipal entre el Total de Población, Índice de Rezago Social y el Presupuesto Municipal, indican que la asignación presupuestal, presenta regionalizaciones que impactan significativamente para la atención de necesidades sociales; lo cual, desde la estadística espacial, bajo el Índice de Moran, representa patrones espaciales de agrupación en los 2,469 municipios de México entre 2010 y 2020.

Palabras Clave: Rezago social, estadística espacial, índice de Moran, presupuesto municipal.

Introducción

El análisis de los municipios de México, bajo la idea de la existencia de innumerables dificultades económicas, políticas, sociales, administrativas y demográficas; refleja que durante varias etapas de la administración pública tanto del gobierno federal como de los gobiernos estatales y municipales, se han asignado fondos a los ayuntamientos y se han considerado como la aproximación más cercana a una función de bienestar social, la cual ha sido llevada a la práctica y mostrado cierto grado de limitaciones en sus niveles de eficiencia y eficacia, sin dejar de lado que las consideraciones en cada orden de gobierno difieren en gran medida entre cada uno de éstos, no de manera estricta en los objetivos, sino más bien en los mecanismos de acción y ejecución de la política social en primer lugar y en segundo lugar a través de las decisiones en que se basa la aplicación de las políticas públicas.

La acción política, ejercida principalmente en el campo económico y social, permite el surgimiento de una nueva técnica de acción en el medio de las finanzas públicas; la cual se puede identificar con el incremento del presupuesto público y por ende el surgimiento de la política financiera, a través de las cuales, las instituciones gubernamentales federales, estatales y municipales, de carácter socioeconómico o sociocultural sufran transformaciones, que reclamen modificaciones en el sistema político, mismas que dependen del nivel de intervencionismo del plano político y que por consiguiente, llevan a la generación de un nivel de relativismo con relación al criterio de rezago social que hasta este momento no es equiparable con la visión de los integrantes del plano político y/o tomadores de decisiones para la atención de necesidades.

La distribución de la riqueza para superar las grandes desigualdades sociales heredadas del pasado y la desigualdad humana, han sido elementos determinantes en la organización económica, política, social y cultural actual, calificada de acuerdo con la estratificación dominante como una sociedad clasista-piramidal de profundos contrastes. Las políticas públicas implementadas para la atención de la población en situación de desigualdad han propiciado que las distinciones separen cada vez más a la población en cada uno de los estratos, esto a pesar de que algunos logren avanzar en alguna de las escalas; de tal forma que resulta preponderante diseñar estrategias para superar este panorama, a través de una función de bienestar basada en proyectos o programas sociales, más acorde a la situación actual.

Las operaciones que las administraciones públicas municipales realizan sobre los fondos públicos para la atención de las necesidades de la población representan un proceso privilegiado, en el que la aplicación de estrategias financieras que atiendan los objetivos económicos y sociales de los municipios juega un papel determinante en el desarrollo. Sin embargo, dada la diversidad de criterios existentes para la identificación, medición y evaluación de los niveles de necesidades sociales, no se han establecido criterios específicos y unificados bajo un mismo enfoque, atendiendo diversas condiciones que impactan directa e indirectamente en el rezago social, y que a su vez se mide desde diversos indicadores. En este sentido, la estimación del rezago social por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), representa un punto de referencia importante; en primer lugar, para las políticas públicas federales, estatales y municipales y, en segundo lugar, dentro de la administración pública para atender y disminuir la desigualdad. Su estimación, se apoya en el estudio del comportamiento de varios indicadores sociales o socioeconómicos como son, el índice de marginación (CONAPO, Consejo Nacional de Población, 2018), los índices de pobreza, desarrollo humano, desarrollo social y rezago social, (CONEVAL, 2018), índice o coeficiente de Gini (World Bank, 2015), por mencionar algunos, los cuales expresan situaciones de desigualdad en la población, pero desde variables y criterios de medición diferentes. Lo que es innegable, es que, en todos esos indicadores, la razón principal para clasificar a un segmento de la población en condiciones de desigualdad es la falta de recursos económicos para la atención de necesidades básicas y/o dotación de servicios, ya sea por parte de los ayuntamientos o de la población.

El presupuesto de egresos es la asignación de recursos para los ayuntamientos, estado y federación, que deben ejecutarse durante un año, con el único fin de atender las necesidades de la población; mediante una programación anual de su ejecución o gasto. Dicha programación, forma parte de las actividades administrativas del ayuntamiento, y se lleva a cabo de manera coordinada, planeada y organizada por la tesorería municipal y con la asesoría y supervisión del alcalde y en su caso, la consulta de los integrantes del cabildo; la planeación de actividades institucionales cumple con calendarios de ejecución de trabajos específicos, así como una ejecución detallada de recursos (INAFED, 2021a).

Tal pareciera que la disminución del rezago social, ejecutada desde la asignación del presupuesto público municipal, no ha alcanzado los objetivos esperados, ya que los cambios en los indicadores socioeconómicos son determinados por el comportamiento de variables demográficas y por el manejo adecuado del presupuesto municipal (Llave, 2018). Empleando los datos del presupuesto municipal para los años 2010, 2015 y 2019, junto con los resultados del índice de rezago social y la población total para los años 2010, 2015 y 2020, se implementaron varios análisis de estadística espacial, con la finalidad de

estimar el tipo de patrón espacial (aleatorio o regional) entre los municipios, clasificando cada uno de los periodos como inicial (2010), intermedio (2015) y final (2019 o 2020).

Metodología

Estudio

De acuerdo con Méndez, Namihira, Moreno y Sosa (2011) por las características de los tipos de estudio, del presente trabajo se clasifica como una revisión de casos, al ser de carácter observacional, retrospectivo, longitudinal y descriptivo. Las unidades de estudio corresponden a los 2,469 municipios de México, de las cuales se integraron los indicadores de Población Total (PT), Presupuesto Municipal (PM) e Índice de Rezago Social (IRS) de carácter cuantitativo, por lo que la investigación será del mismo tipo, asumiendo la existencia de un nivel de incidencia espacial entre los elementos bajo estudio, cumpliendo con aspectos de claridad entre las variables, en la definición y en la delimitación.

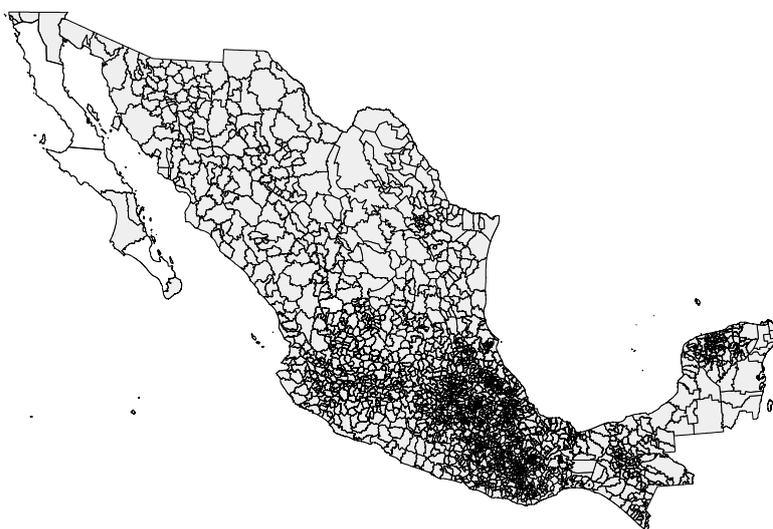


Figura 1. Municipios de México (2,469).

El cumplimiento correcto y planeado de la ejecución del gasto, beneficiará al proyecto de bienestar social propuesto por las administraciones municipales, mediante el desarrollo e implementación de los programas sociales a través de las unidades administrativas y con el aprovechamiento de los insumos humanos, económicos y materiales, clasificando el presupuesto en tres tipos: por objeto del gasto, institucional administrativo y por programas. Las administraciones municipales realizarán su programación municipal de manera integral, flexible, ordenada, participativa y cumpliendo los objetivos (INAFED, 2021b).

El concepto de rezago social tiene sus bases ideológicas en la teoría económica del capitalismo y en la doctrina filosófica del utilitarismo, doctrinas que con el paso del tiempo han sufrido modificaciones y/o ajustes, redimensionando paulatinamente la acción del Estado de bienestar durante las dos etapas desde su aparición; la primera de las etapas, que se basó en la primera economía del bienestar y el utilitarismo cardinalista clásico y la segunda etapa es la basada en la nueva economía del bienestar y el utilitarismo ordinalista.

Parafraseando a Arrow, uno de los "juicios de valor indiscutibles", se refiere a la posibilidad de que, entre los elementos bajo estudio, se pueden llevar a cabo comparaciones de utilidad, evaluadas bajo un esquema de medición ordinal en el que cada uno de los individuos tiene su propia escala de valores y éstas sean comparables entre sí, y no bajo un esquema de medición cardinal que resultaría de sobremanera difícil o imposible. En este sentido, el problema se centra en la maximización de la "función de bienestar social", sujeta a cuatro condiciones posibles (Chávez Presa & Gabriel Budebo, 2000).

1. Dominio ilimitado: la función está definida para una sociedad específica y para un conjunto específico de alternativas y ha de producir un orden o una clasificación social para cualquier perfil derivado de esas alternativas y los municipios.
2. No dictatorial: no debe clasificar sistemáticamente una alternativa sobre otra por el hecho considerarse la mejor opción para un municipio en particular.
3. Regla de unanimidad: la función de bienestar social ha de clasificar la alternativa x sobre la y si todos o la mayoría relativa de los municipios clasifican a x por encima de y (Wilfredo Pareto).
4. Independencia de las alternativas irrelevantes: al estudiar la clasificación de cualquier par de alternativas, las demás (las irrelevantes) no pueden depender del par estudiado.

Variables

El primer indicador (PT) corresponde al dato obtenido mediante eventos censales en los años 2010 y 2020, recopilada por Instituto Nacional de Estadística y Geografía (INEGI) mediante los Censos de Población y Vivienda; el segundo, corresponde al monto estimado por el ejecutivo, aprobado por el H. Congreso de la Unión y publicado en el Diario Oficial de la Federación y el tercero, corresponde a un indicador que las necesidades insatisfechas en materia de variables sociodemográficas relacionadas con educación, salud, servicios básicos en la vivienda y calidad en la vivienda, el cual es estimado por el Consejo Nacional para la Evaluación de la Política de Desarrollo Social (CONEVAL), enmarcado en la Ley General de Desarrollo Social, todos en los diferentes periodos de estudio de 2010, 2015 y 2020, analizando la asociación estadística y espacial de los indicadores.

El tamaño y la composición de la población varía en el tiempo debido a factores de orden biológico, social, cultural, económico y político. El conocimiento de la fisonomía poblacional constituye un elemento de gran importancia en la sociedad, en la medida que relaciona la inserción de la población en las actividades económicas y la disposición y repartición de recursos, constituyéndose en un elemento indispensable para fines de planificación (INEGI, 1997); por lo que se recopilaron las variables de Población total para los periodos indicados, todos levantados por el INEGI; el Presupuesto de Egresos de la Federación se clasifica como uno de los documentos más importantes de las políticas públicas en México, el cual corresponde a la Secretaría de Hacienda y Crédito Público (SHCP) su elaboración, describiendo la cantidad, la forma de distribución y el destino de los recursos públicos de los tres poderes (Ejecutivo, Legislativo y Judicial), de los organismos autónomos, y las transferencias a los gobiernos estatales y municipales (SHCP, 2021), integrado en las variables presupuesto municipal de los años especificados, publicado por el H. Congreso de la Unión; el índice de rezago social es un indicador resumen mediante el cual se clasifican las unidades de referencia (estados,

municipios, localidades o AGEB) dependiendo del nivel de carencias sociales, está integrado por 11 variables, concentradas en cuatro grupos de carencias sociales (educación, salud, servicios básicos y calidad y espacios en la vivienda), especificado en la Ley General de Desarrollo Social y estimado por el CONEVAL, el IRS se recopiló para el periodo de referencia, y cuenta con una variable clasificatoria (Muy Alto, alto, Medio, Bajo y Muy Bajo).

Tabla 1. Indicadores y variables estimadas.

Indicadores	Variables estimadas
Población Total (PT)	Crecimiento 2010-2015
	Crecimiento 2015-2020
Presupuesto Municipal (PM)	Cambio 2010-2015
	Cambio 2015-2019
Índice de Rezago Social (IRS)	Grado 2010-2020
	Cambio 2015-2020

Análisis exploratorio

Con base en un enfoque de costo-beneficio, se estima el crecimiento demográfico y los cambios en el presupuesto municipal y nivel de rezago social para los periodos inicial, intermedio y final, mediante las diferencias para cada uno de los municipios de México estableciendo criterios comparativos, entre un periodo y otro, e identificando la distancia entre el municipio con menor y mayor cambio.

Entre los indicadores de cambio, se establecen comparaciones, considerando que, bajo el principio de atención de las necesidades sociales, los cambios en el rezago social están asociados con el crecimiento demográfico y los niveles de inversión pública.

Con base en un análisis de utilidades, apoyado en la teoría del “utilitarismo”, que considera que las evaluaciones éticas de los estados sociales son las utilidades individuales, se establece que, en la medida en que crece la asignación presupuestal, a contrario sensu debería disminuir el rezago social, incluyendo el efecto del crecimiento demográfico.

Esta teoría involucra tres elementos importantes, el “bienestar basado en la utilidad”, enfoque que se aplica en la estimación de los cambios de los presupuestos municipales, midiendo el nivel de incremento de los montos correspondientes a cada municipio; la “ordenación mediante la suma”, que exige que la información referente a la utilidad de cualquier estado sea evaluada bajo la consideración de la suma total de las utilidades, condición implícita en el incremento constante de montos presupuestales, y por último, el “consecuencialismo” que asume que las obras y acciones llevadas a cabo a partir de la aplicación de los recursos generan un impacto en el actuar de las administraciones municipales, condición que en los municipios se identifica al cambiar su nivel de intensidad de necesidades. Se espera lograr que, a partir del supuesto de comparación municipal de los cambios en el nivel de rezago social y presupuesto asignado, se identificará el nivel de asociación para acercar a la población a los beneficios del desarrollo.

Existen serios contrastes en las asignaciones presupuestales realizadas a los municipios en comparación con el total de población registrada en los diferentes eventos censales, que en algunos casos disminuyeron su tamaño poblacional (Tabla 1); marcando una diferencia del 99.99% con respecto a la población más alta en cada uno de los periodos analizados, esto quiere decir que la población de al menos un municipio, no es equivalente ni al 0.01% del municipio más poblado (Tabla 2), junto con este comportamiento, se registraron casos en los que la disminución de la población alcanzó valores del 42.03% y del 48.13% en los periodos de 2010 a 2015 y de 2015 a 2020 (Tabla 3).

Tabla 2. Diferencias mínima y máxima de la población total municipal.

Población	Mínimo	Máximo	Diferencia
Año 2020	81	1,922,523	1,922,442
Año 2015	87	1,827,868	1,827,781
Año 2010	93	1,815,786	1,815,693

Tabla 3. Diferencias mínima y máxima del cambio estimado en el PM entre periodos.

Diferencia de Población	Mínima	Máxima
Periodo 2015-2020	-48.13%	172.75%
Periodo 2010-2015	-42.03%	318.21%

En términos presupuestales, las diferencias entre los municipios presentan la misma disparidad que en materia demográfica (Tabla 4); caso específico, el municipio con la menor población (81) en el año 2020, recibió un presupuesto en el año 2019 de 2 millones 411 mil 699 pesos, que no representó el presupuesto más bajo del mismo año, ya que fue de un millón 554 mil 495 y registró una población de 296 habitantes, 3.7 veces más que el municipio más pequeño.

Tabla 4. PM asignado en pesos a los municipios por periodo.

Presupuesto municipal	Mínimo	Máximo	Diferencia
Año 2019	1,554,495	8,681,052,746	8,679,498,251
Año 2015	1,504,141	6,356,963,548	6,355,459,407
Año 2010	1,086,749	6,238,269,764	6,237,183,015

Con respecto al IRS, no se registraron cambios significativos en los tres periodos de referencia; a pesar de que hubo un pequeño incremento de municipios, al inicio y final de los diez años, se mantuvieron 152 municipios con nivel muy alto de rezago social y en el caso de los de muy bajo nivel solamente se incrementó un, pasando de 676 a 677 (Tabla 5).

Tabla 5. Diferencias mínima y máxima de la población total,

Grado	2020	2015	2010
Muy bajo	677	673	676
Bajo	893	887	889
Medio	504	496	498
Alto	243	240	241
Muy alto	152	150	152
Total	2,469	2,446	2,456

Estadística espacial

En la estadística, los datos son la herramienta principal en los análisis estadísticos tradicionales y análisis espaciales. Los datos espaciales o geodatos, representan la información con características de georreferenciación que permiten la clara asociación y/o identificación de localización. La referencia de los resultados obtenidos mediante los análisis estadísticos a datos tabulares georreferenciados puede expresarse de manera tradicional y geográfica (Anselin, 1999). El análisis estadístico de datos geográficos es denominado en la literatura científica como análisis exploratorio de datos espaciales (AEDE) y es considerado como una disciplina de la estadística orientada al análisis y procesamiento de datos espaciales o geográficos (geodatos). A través del AEDE se identifica la distribución de los datos, así como, las relaciones entre variables espaciales, destacando que el comportamiento, relación y resultados obtenidos mediante estimaciones por metodologías estadísticas tradicionales puede no reflejar el mismo resultado mediante la asociación espacial, más aún, cuando se trabaja con grandes bases de datos de tipo vectorial o ráster (Chasco, 2014).

Las técnicas del AEDE, combinan el análisis estadístico con el gráfico, que al incorporar el enfoque espacial asociado al comportamiento de los datos, se presentan patrones diferentes, al punto de marcar tendencias, dependencia y/o concentraciones dependientes de la ubicación espacial de los individuos (puntos calientes/fríos o hot spots); por lo tanto, se conciben como una disciplina dentro del más general análisis estadístico, diseñada para el tratamiento específico de los datos geográficos (Chasco, 2003).

El AEDE, se concibe como una disciplina dentro del más general análisis estadístico, diseñada para el tratamiento específico de los datos geográficos e implica la realización de diversos análisis socioespaciales descriptivos, y exploratorios, empleando archivos vectoriales (puntos, líneas o polígonos) de formato shapefile que almacenan datos geográficos, tabulares y topológicos, manejados mediante Sistemas de Información Geográfica (SIG), con los cuales se estimará el Índice de Moran Global (IMG) e Índice de Morán Local (IML), para explicar el comportamiento espacial de las variables de rezago social y presupuesto municipal, identificando el tipo de patrón territorial (aleatorio o regional) que siguen las variables bajo estudio; mediante estos índices, se realizará la estimación de autocorrelaciones espaciales, explicando con el primero (IMG) de manera general como se distribuyen los municipios, ya sea de manera dispersa, aleatoria o agrupada; mientras que el segundo (IML) describe de manera local o individual la significancia estadística de las concentraciones espaciales asociadas, asignándoles una clasificación de acuerdo a la vecindad con las unidades (Alto-Alto, Alto-Bajo, Bajo-Alto, Bajo-Bajo); es decir, ambos indicadores describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o “atípicos espaciales” (“spatial outliers”), descubren esquemas de asociación espacial, agrupamientos (“clusters”) o puntos calientes (“hot spots”) y sugieren estructuras espaciales u otras formas de heterogeneidad espacial (Anselin, 1999).

En este sentido, se identificó que los municipios con los extremos poblacionales en el año 2020 fueron Santa Magdalena Jicotlán, Oaxaca y Tijuana, Baja California, más aún, la mayor parte de los municipios de México (1,855) tienen menos de 35,905 habitantes (75.13%), lo que implica que el país muestra una significativa diferencia en la concentración de población (Figura 2).

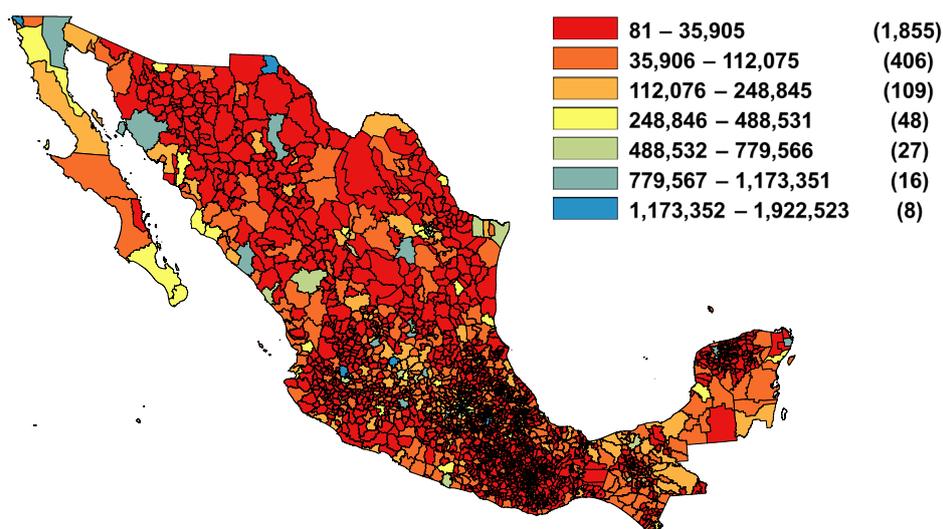


Figura 2. Distribución territorial de la población por municipio, 2020.

Empleando el método de optimización¹ de Jenks o cortes naturales de Jenks sobre el presupuesto municipal para cada periodo y clasificándolos en 5 grupos (Muy bajo, Bajo, Medio, Alto y Muy Alto) se identifica que el PM tampoco ha sufrido cambios significativos en la distribución espacial (Figura 3), pero si en cuanto a los montos (Tabla 6), ya que los montos más bajos de los presupuestos se asignaron aproximadamente al 83% de los municipios, mientras que menos del 1% concentraron los presupuestos más altos.

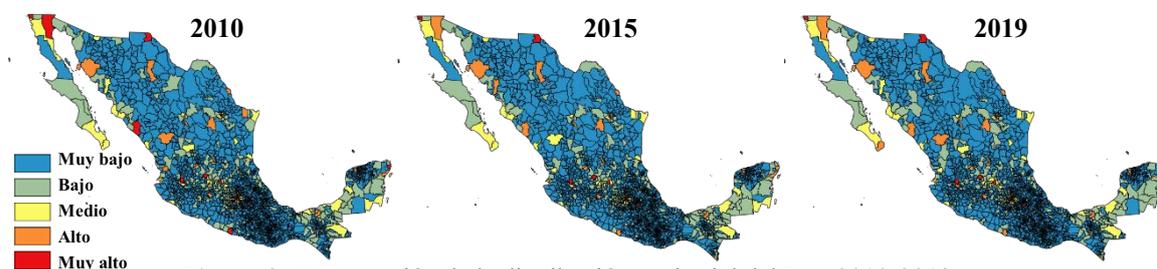


Figura 3. Comparación de la distribución territorial del PM, 2010-2019.

Tabla 6. Clasificación de los municipios por asignación municipal en millones de pesos.

Grupo	2010		2015		2019	
Muy bajo	1.09 - 145.5	83.2%	1.5 - 226.27	83.6%	1.55 - 261.47	82.9%
Bajo	145.5 - 554.19	11.7%	226.27 - 899.79	11.6%	261.47 - 996.05	11.9%
Medio	554.19 - 1,345.51	3.0%	899.79 - 2,051.61	2.9%	996.05 - 2,329.84	2.8%
Alto	1,345.51 - 2,647.49	1.4%	2,051.61 - 3,708.5	1.3%	2,329.84 - 4,445.06	1.8%
Muy Alto	2,647.49 - 6,238.27	0.7%	3,708.5 - 6,356.96	0.6%	4,445.06 - 8,681.05	0.6%

En un contraste espacial del IRS, se aprecia que en los tres periodos la regionalización del rezago se mantuvo casi constante (Figura 4); es decir, la zona con nivel muy alto (152) se ha mantenido, mientras que la clasificación muy baja cuadruplicó esta cantidad en los 3 periodos, situación que es alentadora, y aunado al nivel bajo, aproximadamente el 63% de los municipios de México registran bajos niveles (Tabla 5).

¹ Método de agrupación que reduce la varianza intra clases y la incrementa inter clases.

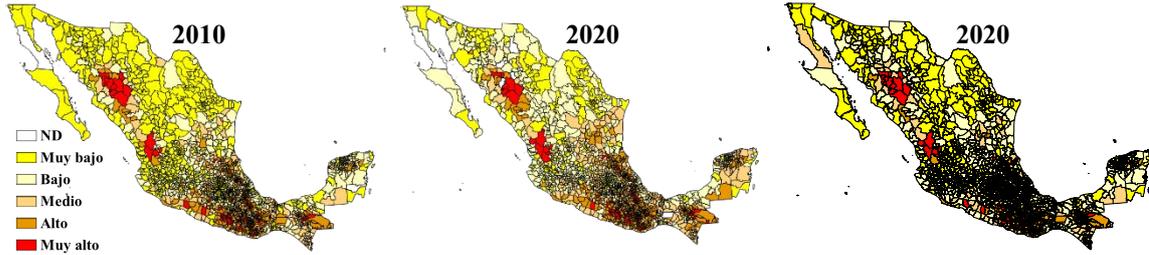


Figura 4. Comparación de la distribución territorial del IRS, 2010-2020.

La estimación de autocorrelación espacial explica el nivel de asociación entre las unidades de medición, indicando su grado de similitud y está directamente relacionada con el planteamiento de la hipótesis nula; en la que se especifica que las unidades se distribuyen de manera aleatoria, y se mide en función de la ubicación espacial de las unidades y la variable bajo estudio (Figura 5). Mediante la autocorrelación se realiza la clasificación de las unidades y la identificación de valores atípicos espaciales, para ello, se estiman el Índice de Moran, puntuaciones z y pseudovalores P, estos últimos, se emplean para identificar la significancia estadística de los valores de índice calculados. Dependiendo del valor P se rechaza o no la hipótesis nula, al ser una estimación del área bajo la curva de una distribución normal, en el caso de rechazar la hipótesis nula, se asumirá que no existe un patrón aleatorio.

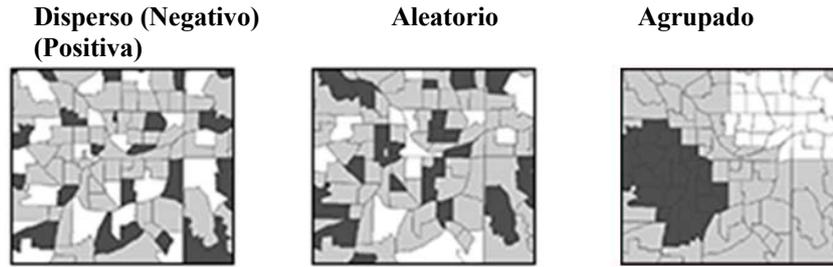


Figura 5. Comparación de la distribución territorial del IRS, 2010-2020.

El estadístico I de Morán Global, para autocorrelaciones espaciales se obtiene mediante:

$$I = \frac{n}{S_0} * \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}$$

Donde:

z_i es la desviación de un atributo de los elementos i donde su promedio es $(x_i - \bar{X})$

w_{ij} es el peso espacial entre los elementos i y j ,

n es el número de elementos

S_0 es el valor agregado de todos los pesos espaciales.

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

Los z_i se estiman como:

$$z_i = \frac{I - E[I]}{\sqrt{V[I]}}$$

Donde:

$$E[I] = \frac{-1}{(n-1)} \quad \text{y} \quad V[I] = E[I^2] - E[I]^2$$

En el caso de que el índice registre un valor positivo, significa que una unidad con valor alto está rodeada de unidades con valores altos, un valor bajo del índice, indica que una unidad con valor bajo está rodeada de valores bajos; en el caso de que el índice registre un valor negativo se considerará como un valor atípico, ya que una unidad está rodeada por unidades con valores diferentes.

Resultados

Los Criterios Generales de Política Económica (CGPE) describen la propuesta del Ejecutivo Federal de los lineamientos de política económica y fiscal, resaltando la importancia de contar con finanzas públicas sólidas y consistentes que aseguren al Estado su cumplimiento de atender los rubros relacionados con la educación, salud, economía, pobreza, servicios, seguridad social, seguridad pública, empleo, entre otros; por lo que impacta de manera directa en los rubros de atención del rezago social, de tal forma que también debería generar cambios positivos en la disminución paulatina de las necesidades sociales; sin embargo, la asignación del PM con base en el IRS refleja que los municipios con mayor nivel de rezago social reciben los montos presupuestales más bajos (Figura 6), lo cual se pudo identificar en por lo menos los 10 años de estudio, condición que se ha acrecentado año tras año, reflejando una mala planeación en la asignación y ejercicio de las finanzas públicas; es decir, se espera un resultado diferente en la disminución del rezago social, pobreza, marginación, y demás indicadores relacionados con la atención de las carencias sociales, realizando las mismas acciones de asignación del PM y ejercicio del mismo por las autoridades municipales.

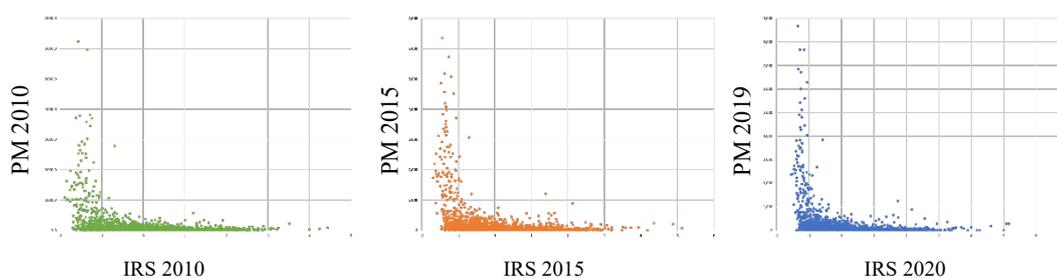


Figura 6. Comparación del PM en mdp contra el IRS por periodo.

El IRS expresa un comportamiento espacial de autocorrelación que indica la presencia de un patrón de agrupación, lo que permite rechazar la hipótesis nula bajo un nivel de significancia $\alpha = 0.01$, contra el que se contrasta el valor P obtenido mediante el IMG que fue de 0.0 para los tres periodos, con valores Z elevados (Tabla 7), esto indica que valores muy grandes o muy pequeños de Z, junto con valores P cercanos a cero ubica la decisión en la zona de rechazo, situación que ocurre en el IRS y en el PM (Figura 7).

Tabla 7. Valor P del IRS y PM por periodo.

Periodo	IRS Valor Z	PM Valor Z
Inicial	190.176893	49.584496
Intermedio	195.441502	60.951401
Final	169.304094	55.358445

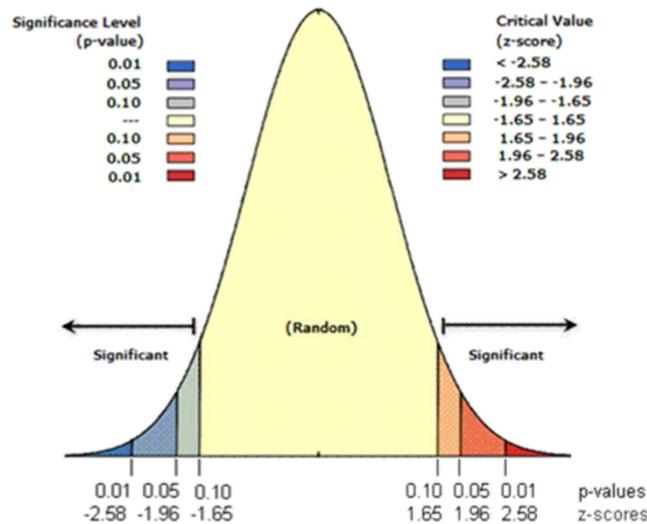


Figura 7. Área bajo la curva con valores P y Z.

Esto se aprecia claramente mediante la Figura 8, que representa las concentraciones territoriales de los municipios y que como se aprecia, las regiones con valores muy altos de rezago social son muy extensas, pero más aún, bajo la consideración de que zonas afines se atraen y zonas diferentes se repelen, así ocurre en este comportamiento, para los periodos inicial, intermedio y final, concentraciones de 847, 853 y 591 municipios respectivamente, registraron valores altos en el IRS y estaban rodeados por municipios con valores muy similares (Alto-Alto); lo mismo en el caso de municipios con valores bajos de rezago social y que sus vecinos tenían características parecidas (Bajo-Bajo).

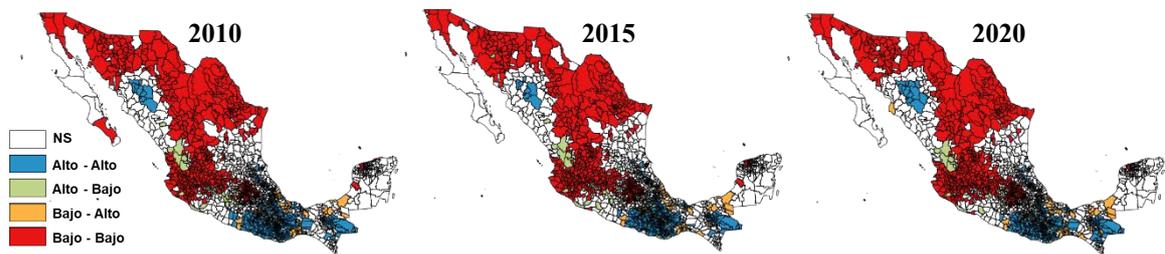


Figura 8. Mapa de autocorrelación espacial del IRS por periodo.

Este comportamiento va más allá de estimaciones estadísticas a las que tradicionalmente recurrimos, ya que involucra contextos territoriales que influyen en el desarrollo o retraso socioeconómico de los municipios, ya que pueden fungir como polos de atracción e influir en la atención de necesidades en las zonas limítrofe.

Por otro lado, la presencia de valores atípicos (Alto-Bajo, Bajo-Alto), 189 en 2010, 199 en 2015 y 181 en 2020, si bien indica una disminución, no representa un logro significativo en la atención de necesidades y mucho menos en la clasificación que se le asigna a nivel nacional, ya que no se considera el efecto espacial.

Tabla 8. Concentración de municipios del IRS por tipo de agrupación y periodo.

IRS	2010	2015	2020
No Significativo (NS)	847	853	906
Alto - Alto	672	623	591
Alto - Bajo	45	47	40
Bajo - Alto	144	152	141
Bajo - Bajo	761	794	791

La concentración de municipios basada en la autocorrelación espacial del PM, también indicó que no siguió un patrón aleatorio; sin embargo, la presencia de registros atípicos y no significativos resalta más que en comparación con el IRS (Tabla 9), lo cual resulta más evidente en la regionalización de los municipios, que para las agrupaciones (Alto- Alto y Bajo-Bajo) se presentan con mayor separación o dispersión grupal (Figura 9).

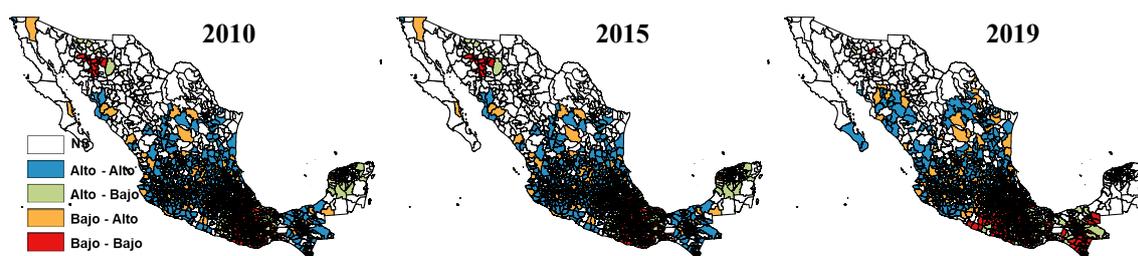


Figura 9. Mapa de autocorrelación espacial del PM por periodo.

En este caso, la influencia de los municipios en las zonas limítrofe cobra una significancia mayor sobre el impacto en el rezago, debido a que la ejecución del gasto municipal no puede transferirse hacia otros, de tal forma que los municipios con altos niveles de rezago y bajo presupuesto que a su vez colindan con municipios con presupuesto alto nivel bajo de rezago, marcarán zonas de alta disparidad, más en la zona limítrofe, considerando que el rezago social se presenta con mayor fuerza en las zonas periféricas.

Tabla 9. Concentración de municipios del PM por tipo de agrupación y periodo.

IRS	2010	2015	2019
No Significativo (NS)	887	975	885
Alto - Alto	532	455	482
Alto - Bajo	151	163	160
Bajo - Alto	149	177	192
Bajo - Bajo	750	699	750

Los cambios en la condición de rezagos social entre 2015 y 2020, asociados al presupuesto asignado, reflejaron avances al lograr una disminución en el 47.23% de los municipios (1,166), a pesar de la disparidad presupuestal, por otro lado, los 12 municipios con los presupuestos más altos que registraban muy bajo IRS, mantuvieron su condición; algo que es importante destacar en materia de eficiencia presupuestal, es que el 70.8% de los municipios que disminuyeron su IRS recibieron presupuestos por debajo de los 261

millones 465 mil 473 pesos (Tabla 10) y más aún, estos mismos, representan el 50% de los municipios; con muy bajo presupuesto.

Tabla 10. Comparativo de municipios por clasificación del PM en 2019 y cambio de posición en el IRS entre 2015 y 2020.

Clasificación del PM 2019	Cambio en el IRS municipal entre 2015-2020				
	Nuevo	Disminuyó	Sin cambio	Incrementó	Total
ND	14	242	217	6	479
Muy bajo	8	825	805	12	1,650
Bajo	1	92	144		237
Medio		6	49		55
Alto		1	35		36
Muy Alto			12		12
Total	23	1,166	1262	18	2,469

Estableciendo un comparativo de los municipios que disminuyeron su IRS con niveles muy bajos de presupuesto, contra la autocorrelación espacial del PM 2019 y del IRS 2020, se identifican regiones afines entre los municipios que disminuyeron su IRS y los que se encontraban auto correlacionados en ese mismo indicador y que a su vez habían resultado no significativos en autocorrelación del PM; esto quiere decir, lo que llevaría a plantear la conjetura de que la disminución del IRS en 2020, es resultado de una correcta, adecuada y acertada administración de recursos (Figura 10).

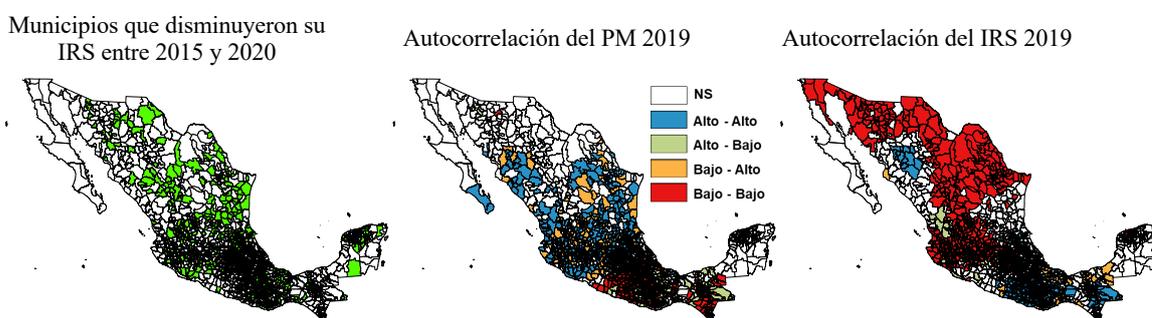


Figura 10. Comparativo municipal de disminución del IRS y autocorrelación del PM y el IRS.

Conclusiones

El presente estudio, muestra que el monto de asignación presupuestal para los municipios se podría considerar un ejercicio eficiente; sin embargo, el nivel de eficiencia puede asociarse a un uso racional, acertado y efectivo, en materia de ejecución del gasto público por parte de cada uno de los ayuntamientos que así lo lograron, ya que en lo que se refiere a la disminución del rezago social se identificó un cambio significativo entre los municipios con niveles presupuestales muy bajos, no así entre los municipios que recibieron montos altos y muy altos, que a su vez concentran la mayor parte del presupuesto de egresos.

Esto se justifica, debido a la identificación de un patrón de concentración de municipios, tanto para el IRS como para el PM, regionalismos que entre las políticas públicas y las desigualdades sociales, se establecen como prioridades de atención a nivel nacional. En materia de desigualdad social, el municipio que no avanza se atrasa; es decir, al no atender las necesidades, éstas se incrementarán con el paso del tiempo, ya que la población día a día genera más necesidades y por ende reclama más y mejores condiciones de vida, aunado al crecimiento demográfico no planeado, que abona en el retraso del desarrollo cuando hay tantas necesidades insatisfechas.

Por lo tanto, en la medida, en que se fije el estándar de atención del rezago social, plenamente vinculado con la asignación presupuestal, será viable desarrollar un proyecto confiable de ejecución del gasto; junto con una serie de actividades interinstitucionales gobierno-sociedad y con el apoyo de análisis estadístico y espacial, que impulsen medidas para la disminución paulatina de las necesidades sociales, y que a su vez genere un crecimiento socioeconómico municipal a un ritmo mayor.

La preparación de dicho proyecto deberá apoyarse en una función de bienestar determinada, la cual exige la definición de diversos aspectos que expliquen de manera clara y precisa la situación en que se encuentran los municipios, nivel de oportunidad para alcanzar los objetivos establecidos, evaluación de la administración de los recursos públicos en periodos anteriores, asertividad en la ejecución del gasto, población beneficiada y valoración de las necesidades sociales a atender, mismas que deberán apearse a los planes de desarrollo municipal y ejecución de los programas sociales marcados en el presupuesto municipal. Entre los principales criterios se encuentra la medición y valoración de costos y beneficios de los programas sociales, consideradas a partir de los resultados obtenidos en torno a la atención de la población rezagada; así mismo, a través de estos criterios se deben identificar algunas de las repercusiones sociales y desventajas de la función de bienestar según sea el caso; aunado a esto, establecer esquemas

Finalmente, urge una transformación de las instituciones gubernamentales federales, estatales y municipales, responsables de la atención de las desigualdades sociales, así como un rediseño o actualización en los criterios de asignación del presupuesto municipal para la atención de las necesidades sociales, ya que, la existencia cierto grado de relativismo con relación a la medición y atención de las desigualdades sociales, provoca una visión confusa o poco clara en los integrantes del plano político y/o tomadores de decisiones para la atención de necesidades, enmarcado por la ausencia de una definición o clasificación única de las desigualdades sociales.

Referencias

- Chávez, J. & Gabriel, M. (2000). Logros y retos de las finanzas públicas en México. Santiago de Chile: CEPAL.
- Chasco, C. (Abril de 2003). Comunidad de Madrid. España: Consejería de Economía e Innovación Tecnológica. Obtenido de <http://www.madrid.org/bvirtual/BVCM005618.pdf>
- Chasco, C. (2014). Análisis estadístico de datos geográficos en geomarketing: el programa GeoDa. (2018). Madrid, España: Departamento de Economía Aplicada. Universidad Autónoma de Madrid.

- Cohen, E., & Tapia. (2001). Gestión de programas sociales en América Latina. Análisis de casos. Santiago de Chile: CEPAL.
- CONAPO. (1999). La situación demográfica de México, 1999. México, D.F.: Consejo Nacional de Población.
- CONAPO. (21 de diciembre de 2018). Consejo Nacional de Población. Obtenido de CONAPO:
http://www.conapo.gob.mx/work/models/CONAPO/Marginacion/Datos_Abiertos/Municipio/Base_Indice_de_marginacion_municipal_90-15.csv
- CONEVAL. (27 de diciembre de 2018). Consejo Nacional de Evaluación de la Política de Desarrollo Social. Obtenido de CONEVAL:
https://www.coneval.org.mx/Medicion/Documents/Pobreza_municipal/Concentrado_indicadores_de_pobreza.zip
- INAFED. (28 de 01 de 2021). Obtenido de Guía Técnica 7. Elaboración y ejercicio de presupuesto de egresos:
http://www.inafed.gob.mx/work/models/inafed/Resource/335/1/images/guia07_elaboracion_y_ejercicio_del_presupuesto_de_egresos.pdf
- INAFED. (07 de 02 de 2021). Obtenido de Guía Técnica 5. La Programación Municipal:
http://www.inafed.gob.mx/work/models/inafed/Resource/335/1/images/guia05_la_programacion_municipal.pdf
- INEGI. (1997). Manual de medidas sociodemográficas. Aguascalientes, Ags.: INEGI.
- Jenks, G. (1967). El concepto de modelo de datos en estadística Mapping, Anuario internacional de cartografía 7:186-190.
- Llave, E. d. (28 de diciembre de 2018). Gaceta Oficial Órgano del Gobierno del Estado de Veracruz de Ignacio de la Llave. Obtenido de Gaceta Oficial:
<http://www.veracruz.gob.mx/gobiernover/gaceta-oficial/>
- Martínez Rodríguez, R. (2000). Teoría de la Elección Pública, Fundamento de las Finanzas Públicas. Xalapa, Ver., México.
- McMaster, R. (1997). In Memoriam: George F. Jenks (1916-1996). Cartography and Geographic Information Systems, 24:1, 56-59, DOI: 10.1559/152304097782438764
- Méndez, I., Namihira, D., Moreno, L., & Sosa. (2011). Protocolos de investigación. Lineamientos para su elaboración y análisis. México, D.F: Editorial Trillas.
- SHCP. (10 de marzo de 2020). Paquete Económico para el Ejercicio Fiscal 2019. Obtenido de <https://www.ppef.hacienda.gob.mx/es/PPEF2019>
- SHCP. (10 de enero de 2021). Presupuesto de Egresos de la Federación. Obtenido de <https://www.pef.hacienda.gob.mx/>
- Vargas Sánchez, J. R., & Velázquez Orihuela, D. (2011). Temas Selectos de Economía. México.
- World Bank, (2015), "A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals", Policy Research Report. Washington, DC: World Bank.

Proyecciones económicas del sector cultural en México: una fuente de crecimiento desaprovechada

Darío Sebastián Pavón Díaz¹, Cecilia Cruz López², José Juan Muñoz León³

¹Facultad de Estadística e Informática. Universidad Veracruzana. d.sebastianpvn@gmail.com

²Facultad de Estadística e Informática. Universidad Veracruzana. cecacruz@uv.mx

³Facultad de Estadística e Informática. Universidad Veracruzana. juanmunoz@uv.mx

Resumen

La cultura es parte fundamental de la vida de cada individuo y de la sociedad en su conjunto y ha sido así desde siglos atrás; hoy en día la cultura también es una industria generadora de riqueza, la cual puede llegar a ser un catalizador del crecimiento económico de los países, aunque esto es algo reciente y no siempre fue pensando así por los grandes economistas. México sufre de una crisis económica lo que se puede ver en el decrecimiento del producto interno bruto, el aumento del desempleo y empleados informales. Esta alarmante crisis nos da la tarea de encontrar nuevas estrategias para la reactivación de la economía; para el crecimiento de una economía es esencial el buscar sectores clave, en los que se tengan una ventajas comparativas, en los cuales enfocarse; por lo que en este proyecto, con datos capturados de la Cuenta Satélite de Cultura que recopila el INEGI año con año, se analizaron los datos del sector cultural de México para poder conocer como se ha comportado, cuál ha sido su rol en la economía, cómo la inversión pública ha afectado su desempeño, qué subsectores tienen un mayor impacto y cómo los resultados que presenta se comparan con los resultados que tiene el sector cultural de Corea del Sur, con datos del Servicio de Información Estadística de Corea (KOSIS), el cual ha sido una fuente de crecimiento para este país en los últimos años; todo esto para conocer si realizar una política económica enfocada a aumentar la inversión designada al sector cultural, con el objetivo de un crecimiento económico, es viable y tendrá los resultados que se esperan. Como resultados, el gasto público explica al PIB cultural, pero para que este impacte de manera positiva, se recomienda que los recursos se destinen a las áreas generales de artesanías y medios audiovisuales; también utilizar estrategias que fueron previamente ejecutadas en Corea del Sur con resultados favorables pueden considerarse para realizarse en México, como lo es el enfoque a la exportación de bienes culturales, ya que los sectores culturales de ambos países tienen similitudes.

Introducción

Thorsby (2001) dice que, aunque la cultura y la economía representan esferas separadas del interés humano, para encontrar el lugar de la cultura dentro de la sociedad se necesita incluir forzosamente aspectos económicos. La cultura podría ser definida de muchas formas sin embargo en general engloba los procesos simbólicos de la sociedad (Giménez, 2005). Para los fines de esta investigación la cultura es considerada como las manifestaciones espirituales y creativas de los humanos en las bellas artes que incluye las prácticas de la música, literatura, pintura, escultura, teatro y cine (Austin, 2000).

La cultura puede ser dividida en tres bloques, el primero está constituido por todas las obras y expresiones artísticas, las cuales son los principales objetos con valor cultural; dentro de este bloque se encuentra el patrimonio, ya sea histórico, arqueológico o documental, así como los museos y bibliotecas que los guardan; las artes plásticas, las artes escénicas y musicales también son parte de este bloque al que se le puede llamar el

sector de la alta cultura. Un segundo sector sería el de las industrias culturales, los cuales producen bienes culturales en grandes cantidades, ya sean obras literarias, musicales o audiovisuales, para la venta y distribución. Finalmente, Herrero (2011), nos dice de la existencia de un tercer bloque el cual engloba a las industrias creativas, en este destacan la creación de software, el diseño y los servicios de publicidad.

La economía es la ciencia que estudia los sistemas que coordinan la producción de bienes y servicios, tanto de los individuos como de la sociedad en su conjunto, para un funcionamiento correcto de la sociedad (Krugman, 2008). La producción es el proceso en donde insumos son transformados para obtener, al final del proceso, un bien. Estos insumos pueden ser o materias primas u otros bienes que ya habían sido previamente producidos, a los cuales se les llaman bienes intermedios ya que volverán a ser utilizados en un proceso productivo (Banco de México, BANXICO, 2020).

México es poseedor de una gran riqueza cultural, pasando desde aspectos prehispánicos, coloniales y modernos, además de los distintos pueblos indígenas que han subsistido desde la época prehispánica (López, 1993). En 2013, más de dos millones de mexicanos conformaban la industria cultural en el país (INEGI, 2013). En 2017 INEGI también reportó que las artes son de los campos de estudio con las tasas emprendedoras más altas. Representando un 2.54% del total de la economía en México, podemos decir que la cultura en México tiene un gran potencial tanto por su historia como por el amplio mercado y personas ocupadas dentro del sector (INEGI, 2018); sin embargo, la globalización, la migración del campo a la ciudad y la urbanización han resultado en una gran pérdida cultural en México. Buenrostro (2013), nos dice que nuestra cultura no solo se ha transformado, sino que también se han integrado elementos de diferentes culturas internacionales. El interés de los mexicanos hacia la cultura es relativamente bajo, siendo de un 52% del total de la población (secretaría de Cultura, 2010).

Gracias a datos de la Cuenta Satélite del INEGI (2018) podemos observar que el porcentaje que representa la cultura dentro del Producto Interno Bruto (PIB) nacional ha disminuido considerablemente en la última década, disminuyendo de un 3.22 % en 2008 a 2.54 % del PIB en 2018. El PIB es uno de los indicadores económicos que se calculan en distintos periodos de tiempo los cuales ayudan a determinar si una economía está teniendo buenos resultados y es uno de los más utilizados a nivel mundial; el PIB recopila los datos de producción de bienes y servicios de consumo final lo cual representa el tamaño de las economías. El PIB es especialmente usado dentro de la macroeconomía, la cual es una rama de la económica que se estudia los agregados económicos, el PIB siendo uno de estos, ya que engloba las cuentas nacionales. La macroeconomía también estudia otros temas como la balanza de pagos, el desempleo, la inflación y el tipo de cambio ya que estos están relacionados con las cuentas nacionales (Ávila, 2004).

El PIB también puede ser medido por medio del gasto, esto es porque todos los bienes que produce una economía se gastan. De Gregorio (2007) nos explica que, al ser medido mediante el gasto, el PIB se escribe de la siguiente manera:

$$Y = C + I + G + XN$$

Y representa el total de la producción, o también llamado demanda agregada; la C representa el consumo, el cual es el gasto final de todos los bienes y servicios que son adquiridos por los hogares o instituciones sin fines de lucro, este representa alrededor de

dos terceras partes del gasto total de la economía; la I que representa a la inversión, la inversión se encuentra en distintas maneras, usualmente son aquellos bienes que se adquieren para ser mantenidos en el futuro, ya que estos se pueden utilizar para la creación de otros bienes, los cuales pueden ser nuevos o como reemplazo, o para ser vendidos en un futuro; la G es el gasto de gobierno es el gasto en bienes y servicios finales, como el de la variable consumo, pero esta vez hechos por el gobierno; sin embargo no hay que confundir esto con el gasto público que se realiza como inversión, la cual se incluye en la variable con el mismo nombre; finalmente XN representa a las exportaciones netas, su cálculo es la diferencia de las exportaciones menos las importaciones, estas se agregan al cálculo de la demanda agregada dado que existen bienes consumidos en el país que no son de origen nacional y no todos los bienes nacionales son consumidos solamente por la población del país (De Gregorio, 2007).

La economía mexicana se encuentra en crisis, en el primer trimestre de 2020 se tenía un Producto Interno Bruto trimestral de -1.2 %, siendo afectado por los mercados internacionales, la dependencia económica al petróleo, la crisis sanitaria por COVID-19 entre otras cosas (INEGI, 2020). Añadiendo a las tasas negativas de producción, en México también existe una tasa de desocupación de 3.5 % mensual en marzo 2020 (INEGI, 2020), 41.9 % de índice de pobreza, y 57.1% de la población económicamente activa se encuentra en situación de empleo informal, por lo que el crecimiento económico es una problemática prioritaria para el país (CONEVAL, 2018). El crecimiento económico es de gran relevancia dentro de la ciencia macroeconómica, mencionada anteriormente. Valenzuela (2008) nos dice que el crecimiento económico es un indicador que explica el aumento la producción agregada o la producción agregada per cápita en un determinado periodo de tiempo de un país, el cual, en la actualidad, es utilizado universalmente como un indicador pensado para trabajar en procesos que lleven a aumentarlo y así la economía se encuentre en mejores condiciones para beneficio de la población. Para el cálculo del crecimiento económico se encuentra la variación del total de producción, de un periodo determinado con otro; la producción también puede ser dividida entre el número de habitantes para obtener el producto agregado per cápita, este es importante ya que muestra una perspectiva diferente a la del país en su conjunto al mostrar cómo se encuentra en promedio la población, aunque este solo da una perspectiva limitada (Banco de México, 2020).

La Organización de las Naciones Unidas para la Educación, Ciencia y la Cultura (UNESCO, 2016) establece que la cultura es un catalizador para el crecimiento económico y México, con su riqueza cultural, podría aprovecharla para reactivar la economía y fortalecer un sector que ofrece gran potencial tanto a corto plazo como a largo plazo. Para ejemplificar los efectos que la cultura tiene en la economía a corto plazo, Seaman (2004), nos dice que la cultura atrae a ciertos tipos de visitantes, ya sea turistas o consumidores locales, mismos que aumentan la demanda agregada y que a su vez desencadenan efectos a otros sectores de la economía. Los eventos culturales atraen ingresos por parte de los consumidores, pero a su vez contrata fuerza laboral que será remunerada con su respectivo salario. Mientras que a largo plazo se pueden formar ciudades creativas, el día de hoy podemos observar claros ejemplos de éstas como Los Ángeles, Nueva York, o Seúl, donde empresas y personas creativas se desplazan a estas ciudades en las cuales las oportunidades dentro del campo cultural son mucho mayores (Florida, 2009).

Una de las grandes tareas de los economistas ha sido encontrar diferentes maneras para que los países alcancen el crecimiento económico, es por eso por lo que muchas escuelas han planteado diferentes teorías de crecimiento económico (Enríquez, 2016). Todas estas teorías pretenden encontrar qué componentes de la economía son las que influyen en el crecimiento económico, para esto se tienen algunos supuestos, también llamados hechos estilizados, los cuales son regularidades que ocurren en todos los sistemas económicos, que facilitan la creación y explicación de las teorías.

La nueva teoría del crecimiento económico (NTC) surge debido a que las teorías neoclásicas que la preceden no lograban explicar el crecimiento a largo plazo sin el uso de la tecnología como una variable exógena, por lo que la NTC incluye la tecnología dentro del mismo modelo (Jiménez, 2011). La NTC relaciona al crecimiento económico con el desarrollo económico de manera muy estrecha a comparación de las teorías neoclásicas; el desarrollo económico estudia al crecimiento, incorporando a las transformaciones que ocurren en la sociedad durante el proceso a su estudio, estas pueden ser dentro de la tecnología, la educación, las relaciones sociales, la estructura productiva, las instituciones, entre otros (Zermeño, 2004). Es por eso por lo que se dice que sin crecimiento económico no se puede dar un desarrollo, donde las personas de una economía logren una mejor calidad de vida. La creación de la NTC se encuentra en los trabajos de distintos autores como Paul Romer (1986;1990), Robert Lucas (1988), Robert Barro (1990) y Sergio Rebelo (1991) y, los trabajos de Arrow (1962) y Frankel (1962) como antecedentes.

Uno de los ámbitos importantes de la nueva teoría de crecimiento económico es su énfasis en la especialización de los diferentes sectores económicos; Ocegueda, Castillo y Varela (2009) explican que la especialización permite aumentar el desempeño económico sin importar el tipo de bien que se produzca. Por lo tanto, se deben identificar sectores clave en los cuales se puedan encontrar ventajas económicas las cuales serán aprovechadas para acelerar el crecimiento.

El sector cultural está conformado por todas aquellas firmas que produzcan bienes y servicios culturales, así como los bienes y servicios culturales que son producidos desde los hogares; este sector se caracteriza por su creación de propiedad intelectual (Yudice, 2002). Citando a Baumon y Bowen. Aguado, Palma y Pulido (2017) nos dan un acercamiento a los inicios del sector cultural, los cuales no se consideraban importantes dentro de la economía hasta inicios de los años sesenta, y concluyen que este tipo de bienes y servicios tienen características únicas que lo diferencian de otros sectores; al empezar a ser considerado importante, el sector empezó a recibir financiamiento público para su mejor funcionamiento. El sector cultural pasó de ser visto como algo que no era de interés para la política económica a un área prestigiosa y especializada.

La política económica es una rama de la ciencia económica que estudia cómo se involucra el Estado, mediante el gasto público, en los procesos económicos y las intervenciones que realiza para alcanzar un objetivo en específico (Ávila, 2004). La NTC incluye a la política económica y el gasto público en ella, esto es importante ya que no en todas las teorías de crecimiento se piensa que las intervenciones de gobierno tienen un efecto en el crecimiento económico y debido a que este trabajo pretende realizar un portafolio de estrategias que impulse el crecimiento económico a través del sector cultural del país, las cuales podrían ser aplicadas mediante la política económica, esta teoría de crecimiento fue escogida como el fundamento teórico de la investigación.

La nueva teoría del crecimiento económico también hace inclusión a las externalidades como una característica de los procesos de producción que influye en el crecimiento económico y a su vez en el desarrollo económico; Duguine (2011) explica que las externalidades ocurren cuando la actividad de un agente económico afecta, ya sea positiva o negativamente, a otro agente externo y esta afectación no es incluida en los precios, alterando el equilibrio, evitando alcanzar un punto óptimo de oferta y demanda. Esto se relaciona con la cultura ya que Herrero (2011) menciona que uno de los efectos que tiene una mayor presencia de la cultura en las vidas de la población en el largo plazo, es una mejora en los niveles de educación y el fomento a los valores y a la creatividad, los cuales son ejemplos de externalidades positivas. Existen grandes academias y escuelas especializadas en todo tipo de artes y en México estas son notables, incluso escuelas no especializadas tanto públicas como privadas, ofrecen clases artísticas y culturales en sus planes de estudio o como actividades extracurriculares. Esto en el largo plazo sensibiliza y prepara a las generaciones siguientes, las cuales tendrán la posibilidad de aliviar muchos de los problemas estructurales del país.

Con México en una crisis económica es necesario plantear estrategias para salir de esta y alcanzar el crecimiento económico; los aportes de la NTC nos dicen que buscar áreas de especialización es una manera de alcanzarlo. Es lógico pensar que, en México, al tener una riqueza cultural tan grande, el sector cultural puede ser una fuente de crecimiento para el país y hacer política económica enfocada para fortalecer al sector cultural y convertirlo en parte vital de la economía del país podría ser un plan de acción pertinente. En primera instancia, debemos conocer detalladamente como se encuentran las cosas; analizar a fondo los resultados económicos que ha tenido el sector cultural en los últimos años, lo que ha hecho el gobierno en materia de política económica y gasto público para apoyar el sector cultural y cómo este apoyo afectó los resultados anteriores. Este análisis no solo servirá para empezar a formular diferentes estrategias con las cuales, mediante la política económica, podemos transformar al sector cultural en un pilar de la economía mexicana, sino que también permitirá a futuras investigaciones crear sus propias estrategias teniendo un antecedente; además también da herramientas e información substancial a aquellos que necesiten tomar decisiones en cuanto a cómo entrar al sector cultural o cómo tener un acercamiento a este desde un punto de vista político o económico.

Anteriormente, Piedras (2005) realizó una de las pocas investigaciones del sector cultural en México por lo cual es un importante antecedente a esta investigación. Aunque se investiga al sector cultura, esta investigación se enfoca en aquellas industrias protegidas por el derecho de autor, buscando su participación en la economía mexicana, en el comercio internacional y el empleo que genera, llegando a la conclusión que el sector cultural es una de las más grandes potencialidades para México a nivel mundial y que es importante que el gobierno impulse el sector.

Es importante señalar cómo internacionalmente el sector cultural ha tenido impacto en la economía de un país, esto se hace aparente al haber una cantidad considerable de investigaciones a nivel internacional que anteceden a esta; una de ellas fue en España, realizada por Herrero (2011), él realiza una investigación para encontrar la contribución de la cultura y las artes al desarrollo económico regional. En este revisa los impactos que tiene el sector a nivel regional tanto a corto y largo plazo. A diferencia de esta investigación, Herrero hace el análisis a nivel regional, en lugar de a nivel nacional, pero

se adentra a lo que es la política económica y las externalidades positivas del sector haciéndolo un antecedente importante de esta investigación.

Buscando investigaciones del sector cultural que se hayan realizado en un área geográfica cercana al país, con contextos políticos y económicos similares, se encontró que en Latinoamérica existen varios antecedentes a esta investigación; dos de ellas fueron hechas en Colombia. Navarro-Hoyos (2012) buscó la aportación de la cultura a la economía, pero en este caso enfocándose en un subsector: las artesanías. Es un antecedente importante para esta investigación debido a que, dentro del sector cultural de México, el INEGI (2020) nos dice que las artesanías son uno de los subsectores más grandes. La investigación concluye que las artesanías enfrentan grandes desafíos, pero a su vez tienen un gran potencial de comercio y exportación.

Finalmente, tenemos la segunda investigación realizada en Colombia, realizada por Ávila y Vargas (2018); es un análisis de las industrias culturales y creativas, en este caso de Bogotá. Siendo muy similar a esta investigación, realiza análisis de los diferentes indicadores económicos, cómo estos se relacionan al crecimiento y desarrollo económico, además de las políticas económicas que han sido aplicado en la región. Su investigación concluye con algunas recomendaciones para el gobierno de Bogotá para como incrementar el presupuesto destinado para el sector.

En esta investigación, dentro del ámbito internacional, se ha tenido un gran interés por el sector cultural de Corea del Sur, siendo un caso de éxito que podría ser replicado en México. En Corea del Sur se tenía una visión muy tradicional acerca del sector cultural, pero después de la crisis asiática de 1997 el gobierno empezó a notar el poder económico del sector cultural, viéndolo como una gran oportunidad para crecer económicamente y además compartir la cultura surcoreana con el mundo, por lo que se empezó a enfocar la inversión pública en el sector cultura en mayor proporción que antes y se transformó a la cultura en un producto de exportación para el mundo globalizado y no solo para consumo local (Fernández, 2014). Estudiar el caso del éxito del sector cultural en Corea del Sur es esencial para la creación de estrategias que se implementarían a través de la política económica para impulsar el sector cultural de México, por lo cual se analizan los datos del sector cultural de Corea del Sur y además se comparan con los de México. Por lo tanto, el objetivo de esta investigación es evaluar los indicadores económicos vinculados al sector cultural mediante análisis estadísticos para generar estrategias que impulsen el crecimiento económico en México.

Metodología

La metodología de esta investigación será dividida en tres etapas: construcción de una base de datos, análisis estadístico y descripción de resultados. Para la construcción de la base de datos, la principal fuente para la obtención de datos del sector cultural de México es la cuenta satélite de cultura del INEGI. Las cuentas satélite son establecidas por las Naciones Unidas para medir el tamaño de los sectores económicos que no son considerados industrias dentro de las cuentas nacionales, dentro de estos sectores se encuentra el sector cultural (Ochoa, 2013). El INEGI (2020), anualmente publica la cuenta satélite del sector cultural y su aportación a la economía de México, esta se realiza mediante recomendaciones hechas por las mismas Naciones Unidas, la OCDE, el Marco para Estadísticas de Cultura de la UNESCO, y la Guía Metodológica para la

implementación de las Cuentas Satélite de Cultura en Iberoamérica del Convenio Andrés Bello. De los datos de la Cuenta Satélite de Cultura de México se recolectaron datos de los años 2008 al 2019 de la oferta y utilización de bienes y servicios por áreas generales y específicas a precios constantes con 2013 como el año base, los puestos de trabajo ocupados totales por áreas generales y específicas. Para los datos de gasto público se recolectaron del Diario Oficial de la Federación de cada año desde 2013 hasta 2019, donde se publica se recapitula el Presupuesto de Egresos de la Federación.

Para la recolección de datos del sector cultural de Corea del Sur, los cuales son de difícil acceso, se utilizó el Servicio de Información Estadística de Corea (KOSIS), complementando con datos del *Economic Statistics System* (ECOS) del Banco de Corea y sitios de prensa que publicaron datos importantes de reportes económicos inaccesibles para el público general; se utilizaron específicamente datos de la encuesta de la industria de los servicios y los censos económicos del 2008 al 2019, como los datos de México; los siguientes apartados fueron los que se integraron a la base de datos, los cuales son similares o equivalentes a los que se encuentran en las áreas específicas de la Cuenta Satélite de Cultura: creación venta y renta de libros, música, instrumentos musicales, videojuegos, fotografía y otros medios audiovisuales; actividades de publicidad, de transmisión y de diseño especializado; actividades fotográficas; managers; escuelas de arte; y otros servicios relacionados a las artes, la creatividad y la recreación.

Finalmente, también se obtuvieron datos de producción y consumo de bienes culturales, de crecimiento económico, el índice de pobreza y la tasa de desempleo de distintos países de la OCDE, Asia y Latinoamérica.

Tabla 1. Tabla de variables para análisis de México.

Nombre de la variable	Tipo de variable	Clasificación de la variable	Escala
Año	Cuantitativa	Discreta	Nominal
Áreas generales	Cualitativa	Nominal	Nominal
Puestos de trabajo	Cuantitativa	Discreta	De razón
Importaciones	Cuantitativa	Continua	De razón
Demanda Intermedia	Cuantitativa	Continua	De razón
Consumo privado	Cuantitativa	Continua	De razón
Consumo de gobierno	Cuantitativa	Continua	De razón
Formación Bruta de Capital Fijo	Cuantitativa	Continua	De razón
Exportaciones	Cuantitativa	Continua	De razón
Gasto público	Cuantitativa	Continua	De razón

Tabla 2. Tabla de variables para comparación entre México y Corea del Sur.

Nombre de la variable	Tipo de variable	Clasificación de la variable	Escala
Año	Cuantitativa	Discreta	Nominal
País	Cualitativa	Nominal	Nominal
Producción	Cuantitativa	Continua	De razón
Aportación del sector cultural a la producción total	Cuantitativa	Continua	De razón
Puestos de trabajo	Cuantitativa	Discreta	De razón
Aportación del sector cultural a los puestos de trabajo totales del sector terciario	Cuantitativa	Continua	De razón
Consumo	Cuantitativa	Continua	De razón

Una vez completada la base de datos, se procede a la etapa de análisis de datos, iniciando con un análisis exploratorio el cual nos permitirá conocer los datos y observar, en una primera instancia, el comportamiento, los componentes y las características de los sectores culturales de México y Corea del Sur. Pero, antes de comenzar los análisis que nos permitirán conocer los datos más a fondo, es necesario cerciorarse que la comparativa entre los sectores culturales de México y Corea del Sur, está fundamentada y tiene sentido realizarla; para esto se realizó un análisis clúster con los datos económicos de los países de la OCDE, Asia y Latinoamérica mencionados en el párrafo anterior.

Siguiendo con el cumplimiento de los objetivos, se utilizarán otros métodos estadísticos: un modelo de regresión para conocer el impacto del gasto público, un análisis de multivariado para crear dos índices, uno para México y otro para Corea del Sur, en el cual se integren los componentes más importantes del sector cultural; con estos índices se hará la comparación de los sectores culturales de ambos países con una prueba de inferencia estadística, esta prueba de inferencia podría ser tanto paramétrica o no paramétrica, por lo cual también se harán las pruebas necesarias para escoger el tipo de prueba de inferencia correcta para los datos. Para la realización de todos estos análisis estadísticos se considera el uso de programas especializados en estadística como R, Minitab, SPSS, entre otros.

Resultados y discusión

El PIB Cultural de México ha crecido en la última década, esto se puede observar en la Figura 1; mientras que, como se ve en la Figura 2, el PIB Cultural de Corea del Sur ha sido bastante irregular en la última década sin mostrar una tendencia clara; cabe mencionar que el cálculo PIB Cultural de Corea del Sur, obtenido de la plataforma del Banco de Corea tiene una metodología diferente a la de la Cuenta Satélite de Cultura del INEGI, es por esto que en los análisis posteriores se utilizan la oferta y utilización de bienes y servicios culturales en lugar del PIB.

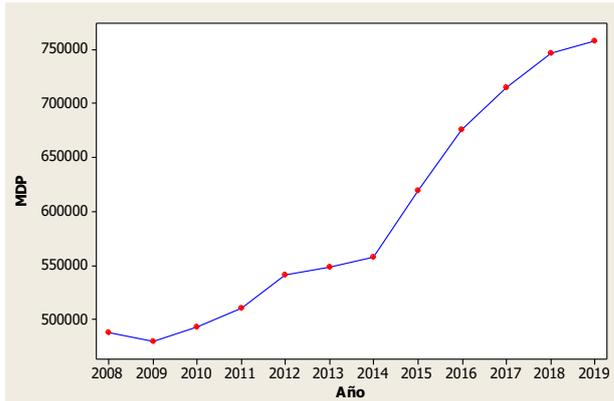


Figura 1. PIB Cultural de México (2008-2019).

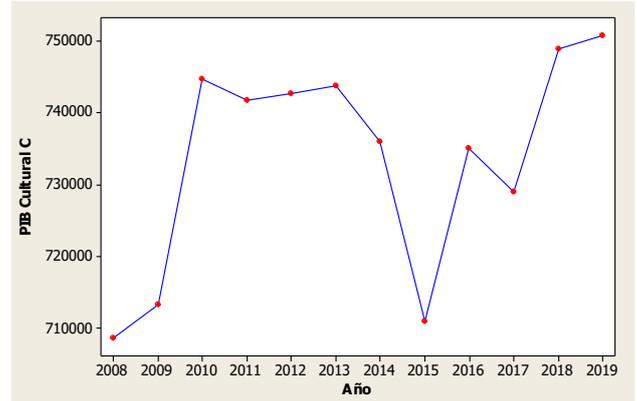


Figura 2. PIB Cultural de Corea del Sur (2008-2019).

La Figura 3 nos muestra en qué proporción el sector cultural de ambos países contribuyen al total de la producción de su país. México después de estar flotando alrededor del 0.035 muestra una tendencia positiva a partir del 2015; mientras que el sector cultural de Corea del Sur aporta cada vez menos al total de la producción con su clara tendencia a la baja.

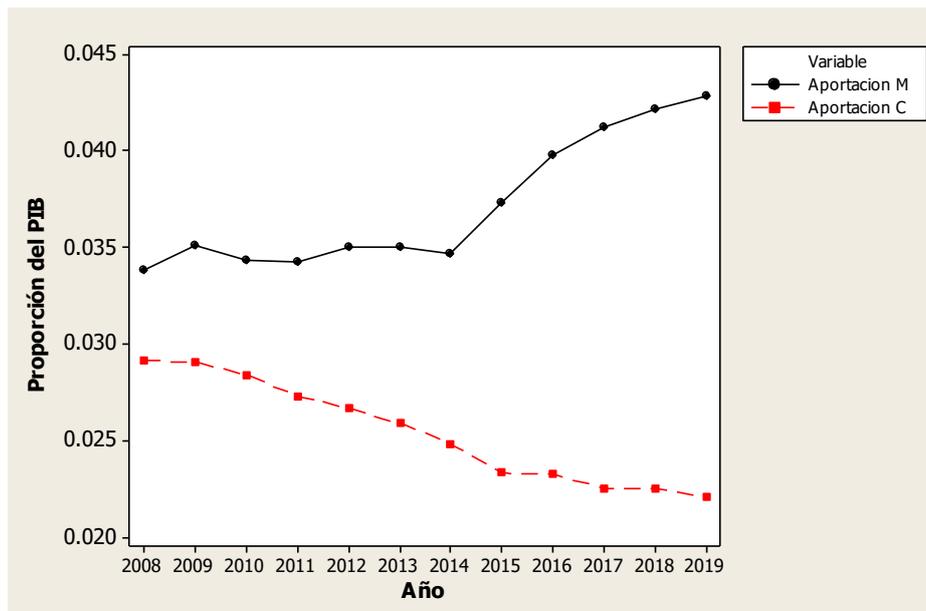


Figura 3. Aportaciones del sector cultural de México y Corea del Sur a sus respectivos "PIB" de 2008 a 2019.

Es importante conocer los resultados de las áreas específicas del sector cultural. Las 5 áreas generales que se muestran en la Figura 4 tienen los peores resultados, esta incluye a la música, las artes visuales, el patrimonio, los libros y las artes escénicas; los tres primeros se han mantenido con los mismos resultados durante una década, mientras que las artes escénicas tienen mejores resultados comparados con la década pasada y los libros peores resultados comparados con la década pasada.

La gráfica nos deja claro que los medios audiovisuales son el pilar del sector cultural de México y que ha crecido exponencialmente durante la década pasada.

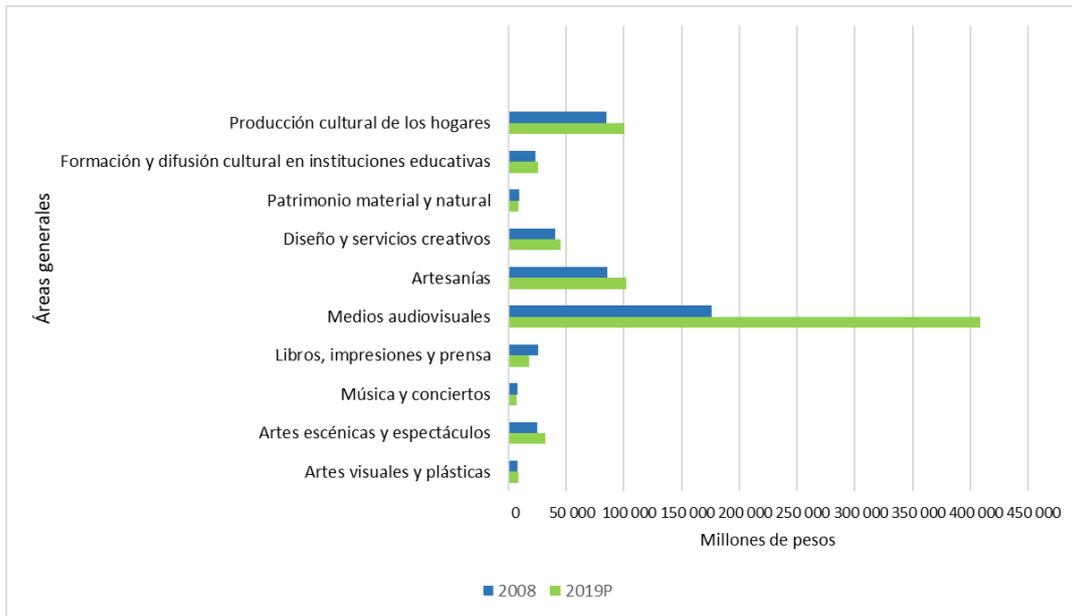


Figura 4. Evolución de áreas generales del sector cultural de 2008 a 2019.

En cuanto a la composición del consumo de bienes y servicios culturales en México, se puede ver en la Figura 5 que la gran mayoría del consumo es realizado por los hogares. La demanda intermedia también representa gran parte de este consumo cultural, la demanda intermedia es aquella en la cual se adquieren bienes o servicios que serán utilizados para la creación de otro bien (INEGI, 2020). Los otros 3 componentes: consumo de gobierno, formación bruta de capital y las exportaciones son parte muy pequeña del consumo cultural a pesar de que las exportaciones han sido vitales para el éxito de los sectores culturales de otros países.

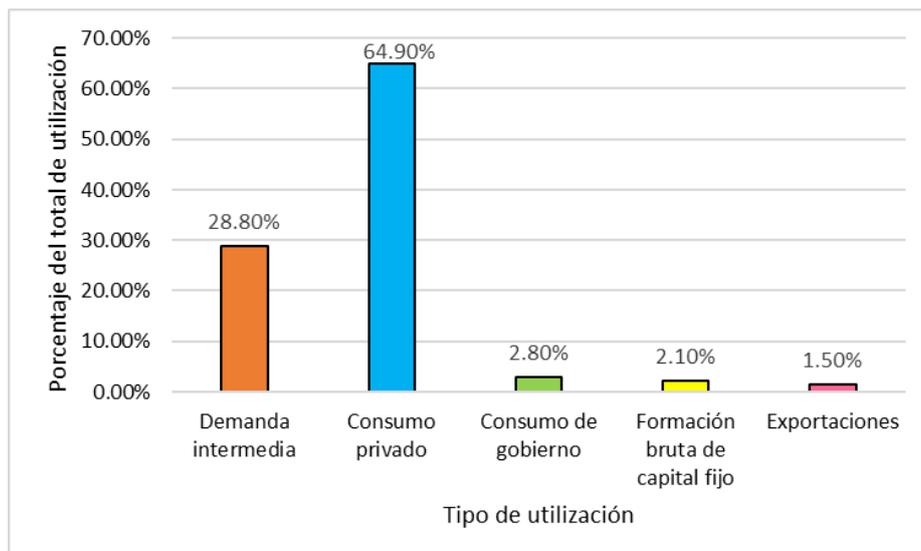


Figura 5. Gráfico de componentes del consumo de bienes y servicios culturales 2019.

Ahora es turno de analizar el gasto de gobierno, lo cual es de suma importancia ya que por este medio es donde las políticas económicas son ejecutadas. Se realizó un modelo de regresión lineal para buscar si existe una relación entre el gasto de gobierno destinado al sector cultural y el PIB cultural.

Tabla 3. Resultados del análisis de regresión PIB Cultura vs Gasto Público.

Predictor	Coefficiente	Estadístico de prueba T	P valor
Constante	1263074	5.94	0.002
Gasto Público	-32.23	-2.85	0.036

Nota. S= 58370.7 R cuadrada = 61.9% R cuadrada (ajustada) = 54.3%

$$PIB\ Cultura = 1263074 - 32.23(Gasto\ Público)$$

Con los resultados del modelo de regresión lineal expuestos en la tabla anterior, podemos observar que, efectivamente, el gasto público es una variable que explica al PIB cultural; esto se puede comprobar dado que con una confiabilidad del 95%, si el valor de P es menor a 0.05, la variable será significativa y, en este caso, el P valor del gasto público es de $0.036 < 0.05$. Con una R cuadrada ajustada del 54.3%, esta regresión es un modelo regular y además es válido ya que se cumplen los supuestos de los residuos, esto se observa a detalle en las Figuras 6 y 7. Para complementar la validez del modelo se hicieron dos pruebas: la de Shapiro-Wilk, que dio como resultado un estadístico de prueba $W = 0.3665$ y un P valor = 0.3665, lo que confirma que se cumple el supuesto de normalidad de los residuos; y la prueba Breusch-Pagan, la cual da como resultado un estadístico de prueba $BP = 1.297$ y un P valor = 0.2548, esto con un grado de libertad, confirmando el supuesto de homocedasticidad.

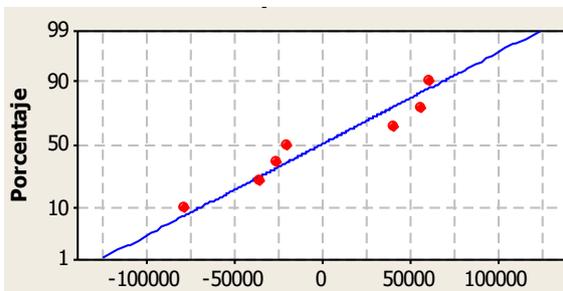


Figura 6. Gráfica de probabilidad normal de los residuos del modelo.

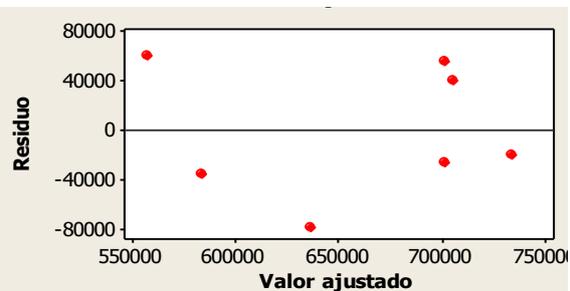


Figura 7. Gráfica de ajuste de los residuos.

Ahora nos queda claro que en México el gasto público ha sido parte clave en los resultados del sector cultural, sin embargo, el coeficiente nos da un resultado sorprendente; sí, el gasto público explica al PIB cultural, pero de forma negativa, esto significa que entre más gasto público se destine al sector cultural, menor será su PIB. Esto parece ser un caso extraño, por lo que es pertinente analizar como interactúa la gestión pública con el sector cultural, los resultados de estos se encuentran en la Figura 8.

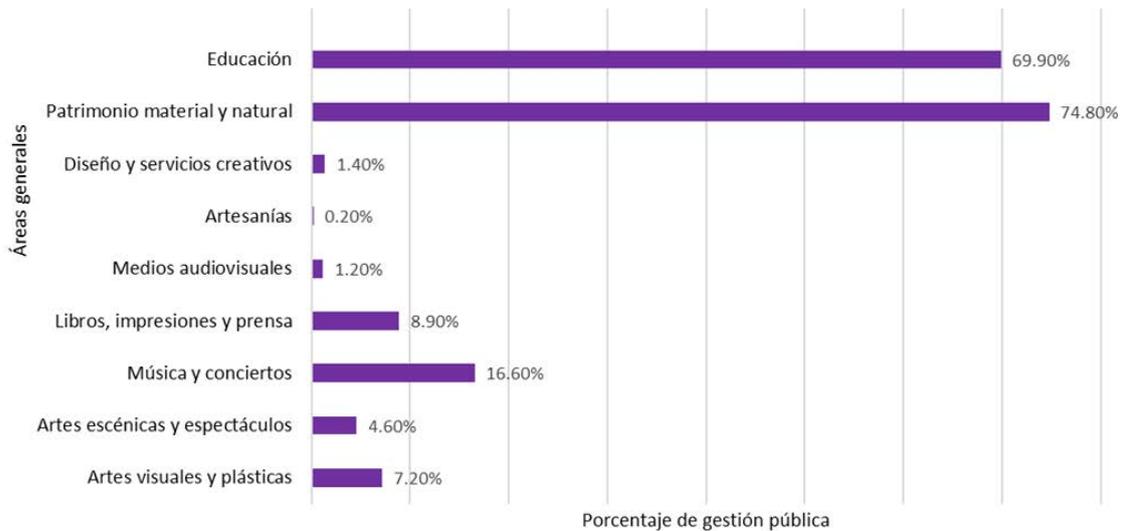


Figura 8. Gráfico de porcentaje de gestión pública en las áreas generales del sector cultural en México.

Análizar como se involucra la gestión pública en el sector cultural, ahora que sabemos que el gasto público afecta en forma negativa a la producción del sector, nos muestra la posibilidad de que el gasto público que se destina al sector cultural, no esté siendo distribuido a las áreas generales adecuadas. Recordando el análisis que se hizo a las áreas generales, los medios audiovisuales y las artesanías son las dos áreas generales con los resultados más altos pero, en cuanto a gestión pública, son las dos áreas generales a las que menos se destina recursos.

Antes de pasar a la comparación entre México y Corea del Sur, es necesario percatarse que esta comparación es válida, por lo cual se construyó un clúster de países de la OCDE, Asia y Latinoamérica, los cuales fueron agrupados de acuerdo a sus indicadores económicos generales como el PIB, la tasa de desempleo y el coeficiente de Gini e indicadores económicos del sector cultural. En la Figura 9, con un color rojo, se observa el primer cluster en el cual se encuentran ambos países: México y Corea del Sur con una similitud de más del 50.86%, lo cual sin duda hace válida la comparación entre estos países.

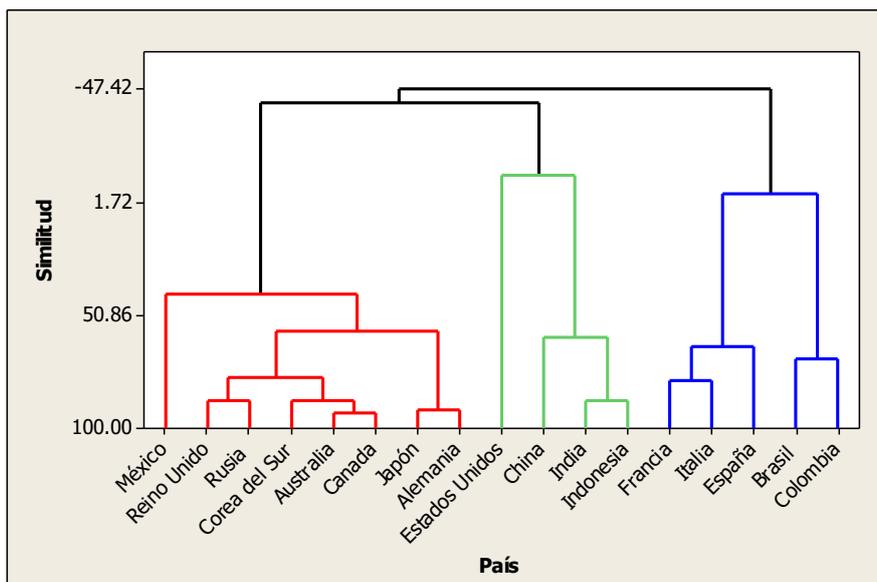


Figura 9. Clúster de países agrupados por indicador económicos generales y del sector cultural.

Comparando la utilización de bienes y servicios culturales de Corea del Sur y México, se puede ver una gran diferencia; no solo Corea del Sur tiene una utilización de bienes y servicios considerablemente mayor a la de México, si no que su crecimiento ha sido mayor, como se observa en la pendiente de la recta (Figura 10).

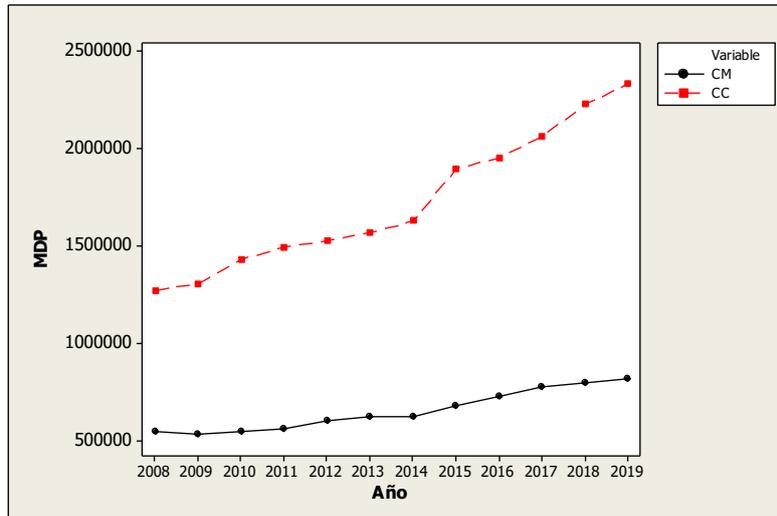


Figura 10. Series de tiempo comparativas de la oferta y utilización de bienes y servicios del sector cultural entre México y Corea del Sur del periodo 2008 a 2019.

Nota. CM es la oferta y utilización de bienes y servicios del sector cultural de México, CC es la oferta y utilización de bienes y servicios del sector cultural de Corea del Sur.

En cuanto al número de trabajadores dentro del sector cultural, también hay grandes diferencias. México tiene considerablemente más trabajadores dentro del sector cultural que Corea del Sur, aunque esto no resulta en una mayor oferta de bienes y servicios culturales. El número de trabajadores dentro del sector cultural de ambos países ha crecido a un ritmo similar en la última década (Figura 11).

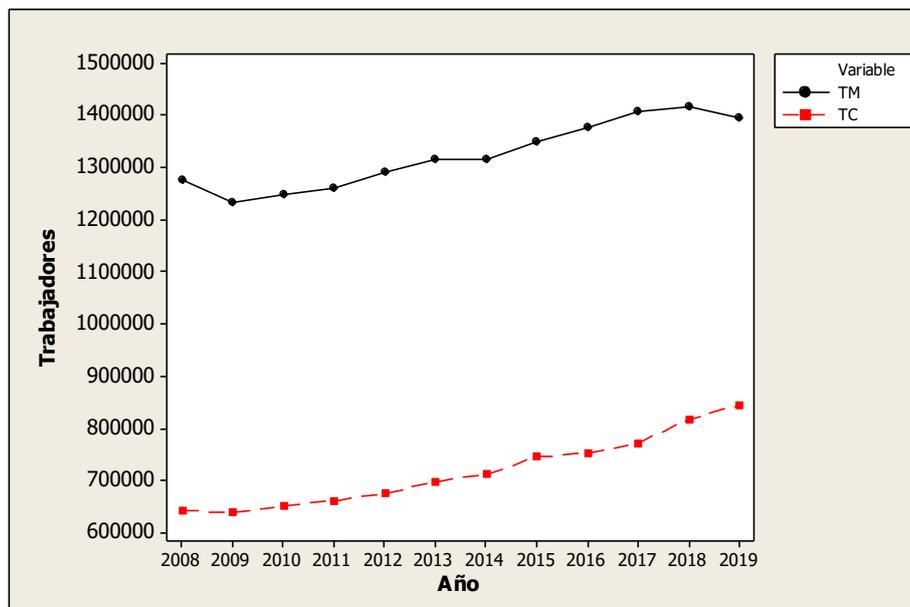


Figura 11. Series de tiempo comparativas del número de trabajadores dentro del sector cultural entre México y Corea del Sur.

Nota. TM es el número trabajadores dentro del sector cultural de México, TC es el número de trabajadores dentro del sector cultural de Corea del Sur.

Para facilitar la comparación del sector cultural de ambos países, se construyeron Índices, uno para cada país, utilizando datos de producción, consumo y empleo del sector cultural; estos índices fueron hechos mediante un análisis de componentes principales, donde solo se toma el primer componente como el índice. La Tabla 4 nos indica que el Índice Sector Cultural México representa 82.9% de los datos originales y la Tabla 5 nos indica que el Índice Sector Cultural Corea del Sur representa 77.8% de los datos originales.

Tabla 4. Resultados del análisis de componentes principales para la creación del Índice: “Sector Cultural México”, realizado con matriz de correlaciones.

Valor propio	4.1432	0.7701	0.0812	0.0047
Proporción	0.829	0.154	0.016	0.001
Acumulada	0.829	0.983	0.999	1.000

Variable	Componente 1
Producción	0.489
Aportación del sector cultural a la producción total	0.477
Puestos de trabajo	0.470
Aportación del sector cultural a los puestos de trabajo totales del sector terciario	-0.268
Consumo	0.489

Tabla 5. Resultados del análisis de componentes principales para la creación del Índice: “Sector Cultural Corea del Sur”, realizado con matriz de correlaciones.

Valor propio	3.8921	0.7992	0.2942	0.0085	0.0060
Proporción	0.778	0.160	0.059	0.002	0.001
Acumulada	0.778	0.938	0.997	0.999	1.000

Variable	Componente 1
Producción	0.264
Aportación del sector cultural a la producción total	-0.495
Puestos de trabajo	-0.487
Aportación del sector cultural a los puestos de trabajo totales del sector terciario	-0.455
Consumo	0.491

Una vez teniendo los índices, se comprobó que estos se distribuyeran normalmente, para así proceder a la comparación haciendo una prueba T. La prueba de normalidad de Ryan-Joiner (similar a Shapiro-Wilk) resultó con un P valor mayor a 0.100 en ambos índices. Con una confiabilidad del 95% podemos decir que ambos índices se distribuyen normalmente (Figura 12 y Figura 13).

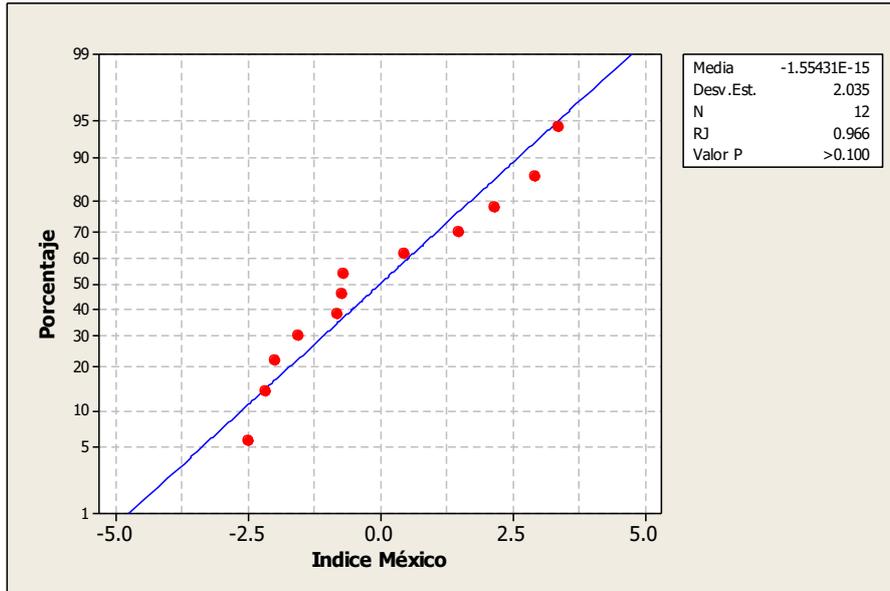


Figura 12. Gráfica de la probabilidad normal del Índice Sector Cultural México.

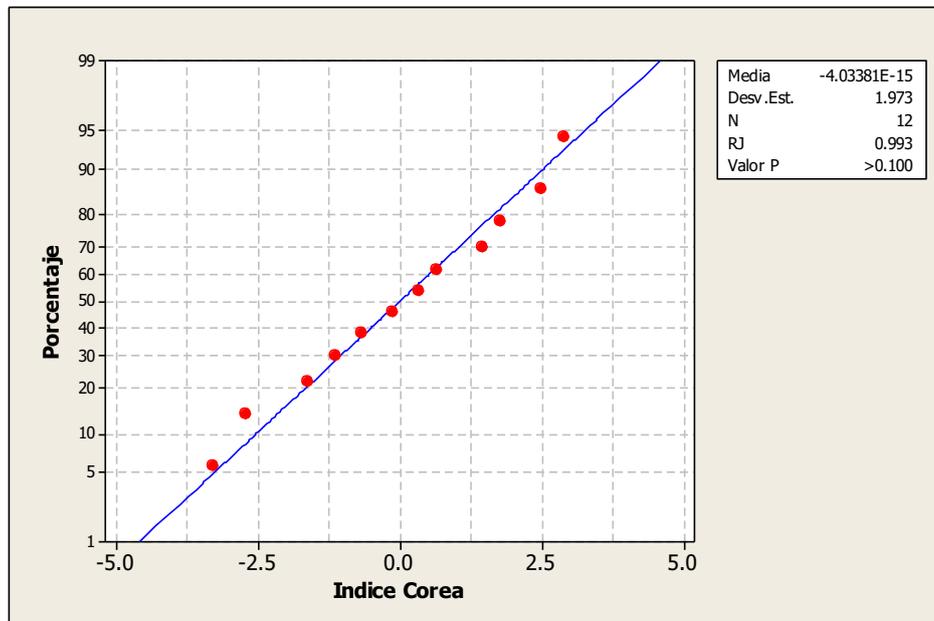


Figura 13. Gráfica de la probabilidad normal del Índice Sector Cultural Corea del Sur.

Finalmente, se realiza la prueba T para comparar los Índices Sector Cultural de ambos países.

$$H_0: \text{Índice México} = \text{Índice Corea del Sur}$$

$$H_1: \text{Índice México} \neq \text{Índice Corea del Sur}$$

Tabla 6. Resultados de la Prueba T e IC de los Índices.

	N	Media	Desviación estándar	Error estándar de la media
Índice México	12	-0.00	2.04	0.59
Índice Corea del Sur	12	-0.00	1.97	0.57

Nota. Diferencia = μ (Índice de México) – μ (Índice de Corea del Sur)

Estimado de la diferencia: 0.000

IC de 95% para la diferencia: (-1.702, 1.702)

Prueba T de diferencia = 0 (vs. no =): Valor T = 0.00 Valor P = 1.000 GL = 21

Con una confiabilidad del 95% y un P valor de 1.000, no podemos rechazar la hipótesis nula y existe suficiente evidencia para decir que el Índice Sector Cultural México y el Índice Sector Cultural Corea son estadísticamente iguales. Esto puede ser visto gráficamente en la figura 14.

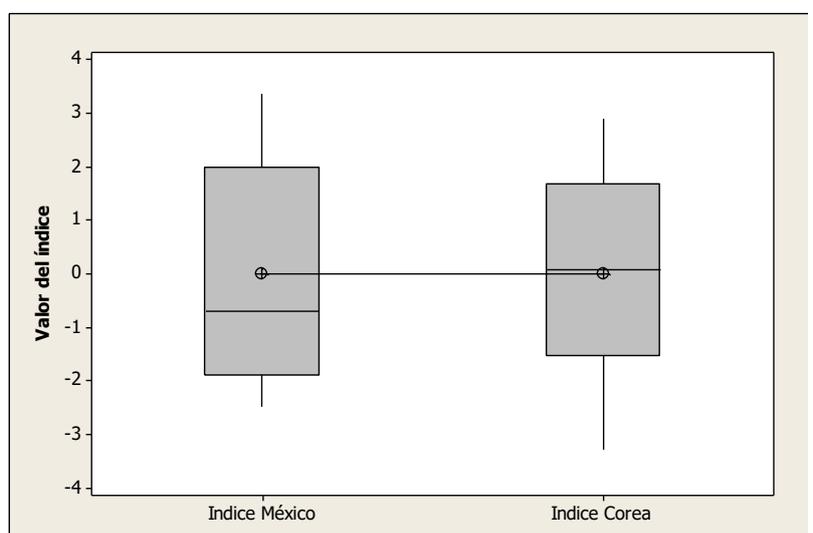


Figura 14. Gráfico de cajas de la comparación entre los Índices del Sector Cultural México y Corea del Sur.

Conclusiones

Con una producción, consumo y número de trabajadores creciendo rápidamente, el sector cultural se encuentra en una buena posición, pero una reacción negativa de la producción de bienes culturales al gasto público destinado al sector cultural confirma que el sector tiene un potencial muy grande dentro de México que no ha sido aprovechado en consecuencia al mal manejo de los recursos. El gobierno debe continuar destinando recursos a la educación cultural y artística, y al patrimonio material y natural, a pesar de ser de las áreas generales que no presentando grandes resultados, debido a que, en su gran mayoría, son de carácter público; además, considero que el gobierno debe aumentar los recursos destinados al sector cultural y que estos recursos se enfoquen a las áreas específicas de artesanías y a los medios audiovisuales, los cuales son las áreas generales que más aportan al sector cultural. Los medios audiovisuales crecieron exponencialmente en México en la última década, en 2019 se consumieron más del doble de bienes y servicios de medios audiovisuales que en 2008 y las artesanías, a pesar de no tener el gran crecimiento de los medios audiovisuales, han sido constantemente una gran parte del sector cultural mexicano y no solo eso, también es de lo más tradicional de la cultura

mexicana y retiene nuestra identidad como país a diferencia de otros subsectores culturales. Estas dos áreas generales, sin mucho apoyo del gobierno, han sobresalido dentro del sector cultural de México y con el apoyo de gobierno, muy posiblemente llegarían a ser parte esencial de la economía del país.

Corea del Sur ha demostrado que la cultura no es algo que se deba quedar en el ámbito local y que los bienes culturales también son un bien de exportación, que no solo trae beneficios económicos, sino que también comparte la cultura de su país con el mundo. México y Corea del Sur pueden parecer países sumamente diferentes, pero con esta investigación vemos que, al menos su sector cultural es muy similar, lo cual nos permite considerar estrategias de política económica dirigidas al sector cultural, que se han aplicado en Corea del Sur y que han sido exitosas, para ser utilizadas aquí en México, como lo es el enfoque a las exportaciones de bienes culturales.

Referencias

- Austin, T. (2000) *Para comprender el concepto de cultura*. UNAP Educación y Desarrollo.
- Ávila, J. (2004). *Introducción a la Economía*. Plaza y Valdés
- Ávila, L. y Vargas, M. (2018). *Análisis de las industrias culturales y creativas en el crecimiento económico de la ciudad de Bogotá durante los años 2010-2016*. Trabajo de Grado. Universidad Católica de Colombia. Facultad de Ciencias Económicas y Administrativas. Programa de Economía. Bogotá, Colombia.
- Buenrostro, M. (2013). *Presentación de Libro "Virgen de la Caridad, Arte, Seda y Oro"*. Entrevista. Consultado en la página web (01 06 2020). <https://www.alcalorpolitico.com/informacion/tradiciones-de-mexico-se-pierden-porque-la-gente-prefiere-las-de-otros-paises-marco-buenrostro-116230>
- Banco de México (2020) *Banxico Educa, Crecimiento PIB*. Consultado en la página web (05 12 2020). <http://educa.banxico.org.mx/economia/crecimiento-pib.html>
- Bank Of Korea (2021) *Economic Statistics System*. Consultado en la página web (22 01 21) http://ecos.bok.or.kr/flex/EasySearch_e.jsp
- Consejo Nacional de Evaluación de la Política de Desarrollo Social CONEVAL (2018) *Medición de la Pobreza*. Consultado en la página web (30 05 2020) <https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza-2018.aspx>
- De Gregorio, J. (2007). *Macroeconomía. Teoría y Políticas*. Pearson-Education.
- Duguine, H. (2011). *Economía del Turismo: externalidades en el sector turístico*. Notas en Turismo y Economía. Año II. Nro. III. 2011. ISSN 1853-1504
- Enriquez, I. (2016). *Las teorías del crecimiento económico: notas críticas para incursionar en un debate inconcluso*. rldonline n.25 pp. 73-125.
- Fernández, P. (2014). Corea del Sur: sinergia de industrias culturales y turismo. *Revista Geográfica del Sur*, 4(5), 95-111.
- Florida, R. (2009). *Las ciudades creativas*. Barcelona: Ediciones Paidós Ibérica.
- Giménez Montiel, G. (2005). *Teoría y análisis de la cultura*. Colección Intersecciones. CONACULTA
- Herrero Prieto, L. C. (2011). La contribución de la cultura y las artes al desarrollo económico regional. *Investigaciones Regionales*, 19, 177-202.
- INEGI (2017). *Medición de la Economía Informal*. Consultado en la página web (24 05 2020). <https://www.inegi.org.mx/app/indicadores/?ind=6207067228&tm=6>

- INEGI (2018) *Cuenta Satélite: Cultura*. Consultado en la página web (24 05 2020).
<https://www.inegi.org.mx/temas/cultura/>
- Jiménez, F. (2011). *Crecimiento Económico: Enfoques y Modelos*. Libros PUCP / PUCP Books, Fondo Editorial - Pontificia Universidad Católica del Perú, edition 1.
- Lopez, E. (1993). El patrimonio cultural de México. Fondo de Cultura Económica.
- Navarro-Hoyos, S. (2012). *La artesanía como industria cultural: desafíos y oportunidades*. Pontificia Universidad Javeriana.
- Ocuegueda, J., Castillo, R. y Varela, R. (2009). Crecimiento regional en México: especialización y sectores clave. *Prob. Des*, 40(159), 61-84.
- Ochoa, A. (2013). *Economía del turismo*. Fundación Universitaria Camaf
- Piedras, E. (2005). *¿Cuánto vale la cultura? Contribución económica de las industrias protegidas por el derecho de autor en México*. Sociedad de Autores y Compositores de Música, Sociedad General de Escritores de México, Consejo Nacional para la Cultura y las Artes, Cámara Nacional de la Industria Editorial Mexicana
- Seaman, B. (2004). El impacto económico de la cultura. Manual de economía de la cultura, *Fundación Autor, Madrid*, 224-231.
- Secretaria de Cultura (2010). *Encuesta Nacional de Hábitos, Prácticas y Consumos Culturales*. Consultado en la página web (27 05 20)
https://www.cultura.gob.mx/encuesta_nacional/#.XtmicDpKhPY
- Secretaria de Hacienda y Crédito Público (2020). *Presupuesto de Egresos de la Federación*. Consultado en la página web (10 11 20)
<http://www5.diputados.gob.mx/index.php/camara/Programacion-Presupuesto-y-Contabilidad/Presupuesto-asignado>
- Statistics Korea (2021) *Service Industry Survey*. Consultado en la página web (22 01 21)
https://kosis.kr/eng/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ETITLE&parmTabId=M_01_01
- Thorsby, D. (2001). *Economía y cultura*. Ediciones AKAL, Sydney.
- Valenzuela, J. (2008). El crecimiento económico: concepto, determinantes inmediatos y evidencia empírica. Aportes, *Revista de la Facultad de Economía*, BUAP, Año XIII, Números 38-39.
- Yudice, G. (2002). *El recurso de la cultura*. University of Miami
- Zermeño, F. (2004). *Lecciones de desarrollo económico*. Plaza y Valdés.

Análisis espacial de la pobreza en el municipio de Xalapa-Enríquez de 2010 - 2020

Irma Janeth Prado Méndez¹, Julia Aurora Montano Rivas²

¹ Facultad de Estadística e Informática, Universidad Veracruzana. pirmajaneth@gmail.com

² Facultad de Estadística e Informática, Universidad Veracruzana. julmontano@uv.mx

Resumen

La pobreza en Xalapa es un problema social muy extenso, pero a medida que se tiene un mayor nivel de desglose geográfico, se obtiene información más concreta para tomar decisiones de intervención y lograr un mejor desarrollo social. Con respecto a lo anterior, se expone una propuesta de construcción del índice de desarrollo social a nivel colonia para Xalapa con información de los Censos de Población y Vivienda de los años 2010 y 2020, de los principales resultados por manzana, se utilizan las variables que dan evidencia de las carencias, dicha propuesta está basada en el Método de Medición de la Pobreza (MMIP) y de la propuesta del Consejo de Evaluación del Desarrollo Social de la Ciudad de México. Por consiguiente, la propuesta del índice se realiza mediante la utilización de las técnicas multivariantes, tales como: componentes principales y componentes principales con Kernels, se explora la utilidad de los métodos espaciales en el estudio de la pobreza a través del desarrollo social; permitiendo un análisis más completo de la dimensión espacial de los fenómenos sociales. La importancia de los resultados radica en que permiten entender el grado de concentración espacial de la pobreza, al identificar específicamente las zonas geográficas del municipio de Xalapa donde el desarrollo social es alto, los factores explicativos de la pobreza que pueden tener efectos diferenciados y comparar los resultados generados a partir del índice y análisis espaciales correspondientes a los años 2010 y 2020.

Palabras clave: Desarrollo social, Índice, Análisis espacial.

Introducción

Una persona se encuentra en situación de pobreza cuando tiene al menos una de las seis carencias sociales (rezago educativo, acceso a servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda y acceso a la alimentación) y su ingreso está por debajo de la línea de pobreza por ingresos (CONEVAL, 2020).

De acuerdo con los informes de CONEVAL mediante la pobreza se mide el desarrollo que han tenido los municipios y entidades subdividiendo así la pobreza en moderada, extrema, vulnerable por carencia social y vulnerable por ingresos de donde se enmarco lo siguiente:

Se considera en pobreza a la población con al menos una carencia social e ingreso inferior al valor de la línea de pobreza por ingresos, antes línea de bienestar y que se refiere al valor total de la canasta alimentaria y de la canasta no alimentaria por persona al mes.

Se considera en pobreza moderada a la población que, a pesar de estar en situación de pobreza, no cae en la categoría de pobreza extrema. Se considera en pobreza extrema a la población con tres o más carencias sociales e ingreso inferior al valor de la línea de pobreza extrema por ingresos, antes línea de bienestar mínimo que se refiere al valor de la canasta alimentaria por persona al mes. Se considera en vulnerabilidad por carencias sociales a la población con un ingreso superior a la línea de pobreza por ingresos, antes línea de bienestar, pero con una o más carencias sociales. Se considera en vulnerabilidad por ingresos a la población que no padece ninguna carencia social, pero su ingreso es igual o inferior a la línea de pobreza por ingresos, antes línea de bienestar (CONEVAL, 2020).

El desarrollo social es un proceso que implica un mejoramiento de bienestar y condiciones de vida en la población encaminado por el desarrollo económico. Sin duda es un tema extenso e incluye una gran gama de factores que afectan directa o indirectamente el desarrollo social. De acuerdo con James Midgley (1995) el desarrollo social es “un proceso de promoción del bienestar de las personas en conjunción con un proceso dinámico de desarrollo económico”.

Es eminente que en países potencias el desarrollo social es alto, sin embargo, para los demás el desarrollo social es menor; dado que las condiciones económicas son carentes. Al estudiar detalladamente este tema se puede dar cuenta que no solo existen diferencias entre países, si no también internamente las condiciones económicas y sociales son diferentes.

Hace algunos años se han venido desarrollando metodologías de análisis espacial, las cuales permiten analizar los agentes de estudio con la ubicación geográfica aportando así un mejor análisis de la situación de dependencia espacial, heterogeneidad y muestra una visión clara del objeto de estudio en cada espacio específico.

Un claro ejemplo de estos análisis con métodos de estadística espacial es el estudio realizado por (Pérez, 2005) quien con base en los indicadores de necesidades básicas insatisfechas (NBI) y de calidad de vida (ICV), realiza un análisis espacial de la pobreza en Colombia para establecer si existe evidencia de dependencia espacial con las medidas de pobreza a nivel departamentos y municipios; así como explicar la pobreza en Colombia destacando las características geográficas. Además de mostrar que estos estudios son de gran utilidad para identificar áreas de atención para el abatimiento de pobreza y permiten dar claridad de situaciones específicas para planear y ejecutar políticas públicas específicas para cada espacio geográfico.

Otro claro ejemplo de la utilización de estos métodos y técnicas de análisis en estudios de situaciones sociales es el mostrado en el artículo Alcances y límites de los métodos de análisis espacial para el estudio de la pobreza urbana elaborado por Landy L. Sánchez-Peña (2012) quien pone en claro la poca atención que se le brinda a la dimensión espacial en fenómenos sociales, además de que muestra una revisión de la dependencia y la heterogeneidad espacial mediante la exploración de tres aproximaciones: análisis exploratorio de datos espaciales, modelos de regresión espacial autorregresivos y regresión geográficamente ponderada, haciendo uso de datos del instituto Nacional de Estadística y Geografía (INEGI) a nivel de Área Geográfica Básica (AGEB).

En el transcurso de los años el territorio que contempla el municipio de Xalapa ha comenzado a crecer y a generar un desarrollo, pero una parte de su población se rezaga. Según Guerra “puede tener un origen multifactorial que abarque aspectos como una deficiente redistribución de los ingresos, la concentración de poder en unos cuantos, la falta de políticas públicas incluyentes, corrupción, etc.” (Osorno, 2012). Por lo que es importante reconocer el problema cuando éste se manifieste y sobre todo identificar las áreas con menor desarrollo social, que es donde se concentran las necesidades básicas insatisfechas, ya sea por sus características sociodemográficas, su localización, o bien por ambas. Hacer esto resulta fundamental para evitar que esto mismo se convierta, en un freno para el crecimiento y el desarrollo del municipio.

Por lo anterior, surge el interés de presentar una propuesta de medición de la pobreza mediante un índice de Desarrollo Social aprovechando el potencial de la información de los censos como lo es el Censo de Población y Vivienda de donde se extraen variables que dan evidencia de las dimensiones de: adecuación energética, bienes durables, adecuación sanitaria, de conectividad, seguridad social, servicio médico, calidad y espacio en la vivienda, y privación educativa del hogar. Para hacer la obtención del índice se propone utilizar las técnicas multivariantes de componentes principales y componentes principales con Kernels para calcular y conocer la variabilidad explicada por los índices e identificar las colonias y sus niveles de pobreza.

Las herramientas y métodos de análisis espacial permiten evaluar la existencia de relaciones socioeconómicas entre los agentes y, la relación de estos con su ubicación geográfica.

Los estudios poblacionales han puesto poca atención a la dimensión espacial de los fenómenos sociales, por lo que el análisis espacial permite realizar aproximaciones para dar cuenta del comportamiento de la dependencia y la heterogeneidad del rezago o desarrollo social; por consiguiente en este estudio se realiza una descripción gráfica y estadística con el fin de verificar la existencia de patrones en los datos; se explica el comportamiento espacial en el indicador de desarrollo social, este permite entender el grado de concentración espacial de la pobreza, identificar precisamente las zonas geográficas del municipio de Xalapa en donde el desarrollo social es alto, los posibles efectos de propagación y cómo los factores explicativos de la pobreza mediante el desarrollo social pueden tener efectos diferenciados.

La importancia de la información y los resultados que se generen radica en la utilización de la evidencia presentada para el diseño de políticas públicas, así como programas dirigidos de desarrollo social y abatimiento de la pobreza.

Metodología

Alcance de la investigación

El trabajo tiene como área de estudio el municipio de Xalapa que se encuentra en la entidad de Veracruz, es de interés el análisis en esta área dado que este municipio es muy diverso en cuestión de pobreza lo que para el estudio además de que no cuenta con estudios a niveles de desgloses específicos para precisar áreas de atención para el abatimiento de la pobreza ; colinda al norte con los municipios de Banderilla, Jilotepec y Naolinco; al este con los municipios de Naolinco y Emiliano Zapata; al sur con los

municipios de Emiliano Zapata y Coatepec; al oeste con los municipios de Coatepec, Tlalnahuayocan y Banderilla.



Figura 1. Área de estudio. Fuente: elaboración propia.

El estudio es descriptivo de tipo retrospectivo y longitudinal, ya que se analizó el comportamiento de una serie de variables para construir un índice en el año 2010 y en el año 2020, para describir la situación de cambio en un periodo definido.

Consideraciones generales

Los datos fueron recuperados de la información del Censo de Población y Vivienda de 2010 y 2020 que realiza el Instituto Nacional de Estadística y Geografía (INEGI). En un principio se pretendía trabajar con los microdatos a nivel vivienda, pero debido a la duración del proceso de entrega de los datos por el INEGI, se utilizaron la información de los principales resultados por manzana del censo, también se tenía la idea de trabajar con los datos a nivel manzana y AGEB, sin embargo, para efecto de una mejor aportación y facilitación de entendimiento de los resultados, los datos se utilizaron a nivel manzana y se agruparon por colonia. Es importante mencionar que el nivel de desglose se considerara como delimitación espacial para el estudio; ya que este espacio está constituido por información de viviendas, hogares y población.

La información con la que se trabaja “Principales resultados por AGEB y manzana ” del Censo de Población y Vivienda 2010 y 2020 que publica INEGI, no incluye las manzanas rurales, aunque son en éstas en donde podría presentarse un menor de desarrollo social, esto no significa que no valga la pena analizar estas manzanas, pero en este caso para aprovechar la información que brinda INEGI se trabajó solo con las manzanas urbanas, en futuros estudios será importante encontrar forma de incluir aquellas que se localizan en zonas rurales.

Los datos se trabajaron para el año 2010 y 2020 dado que el censo se realiza cada diez años y es de interés este periodo debido a que los Objetivos del Desarrollo Sostenible (ODS) fueron propuestos en el 2015 por lo que se vuelve interesante ver el cambio mediante los primeros cinco años después de la implementación de las estrategias de desarrollo social.

De un total de 5,842 manzanas urbanas en 2010, sólo fue posible analizar 5,106 debido a que el resto presentaba datos faltantes o la manzana no contaba con información; en 2020 de 6732 manzanas solo se trabajó solo con 6055 manzanas. Posteriormente debido a la complejidad que se presentó al realizar el análisis por manzana para componentes principales con Kernel se optó por trabajar los datos para componentes principales y componentes principales con Kernels los datos a nivel manzana de donde para 2010 se trabajó con 377 colonia y para 2020 390 colonias.

Diseño estadístico

Variables o categorías de análisis

Se utilizaron las variables que dan evidencia de la adecuación energética, bienes durables, adecuación sanitaria, de conectividad, seguridad social, servicio médico, calidad y espacio en la vivienda, privación educativa del hogar. Para construir las siguientes dimensiones indicadoras que dan pie al índice de desarrollo social.

- Calidad y espacio disponible en la vivienda
- Acceso a electricidad
- Bienes durables
- Adecuación sanitaria
- Acceso a seguridad social y servicio medico
- Desarrollo educativo

Debido a la falta de variables para medir el ingreso económico solo se considera la variable de población ocupada como variable que evidencia un ingreso.

Tabla 1. Descripción de variables y dimensiones.

Dimensión	Nombre de la variable	Descripción
Calidad y espacio disponible en la vivienda	VPH_PISODT	Viviendas particulares habitadas con piso de material diferente de tierra
	VPH_3YMASC	Viviendas particulares habitadas con 3 cuartos y más
	VPH_2YMASD	Viviendas particulares habitadas con dos dormitorios y más
	PRO_OCUP_C	Promedio de ocupantes por cuarto en viviendas particulares habitadas
Acceso a electricidad	VPH_C_ELEC	Viviendas particulares habitadas que disponen de luz eléctrica
Bienes durables	VPH_TV	Viviendas particulares habitadas que disponen de televisor
	VPH_REFRI	Viviendas particulares habitadas que disponen de refrigerador
	VPH_LAVAD	Viviendas particulares habitadas que disponen de lavadora
	VPH_PC	Viviendas particulares habitadas que disponen de computadora
Adecuación sanitaria	VPH_AGUADV	Viviendas particulares habitadas que disponen de agua entubada en el ámbito de la vivienda
	VPH_EXCSA	Viviendas particulares habitadas que disponen de excusado o sanitario
	VPH_DRENAJ	Viviendas particulares habitadas que disponen de drenaje
Acceso a seguridad social y servicio médico	PDER_SS	Población derechohabiente a servicios de salud
Desarrollo educativo	GRAPROES	Grado promedio de escolaridad
Ingreso	POCUPADA	Personas de 12 a 130 años de edad que trabajaron o que no trabajaron, pero si tenían trabajo en la semana de referencia

Análisis de los datos

Construcción del indicador

Para obtener los resultados del índice de Desarrollo Social se obtuvieron las variables de acuerdo con la metodología utilizada por Consejo de Evaluación del Desarrollo Social de la Ciudad de México, ya que se basa en el Método de Medición Integrada de la Pobreza. Dicha información fue procesada y analizada mediante componentes principales para obtener un índice, pero primero se estudió el comportamiento por cada dimensión.

Posteriormente para lograr una mayor explicación de la variabilidad de la información mediante el índice se propuso obtener el indicador mediante la aplicación de componentes principales con Kernels basándose en la propuesta Análisis de Componentes Principales con Kernels: Una Propuesta de Mejora del Kernel que consiste en “determinar un kernel óptimo dentro de un conjunto de kernels, donde el criterio de optimalidad es la proporción de varianza explicada proporcionada por el kernel”. (Montano & Juárez, 2013)

Análisis espacial

El análisis se realiza con ayuda de técnicas de análisis espacial, las cuales permiten establecer si el desarrollo social en las manzanas se distribuye en forma aleatoria o si, existe evidencia de algún tipo de dependencia espacial entre los datos. Se realizó una descripción gráfica y estadística con el fin de establecer la existencia de patrones de comportamiento en los datos utilizados.

Mediante esta etapa fue posible establecer la existencia de características espaciales en el indicador de desarrollo social, como valores espaciales atípicos o de la existencia de clusters. Posteriormente se realizó un análisis local a través de indicadores de asociación espacial local (LISA). Específicamente con el índice Moran

Por consiguiente, se exploró la utilidad de los métodos espaciales en el estudio de la pobreza a través del desarrollo social; permitiendo un análisis más completo de la dimensión espacial de los fenómenos sociales.

Resultados y discusión

Descripción de la situación por dimensión

En 2010, el estado de Veracruz presento el 57.6% de pobreza y en 2015 presento el 57.3%, para el caso del municipio de Xalapa en 2010 el porcentaje de pobreza fue de 34.2, pero para el 2015 aumento 3.4 %; según datos del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, 2015). Este caso indica que de manera general podemos ver reducción de pobreza en el estado de Veracruz, pero cuando empezamos a ver la situación con mayor nivel de desglose geográfico se empieza a evidenciar las discrepancias territoriales en cuestión de pobreza.

La pobreza en la ciudad de Xalapa es un fenómeno social que se ve afectado por diversos factores en este estudio en particular se realizó la propuesta mediante la medición de 7 dimensiones que dan evidencia de las necesidades básicas insatisfechas. Como primera fase se realizó un análisis descriptivo para conocer de manera específica la diferencia por variables para el año 2010 y 2020.

Tabla 2. Descripción calidad y espacio disponible de la vivienda.

Calidad y espacio disponible en la vivienda				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de viviendas particulares habitadas con piso de material diferente de tierra	76.63	16.40	97.51	9.55
Porcentaje de viviendas particulares habitadas con 3 cuartos y más	60.38	21.74	78.21	23.20
Porcentaje de viviendas particulares habitadas con dos dormitorios y más	51.30	21.60	66.80	23.95
Promedio de ocupantes por cuarto en viviendas particulares habitadas	0.96	0.40	0.85	0.31

Para la dimensión de calidad y espacio disponible en la vivienda el cambio más notorio fue en el porcentaje de viviendas particulares habitadas con piso de material diferente de tierra que en el 2020 tuvo un incremento, también en promedio la cantidad de ocupantes por cuarto en las viviendas disminuyó.

Tabla 3. Descripción de acceso a electricidad.

Acceso a electricidad				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de viviendas particulares habitadas que disponen de luz eléctrica	78.59	15.69	99.33	4.13

Para el indicador que da evidencia de la dimensión de acceso a electricidad tuvo un incremento de 20.74 puntos porcentuales.

Tabla 4. Descripción de bienes durables.

Bienes durables				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de viviendas particulares habitadas que disponen de televisor	76.78	15.86	92.87	10.92
Porcentaje de viviendas particulares habitadas que disponen de refrigerador	68.73	18.69	89.28	15.73
Porcentaje de viviendas particulares habitadas que disponen de lavadora	52.61	21.86	72.90	22.51
Porcentaje de viviendas particulares habitadas que disponen de computadora	33.48	24.47	49.11	29.28

En la dimensión de bienes durables todas variables tuvieron un incremento, sin embargo, el porcentaje de viviendas particulares habitadas que disponen de computadora fue en el que las manzanas en promedio solo incrementaron 15.63

Tabla 5. Descripción de adecuación sanitaria.

Adecuación sanitaria				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de viviendas particulares habitadas que disponen de agua entubada en el ámbito de la vivienda	76.30	18.93	98.82	6.61
Porcentaje de viviendas particulares habitadas que disponen de excusado o sanitario	78.55	15.66	98.83	5.65
Porcentaje de viviendas particulares habitadas que disponen de drenaje	78.14	15.97	99.31	3.86

De manera descriptiva podemos notar que los promedios por manzana incrementaron en 2020 para esta dimensión.

Tabla 6. Descripción de acceso a seguridad y servicio médico.

Acceso a seguridad social y servicio medico				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de población a derechohabiente a servicios de salud	58.60	17.32	72.27	14.36

En la tabla 6 se pudo observar un incremento el promedio por manzana en el porcentaje de población derechohabiente a servicios de salud.

Tabla 7. Descripción de desarrollo educativo.

Desarrollo educativo				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Grado promedio de escolaridad	9.09	4.16	11.34	2.44

En la tabla 7 podemos notar que el grado promedio de escolaridad en promedio por manzana tuvo un aumento en el año 2020 con respecto al 2010.

Tabla 8. Descripción de ingreso.

Ingreso				
Variable	2010		2020	
	Promedio por manzana	Desviación estándar	Promedio por manzana	Desviación estándar
Porcentaje de personas de 12 a 130 años que trabajaron o que no trabajaron, pero si tenían trabajo en la semana de referencia	42	9.43	51.11	9.38

La variable que se contempló para evidenciar la dimensión de ingreso también mostró un cambio en el promedio por manzana en el 2020.

Análisis de componentes principales (ACP)

Como segunda fase de esta propuesta de medición de la pobreza se propuso obtener el índice de desarrollo social mediante la técnica de componentes principales de donde se obtuvieron los siguientes resultados:

Para el año 2010 se observó que con dos componentes estaríamos explicando el 70.19 por ciento de la variabilidad de la información que es lo recomendable pero como para obtener el índice de desarrollo social solo interesa el primer componente, aunque apenas y explico el 54.15 % de la variabilidad de los datos; para el año 2020 se analizó que se necesitaban 6 componentes para lograr explicar el 74 % de la variabilidad de los datos y como el objetivo era construir el índice con el primer componente pero solo alcanzo a explicar el 37 % de la variabilidad lo que no es muy bueno dado que se estaría perdiendo mucha información y el índice no sería muy confiable.

Dada las limitantes mencionadas, la dimensionalidad de la información con la que se trabaja a nivel manzana se optó realizar el análisis a nivel colonia lo cual también facilitaría una mayor comprensión de los lectores y ubicación del comportamiento del objeto de estudio.

Análisis componentes principales a nivel colonia

- Análisis del año 2010

Tabla 9. Importancia de los componentes.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Desviación estándar	3.016	1.718	0.973	0.907	0.639	0.523	0.374	0.351
Proporción de varianza	0.607	0.197	0.063	0.055	0.027	0.018	0.009	0.008
Proporción acumulada	0.607	0.804	0.867	0.922	0.949	0.967	0.976	0.984

Por medio del análisis para las colonias del 2010 se mostró que con dos componentes se estaría explicando más del 80 % de variabilidad de los datos, lo cual es aceptable, sin embargo, el fin es obtener el índice con el primer componente que solo explico el 60.7% de variabilidad de los datos.

Tabla 10. Importancia de las variables dentro del primer componente.

Variable	Componente 1
PVPH_PISODT	0.305
PVPH_3YMASC	0.290
PVPH_2YMASD	0.278
PRO_OCUP_C	-0.191
PVPH_C_ELEC	0.276
PVPH_TV	0.292
PVPH_REFRI	0.320
PVPH_LAVAD	0.286
PVPH_PC	0.219
PVPH_AGUADV	0.281
PVPH_EXCSA	0.272
PVPH_DRENAJ	0.281
PPDER_SS	0.213
PPOCUPADA	0.172
GRAPROES	

Dentro del componente se observó que las variables que menos explicaban el componente fue el promedio de ocupantes por cuarto, Grado promedio de escolaridad, población ocupada, porcentaje de población derechohabiente a servicios de salud y porcentaje de viviendas particulares habitadas que disponen de computadora, cabe mencionar que no se eliminaron la primeras 3 variables mencionadas dado que era la única variable con la que se contaba que diera evidencia de esa dimensión y la última se consideró por la importancia que tiene para el desarrollo social dentro de un ámbito profesional y académico.

- Análisis del año 2020

Tabla 11. Importancia de los componentes.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Desviación estándar	2.914	1.311	0.998	0.941	0.856	0.748	0.658	0.544
Proporción de varianza	0.566	0.115	0.066	0.059	0.045	0.037	0.029	0.020
Proporción acumulada	0.566	0.681	0.747	0.806	0.851	0.888	0.917	0.937

Mediante este análisis se observó que con tres componentes estaríamos explicando el 74.7 por ciento de la variabilidad de la información que es lo recomendable, pero como para obtener el índice de desarrollo social solo interesa el primer componente, aunque apenas y explico el 56.6 % de la variabilidad de los datos.

Tabla 12. Importancia de las variables dentro del primer componente.

Variable	Componente 1
PVPH_PISODT	0.254
PVPH_3YMASC	0.309
PVPH_2YMASD	0.292
PRO_OCUP_C	-0.315
PVPH_C_ELEC	0.184
PVPH_TV	0.251
PVPH_REFRI	0.321
PVPH_LAVAD	0.313
PVPH_PC	0.312
PVPH_AGUADV	0.211
PVPH_EXCSA	0.251
PVPH_DRENAJ	0.140
PPDER_SS	0.173
PPOCUPADA	
GRAPROES	0.311

Para el primer componente se observó que algunas variables de menor peso cambiaban para el 2020 con respecto al año 2010, las cuales fueron: porcentaje de población derechohabiente a servicios de salud, porcentaje de población ocupada, porcentaje de viviendas particulares habitadas que disponen de drenaje y la variable promedio de ocupantes por cuarto obtuvo un peso negativo.

Análisis de componentes principales con Kernels

Considerando los resultados obtenidos y analizando el principal objetivo de la generación de un índice que es que explique la mayor variabilidad posible de la información y pueda dar mayor evidencia de lo certero no se realizó el cálculo del índice mediante este análisis por los escasos porcentajes de variabilidad que explicaba el primer componente para ambos años y la recomendación es que explique al menos el 70%. Es por ello que como tercera fase de la propuesta del estudio se decidió aplicar una extensión: componentes principales con kernels (ACPK) con la finalidad de aumentar la variabilidad explicada por el primer componente utilizando una transformación no lineal de donde se obtuvo lo siguiente:

Análisis ACPK para las colonias de 2010

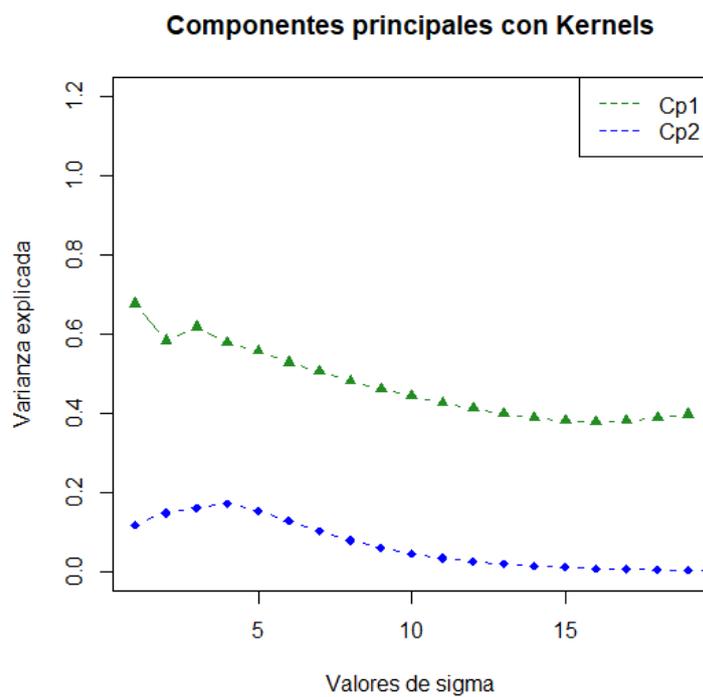


Figura 2.- Comportamiento de la varianza explicada por valor del parámetro sigma.

En este caso el comportamiento de la varianza explicada solo alcanza el 67% para una sigma igual a 1, es decir, ACPK solo mejora un 7% el porcentaje de varianza explicada proporcionado por ACP.

Análisis ACPK para las colonias de 2020

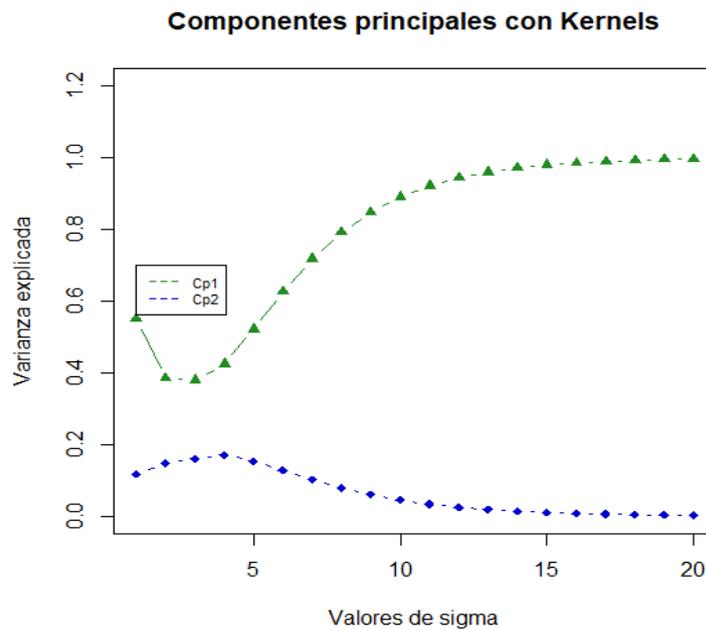


Figura 3. Comportamiento de la varianza explicada por valor del parámetro sigma para el año 2020.

Nótese que los porcentajes de varianza explicados por el primer componente es muy alto con respecto a los resultados proporcionados por el ACP conforme aumenta el valor de sigma. En el valor 20 de sigma podemos observar que aportó la máxima varianza explicada de 99.60%.

Por lo que con esta técnica se procedió a realizar el cálculo del IDS para ambos años siendo la mejor combinación que permite una mayor explicación de la variabilidad de los datos de las dimensiones por colonias.

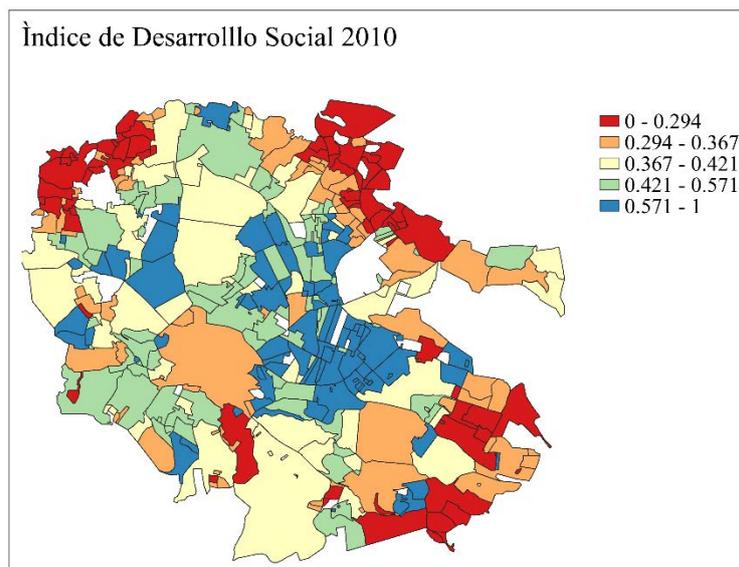


Figura 4. Índice de desarrollo social para colonias en 2010.

Se observo que en año 2010 el Índice de Desarrollo Social es bajo para los alrededores del municipio lo que estaría indicando que con la expansión urbana las zonas que se ven afectadas son las zonas extremas que están en plena construcción.

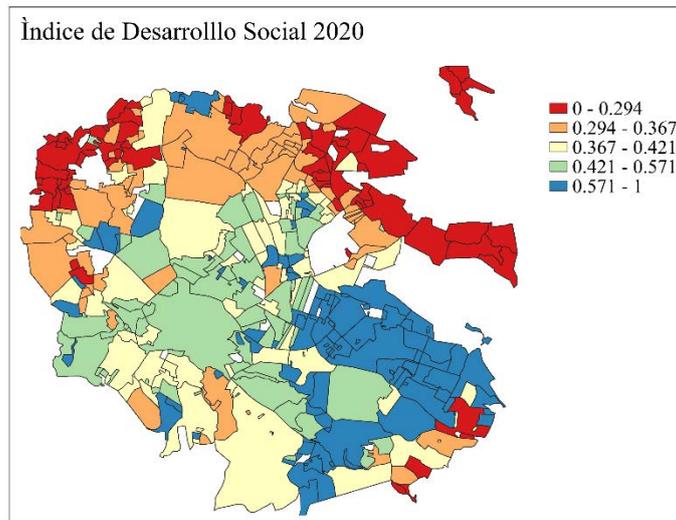


Figura 5. Índice de desarrollo social para colonias en 2020.

Para el año 2020 el comportamiento del desarrollo social es muy marcado tanto que se observó que la zona norte en las colindancias del municipio con Naolinco, Jilotepec, Banderilla, Tlalnelhuayocan el desarrollo social es muy bajo, es punto de atención para los programas de desarrollo sociales que propone el gobierno para atender las necesidades básicas insatisfechas de esta zona. También es importante destacar que para el análisis del año 2020 se contemplaron 390 colonias cuando en el año 2010 solo se registro información para 377 colonias.

Por último, se propuso medir la asociación espacial mediante el índice de Moran con el fin de analizar las agrupaciones de la información con su respectiva significancia, logrando identificar las unidades territoriales que se agrupan.

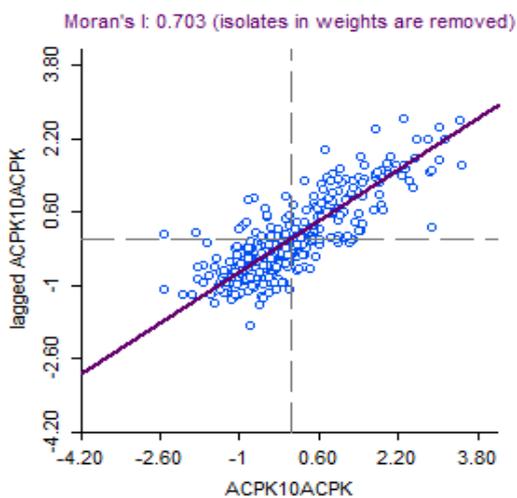


Figura 6. Asociación espacial 2010.

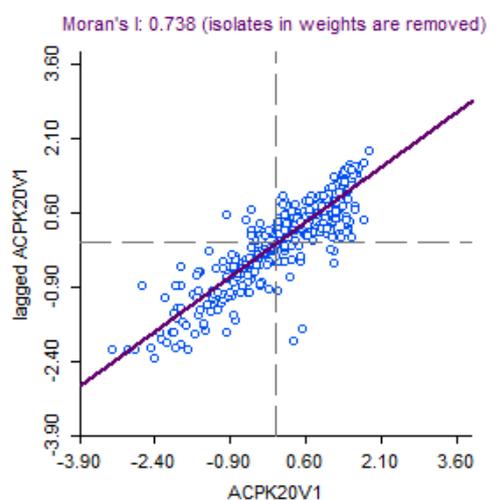


Figura 7. Asociación espacial 2020.

El valor de 0.7 para ambos años indica que existe una agrupación alta de valores del IDS de acuerdo con el espacio. Es importante destacar que en las dos fechas la mayoría de las colonias se concentra en los grupos de valores altos de índice rodeados de valores alto y valores bajos rodeados de valores bajos, aunque para el 2010 se encuentran concentradas más colonias en los grupos de valores bajos.

Conclusión

En este trabajo se ha presentado de manera breve una propuesta de medición de una gran situación social como lo es la pobreza, medida desde un índice de desarrollo social analizado mediante la utilización de técnicas que den una mayor explicación de la variabilidad de la situación como lo fue la técnica de ACPK.

Se logro mostrar evidencia de la situación espacial desde técnicas espaciales de indicadores locales de asociación espacial destacando que el desarrollo social se ve influenciado por el espacio dado que mostró un índice de Morgan alto, es interesante destacar la importancia de la utilización de estas técnicas en situaciones sociales.

La situación espacial de la pobreza que se presenta a nivel colonia en municipio de Xalapa, a partir de un índice de desarrollo social podemos decir que es diversa, pero presenta un patrón de comportamiento influenciado por el tiempo y espacio. Para el año 2010 parecía que la periferia mostraba una pobreza alta, lo cual podría ser causa de la expansión urbana, sin embargo, para 2020 se logró ver un avance en el desarrollo social e identificar un foco de atención en las colindancias del norte y noroeste del municipio donde siguen existiendo carencias. Solo por mencionar algunas colonias que se ha evidenciado con menos desarrollo social son: Miradores de San Roque, Ignacio Zaragoza, Luis Donald Colosio, Potrerillo, Reserva del Tronconal, Brisas del Sedeño, Independencia, Nuevo Sol, La Hermita, Lomas del Sedeño, Fraternidad Antorchista, Comercio, Convergencia, Valle del Tronconal, Diamante, Dante Delgado Ranuro, Insurgentes, Del Cerrillo, La Barrera, Dolores Hidalgo, El Nacimiento, Plan de Ayala, Constitución del Sumidero, Solidaridad, Las Trancas, La Palma, San Judas Tadeo, Loma Bonita, Arroyo Zarco, Del Centro, Campamento 24 de junio, FREDEPO, La Tranca, Reserva Territorial, Lomas de los Castillos, Miradores de las Flores, Altamira, Bosques del Sumidero, Encinos, Vicente Guerrero, Valle de los Pinos, UCISVER, Arboledas del Tronconal, Lealtad Institucional, Manantiales, UNE-PRI y Las Torres.

Los resultados de esta investigación destacan la utilización de técnicas estadísticas multivariante en situaciones sociales, así se centra en la innovación de nuevas técnicas y herramientas para este tipo de estudio sin embargo se destaca la implicación que éstas requieren para trabajar grandes cantidades de datos para realizar estudios con niveles de desagregación de la información tan específica como lo fue en un principio al trabajarlo a nivel manzana.

Referencias

- CONEVAL. (2020). *Consejo Nacional de Evaluación de la Política de Desarrollo Social*. Obtenido de Informe de pobreza y evaluación 2020. Veracruz: https://www.coneval.org.mx/coordinacion/entidades/Documents/Informes_de_pobreza_y_evaluacion_2020_Documentos/Informe_Veracruz_2020.pdf
- CONEVAL. (2020). *Consejo Nacional de Evaluación de la Política de Desarrollo Social*. Obtenido de Pobreza Municipal: https://www.coneval.org.mx/coordinacion/entidades/Veracruz/Paginas/pobreza_municipal2015.aspx
- EVALUA, C. d. (2011). *EVALUÀ*. Obtenido de Índice de Desarrollo Social de las Unidades Territoriales del Distrito Federal/IEDS (Delegaciones-Colonias-Manzanas): <http://data.evalua.cdmx.gob.mx/medicion-unidades-territoriales.php>
- Julia Aurora Montano Rivas, S. F. (Febrero de 2013). *Repositorio institucional Universidad Veracruzana*. Obtenido de <https://cdigital.uv.mx/handle/123456789/36349>
- Osorno, I. D. (2012). *ANÁLISIS ESTADÍSTICO DEL REZAGO*. Obtenido de Tesis de especialización, Universidad Veracruzana: <https://www.uv.mx/eme/files/2012/11/Analisis-estadistico-del-rezago-social-y-economico-en-Xalapa-Ver.-2010.pdf>
- Pérez, G. J. (Enero de 2005). *Banco de la Republica Colombia*. Obtenido de Banco de la republica Colombia: <https://www.banrep.gov.co/sites/default/files/publicaciones/archivos/DTSER-54.pdf>
- Peña, L. L. (2012). Alcances y límites de los métodos de análisis espacial para el estudio de la pobreza urbana. Papeles de Población.

Pronóstico del desempleo en México: una aplicación de series de tiempo multivariadas

Darío Alejandro Segovia Hernández¹, Eliseo Gabriel Argüelles², Cecilia Cruz López³

¹Especialización en Métodos Estadísticos. zS20000594@estudiantes.uv.mx

²Facultad de Estadística e Informática, Universidad Veracruzana. egabriel@uv.mx

³Facultad de Estadística e Informática, Universidad Veracruzana. cecacruz@uv.mx

Resumen

Anticipar la evolución de la tasa de desempleo es de suma relevancia para la conducción de la política económica, ya que permite al gobierno adoptar medidas oportunas para minimizar los efectos sobre la economía. Con tal propósito, la presente investigación estima un modelo econométrico que incluye un índice de flexibilidad laboral numérica, el cual, permite realizar una simulación que reproduce con gran capacidad la evolución de la tasa de desempleo. Los resultados sugieren que, de mantenerse invariables los esquemas de contratación y despido ante los efectos de la crisis COVID-19, la tasa de desempleo se verá incrementada.

Palabras Clave: Desempleo, flexibilidad laboral, series de tiempo, vector de corrección de error.

Introducción

Los pronósticos se realizan con la finalidad de orientar las decisiones en diversas áreas. Para pronosticar una variable se debe construir un modelo y estimar sus parámetros usando datos históricos, es decir, logrando una caracterización estadística de los enlaces entre el presente y el pasado. En forma simple se usa el modelo estimado del pronóstico para extrapolar los datos observados (Diebold 2001).

Una de las variables que ha sido objeto de especulación a raíz de los efectos del impacto de la crisis COVID-19 en México ha sido el desempleo, la cual se define como aquella situación en la que uno o varios individuos que forman parte de la población en edad de trabajar y con disposición de hacerlo no tienen una ocupación remunerada; es decir, no desempeñan actividad económica alguna (BANXICO, 2021).

Lo anterior se debe a que los efectos de la crisis se percibieron primordialmente sobre la actividad económica mexicana, afectando un mercado laboral ya debilitado por brechas en materia de acceso y calidad del empleo. La estimación de empleos en riesgo de verse afectado por la pandemia rebasa los 24 millones, lo que representa el 44% del empleo total en México (CEPAL, 2020).

La flexibilidad laboral, como la velocidad con la que los mercados laborales se adaptan a los cambios y fluctuaciones de la economía y la sociedad (OCDE, 1986), ha sido señalada como uno de los factores que agudizaron los efectos de la crisis sobre el mercado laboral. Ya que, como señalan Chiquiar y Ramos-Francia (2010) y Ros (2013), las medidas de flexibilidad laboral implementadas en México han sido insuficientes, debido a que el mercado laboral mexicano presenta diversas restricciones que han derivado en rigidez laboral.

Un mercado laboral flexible permite a los empleadores realizar cambios debido a problemas de oferta y demanda, el ciclo económico y otras condiciones del mercado, mientras que, un mercado que presenta rigidez, este sujeto a más reglas y regulaciones, lo cual conlleva a una serie de desajustes económicos (Ibarra y González, 2010). De tal manera que, no es extrañar, que el mercado laboral mexicano no pudiese sobreponerse a la crisis COVID-19.

La lenta recuperación de la economía mexicana ha generado expectativas variadas sobre la evolución de la tasa de desempleo, generando discrepancias sobre la conducción de la política económica y los esquemas laborales derivados de las medidas de contingencia de la crisis COVID-19.

Es así como, esta investigación se centra en las implicaciones macroeconómicas de las medidas de flexibilidad laboral sobre la tasa de desempleo en México. Por tal motivo, se tiene como objetivo determinar el impacto de mantener invariable los esquemas de contratación y despido sobre la tasa de desempleo de México. Se pretende proponer un modelo para el pronóstico de la tasa de desempleo, a partir de la comprensión de la relación existente entre la tasa de desempleo y el índice de flexibilidad laboral numérica a partir de la metodología de Johansen (1998).

Metodología

Datos

Se utilizaron dos series de tiempo con periodicidad trimestral, tomando como periodo de análisis el primer trimestre de 1998 al último trimestre de 2020. Las series de tiempo se obtuvieron de la base de información del Instituto Nacional de Estadística y Geografía (INEGI) y de la Secretaría de Trabajo y Previsión Social (STPS).

Así, se presentan las siguientes variables macroeconómicas en el modelo econométrico:

- *Tasa de desempleo (td)*. Se utiliza la tasa de desempleo de la población urbana de las 32 ciudades con el objetivo de evitar encadenar series y utilizar series homogéneas. Se obtiene la tasa de desempleo con periodicidad trimestral de acuerdo con información obtenida del INEGI presentada en unidades porcentuales con una longitud de la serie de 1998T1-2020T4.
- *Índice de flexibilidad laboral numérica (fl)*. El índice de flexibilidad laboral numérica representa la facilidad de contratación y despido en el sector formal como la relación de los empleos eventuales respecto al empleo total formal, tal que nos permita conocer cómo ha cambiado la tasa de empleos eventuales en el sector formal². Las series se obtienen a partir de la base de datos de la STPS, con una periodicidad de 1998T1-2020T4, así el índice de flexibilidad laboral se obtiene en puntos porcentuales con la misma periodicidad.

² El empleo eventual se considera como todos aquellos cotizantes del IMSS que trabajan bajo esquemas de contratación definida o temporal y el empleo formal como el total de los cotizantes en el Instituto Mexicano del Seguro Social.

El modelo por desarrollar se estima con las variables en logaritmos para reducir escalas, y así obtener las elasticidades de manera directa (Figura 3). Se referirá al logaritmo natural de la tasa de desempleo como $\ln(td)$ y al logaritmo natural del índice de flexibilidad numérica como $\ln(fl)$.

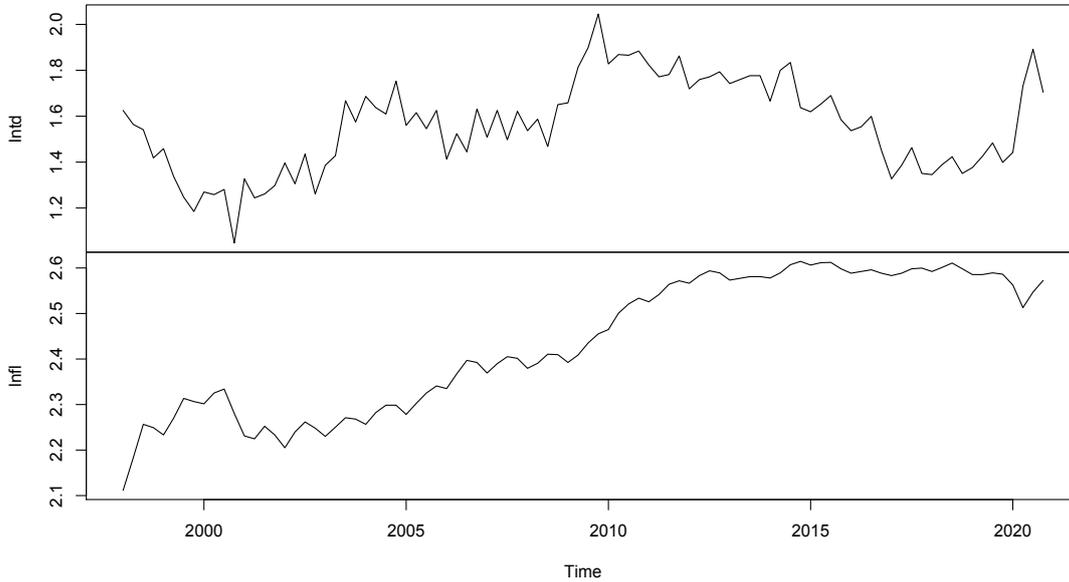


Figura 1. Variables del modelo en logaritmos.

Especificación del modelo

Las series de la Figura 1 presentan un componente de tendencia, sin determinar si es estocástica o determinística, lo que nos permite sugerir que las series presentan un orden $I(1)$. De tal forma que, las series no presentan consistencia en su media y/o varianzas a través del tiempo.

Por lo anterior, cabe la posibilidad de que el proceso generador de información se vea sesgado ante la posibilidad de correlación de errores. Para solucionar este problema, se sugiere una prueba de cointegración, la cual permite encontrar una combinación lineal entre dos o más vectores, tal que, la resultante sea un vector de media cero y varianzas constantes.

La metodología de cointegración fue desarrollada por Engle y Granger (1987), a través de una prueba bivariada, en la que, se plantea que la variable dependiente y_t y la independiente x_t tienen el mismo orden de integración, de tal forma que de la diferencia entre estas exista un vector cointegrante (A) que represente una combinación lineal que genera un vector z_t de orden $I(0)$ a fin de que $z_{t-i} \sim N(0, \sigma^2)$. Así, la cointegración bivariada se representa bajo la siguiente especificación:

$$z_t = x_t - \alpha_0 y_t ; A = \begin{pmatrix} 1 \\ -\alpha_0 \end{pmatrix} \tag{1}$$

La metodología de vectores autorregresivos propuesta por Sims (1980) se basa en un modelo econométrico sustentado como un sistema de ecuaciones dinámico, haciéndola una técnica adecuada para modelar dos o más variables. Sin embargo, al tener series de orden de integración $I(1)$, es necesario estimar un vector de cointegración a partir de la prueba de cointegración multivariada propuesta por Johansen (1998).

En el caso de las series de la tasa de desempleo y el índice de flexibilidad laboral numérica, debido a que todas las series presentan un orden de integración $I(1)$, se aplicó el procedimiento de cointegración de Johansen (1988) a fin de encontrar un vector de cointegración con el modelo de ausencia de intercepción y tendencia³. Por lo anterior, se modela un vector de corrección de error que, bajo los criterios de Aikake y Schwartz, se determina un orden óptimo de rezagos $k = 4$. Además, se incluye una variable exógena dicotómica para corregir las atipicidades del modelo, el cual se presenta con la siguiente expresión:

$$\Delta x_t = \sum_{i=1}^4 r_i \Delta x_{t-i} + \pi x_{t-1} + \emptyset D_t + \varepsilon_{mt} \quad (2)$$

Donde $\pi = BA'$, $\varepsilon_{mt} \sim N(\mu, \sigma^2)$, $\Delta x_t \sim N(0, \sigma^2)$. Reescribiendo en su forma estructural:

$$\Delta x_t = r_1 \Delta x_{t-1} + r_2 \Delta x_{t-2} + r_3 \Delta x_{t-3} + r_4 \Delta x_{t-4} + BA' x_{t-1} + \emptyset D_t + \varepsilon_{mt} \quad (3)$$

Cuya forma estructural primitiva se expresa de la siguiente manera:

$$\begin{aligned} \begin{pmatrix} \Delta \ln td_t \\ \Delta \ln fl_t \end{pmatrix} &= \begin{pmatrix} y_{1,11} & y_{1,12} \\ y_{1,21} & y_{1,22} \end{pmatrix} \begin{pmatrix} \Delta \ln td_{1t-1} \\ \Delta \ln fl_{2t-1} \end{pmatrix} + \begin{pmatrix} y_{2,11} & y_{2,12} \\ y_{2,21} & y_{2,22} \end{pmatrix} \begin{pmatrix} \Delta \ln td_{1t-2} \\ \Delta \ln fl_{2t-2} \end{pmatrix} \\ &+ \begin{pmatrix} y_{3,11} & y_{3,12} \\ y_{3,21} & y_{3,22} \end{pmatrix} \begin{pmatrix} \Delta \ln td_{1t-3} \\ \Delta \ln fl_{2t-3} \end{pmatrix} \\ &+ \begin{pmatrix} y_{4,11} & y_{4,12} \\ y_{4,21} & y_{4,22} \end{pmatrix} \begin{pmatrix} \Delta \ln td_{1t-4} \\ \Delta \ln fl_{2t-4} \end{pmatrix} \\ &+ \begin{pmatrix} -B_1 \\ B_2 \end{pmatrix} (\hat{\alpha}_0 \quad - \hat{\alpha}_1)' \begin{pmatrix} \ln td_{t-1} \\ \ln fl_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 1,1 \\ 1 & 1,2 \end{pmatrix} D + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \end{aligned} \quad (4)$$

De la forma primitiva se obtiene un sistema de matrices, donde las variables endógenas son: el logaritmo de la tasa de desempleo ($\ln td_t$) y el logaritmo del índice de flexibilidad laboral ($\ln fl_t$), y se incluye en el modelo una variable exógena de ajuste (D). No obstante, los parámetros c , y , \emptyset no tiene lectura económica, debido a que son coeficientes de estimación. De tal manera que, el análisis económico se realiza a partir del vector de cointegración definido por $z_{t-i} = A' x_{t-i}$, donde $i = 1, 2, 3, \dots, n$ tal que $z_{t-i} \sim N(0, \sigma^2)$. Por último, el vector de innovaciones conformados por los ε_{mt} , tiene que cumplir con las condiciones de estabilidad en $\varepsilon_{mt} \sim N(\mu, \sigma^2)$.

³ Para más información respecto al procedimiento cointegración Johansen, véase Montenegro (1988; 231-254).

Es importante destacar que un modelo de vector de corrección de error se puede entender como un modelo de vector autorregresivo, en diferencia que permite al modelo estabilidad, tal que, incorpora un vector de corrección de error incorporado en π . Ya que $\pi = BA'$, el vector A' tiene la importancia de guardar el efecto de la combinación lineal de las tendencias comunes entre las variables, conocido como el vector cointegrante, que al ser multiplicado por el vector x_{t-1} , se obtiene el vector de cointegración z_{t-i} , es decir, $z_{t-i} = A'x_{t-i}$ donde $i = 1, 2, 3, \dots, n$ y $z_{t-i} \sim N(0, \sigma^2)$.

De esta manera, se puede entender al vector z_{t-i} como el equilibrio a largo plazo de las variables endógenas del modelo, y a B como el vector que expresa la incidencia de cada variable. Las dimensiones de los vectores A y B son de $M \times m$ y $m \times M$ respectivamente, donde M es el número de vectores de cointegración y m es el número de variables en el sistema.

De acuerdo con Montenegro (2010: 271), el algoritmo de Johansen (1988) se basa en la maximización de una función de verisimilitud concentrada⁴, con la restricción de $\pi = BA'$ y que π debe cumplir con la condición $0 < [rango(\pi) = M] < m$, donde m representa el número de variables endógenas del modelo. Para una correcta especificación de cointegración, M es el número de relaciones de equilibrio entre las variables.

Dos aspectos del modelo que son necesarios abordar son: el mecanismo de corrección de error como elemento equilibrador y la estabilidad dinámica del sistema. Para ello se utiliza la ecuación de cointegración bivariada y se despejan las variables en diferencia con el uso del operador de rezago L , que para este caso va del 1 al 4. De esta manera, la ecuación se expresa de la siguiente manera:

$$(I - L^1 r_1 - L^2 r_2 - L^3 r_3 - L^4 r_4) = BA'x_{t-1} + \emptyset D_t + \varepsilon_{mt} \quad (5)$$

Si $\Omega = I - L^1 r_1 - L^2 r_2 - L^3 r_3 - L^4 r_4$, entonces:

$$\Omega \Delta x_t = BA'x_{t-1} + \emptyset D_t + \varepsilon_{mt} \quad (6)$$

$$\Delta x_t = \Omega^{-1} (BA'x_{t-1} + \emptyset D_t + \varepsilon_{mt}) \quad (7)$$

$$\Delta x_t = \frac{1}{\det(\Omega)} \Omega^{cof} (BA'x_{t-1} + \emptyset D_t + \varepsilon_{mt}) \quad (8)$$

Si u_{mt} son los errores no corregidos con la variable exógena D_t y su parámetro \emptyset , entonces $u_{mt} = \emptyset D_t + \varepsilon_{mt}$, tal que, la distribución de $u_{mt} = N(0, \sigma^2)$. Por lo tanto:

$$\Delta x_t = \frac{1}{\det(\Omega)} \Omega^{cof} (BA'x_{t-1} + u_{mt}) \quad (9)$$

Ampliando matricialmente la ecuación (6), se obtiene:

$$\begin{pmatrix} \Delta \ln td_t \\ \Delta \ln fl_t \end{pmatrix} = \frac{1}{\det(\Omega)} \Omega^{cof} \left[\begin{pmatrix} -B_1 \\ B_2 \end{pmatrix} (1 \hat{\alpha})' + \begin{pmatrix} \ln td_{t-1} \\ \ln fl_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \right] \quad (10)$$

⁴ Para la demostración matemática del algoritmo de Johansen de una manera sintética, véase Montenegro (2010; 270-282).

Cabe señalar que en la anterior expresión el vector de cointegración esta normalizado sobre $\hat{\alpha}_0$, tal que $A'/\hat{\alpha}_0$. Para obtener el vector de cointegración z_{t-1} se requiere multiplicar el vector cointegrante A' por el vector variable x_{t-1} , tal que $z_{t-1} = A'x_{t-1}$. Por lo tanto:

$$\begin{pmatrix} \Delta \ln td_t \\ \Delta \ln fl_t \end{pmatrix} = \frac{1}{\det(\Omega)} \Omega^{cof} \left[\begin{pmatrix} -B_1 \\ B_2 \end{pmatrix} (\ln td_{t-1} + \hat{\alpha} \ln fl_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \right] \quad (11)$$

Sustituyendo las diferencias y despejando:

$$\begin{pmatrix} \ln td_t \\ \ln fl_t \end{pmatrix} - \begin{pmatrix} \ln td_{t-1} \\ \ln fl_{t-1} \end{pmatrix} = \frac{1}{\det(\Omega)} \Omega^{cof} \left[\begin{pmatrix} -B_1 \\ B_2 \end{pmatrix} (\ln td_{t-1} + \hat{\alpha} \ln fl_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \right] \quad (12)$$

$$\begin{pmatrix} \ln td_t \\ \ln fl_t \end{pmatrix} = \begin{pmatrix} \ln td_{t-1} \\ \ln fl_{t-1} \end{pmatrix} + \frac{1}{\det(\Omega)} \Omega^{cof} \left[\begin{pmatrix} -B_1 \\ B_2 \end{pmatrix} (\ln td_{t-1} + \hat{\alpha} \ln fl_{t-1}) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \right] \quad (13)$$

Se define al vector cointegrante A' como aquel que mantendrá la relación en el largo plazo con los signos especificados, tal que en el largo plazo $i = n$ de la ecuación en el vector de cointegración z_{t-i} los signos se mantienen para todo el periodo desde $i = 1, 2, 3, \dots, n$, tal que z_{t-i} se despeja sobre $\ln(td_{t-i})$ quedando la relación a largo plazo: $\ln(td_{t-i}) = \hat{\alpha}(\ln(fl_{t-i}))$.

No obstante, de presentarse una desviación en el primer periodo tal que, $\ln td_{t-1} > \hat{\alpha} \ln fl_{t-1}$, es entonces la importancia del vector B como mecanismo de corrección de error, ya que tiene la capacidad de mantener el equilibrio sobre el vector de cointegración dinámicamente y fijar estabilidad en el sistema, por lo que en cualquier caso es necesario que $-1 < B < 0$, ya que el efecto de disminución reducirá $\ln td_{t-1}$ y en la siguiente observación será menor el efecto de retroalimentación del sistema sobre el vector de cointegración, y el proceso continuará hasta que las variables retomen la igualdad. Por lo tanto, B siempre debe de ser negativo para equilibrar el vector de cointegración y $|B| < 1$ para que el sistema tenga estabilidad.

Estabilidad dinámica y convergencia del sistema

En cuanto a la estabilidad del sistema, retomamos la ecuación (8) y multiplicamos

$$\Delta x_t = \frac{1}{\det(\Omega)} \Omega^{cof} B A' x_{t-1} + \frac{1}{\det(\Omega)} \Omega^{cof} u_t \quad (14)$$

Como tenemos un vector de corrección de error (k), todo proceso autorregresivo tiene la propiedad de invertibilidad, tal que, un proceso de vectores autorregresivos se convierte en un proceso vectorial de medias móviles infinito. Para esto, ampliamos el número de vectores de innovaciones en un proceso infinito y obtenemos:

$$\begin{aligned}
 \Delta x_t = & \frac{1}{\det(\Omega)} \Omega^{cof} B A' x_{t-1} + I u_t + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right) u_{t-1} \\
 & + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^2 u_{t-2} + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^3 u_{t-3} + \dots + \\
 & + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^n u_{t-n}
 \end{aligned} \quad (15)$$

Dado que el proceso infinito en el vector de las innovaciones es una sucesión geométrica, se aplica un límite para determinar el valor total de la progresión. Pueden existir dos casos, uno, es que la serie sea convergente y el límite sea igual a una constante:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} & \left[I u_t + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right) u_{t-1} + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^2 u_{t-2} \right. \\
 & \left. + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^3 u_{t-3} + \dots + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^n u_{t-n} \right] \\
 & = I \left(I - \frac{1}{\det(\Omega)} \Omega^{cof} \right)^{-1} u_t = I \Pi^{-1} u_t = \begin{pmatrix} \mu_{u1t} \\ \mu_{u2t} \end{pmatrix} = \mu_{ut}
 \end{aligned} \quad (16)$$

O bien, que el sistema sea divergente hasta infinito.

$$\begin{aligned}
 \lim_{n \rightarrow \infty} & \left[I u_t + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right) u_{t-1} + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^2 u_{t-2} \right. \\
 & \left. + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^3 u_{t-3} + \dots + \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^n u_{t-n} \right] \\
 & = \infty
 \end{aligned} \quad (17)$$

Para cumplir con las condiciones de estabilidad dinámica se tiene que garantizar que el $\det(\Omega) \neq 0$ para que la matriz Ω sea regular y se pueda invertir, además, que el polinomio definido por $\det(\Omega)$, donde $\Omega = I - L^1 r_1 - L^2 r_2 - L^3 r_3 - L^4 r_4$, las raíces características reales definidas por L deben ser $\lambda < |1|$, o en los casos que las raíces complejas, el coeficiente que acompaña a la raíz imaginaria debe ser menor que uno en términos absolutos para que el módulo en la ecuación de Euler sea $\lambda < |1|$. Además, con el objetivo de que los elementos de $\frac{1}{\det(\Omega)} \Omega^{cof}$ sean menores que 1 en términos absolutos, la matriz es convergente y cumple con la condición:

$$\begin{aligned}
 I > \left| \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right) \right| > \left| \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^2 \right| > \left| \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^3 \right| > \dots \\
 > \left| \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^n \right|
 \end{aligned} \quad (18)$$

Donde $\left| \left(\frac{1}{\det(\Omega)} \Omega^{cof} \right)^n \right| = 0$

Si no se cumple con estas condiciones el sistema no tendrá estabilidad dinámica. Por lo tanto, una forma de comprobar la convergencia es a través del vector de innovaciones. Para esto, utilizamos el proceso vectorial de medias móviles definido por:

$$\begin{aligned}
 Iu_t + \left(\frac{1}{\det(\Omega)}\Omega^{cof}\right)u_{t-1} + \left(\frac{1}{\det(\Omega)}\Omega^{cof}\right)^2 u_{t-2} \\
 + \left(\frac{1}{\det(\Omega)}\Omega^{cof}\right)^3 u_{t-3} + \dots + \left(\frac{1}{\det(\Omega)}\Omega^{cof}\right)^n u_{t-n}
 \end{aligned}
 \tag{19}$$

De tal manera que, para determinar convergencia, se le aplica un choque de una desviación estándar al vector de las innovaciones en uno de sus elementos para el periodo $i = 0$, tal que este lo saque del equilibrio en su media u_{ut} . Siendo u_t un proceso infinito, para que este regrese a su equilibrio, es necesario que se cumpla con la condición de estabilidad, para lo que esta matriz regresará al vector de innovaciones a su equilibrio en el tiempo n , tal que ante el choque en el vector de las innovaciones en $i = 1$, entonces:

$$\lim_{n \rightarrow \infty} \left[\sum_{i=0}^{\infty} \left(\frac{1}{\det(\Omega)}\Omega^{cof}\right)^i u_{t-i} \right] = \mu_{ut}
 \tag{20}$$

Por lo anterior, resulta relevante el análisis del impulso-respuesta en la estimación econométrica, ya que nos permite determinar convergencia a través de las condiciones de estabilidad en los parámetros, para así ver la respuesta de las demás variables en las interacciones del sistema, además de, verificar los signos en el vector de cointegración.

Resultados

Se procede a determinar el orden de integración de las series a partir de la prueba de Dickey Fuller aumentada, determinando que las series son de orden de integración 1, es decir que, las series no son estacionarias. Los resultados se muestran en la Tabla 1.

Tabla 1. Prueba de raíz unitaria ADF.

Variable	T-value	Prob
$\ln(td)$	-2.4032*	0.0410*
$\Delta \ln(td)$	-2.5429*	0.0352*
$\ln(fl)$	-3.2051*	0.09193
$\Delta \ln(fl)$	-3.8262*	0.02112*

Nota: *Rechazo de raíz unitaria

Se procede a realizar cointegración multivariada por el método de Johansen (1988). Se encuentra un vector de cointegración con el modelo sin tendencia ni intercepto para el vector de autocorrelación y el vector de cointegración, obteniendo estadísticos de la traza significativos. Se utilizan los criterios de Aikake y Schwartz para determinar un óptimo de cuatro rezagos, obteniendo una ecuación doble logarítmica en el vector de cointegración que se despeja para obtener el siguiente modelo:

$$\ln(td_t) = 1.4079 \ln(fl_t) + e_t \quad (21)$$

El modelo cumple las pruebas de especificación, y tiene un sentido económico debido a que el vector de cointegración se define como una ecuación de largo plazo. Además, el mecanismo de cointegración de error funciona como el elemento equilibrador en el vector de cointegración.

Los resultados muestran que, al expresar la ecuación a largo plazo en logaritmos, el vector de cointegración muestra en sus coeficientes las elasticidades que determinen la sensibilidad entre variables. La elasticidad de la flexibilidad laboral al desempleo estima que, ante un crecimiento del índice de flexibilidad laboral numérica en 1%, la tasa de desempleo crece 1.4079 puntos porcentuales.

Debido a lo antes expuesto, la elasticidad significativamente alta ratifica la resolución de Loria, Ramírez y Salas (2015), quienes mostraron que, a partir de contratos temporales en la economía, se ha aumentado considerablemente la tasa de desempleo. De tal manera que, las medidas de flexibilidad laboral implementadas en México han tenido efectos contraproducentes sobre el desempleo.

Por su parte, el mecanismo de corrección de error precisa ser un mecanismo equilibrador a desviaciones en el vector de cointegración. No obstante, dado que el mecanismo de corrección de error presenta un valor bajo, el ajuste de recuperación del equilibrio es pausado.

Se corroboran los signos del vector de cointegración a través de un análisis impulso-respuesta, en el que se confirma los signos del vector de cointegración y se obtiene evidencia de los efectos más relevantes del sistema. En primer lugar, se puede apreciar que la variable tasa de desempleo se explica a si misma (Figura 2), lo cual se debe a un efecto de histéresis económica. En segundo lugar, se apreciar que, ante un choque del índice de flexibilidad laboral numérica, la tasa de desempleo aumenta (Figura 3).

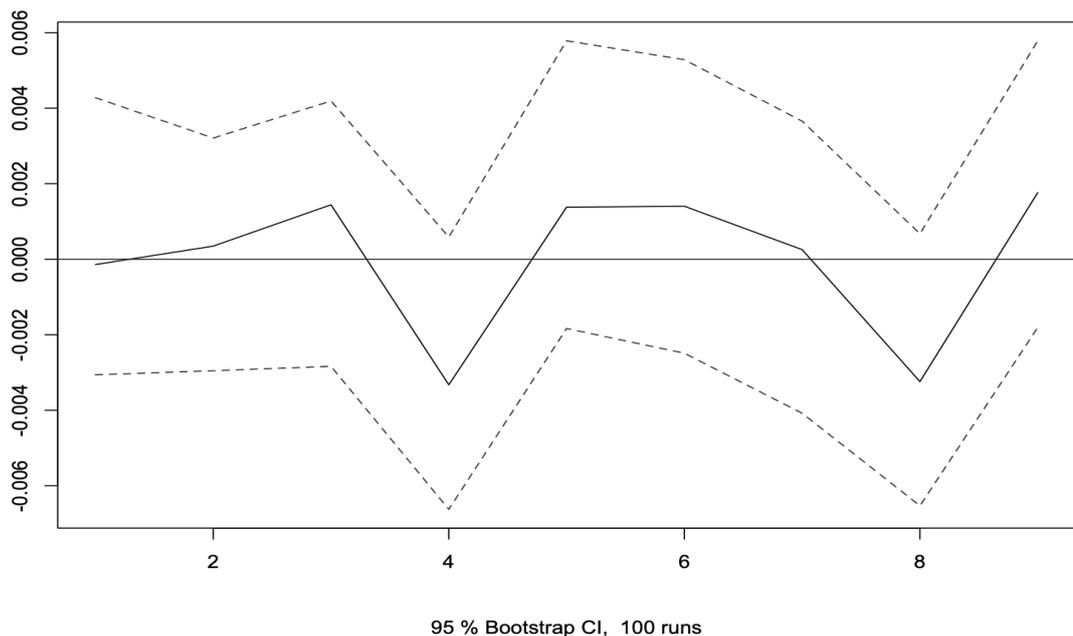


Figura 2. Respuesta de ln(td) a ln(td).

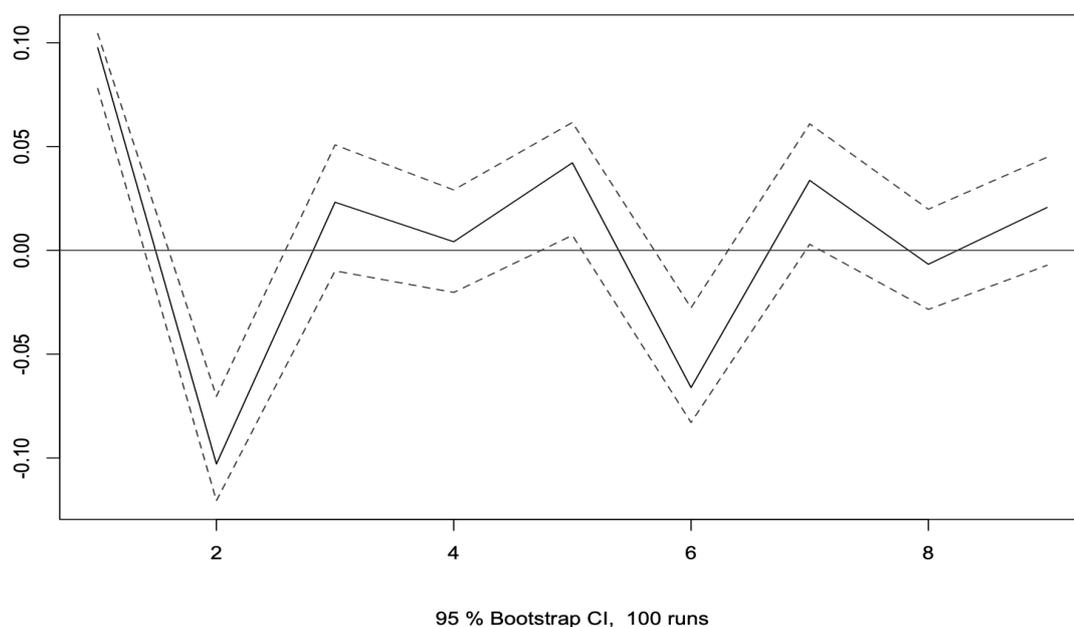


Figura 3. Respuesta de $\ln(td)$ a $\ln(fl)$.

De lo anterior, se infiere que las medidas de flexibilidad laboral implementadas en México han funcionado como una barrera para la incorporación de nuevos elementos al mercado laboral, promoviendo una condición de desempleo que se ve acentuada ante el crecimiento de la población económicamente activa.

La descomposición de varianza de Cholesky (Tabla 2) corrobora que la variable desempleo es explicada por sí misma y por la variable flexibilidad laboral. Por otro lado, la variable flexibilidad laboral solo se explica a través de la variable desempleo, siendo un aspecto relevante el hecho de que no se explique a sí misma.

Tabla 2. Descomposición para td .

Periodo	td	fl
1	1.0000	0.0000
5	0.8711	0.1288
10	0.7654	0.2345
15	0.7248	0.2751
20	0.7091	0.2908
25	0.6999	0.3000
30	0.6952	0.3047

Tabla 3. Descomposición para fl .

Periodo	td	fl
1	0.0014	0.9985
5	0.0408	0.9591
10	0.0941	0.905
15	0.1166	0.8833
20	0.1446	0.8553
25	0.1499	0.8500
30	0.1581	0.8418

Para complementar el análisis se realiza la prueba de causalidad de Granger, la cual nos permite determinar si una serie temporal puede predecir a otra. Los resultados mostraron que las presentan causalidad entre sí (Tabla 3). Recordemos que la causalidad de Granger no debería entenderse como una situación determinística de causa-efecto, sino más bien como un precedente (Kirchgässner & Wolters, 2007).

Tabla 3. Prueba de causalidad de Granger.

Variable dependiente	Variable causal	Chi-sq	df	Prob
Δtd	Δfl	24.2119	4	0.0000*
Δfl	Δtd	39.1129	4	0.0000*

Nota: *Aceptación de causalidad en el sentido de Granger.

Las pruebas confirman que el modelo de corrección de errores vectoriales de orden 4 es robusto, de tal manera que, es capaz de reproducir la evolución histórica de las series temporales. Se estima la tasa de desempleo para 5 años en adelante, los resultados se muestran en la Figura 4.

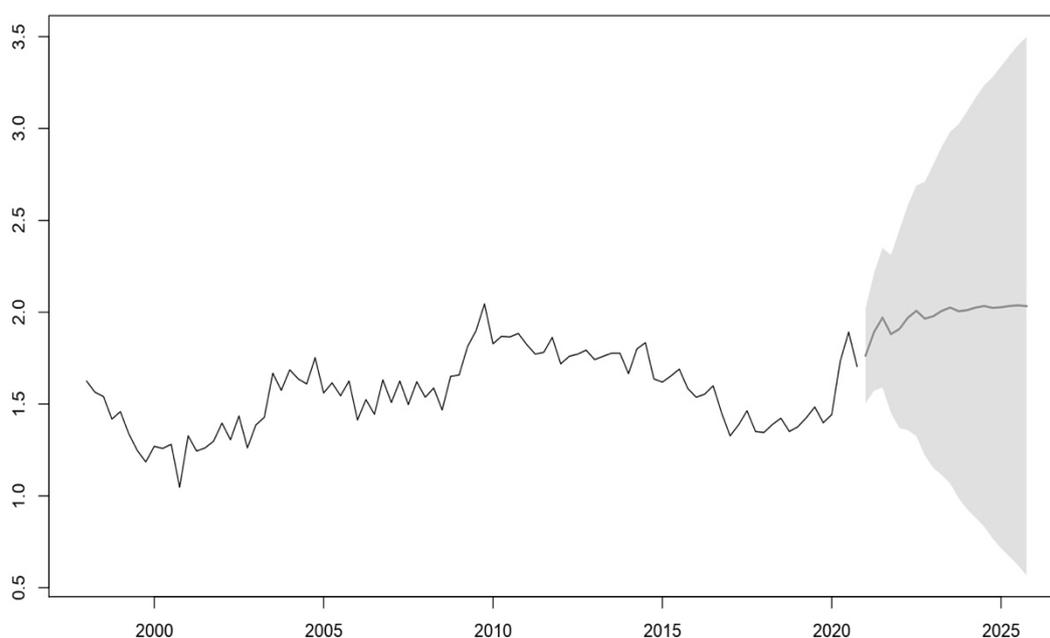


Figura 4. Pronóstico de la tasa de desempleo.

Se observa que los datos pronosticados de la tasa de desempleo siguen el patrón de los datos trimestrales, por lo que los niveles de desempleo mantendrán una tendencia positiva. Lo anterior sugiere que, de mantenerse invariables los esquemas de contratación y despido, la tasa de desempleo se verá incrementada.

Conclusiones

En la presente investigación, se analizó la relación de la flexibilidad laboral con la tasa de desempleo para México en el periodo 1998T1-2020T4, con la finalidad de buscar evidencia empírica a favor de la hipótesis de que los esquemas de contratación y despido tienen una relación de causalidad con los niveles de desempleo.

Se estimó un modelo de vectores de corrección de error de orden 4 que incorpora la tasa de desempleo y un índice de flexibilidad laboral numérica que representa la facilidad de contratación y despido a partir de los contratos temporales. Se prueba que ante un crecimiento del índice de flexibilidad laboral numérica en 1%, la tasa de desempleo crece 1.4079 puntos porcentuales.

El análisis impulso-respuesta y la prueba de causalidad de Granger señalan un efecto contraproducente de las medidas de flexibilidad laboral sobre los niveles de desempleo, corroborando los resultados de Loria, Ramírez y Salas (2015), quienes señalan que la flexibilidad laboral numérica ha actuado como un mecanismo ahorrador de mano de obra que acompañado de tasas de crecimiento económico moderadas, ha promovido un entorno de austeridad en la generación de empleos formales que ha actuado como contención para la inserción de nueva fuerza laboral en el mercado formal de trabajo y esta se convierte en un remanente acumulativo que ha engrosado la tasa de desempleo en México.

Por lo tanto, la relación observada explicaría los niveles de desempleo registrados en la crisis COVID-19 como una marcada rigidez laboral en el ajuste de personal, debido a que el crecimiento de los contratos eventuales es menor que el de población activa. Este escenario es coherente con economías que se distinguen por bajos niveles de productividad y competitividad, debido a la insuficiente inversión pública y privada, al retraso tecnológico, a una excesiva regulación burocrática y a un severo déficit de infraestructura (Tapia, 2009).

El pronóstico de la tasa de desempleo sugiere que, si bien el país ha tomado algunas medidas para contener el impacto de la pandemia en el mercado laboral, de mantenerse invariables los esquemas de contratación y despido, los niveles de desempleo seguirán una tendencia positiva.

Referencias

- BANXICO. (2021). Glosario. Recuperado de http://educa.banxico.org.mx/recursos_banxico_educa/glosario.html
- CEPAL (2020), Enfrentar los efectos cada vez mayores del COVID-19 para una reactivación con igualdad: nuevas proyecciones, Santiago, julio 2020.
- Chiquiar, D. & Ramos-Francia, M. (2009). Competitividad y crecimiento de la economía mexicana. Banco de México: Documentos de Investigación No. 11.
- Diebold, F. & Kilian, L. (2001). Unit root tests are useful for selecting forecasting model. *Journal of Business and Economic Statistics*, 18, 265–273.
- Engle, R. & Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.
- Ibarra, M & González, L. (2010). La flexibilidad laboral como estrategia de competitividad y sus efectos sobre la economía, la empresa y el mercado de trabajo. *Revista UNAM: Contaduría y Administración*, 231.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2), 231-254.
- Kirchgässner G. & Wolters J. (2007). *Granger Causality. In: Introduction to Modern Time Series Analysis*. Springer, Berlin, Heidelberg.

- Loría, E., Ramírez, E., & Salas, E. (2015). La Ley de Okun y la flexibilidad laboral en México: un análisis de cointegración, 1997Q3-2014Q1. *Contaduría y administración*, 60(3), 631-650.
- Montenegro A. (2010). *Análisis de series de tiempo*. Bogotá, Colombia: Universidad Javeriana.
- OCDE (1986). *Labour market flexibility. A report by a high-level group of experts to the Secretary General*. Paris, May
- Ros, J. (2013). *Algunas tesis equivocadas sobre el estancamiento económico de México*. 1.a ed. México: UNAM-Colegio de México
- Tapia, E. (2009). *La Ley de Okun: un enfoque regional para México, 1987-2006*. México: UNAM, Facultad de Economía.

Desigualdad y crecimiento económico: evidencia de la hipótesis del colesterol para México del 2000-2019

José Manuel Mendoza Rivera¹, Ernesto Peredo Rivera², Josefa Carolina Fortuno Hernández³

¹Facultad de Estadística e Informática, Universidad Veracruzana. j_manuelmr@hotmail.com

²Colegio de Postgraduados, Campus San Luis Potosí, peredo.ernesto@colpos.mx

³IIESES, Universidad Veracruzana. jfortuno@uv.mx

Resumen

El efecto de la desigualdad sobre el crecimiento económico ha demostrado empírica y teóricamente ser positivo, pero también negativo. Para aclarar esta ambivalencia recurrimos a la hipótesis del colesterol, la cual separa a la desigualdad en: desigualdad de oportunidades y desigualdad de esfuerzos, cuyos efectos son negativos y positivos para el crecimiento dependiendo del tipo de desigualdad. La hipótesis del colesterol ha sido probada para Estados Unidos, así como en un conjunto de países, sin embargo, no es el caso de México, es por ello, que la presente investigación propone analizar la hipótesis del colesterol en México considerando el periodo de 2000 a 2019. Para ello se usaron datos censales con los cuales se calcularon los índices de desigualdad, así como sus componentes para cada entidad federativa. Los modelos de regresión: pooled OLS, con efectos fijos y efectos aleatorios, dan como resultado que a nivel nacional no se cumple la hipótesis del colesterol, pues los efectos parecen contrarios a lo planteado. Dividimos el país en dos grupos de acuerdo con el nivel de ingresos, los componentes de desigualdad se comportan de manera distinta a lo nacional, dando un nuevo panorama de los efectos de la desigualdad según el nivel de ingresos de las entidades de estudio.

Palabras clave: desigualdad, crecimiento económico, hipótesis del colesterol, desigualdad de oportunidades, desigualdad de esfuerzos.

Introducción

La hipótesis del colesterol como un símil del efecto del colesterol en el cuerpo humano al efecto de la desigualdad en el crecimiento económico de una nación, habiendo un tipo de colesterol bueno y uno malo, y un tipo de desigualdad benéfica y otra perjudicial, así fue denominada por Marrero y Rodríguez (2019), donde a partir de la teoría de desigualdad de oportunidades aportan al debate del efecto de la desigualdad en el crecimiento económico.

La teoría de desigualdad de oportunidades de Roemer (1998), expone que la desigualdad está compuesta por una desigualdad debida a circunstancias, esfuerzos y otros factores, como la suerte. Donde, las circunstancias son elementos económicamente exógenos, fuera del control y responsabilidad de los individuos, por ejemplo: su sexo, lugar de origen y antecedentes parentales, las circunstancias originan desigualdad de oportunidades. Mientras que, los esfuerzos son elementos dentro del control y responsabilidad de los individuos, esto corresponde a las horas dedicadas a estudiar o trabajar, por mencionar un ejemplo, generando así desigualdad de esfuerzos. Otros elementos como la suerte originarían algún tipo de desigualdad residual.

Para Marrero y Rodríguez (2013, 2019) y Marrero et al., (2016) los componentes de la desigualdad deben tener un efecto diverso sobre el crecimiento económico, la desigualdad de oportunidades es perjudicial y la desigualdad de esfuerzos benéfica, resultando un efecto general positivo o negativo, según sea el componente predominante en el contexto de estudio. De allí que, en la múltiple literatura del efecto de la desigualdad sobre el crecimiento, se hayan observado resultados no unánimes (Voichosky, 2009).

Aunque la hipótesis del colesterol es una vía prometedora para develar el verdadero efecto de la desigualdad en el crecimiento económico y a su vez promover la igualdad de oportunidades, la hipótesis necesita ser probada bajo otras condiciones, como en países en desarrollo, donde la distribución de factores de desigualdad es distinta, por ejemplo, México, un país de alta desigualdad de oportunidades (Monroy-Gómez-Franco y Corak, 2019), de bajo crecimiento de largo plazo (Bleynat et al., 2020), y donde solo ha sido probada a medias (Parra, 2019).

Por lo anterior, el objetivo de este capítulo es analizar la hipótesis del colesterol para México. Para ello, se consideraron ciertos objetivos particulares, *i)* estimar la desigualdad de oportunidades y de esfuerzos de las entidades federativas de México del 2000 al 2019; *ii)* Obtener el conjunto de variables independientes restantes a la desigualdad de oportunidades y desigualdad de esfuerzos, así como la dependiente, tasa de crecimiento del Producto Interno Bruto por Entidad Federativa (PIBE) y para el periodo del 2000 al 2019; *iii)* Analizar la hipótesis del colesterol mediante métodos de regresión de datos panel; *iv)* Ejecutar pruebas estadísticas para probar la robustez de los resultados.

De revelarse el efecto variado de los componentes de la desigualdad en el crecimiento, se podrían efectuar medidas de política pública que promuevan la igualdad de oportunidades y la libre elección de los individuos, de tal forma, que promoviendo el desarrollo se promueve el crecimiento. Además de esta Introducción, este capítulo se compone de las siguientes tres partes: Métodos, en el cual exponen los datos y el modo de obtenerlos, así como los modelos de regresión utilizados para analizar la hipótesis. Resultados, en esta sección se describe lo que los modelos y pruebas de hipótesis revelan y, por último, la Conclusión y Discusión.

Metodología

Descripción de la base de datos

Se construyó un panel de datos con 10 variables para los 32 estados de la República Mexicana durante 4 quinquenios, logrando cubrir el periodo que va del 2000 al 2019, por lo que contamos con 128 observaciones por cada unidad de estudio, los datos fueron tomados y contruidos con las bases disponibles por el Instituto Nacional de Estadística y Geografía (INEGI) y la Secretaría de Economía.

Se tienen tres medidas de desigualdad, fueron basadas en la variable independiente educación (tomada de los Censos de Población y Vivienda del 2000 y 2010, el II Censo de Población y Vivienda del 2005 y Encuesta Intercensal del 2015, disponibles por el INEGI), ya que la variable ingreso (variable tradicional en la literatura) no se encuentra registrada longitudinalmente para padres e hijos en México, el utilizar educación como variable de resultado, no daña los objetivos, dado que los ingresos son una variable muy

irregular en el tiempo, por lo que debe medirse en un largo periodo de tiempo y sobre una edad de mayor estabilidad económica de los individuos.

La educación es una variable que después de cierta edad tiende a estancarse, por lo que es más cercana a los ingresos permanentes. La educación es seleccionada de padres e hijos que viven en el mismo hogar, y solo de individuos mayores de 25 años, de esta manera tratar de evitar un posible sesgo, de incluir individuos que aún siguen estudiando, esto es similar a lo que hace Solís (2010) y Moreno (2017).

En los hogares que registraron tanto padre y madre, se seleccionó el de más años de educación, de esta manera pensamos que el individuo que tiene más estudios influye de manera preponderante en el resultado educativo de sus hijos y, se hace más grande la muestra de individuos dentro de cada estado.

Se obtuvo la *desigualdad educativa (DE)*, calculada a través de la varianza de la distribución de los años de escolaridad de la población de cada estado y año, se eligió la varianza como herramienta para calcular desigualdad, debido a homologar la herramienta con nuestros siguientes índices, los cuales requieren permitir la descomposición entre componentes de la desigualdad. La *Desigualdad de oportunidades (IDO)* y *desigualdad de esfuerzos (IDE)* tuvieron como variable de resultado los años de escolaridad, variable que se calculó a partir del método *ex ante* de Ferreira y Gignoux (2011b). Al respecto, hay 5 variables circunstanciales extraídas de los datos censales mencionados párrafos atrás:

- Lugar de origen: urbano o rural.
- Género: hombre o mujer.
- Etnicidad: es indígena o no indígena, esta variable al registrar si el individuo habla algún dialecto.
- Educación de los padres: tomando valores desde cero, hasta el máximo número de años estudiados.
- Índice *proxy* de riqueza: construido a partir del Método de Correspondencias Múltiple (MCM), con variables de bienes, servicios y hacinamiento del hogar, y construido con apoyo del software Stata 13 (pueden consultar estas variables en el Anexo 1 y el grado explicativo de cada índice en el Anexo 2).

La variable principal o de respuesta, fue la *Tasa de Crecimiento del Producto Interno Bruto (TC)* calculada con base en la variación en el Producto Interno Bruto Estatal i , en el quinquenio j a $j+1$.

Las variables de control son la *Tasa de desempleo (TD)*, calculada a partir de datos de la Encuesta Nacional de Ocupación y Empleo y la Encuesta Nacional de Empleo Urbano. *Educación promedio (ESCOPE)*, considerada a partir del promedio de años de escolaridad por entidad y año inicial del periodo, y utilizada para asimilar un grado *proxy* de capital humano por entidad.

La variable *Producto Interno Bruto por Entidad (PIBE)*, fue tomada del Sistema de Cuentas Nacionales de México (SCNM) y es denotada como inicial y per cápita, por ser una medida del PIBE al inicio de cada quinquenio y dividida por la población total del mismo año. También se cuenta con las variables, *Inversión Extranjera Directa (IED)* e *Inversión Pública (INVPU)*, estas se expresaron como la proporción de cada una tanto

por estado y como para cada quinquenio, la primera se obtuvo de datos de la Secretaría de Economía y la segunda de la información de los gobiernos recabado por INEGI.

Una característica para señalar es que se dividió el panel de datos en dos, tomando como referencia el promedio del PIBE entre 2000 y 2015, categorizando como entidades de menores ingresos a las entidades menores a la mediana de ingresos (\$110, 350 pesos anuales por individuo) y como entidades de mayores ingresos a aquellas que estén por arriba de la mediana. Esto pensando que la hipótesis del colesterol podría ser distinta de acuerdo con el nivel de ingresos.

Índice de desigualdad de oportunidades (IDO)

El índice IDO fue calculado por el método paramétrico *ex ante* de Ferreira y Gignoux (2011 a y b), el cual se puede resumir de la siguiente forma: Si se tiene una variable de resultado G , en este caso escolaridad en años, y un conjunto de cinco variables circunstancias C , dichas variables se relacionarán en la forma de la ecuación 1.

$$G = f(C) \quad (1)$$

Para resolver por mínimos cuadrados ordinarios (MCO), la ecuación 1 se convierte en un modelo de la siguiente forma:

$$G = \Psi C + \varepsilon \quad (2)$$

Donde, ΨC mide la dispersión encontrada de la desigualdad debida a las circunstancias y separa la desigualdad de esfuerzos contenida en los residuos del modelo. Por lo que el índice de desigualdad debido a circunstancias o índice de desigualdad de oportunidades relativo se calcula de la siguiente manera.

$$IDOr = \frac{I(\Psi C)}{I(G)} \quad (3)$$

Donde, $I(\Psi C)$ es la desigualdad de oportunidades absoluta y $I(G)$ es la desigualdad general.

Para facilitar el cálculo del IDO, Ferreira y Gignoux (op.cit.) utilizan el valor del ajuste del modelo de regresión 2, es decir el coeficiente de determinación múltiple R^2 , y demuestran que es igual al índice calculado por los tres pasos descriptos.

Índice de desigualdad de esfuerzos (IDE)

Para calcular el índice de desigualdad de esfuerzos, se consideró que la desigualdad general $I(G)$ está compuesta únicamente de desigualdad debida a circunstancias IDO y desigualdad debida a esfuerzos IDE, por lo que se puede expresar la siguiente igualdad.

$$I(G) = IDO + IDE \quad (4)$$

Entonces, si el índice $I(G)$ se toma como el 100 % de desigualdad, podemos expresar la ecuación 4 como:

$$IDE = 1 - IDOr \quad (5)$$

Así, con la ecuación 5 podemos obtener el porcentaje de desigualdad debida a los esfuerzos.

Análisis estadístico

El análisis estadístico se realizó con apoyo del software estadístico y de programación R, versión 4.0.3. Se ejecutaron modelos de regresión para datos panel: *Pooled OLS* (agrupada MCO), efectos fijos y efectos aleatorios, lo cual requirió de la librería “plm”. Como también se buscó encontrar un mejor modelo de regresión por medio de un algoritmo de selección de variables *Backward*, esto para el modelo *pooled*, se utilizó la librería “MASS”. Para probar los supuestos de los residuos en los modelos de regresión agrupados, se utilizaron las librerías “nortest” y “lmtest”. A continuación, describimos en que consiste cada método de regresión y las pruebas de hipótesis para seleccionar el mejor modelo.

Regresión pooled OLS o regresión agrupada

Este método permite suponer que los efectos por entidad son semejantes entre las mismas, de esta forma se planteó el siguiente modelo 6:

$$TC_{ij} = \beta_0 + \beta_1 \log(PIBE_{ij}) + \beta_2 \log(IDO_{ij}) + \beta_3 \log(IDE_{ij}) + \beta_4 \log(TD_{ij}) + \beta_5 \log(ESCOPRO_{ij}) + \beta_6 \log(INVPU_{ij}) + \beta_7 \log(IED_{ij}) + \varepsilon_{ij} \quad (6)$$

Donde, TC_{ij} agrega las dimensiones de la entidad i y del tiempo t , sin tomar en cuenta las particularidades de cada entidad. Lo que se espera obtener de este modelo son los parámetros β_i , que serán los mejores estimadores, lineales e insesgados (MELI). Además, se ejecutó un algoritmo de selección de variables *Backward* que eliminara las regresoras si no son necesarias, pero conservando el IDO y el IDE . El modelo propuesto por el algoritmo *Backward* fue:

$$TC_{ij} = \beta_0 + \beta_1 \log(PIBE_{ij}) + \beta_2 \log(IDO_{ij}) + \beta_3 \log(IDE_{ij}) + \beta_4 \log(TD_{ij}) + \beta_5 \log(INVPU_{ij}) + \varepsilon_{ij} \quad (7)$$

Este modelo de regresión agrupada debe cumplir los supuestos de los residuos, y sus resultados indicarán la falta de un método distinto para el modelo general.

Regresión con efectos fijos

Con este método, los efectos individuales son importantes para estimar el modelo general, pues tendrán un efecto diverso inmerso en la constante (Liviano y Pujol, 2013). El modelo para calcular los estimadores es el siguiente:

$$TC_{ij} = \beta^0 + \mu_i + \beta_1 \log(PIBE_{ij}) + \beta_2 \log(IDO_{ij}) + \beta_3 \log(IDE_{ij}) + \beta_4 \log(TD_{ij}) + \beta_5 \log(ESCOPRO_{ij}) + \beta_6 \log(INVPU_{ij}) + \beta_7 \log(IED_{ij}) + \varepsilon_{ij} \quad (8)$$

Donde $\beta_0 = \beta^0 + \mu_i$ donde β_0 contiene a β^0 como término común y a μ_i como efectos individuales. El parámetro μ_i , tal y como Liviano y Pujol (2013) observan, es un parámetro

distinto al término de error, esto permite que la condición de ortogonalidad $E(x_{ij}u_i) = 0$ se siga cumpliendo. Para calcular μ_i se puede usar lo que se conoce como *within groups (WG)* o *least squares dummy variables (LSDV)*, los cuales otorgan las siguientes características, al modelo general:

1. El estimador WG/ LSDV es consistente tanto para $E(x_{ij}u_i) = 0$ como para $E(x_{ij}u_i) \neq 0$, ya que las transformaciones internas eliminan los efectos individuales por entidad.
2. Si se cumple que $E(x_{ij}u_i) = 0$ el estimador WG/ LSDV será menos eficiente a medida que incrementa el número de entidades.
3. Si se cumple que todas las regresoras estén correlacionadas con el parámetro de efectos fijos, el estimado WG/ LSDV será eficiente.

Regresión con efectos aleatorios

El método de regresión con efectos aleatorios o *random effects generalized least squares (RE/GLS)* lo utilizaremos para elaborar inferencias de la población a partir de muestras, sin hacer observaciones individuales. Este método consiste en:

$$TC_{ij} = +\beta_1 \log(PIBE_{ij}) + \beta_2 \log(IDO_{ij}) + \beta_3 \log(IDE_{ij}) + \beta_4 \log(TD_{ij}) + \beta_5 \log(ESCOPRO_{ij}) + \beta_6 \log(INVPU_{ij}) + \beta_7 \log(IED_{ij}) + \gamma_i + \varepsilon_i \quad (9)$$

Donde el término ε_{ij} contiene a $\gamma_i + \varepsilon_i$, el primero como una perturbación aleatoria por individuo y la segunda como perturbación aleatoria del modelo en conjunto. γ_i va a cumplir que es constante en el tiempo, homocedástica, no autocorrelacionada y con media cero, esto es similar al término aleatorio del modelo general.

Pruebas estadísticas para los modelos

Para elegir el mejor modelo entre el modelo de regresión agrupado y el de regresión con efectos fijos, se elabora una prueba F de Fischer, esta nos indica si hay suficiente variación entre entidades (hipótesis alterna, el modelo de efectos fijos es mejor) comparado a si hay nula variación (hipótesis nula, el modelo agrupado es mejor). Si ya se ejecutó un modelo agrupado, también se puede verificar si hace falta realizar una regresión con efectos fijos, esto se puede probar mediante una prueba de multiplicadores de Lagrange, donde la hipótesis nula es que no hay suficiente variación entre entidades y la alterna que si la hay.

Resultados

Todas las entidades federativas

Se ejecutó el modelo 6 por medio de mínimos cuadrados ordinarios o Pooled OLS, los resultados pueden ser observados en la Tabla 1A. De las variables de interés, solo el IDE tiene un coeficiente significativo al 99% de confianza, y el signo contradice el resultado esperado y sugerido por la hipótesis del colesterol, la desigualdad de esfuerzos impacta de manera negativa al crecimiento económico, mientras que, la desigualdad de oportunidades lo hace de manera positiva, aunque el coeficiente no sea significativo.

Como el modelo mediante MCO no tiene un buen ajuste, R^2 de 0.218 y R ajustado aún más pequeño y además al verificar los supuestos de los residuos, no cumplen con normalidad, hay heterocedasticidad y autocorrelación, se buscó reducir el número de regresores haciendo uso del algoritmo de selección de variables Backward, el cual eliminó de la regresión pooled a la variable ESCOPRO y a la IED, quedando solo cinco variables independientes.

El modelo resultante se observa en la Tabla 1B. Si bien el ajuste general no mejora mucho, la significancia de los regresores aumenta, la variable IDO sigue siendo positiva, aunque sin ser significativa. El IDE es altamente significativo, aunque el coeficiente en términos del logaritmo es muy pequeño, por cada unidad de incremento en el IDE, el crecimiento se reduce en un -0.047%. Este modelo parece prometedor, aunque sigue sin cumplir los supuestos de los residuos.

Para tratar de superar los problemas en los supuestos, se ejecutó el modelo 8 y 9 mediante efectos fijos y efectos aleatorios (Tabla 1C y D). Ambos modelos sugieren al IDE con signo negativo y significativo con al menos un alfa de 0.05. Mientras que, el IDO no es significativo en ningún caso, aunque es positivo, contradiciendo la hipótesis del colesterol.

Se comparó entre modelos mediante una prueba F de Fischer, para identificar si los efectos individuales son importantes ($F_{31, 89} = 1.19$, $p = 0.26$), si esto se cumple entonces el modelo de efectos fijos es mejor al pooled, por lo que la hipótesis nula de efectos insignificantes no se rechaza y el modelo de regresión Backward es mejor que el de efectos fijos.

Para comprobar si el modelo de efecto aleatorios era necesario o no, se aplicó una prueba de multiplicadores de Lagrange (normal= -0.4318, $p = 0.67$), por lo que no se rechaza la H_0 de efectos insignificantes, y así el modelo pooled sigue siendo el mejor.

Algo a destacar en nuestros modelos, es que el PIBE siempre fue negativo y significativo con alfa de 0.05, a excepción de la regresión con efectos fijos, esto sugiere convergencia entre entidades, tal y como la literatura del crecimiento económico lo explica. Otras de las variables a resaltar son: la tasa de desempleo, siempre positiva y algunas veces significativa, y la inversión pública siempre negativa y a veces significativa.

Entidades de menores ingresos

Se analizó la hipótesis por nivel de ingresos para verificar si la hipótesis del colesterol se comportaba de manera distinta de acuerdo distintos niveles de ingreso. El modelo de regresión agrupado ampliado (Tabla 2A), vuelve a sugerir coeficientes contrarios a lo esperado por la hipótesis, aunque solo es significativo el parámetro del IDE al 95 % de confianza. El ajuste sigue siendo muy bajo, aunque sí se cumplen los supuestos de normalidad y homocedasticidad de los residuos.

Tabla 1. Regresiones para todas las entidades.

COEFICIENTES								
Columna	A		B		C		D	
	Pooled OLS		Pooled OLS Backward		Efectos fijos		Efectos aleatorios	
	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor
Intercepto	16.85	0.004 **	18.75	3.7e-05 ***			16.85	0.004 **
Log (PIBE)	-0.75	0.017 *	-0.65	0.0127 *	-2.25	0.189	-0.75	0.016 *
Log (IDO)	1.22	0.107	1.04	0.147	2.14	0.133	1.22	0.104
Log (IDE)	-4.14	0.002 **	-4.71	4.2e-05 ***	-4.36	0.026 *	-4.14	0.002 **
Log (TD)	0.50	0.129	0.60	0.008 **	1.19	0.009 **	0.50	0.126
Log (ESCOPRO)	0.62	0.759			-4.81	0.141	0.62	0.759
Log (INVPU)	-0.44	0.0385 *	-0.37	0.0357 *	-0.17	0.614	-0.44	0.037 *
Log (IED)	0.14	0.478			0.06	0.895	0.14	0.477
Códigos de significancia.: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘ ’ 0.1 ‘ ’ 1								
	R ² múltiple: 0.218		R ² múltiple: 0.212		R ² múltiple: 0.164		R ² múltiple: 0.218	
	R ² ajustado: 0.171		R ² ajustado: 0.180		R ² ajustado: -0.192		R ² ajustado: 0.171	
	F: p-valor: 9.6e-05		F: p-valor: 1.9e-05		F: p-valor: 0.021		Chi ² : p-valor: 2.3e-05	

Notas: Pooled OLS se refiere a las regresiones agrupadas o de MCO, Pooled OLS Backward se refiere al modelo de regresión que el algoritmo Backward nos sugiere. La fila columna se utilizó para mostrar el nombre de la columna que corresponde a cada tipo de regresión.

Al ejecutar la regresión propuesta por el algoritmo Backward (Tabla 2B) se vuelven a encontrar los mismos resultados que la regresión ampliada, solo que ahora el nivel de significancia de los coeficientes se pierde y el ajuste del modelo resulta no tan adecuado, pues aún es baja la R^2 de 0.161 y un R^2 ajustado de 0.089, aunque los supuestos de los residuos se siguen cumpliendo, esto quiere decir, que el modelo sugerido por el algoritmo Backward para toda la muestra no es el mejor para las entidades de menores ingresos.

En la Tabla 2C, se presentan los resultados de la regresión con efectos fijos, los coeficientes siguen contradiciendo la hipótesis del colesterol para las entidades de menores ingresos, con los mismos niveles de significancia que con MCO. Solo que ahora el modelo aumenta su bondad de ajuste y otras variables como la IED ganan significancia, impactando de manera positiva al crecimiento y el PIBE, mostrando un signo negativo altamente significativo, que demuestra la convergencia de entidades de este nivel de ingresos.

Los resultados del modelo de efectos aleatorios siguen los mismos resultados de los modelos anteriores, solo que ahora el PIBE pierde significancia y la INVPU gana. La prueba F que compara el modelo reducido pooled con el de efectos fijos ($F_{17, 41} = 1.96$, $p = 0.04$), expresa que en este caso el modelo de efectos fijos es mejor, pues los efectos individuales sí son distintos e importantes.

La prueba de multiplicadores de Lagrange, que verifica si el modelo de efectos aleatorios es mejor que el de efectos fijos (normal = -1.74, $p = 0.96$), no rechaza la H_0 , de que los efectos aleatorios son insignificantes, y el modelo de efectos fijos es el mejor.

Entidades de mayores ingresos

Las regresiones para entidades de mayores ingresos están en la Tabla 3, en la regresión ampliada (A) los coeficientes de IDO e IDE son negativos, lo que supone que el efecto de la desigualdad es perjudicial en las entidades de mayores ingresos, aunque el único coeficiente significativo sigue siendo el IDE. El ajuste del modelo solo cubre un 30 % y los supuestos del modelo no se cumplen.

Si se observa la regresión sugerida por el algoritmo Backward (Tabla 3B), el ajuste no mejora, pero los coeficientes y su significancia cambian, los coeficientes contradicen la hipótesis del colesterol, aunque solo el IDE es significativo con un alfa de 0.05. Por cada aumento en una unidad del índice, el crecimiento disminuye en un 0.05 %. Este modelo tampoco cumple con los supuestos de los residuos, por lo que no es el mejor modelo para el problema.

Se ejecutó el modelo de efectos fijos (Tabla 3C) y a excepción de la tasa de desempleo que es significativa al 90 % de confianza, todos los demás coeficientes no son significativos, de hecho, el estadístico F, expresa que el modelo no tiene coeficientes distintos de cero. Este tampoco es el mejor modelo para representar el problema. Si se compara el modelo de efectos fijos con el modelo de regresión pooled ampliado ($F_{15, 41} = 1.04$, $p = 0.44$), que indicó que el mejor modelo es el de regresión agrupada ampliada.

Tabla 2. Regresiones para entidades de menores ingresos.

COEFICIENTES								
Columna	A		B		C		D	
	Pooled OLS		Pooled OLS Backward		Efectos fijos		Efectos aleatorios	
	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor
Intercepto	19.44	0.117	6.34	0.584			19.44	0.112
Log (PIBE)	-0.73	0.501	0.12	0.890	-9.73	0.000 ***	-0.73	0.499
Log (IDO)	0.70	0.450	1.14	0.227	0.31	0.014 *	0.70	0.447
Log (IDE)	-3.08	0.028 *	-3.38	0.020 *	-4.36	0.014 *	-3.08	0.024 *
Log (TD)	0.41	0.327	0.38	0.192	0.13	0.817	0.41	0.323
Log (ESCOPRO)	-1.53	0.482			2.12	0.526	-1.53	0.480
Log (INVPU)	-0.80	0.002 **	-0.31	0.090.	-0.58	0.110	-0.80	0.001 **
Log (IED)	0.65	0.012 *			1.25	0.010 **	0.65	0.010 **
Códigos de significancia.: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘ ’ 0.1 ‘ ’ 1								
	R ² múltiple: 0.260		R ² múltiple: 0.161		R ² múltiple: 0.384		R ² múltiple: 0.260	
	R ² ajustado: 0.168		R ² ajustado: 0.089		R ² ajustado:0.054		R ² ajustado: 0.168	
	F: p-valor: 0.014		F: p-valor: 0.064		F: p-valor: 0.004		Chi ² : p-valor: 0.006	

Notas: Pooled OLS se refiere a las regresiones agrupadas o de MCO, Pooled OLS Backward se refiere al modelo de regresión que el algoritmo Backward nos sugiere. La fila columna se utilizó para mostrar el nombre de la columna que corresponde a cada tipo de regresión.

Los resultados que se obtuvieron con efectos aleatorios (Tabla 3D), indican que este método arroja un ajuste similar a la regresión agrupada, y los coeficientes estimados son iguales, solo se diferencian en que tiene coeficientes con mayor significancia. Y se aprecia que los componentes de la desigualdad son ambos negativos para el crecimiento económico.

Al realizar la prueba de multiplicadores de Lagrange para verificar si el modelo de efectos aleatorios es necesario (normal = -0.58428, $p = 0.72$), no se rechaza la hipótesis nula de que los efectos aleatorios son insignificantes, por lo que el modelo de regresión agrupada, una vez más, es el mejor modelo para el problema, aún con sus detalles en los supuestos, lo cual se debe manejar con cuidado.

Conclusiones

Discusión

El objetivo de esta investigación fue analizar la hipótesis del colesterol en México, está por medio de descomponer la desigualdad general expresada en años de escolaridad (educación), en desigualdad de oportunidades y desigualdad de esfuerzos. Se calculó una tasa de crecimiento del PIB y variables de control para los quinquenios del 2000-2019. Se ejecutaron modelos de regresión de datos panel y se eligió el modelo más robusto.

Los resultados del análisis de la hipótesis del colesterol en México son en su mayoría contrarios a lo planteado por Marrero y Rodríguez (2013 y 2019), quienes plantean que la desigualdad de oportunidades es el componente negativo para el crecimiento económico, mientras que la desigualdad de esfuerzos es el componente positivo. En esta investigación, a nivel nacional el IDE es negativo y significativo, mientras que el IDO es positivo y no significativo, un resultado similar al de Ferreira et al., (2018), quienes no encuentran evidencia de la hipótesis del colesterol en el ámbito internacional.

Los estados mexicanos con menor nivel de ingreso contradicen la hipótesis del colesterol con un nivel de significancia aceptable para el IDE y un coeficiente no significativo para el IDO; para los estados con mayores ingresos, ambos tipos de desigualdad son perjudiciales y nuevamente solo el IDE es significativo a un alto nivel de confianza.

De acuerdo con Mohammadi et al., (2018) los resultados de acuerdo con el nivel de ingresos si varían, manifestándose en un efecto negativo en el largo plazo en países desarrollados y un efecto negativo en el corto plazo para países en desarrollo, como nuestra investigación captura efectos de largo plazo, el efecto de corto plazo esperado no está revelado aún.

En Parra (2019) se investigó el efecto de un índice proxy de desigualdad de oportunidades (movilidad social intergeneracional) en el crecimiento económico de México del periodo 2010-2016, y aunque encuentran que la desigualdad de oportunidades es perjudicial para el crecimiento se incurre en un supuesto no probado, el hecho de que la movilidad social sea un proxy inverso de la desigualdad de oportunidades, por lo que los resultados se deben tomar con precaución.

Tabla 3. Regresiones para entidades de mayores ingresos

COEFICIENTES								
Columna	A		B		C		D	
	Pooled OLS		Pooled OLS Backward		Efectos fijos		Efectos aleatorios	
	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor	Estimado	P-valor
Intercepto	41.55	0.017 *	26.34	0.000 ***			41.55	0.014 *
PIBE	-1.22	0.012 *	-1.24	0.010 **	-0.83	0.738	-1.22	0.009 **
IDO	-0.37	0.783	0.83	0.437	3.84	0.228	-0.37	0.781
IDE	-7.62	0.010 **	-4.64	0.014 *	-3.17	0.541	-7.62	0.007 **
TD	0.85	0.147	0.60	0.086.	1.48	0.052.	0.85	0.140
ESCOPRO	-2.46	0.624			-7.35	0.287	-2.46	0.622
INVPU	-0.16	0.670	-0.39	0.251	0.06	0.909	-0.16	0.668
IED	-0.54	0.140			-0.66	0.486	-0.54	0.134
Códigos de signif.: 0 '****' 0.001 '**'								
0.01 '*'								
0.05 ' ' 0.1 ' ' 1								
	R ² múltiple: 0.307		R ² múltiple: 0.277		R ² múltiple: 0.208		R ² múltiple: 0.308	
	R ² ajustado: 0.220		R ² ajustado: 0.215		R ² ajustado: -0.217		R ² ajustado: 0.221	
	F: p-valor: 0.003		F: p-valor: 0.002		F: p-valor: 0.181		Chi ² : p-valor: 0.001	

Notas: Pooled OLS se refiere a las regresiones agrupadas o de MCO, Pooled OLS Backward se refiere al modelo de regresión que el algoritmo Backward nos sugiere. La fila columna se utilizó para mostrar el nombre de la columna que corresponde a cada tipo de regresión.

A su vez, la investigación de Parra se podría considerar un análisis para el corto plazo, de allí que se asemeje a lo encontrado por Mahammadi et al., (2018). Para esta investigación se calculó índices diseñados para medir propiamente desigualdad de oportunidades y se consideró un periodo más grande que permite tener una visión de largo plazo, de allí que los resultados de esta investigación si sean un nuevo hallazgo de los componentes de la desigualdad en un país en desarrollo.

Mas aun, como menciona Ferreira et al., (2011b), el IDO calculado está subestimado, mientras que el IDE esta sobreestimado, al incluir un conjunto no exhaustivo de variables circunstancias, lo que indica que quizá los modelos de regresión no captaron el efecto total. Y como prueba, Carranza (2020) calculó el IDO en el límite superior (contrario al método de Ferreira y Gignoux, 2011a y b) y encuentra que los efectos del IDO en el crecimiento son altamente significativos y positivos. De esta manera, los índices pueden estar mostrando un efecto parcial.

Otra condición para resaltar es que los modelos de regresión no siempre dicen lo mismo (Ferreira et al., 2018), por lo que los resultados fueron correctos desde el modelo más robusto seleccionado.

El hallazgo del IDE es nuevo en la literatura, no se había analizado el efecto de este en el crecimiento de un país en desarrollo. Investigaciones futuras, con índices IDO e IDE más cercanos a los verdaderos, y haciendo uso de modelos de regresión homogéneos a estos y de previos autores, podrían probar la hipótesis en los países en desarrollo.

Conclusión

La hipótesis del colesterol no se cumple en México para el periodo 2000-2019, pues el índice de desigualdad de esfuerzos fue negativo para el crecimiento, mientras que el de desigualdad de oportunidades fue positivo, pero no estadísticamente significativo. Cuando se desagrega el país de acuerdo con el nivel de ingresos, la hipótesis sigue sin demostrarse, pero las desigualdades son negativas para el crecimiento de las entidades de mayores ingresos, mientras que en las entidades de menores ingresos se mantienen los resultados internacionales.

Referencias

- Bleynat, I., Challú, A. and Segal, P. (2020). Inequality, living standars, and growth: two centuries of economic development in Mexico. *The Economic Historic Review*. 0 (0): 1-27.
- Carranza, R. (2020). Inequality of Outcomes, Inequality of Opportunity, and Economic Growth. *Society for the Study of Economic Inequality, Working Paper Series* 534.
- Ferreira, F. and Gignoux, J. (2011a). The measurement of educational inequality: Achievement and opportunity. *World Bank Economic Review*, 5873: 1-41.
- Ferreira, F. and Gignoux, J. (2011b). The measurement of inequality of opportunity: theory and an application to Latin America. *The Review of Income and Wealth*. 57(4): 622-657.

- Ferreira, F., Lakner, C., Lugo, M. and Özler, B. (2018). Inequality of opportunity and economic growth: how much can cross-country regressions really tell us? *Review of Income and Wealth*. 64(4), 800-827.
- Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45 (1): 1-28.
- Liviano, D. y Pujol, M. (2013). *Modelos econométricos avanzados en R*. España, Universidad Abierta de Cataluña.
- Marrero, G. and Rodríguez, J. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, 104, 107-122.
- Marrero, G. and Rodríguez, J. (2019). Inequality and growth: The cholesterol hypothesis. *Working paper Society for the study of economic inequality*, 501.
- Marrero, G., Rodríguez, J. and Van der Weide, R. (2016). Unequal opportunity, unequal growth. *Bank Policy Working Paper*, 7853.
- Mohammadi, H. Hamid, Moarefi, A. and Aghasafari, H. (2018). Opportunity inequality and its effects on economic growth and per capita GDP. *Iranian Journal of Economic Studies*. 6(2): 195-217.
- Monroy-Gómez-Franco, L. and Corak, M. (2019). A Land of Unequal Chances: Social Mobility and Inequality of Opportunity Across Mexican Regions. *Working Paper*. Presentado en *Society for the Study of Economic Inequality*. ECINEQ.
- Moreno, H. (2017). Determinantes de la movilidad educativa intergeneracional y políticas públicas para promoverla. *Documento de trabajo 7, Centro de Estudios Espinoza Yglesias (CEEY)*.
- Parra, S. (2019). *Desigualdad de oportunidades y crecimiento económico en México*. [Tesina de maestría del Centro de Investigación y Docencia Económicas].
- Roemer, J. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.
- Solís, P. (2010). La desigualdad de oportunidades y las brechas escolares. En *Los grandes problemas de México, educación*. Coord. A. Arnaut y S. Giorguli. VII: 599-622.
- Voitchosky, S. (2009). Inequality and growth, in B. Nolan, W. Salverda and T. Smeeding (Eds.). *The Oxford Handbook of Economic Inequality*. Oxford University Press.

Anexos
Anexo 1. Variables utilizadas para índice de riqueza.

BIEN, SERVICIO, HACINAMIENTO	AÑO			
	2000	2005	2010	2015
Pisos	X	X	X	X
Disponibilidad de agua	X	X	X	X
Drenaje	X	X	X	X
Electricidad	X	X	X	X
Tenencia de la vivienda	X	X	X	X
Radio	X		X	X
Televisión	X	X	X	X
Videocasetera	X			
Licuada	X			
Estufa			X	
Refrigerador	X	X	X	X
Lavadora	X	X	X	X
Teléfono	X		X	X
Celular			X	X
Internet			X	X
Boiler	X		X	X
Auto propio	X		X	X
Computadora	X	X	X	X
Horno				X
Servicio de TV de paga				X
Hacinamiento	X	X	X	X

Nota. La X significa que el ítem se encuentra disponible en la base de datos de los Censos, y encuestas intercensales ocupados en esta investigación.

Anexo 2. Porcentaje de inercia explicada por los factores del ACM.

PORCENTAJE DE INERCIA EXPLICADA POR LOS PRIMEROS FACTORES DEL ACM	Año			
	2000	2005	2010	2015
Dimensión 1	94.93	92.41	91.21	91.55
Dimensión 2	1.06	0.28	2.42	2.54
Dimensión 3	0	0	0	0.01

Nota: La dimensión 1 que se utilizó para el índice de riqueza siempre capturó más del 90% de varianza de los datos de Tabla 1 de este anexo.

Algoritmos clasificatorios para estudiar la relación entre rendimiento escolar y uso del teléfono inteligente en estudiantes universitarios de álgebra lineal

Irving Dennier Alatorre Barillas¹, José Juan Muñoz León², Cecilia Cruz López³, Ángel Juan Sánchez García⁴.

¹Facultad de Estadística e Informática. Universidad Veracruzana. dennierbarillas@gmail.com

²Facultad de Estadística e Informática. Universidad Veracruzana. juanmunoz@uv.mx

³Facultad de Estadística e Informática. Universidad Veracruzana. cecacruz@uv.mx

⁴Facultad de Estadística e Informática. Universidad Veracruzana. angesanchez@uv.mx

Resumen

Es incuestionable que el uso correcto de la tecnología en la educación tiene beneficio en distintas áreas del conocimiento. En educación matemática universitaria algunos de estos beneficios son tangibles y otros no tan claros. Son muchos los estudios que reportan relación entre el uso de algunas aplicaciones en el Teléfono Inteligente (TI) y los aprendizajes esperados. No obstante, parece que esta relación no siempre es positiva tal y como se reporta en este estudio. Se presenta un análisis del impacto en el rendimiento escolar que causa el uso del teléfono inteligente en actividades académicas de estudiantes de un curso universitario de Álgebra Lineal, generado bajo análisis de componentes principales, correlación lineal y regresión estadística de distintas variables analizadas que involucran el uso del TI. El resultado de ello muestra que no existe relación con el uso de este dispositivo y el aprovechamiento académico en este contexto. Además, se presentan perfiles predictores de los estudiantes, generados bajo dos técnicas de clasificación: clúster y árboles de decisión. El profesor y el tipo de escuela de procedencia son las características predominantes en ambos perfiles.

Palabras clave: Análisis multivariante, Análisis clúster, Árboles de decisión.

Introducción

La educación es un proceso a través del cual los seres humanos incorporan conocimientos nuevos, ya sea habilidades, creencias, valores o hábitos, los cuales son adquiridos por transmisión directa de otros individuos quienes son los responsables de transmitir el conocimiento, comúnmente conocidos como docentes, utilizando para ello distintos métodos, la educación es la formación práctica y metodológica que se le da a una persona en vías de su desarrollo y crecimiento. En este proceso se le provee de herramientas y conocimientos esenciales al individuo para que pueda ser capaz de ponerlos en práctica en la vida cotidiana. Para los fines de la presente investigación, se consideran dos variantes de este proceso llamado educación, una en la que el docente opta por transmitir la información de acuerdo al método tradicional, la cual de acuerdo a Freedman, Pisani, Purves, y Adhikari (2010) la educación tradicional es “aquella en donde las enseñanzas son repetidas a los alumnos de generación en generación sin promoverles a utilizar las herramientas a su alcance”, y otra variante en donde la tecnología figura como una de los ejes principales, donde al alumno se le invita a la exploración de herramientas digitales las cuales representan medios de

aprendizajes para fines específicos, los cuales facilitan la incorporación de nuevos conocimientos optimizando los esfuerzos, conocida como educación tecnológica.

Cada uno de los últimos siglos, fue testigo del nacimiento de nuevas tecnologías, las cuales fueron dominantes en su momento. En el siglo XVIII, surgieron los grandes inventos mecánicos, mismos que jugaron un papel importante en la revolución industrial. En el siglo XIX fue el momento de la máquina de vapor, durante el siglo XX fue la obtención, el procesamiento y la distribución de la información, en esas épocas pudimos observar la implementación de redes mundiales de telefonía, la invención de la radio y la televisión y el crecimiento sin precedentes de la industria de la computación.

Como resultado del rápido progreso tecnológico, estas áreas están convergiendo de una manera acelerada y las diferencias entre la recolección, transportación, almacenamiento y procesamiento de datos está desapareciendo rápidamente.

Junto con la evolución de las redes de comunicación, al crearse la famosa red de redes, mejor conocida como Internet ha supuesto un cambio de paradigma en la vida del hombre al permitirle estar conectado en tiempo real con otros usuarios y con una serie de procesos institucionales, viniendo también a cambiar los procesos pedagógicos debido a las potencialidades que poseen en el plano de la enseñanza y la educación.

Es por ello que resulta importante hoy en día que tanto los docentes como los estudiantes conozcan algunas de las herramientas virtuales que poseen ciertas utilidades y fines pedagógicos que faciliten el proceso enseñanza-aprendizaje; estas herramientas se pueden encontrar en ambientes virtuales aptos tanto para computadoras como para TI's, y uno de los retos en la actualidad es averiguar si la incorporación de estas tecnologías refleja una relación positiva en su aplicación en el campo de la educación en general.

Este trabajo centra su atención en la integración del TI aplicado a las actividades académicas, por ser uno de los dispositivos más utilizados por estudiantes universitarios, el cual llevan consigo en todo momento por sus características y beneficios que este ofrece a sus actividades diarias, tal es el caso de los alumnos de la Licenciatura en Estadística de la Universidad Veracruzana, alumnos con los cuales se desarrolló el presente trabajo de investigación, el cual busca establecer si existe relación entre el uso del teléfono inteligente en actividades académicas con el rendimiento escolar.

Se aborda una problemática actual en el proceso de enseñanza-aprendizaje no solo en niveles universitarios, sino en todos los niveles del sistema de educación en México, en el área de estudios de las matemáticas donde habitualmente existen dificultades por parte de los alumnos para su entendimiento y aplicación, de resultar positiva la relación entre el uso del teléfono inteligente, será un factor importante para poder propiciar un cambio en la manera tradicional en la que los docentes de nivel superior realizan su práctica docente e incorporen estas nuevas tecnologías en el proceso de enseñanza aprendizaje de las matemáticas a nivel superior.

Parte importante del quehacer docente es conocer las características cualitativas de los alumnos que recibe para impartirles un nuevo curso, algunas de las cuales suelen ser un factor importante en su desarrollo académico, en esta investigación, se construyen perfiles predictores de los alumnos, para poder ubicarlos en tres niveles posibles de desempeño, destacado, regular o malo, de acuerdo a la clasificación de las características cualitativas con las que ingresan al curso de álgebra lineal, tomando en cuenta también aquellas que tienen que ver con el uso del TI, esto con el objetivo de darle a los docentes de la asignatura herramientas útiles para poder modificar sus estrategias de enseñanza en base a estos perfiles que son propios de cada estudiante. De acuerdo con Barrios (2020) no existe una relación directa en los estudiantes universitarios de álgebra lineal 1 de la Universidad Veracruzana, entre el uso del teléfono inteligente y su aprovechamiento académico, situación que se ve sustentada según con Arias, Lugo y Morales (2016) quienes aseguran que el uso que se le dan a estos dispositivos no se está transversalizando en el ámbito educativo y en gran medida se debe a que los docentes no tienen una metodología para vincular los dispositivos con la clase y así poder utilizarlos de manera apropiada como una herramienta didáctica y de aprendizaje. En este mismo sentido, de acuerdo con Oliva (2014) se puede considerar al celular como el principal enemigo actual del aprendizaje, pues su inadecuado uso interfiere en el avance escolar, ya que los estudiantes prefieren dar prioridad a la tecnología de vanguardia que tienen en su teléfono celular que a lo que se explica en clases.

Sin embargo, en contraposición a lo anteriormente expuesto, Cisneros y Robles (2017) establecen que el uso del TI se ha evolucionado pasando de ser un dispositivo solo para hacer llamadas y buscar el estatus entre sus iguales, a una herramienta para comunicarse, buscar información y aprender; en el mismo sentido, de acuerdo con Basantes, Naranjo, Gallegos y Benítez (2017) el uso de los dispositivos móviles potencia la interacción dentro y fuera del aula estimulando la exploración, la comunicación, el pensamiento crítico y reflexivo, situación que debe de reflejarse de manera directa en el aprovechamiento académico de los estudiantes, conforme a la revisión teórica y exploratoria sobre el celular como mediador en el proceso de enseñanza aprendizaje se puede concluir, en una primera aproximación, que este dispositivo juega y jugará un papel importante como mediador educativo, en virtud de que a través del “aprendizaje móvil” centrado en el estudiante, se privilegia la movilidad y el contexto de aprendizaje. De acuerdo con Figueroa (2016) de estar contemplado el uso del TI como parte de la planificación curricular de la universidad, las competencias adquiridas por parte de los estudiantes serían mucho mayores.

Para estudiar la relación entre el uso del TI y el rendimiento escolar en estudiantes universitarios de álgebra lineal se presentan resultados de análisis clásicos en términos de relación causa-efecto, así como los derivados de la aplicación de algoritmos clasificatorios. Las técnicas clásicas de estadística se han constituido a través de los años en una herramienta útil para la toma de decisiones en diferentes áreas del conocimiento, en esta investigación se aplicarán técnicas estadísticas para generar perfiles y clasificar a estudiantes universitarios en virtud de su uso del TI y del rendimiento escolar en la experiencia educativa ya referida, buscando asignar una o varias categorías a la población objetivo (Tan, Steinbach, Kumar, 2006) para predecir su aprovechamiento en un curso en donde suele presentarse un alto grado de reprobación y un bajo rendimiento académico.

Para poder encontrar patrones significativos o similitudes en algún conjunto de datos de una base de información, existen herramientas estadísticas que conocemos como algoritmos clasificadores, los cuales son propios de la estadística clásica, pero no solo existen esas herramientas, también se cuenta con las herramientas propias que provienen de la inteligencia artificial, en este sentido se reportan los hallazgos encontrados al aplicar dos algoritmos clasificatorios para poder generar los perfiles predictivos de los alumnos que los ubiquen en un posible nivel de aprovechamiento. De estos dos algoritmos, uno de ellos es propio de la estadística clásica, conocido como análisis clúster y otro propio de la Inteligencia Artificial, conocido como árboles de decisión, los cuales serán comparados para establecer ventajas y desventajas al realizar esta tarea.

De acuerdo con Solarte y Ocampo (2009) el análisis clúster permite agrupar variables, en función de la similitud existente entre ellos. Usamos los conglomerados jerárquicos. Los datos dentro de cada conglomerado son similares entre sí (alta homogeneidad interna) y diferentes a los objetos de los otros conglomerados (alta heterogeneidad externa) y los árboles de decisión son una técnica de predicción que se emplea en el campo de inteligencia artificial, donde a partir de una base de datos se construyen diagramas de construcción lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren en forma repetitiva para solución de un problema.

Resulta común confundir a los algoritmos que nos brinda la estadística clásica con los de la inteligencia artificial, para muchos estudiosos del tema, los algoritmos de la estadística clásica son los antecesores de los algoritmos de la inteligencia artificial, aunque ambos tienen el mismo objetivo, que es el agrupamiento de los datos en base a características o perfiles coincidentes, existen diferencias marcadas entre ambas corrientes de técnicas clasificatorias. De acuerdo con Beltrán (2016), la diferencia entre ambas consiste en que las técnicas de la Inteligencia Artificial construyen el modelo de manera automática mientras que las técnicas estadísticas clásicas necesitan ser manejadas y orientadas por un estadístico profesional. Las técnicas de la inteligencia artificial permiten ganar tanto en su ejecución, como en manejabilidad e incluso en tiempo de trabajo, la posibilidad de realizar uno mismo sus propios modelos sin necesidad de contratar ni ponerse de acuerdo con un estadístico, proporciona gran libertad a los usuarios profesionales.

Metodología

La población sujeta a estudio son estudiantes de primer periodo inscritos en el curso de Álgebra Lineal Aplicada a la Estadística, ofertada por la Licenciatura en Estadística de la Universidad Veracruzana. Se construyeron y aplicaron instrumentos para recopilar información de cada uno de los estudiantes sobre el uso del TI incluyendo dimensiones de uso para la comunicación, la salud, entretenimiento, deportes y educación.

Se trabajó con dos grupos de alumnos, un grupo experimental y otro de control, el grupo experimental está compuesto por 67 de alumnos divididos en dos horarios diferentes, es en este grupo en donde se promovió el uso del TI para la realización de las actividades académicas. El grupo de control está compuesto por 33 alumnos en un solo horario, es en este grupo donde no se promovió el uso del TI.

Del análisis estadístico

Para determinar la relación del uso del teléfono inteligente con el aprovechamiento académico en matemáticas se realizaron dos indicadores del uso del teléfono: uno en donde se plasmaron las 6 dimensiones ya mencionadas, las cuales incluyen el uso del teléfono inteligente en las actividades matemáticas, mientras que el segundo es el uso general del TI sin incluir el uso que le dan para matemáticas.

Por medio de estos indicadores se realizó el análisis de correlación para ver si las variables de interés estaban o no relacionadas, utilizando el coeficiente de correlación de Pearson o R de Pearson, el cual está dentro del intervalo cerrado $[-1,1]$ donde 1 establece una relación fuerte y positiva, -1 una relación fuerte pero negativa, y cero quiere decir que no existe relación lineal entre las variables.

Para obtener los perfiles predictores se aplicaron dos algoritmos clasificatorios para poder establecer las principales características dominantes de los alumnos ubicados en cada uno de los tres niveles de desempeño de acuerdo al promedio final alcanzado en el curso de álgebra lineal, a cada grupo perteneciente a cierto nivel de aprovechamiento se le aplicó un análisis clúster (algoritmo clasificatorio de la estadística clásica) bajo la técnica del vecino más cercano ya que se están buscando similitudes entre los individuos de la base de datos y un análisis clasificatorio bajo la técnica de árboles de decisión (algoritmo perteneciente a la Inteligencia Artificial).

Los análisis preliminares y definitivos se realizaron en el software estadístico SPSS *Statistics* (versión 23) y R-Project (versión 3.5.1), respectivamente.

Resultados

De la relación del uso del teléfono inteligente con el aprovechamiento académico

En primera instancia se obtiene el coeficiente de correlación de Pearson y el indicador de uso general del teléfono inteligente, el cual como ya se mencionó anteriormente un resultado de 0 indica relación nula entre las variables analizadas y un resultado ± 1 indica una relación fuerte ya sea positiva o negativa. En el siguiente gráfico se muestra la matriz de correlación entre el promedio obtenido de los estudiantes y el uso del teléfono inteligente.

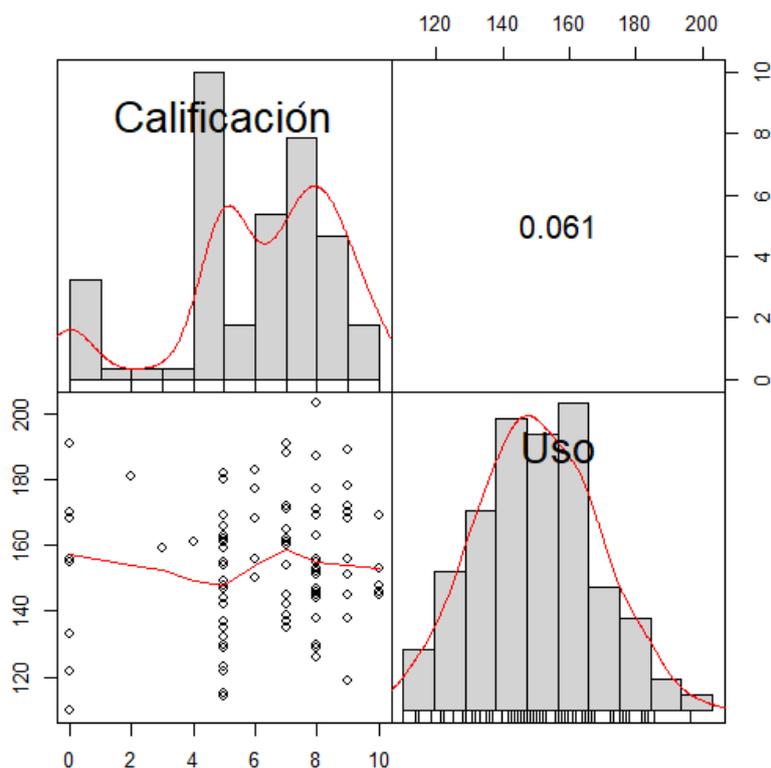


Figura 1. Matriz de correlación entre la calificación y el uso del teléfono inteligente.

En la Figura 1 se puede observar que el coeficiente de correlación entre el uso del teléfono inteligente y la calificación final obtenida por los alumnos es de 0.061, lo cual indica que existe relación nula entre estas dos variables.

Al observar que no es posible establecer un nivel de correlación entre el uso del teléfono inteligente y el aprovechamiento académico, se realiza el mismo análisis anterior, pero ahora aislando cada una de las dimensiones que involucran el uso del teléfono inteligente, en la tabla 1 se puede observar los distintos coeficientes de correlación obtenidos entre cada dimensión del uso del teléfono inteligente y el nivel de aprovechamiento obtenido por los alumnos.

Tabla 1. Coeficiente de correlación por dimensión.

Variable	Relación
Comunicación - Calificación	-0.0405
Entretenimiento - Calificación	0.143
Deportes - Calificación	0.0139
Salud - Calificación	0.0471
Educativas – Calificación	0.048
Matematicas - Calificación	0.055

Al analizar la Tabla 1 es posible observar que no existe en ninguna de las dimensiones un nivel de correlación significativo ya que todos los índices encontrados son cercanos a 0 lo que indica que no existe relación entre las variables y el aprovechamiento académico expresado a partir de la variable calificación. Dicha independencia entre el rendimiento

académico y las variables de uso del teléfono inteligente se ve reflejado tanto agrupando todas las dimensiones como evaluando el impacto en el rendimiento escolar de manera independiente.

Al demostrar que no existe relación entre las variables, se tiene la certeza de que no se presentará un problema de colinealidad al estimar un modelo de regresión; se estima el modelo para dar un mayor sustento a los resultados encontrados en el análisis de correlación, para llevar este análisis de regresión se emplean las variables mostradas en la Tabla 2:

Tabla 2. Variables empleadas en el modelo de regresión.

Variable	Relación
Variable dependiente (Y)	Calificación
Variable Independiente #1 (X_1)	Sexo
Variable Independiente #2 (X_2)	Tratamiento
Variable Independiente #3 (X_3)	Uso
Variable Independiente #4 (X_4)	MI
Variable Independiente #5 (X_5)	MF

Con el análisis de regresión se pretendió explicar que la calificación depende en gran manera al tipo de tratamiento que pertenece (grupo experimental o grupo de control), al sexo del individuo, uso que le dan al TI en actividades generales (Uso), el uso matemático que traían al inicio del periodo (MI) y el uso matemático que adquirieron al final del periodo (MF).

En un primer momento se realizaron los diagramas de dispersión entre cada una de las variables independientes (Uso, MI y MF) con la variable dependiente (Calificación) con el objetivo de determinar si existe algún tipo de relación lineal o no lineal entre ellas.

En la Figura 2 se pueden observar los diagramas de dispersión en donde se demuestra que las variables independientes no están asociadas bajo ninguna relación lineal o no lineal con la variable dependiente, en esta figura está representada toda la base de datos (el grupo de control y el experimental); esto da sustento a los resultados del análisis de correlación realizado anteriormente.

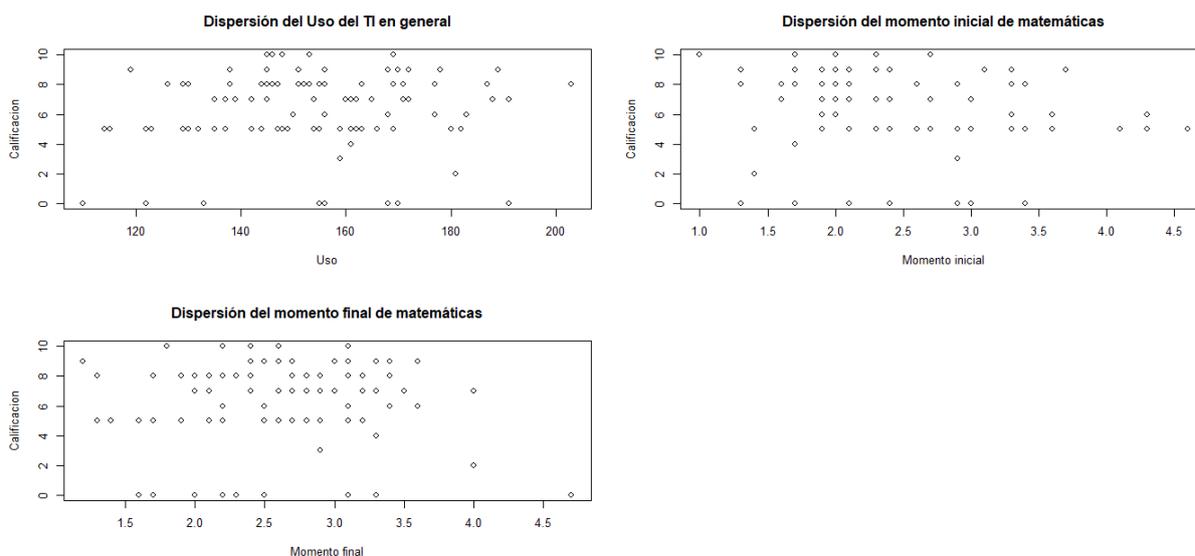


Figura 2. Diagramas de dispersión.

Uso del Teléfono Inteligente en actividades generales, el uso matemático que traían al inicio del periodo (MI) y el uso matemático que adquirieron al final del periodo (MF). Al observar que no están asociadas las variables se propone el siguiente modelo de regresión:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Donde las variables aplicadas son las correspondientes de la tabla 2, los coeficientes de regresión se calcularon en el programa R Project (versión 3.5.1), los cuales se presentan en la Tabla 3 donde se observa que la única variable que condiciona a la variable dependiente calificación es el uso recreativo que se le da al teléfono inteligente con valor de probabilidad de 0.05.

Tabla 3. Resultados de los coeficientes de regresión estimados para para toda la población.

Coeficientes de regresión estimados para toda la población				
Variable	Estimador β_1	Error Estandar	Valor de t	Valor de p
Sexo	0.28450	0.52447	0.542	0.5888
Tratamiento	0.61770	0.55353	1.116	0.2673
Uso	-0.04232	0.02174	-1.947	0.0546
MI	-0.0447	0.05128	-0.872	0.3855
MF	0.01596	0.02141	0.746	0.4577

Con base en el resultado encontrado, se procede a realizar el modelo de regresión simple con la variable Uso, los resultados se muestran en la siguiente tabla, en donde se puede observar mediante la estimación, que la variable Uso no influye en la calificación obtenida por los estudiantes.

Tabla 4. Resultados del coeficiente de regresión para la variable Uso.

Coeficiente de regresión estimado para toda la población				
Variable	Estimador β_1	Error Estandar	Valor de t	Valor de p
Uso	0.00837	0.01372	0.610	0.5433

De acuerdo con los resultados anteriores podemos decir que el teléfono inteligente no condiciona el aprovechamiento académico de la población bajo estudio, sin embargo, se procedió a realizar el mismo análisis de regresión lineal anteriormente expuesto, pero ahora de manera independiente para los grupos experimental y de control en los que se secciono la base de datos.

Se presentan los resultados para ambas secciones a continuación.

Los resultados para la sección de control (donde no se promovió el uso del teléfono inteligente en las actividades académicas) se presentan en la Tabla 5 donde se muestran los estimadores de la recta de regresión para el grupo control, donde ninguno aporta gran relevancia al modelo.

Tabla 5. Resultados de coeficientes de regresión estimados para la sección de control.

Coefficientes de regresión estimados para el grupo de control				
Variable	Estimador β_1	Error Estandar	Valor de t	Valor de p
Sexo	-0.16856	1.15905	-0.145	0.885
Uso	0.02220	0.04153	0.535	0.597
MI	-0.41923	0.98450	-0.426	0.673
MF	-0.02635	0.93068	-0.028	0.978

Los resultados para la sección experimental se presentan en la Tabla 6 donde se muestran los estimadores de la recta de regresión para el grupo experimental, donde ninguno aporta gran relevancia al modelo.

Tabla 6. Resultados de coeficientes de regresión estimados para la sección experimental.

Coefficientes de regresión estimados para el grupo de control				
Variable	Estimador β_1	Error Estandar	Valor de t	Valor de p
Sexo	0.38318	0.59906	0.640	0.5248
Uso	0.01542	0.02617	0.589	0.5580
MI	-0.75298	0.41306	-1.823	0.0731
MF	-0.10506	0.70109	-0.150	0.8814

De la construcción de los perfiles predictores de los alumnos bajo las técnicas de los algoritmos clasificatorios

En un primer momento a la base de datos en general se le aplicaron dos algoritmos clasificatorios, análisis clúster y arboles de decisión, pertenecientes a la estadística clásica y a la inteligencia artificial respectivamente, con el fin de conocer características generales de la población bajo estudio y poder encontrar similitudes y diferencias entre los posibles grupos de alumnos que pudieran formarse, sin embargo, al someter a la base de datos a estos análisis estadísticos no arrojaron resultados significativos bajo los cuales se pudiera realizar un análisis formal para poder establecer características predictivas enfocadas al aprovechamiento académico de los alumnos de la clase de Álgebra Lineal 1, en consecuencia se procedió a dividir la base de datos en tres grupos en función de la calificación final obtenida por los alumnos al concluir el curso, se presentan en la Tabla 7 los niveles de desempeño establecidos para la población de estudio.

Tabla 7. Niveles de desempeño en base a la calificación final obtenida por los alumnos.

Niveles de desempeño	
Nivel	Rangos de Promedio
Destacado	10 - 9
Regular	8-6
Malo	5-0

Una vez que se establecieron estos niveles de desempeño se aplicaron algoritmos clasificatorios a cada una de las secciones de la base de datos de manera independiente, con el objetivo de establecer las características principales que representan a los alumnos pertenecientes a cada uno de estos niveles de desempeño.

Se desarrolló un estudio clasificatorio con todas las variables de la base datos incluyendo aquellas que involucran el uso del teléfono inteligente, donde se pudo observar de igual manera que no tiene relación alguna con el rendimiento académico de los estudiantes, por lo que se optó por excluirlas en la aplicación de estos algoritmos clasificatorios para eliminar el ruido que pudiera ocasionar en los resultados al buscar los perfiles predictores de los alumnos bajo estudio, en este sentido las variables con las que se trabajaron fueron: Sexo, Sección (grupo al que pertenece cada estudiante y de acuerdo a la sección de pertenencia, el alumno tomo clase con uno de los dos docentes que imparten la materia de álgebra lineal, existen tres secciones, la sección 1 y 2 son atendidas por un docente a quien llamaremos a partir de este momento como Docente A y la sección 3 recibió por parte del Docente B), Escuela, Lugar de procedencia (agrupados en Local en sus estudios previos para quienes radicaron y estudiaron en Xalapa; Regional: estudiantes que radicaron y estudiaron en municipios fronterizos o en las cercanías a Xalapa; Estatal: resto de municipios del estado; Nacional: otros estados de la República y Extranjeros: estudiantes provenientes de otros países), y Edad.

Estas variables fueron codificadas para su tratamiento bajo las técnicas estadísticas antes mencionadas.

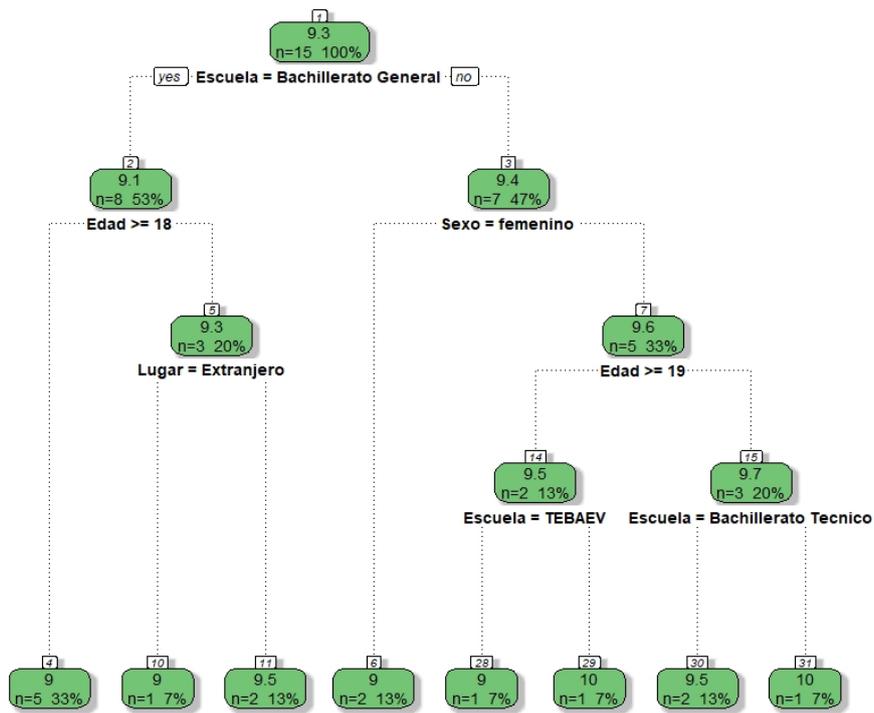
Tabla 8. Códigos empleados en las diferentes variables analizadas al aplicar algoritmos clasificatorios.

Codificación de las variables aplicadas en los algoritmos clasificatorios		
Variable	Indicadores	Código
Sexo	Mujer	1
	Hombre	2
Sección	Sección 1	1 (Docente A)
	Sección 2	2(Docente A)
	Sección 3	3(Docente B)
Escuela	Bachillerato General	1
	Bachillerato Técnico	2
	TEBAEV	3
Lugar	Local	1
	Regional	2
	Estatad	3
	Nacional	4
	Extranjero	5
Edad	17 – 23	17 - 23

A continuación, se muestran los resultados obtenidos de la aplicación de los algoritmos clasificatorios en los 3 niveles de desempeño de la base de datos atendiendo su desempeño académico durante el curso referido.

Desempeño destacado

En este nivel de desempeño se consideran aquellos alumnos que obtuvieron una calificación final entre 9 y 10, la integran 18 estudiantes que cumplen esta característica, se muestra a continuación el análisis desarrollado con la técnica de árboles de decisión de la inteligencia artificial, el cual culmina con la generación del diagrama mostrado en la Figura 3, en donde la agrupación más representativa son aquellos alumnos que provienen de un bachillerato general y tienen 18 años o más.



Rattle 2021-mar.-12 14:16:16 52229

Figura 3. Árbol de decisión correspondiente al nivel de destacado.

De igual manera se realizó el estudio bajo la técnica de análisis clúster de donde se obtienen el diagrama de témpanos representado en la Figura 4 en el que se puede observar de manera inmediata el grupo de estudiantes más representativos del grupo de destacados. En la Figura 5 se puede observar el dendograma en el cual se visualiza como es que esta técnica va agrupando a los individuos en base a sus características similares y tiene la particularidad que permite identificar cuales individuos pertenecen a determinado clúster.

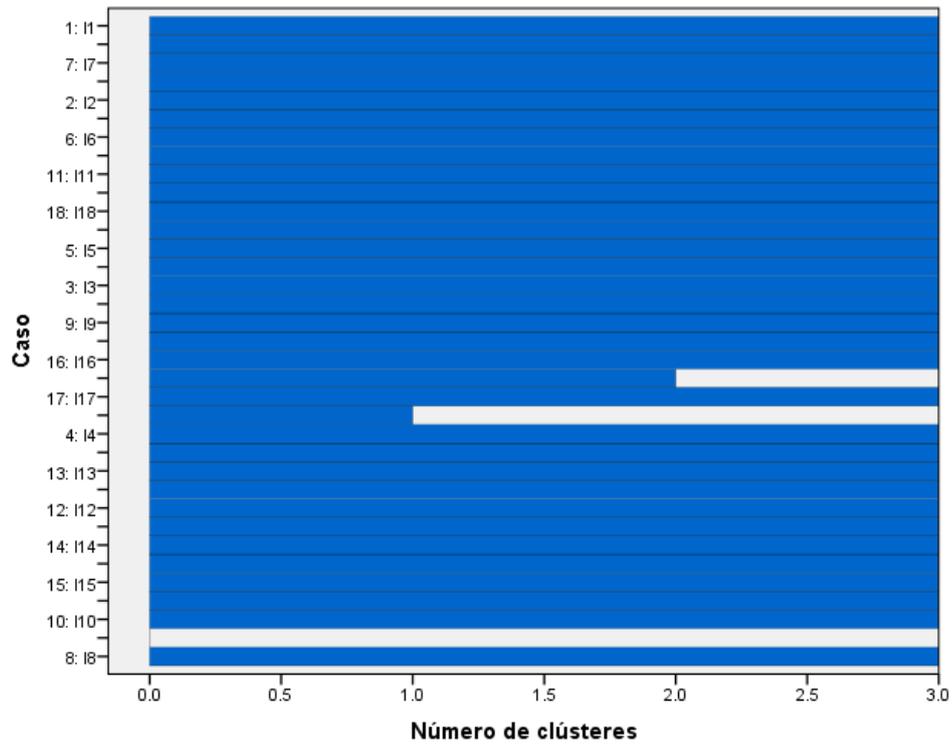


Figura 4. Diagrama de témpanos de la población con nivel de aprovechamiento destacado.

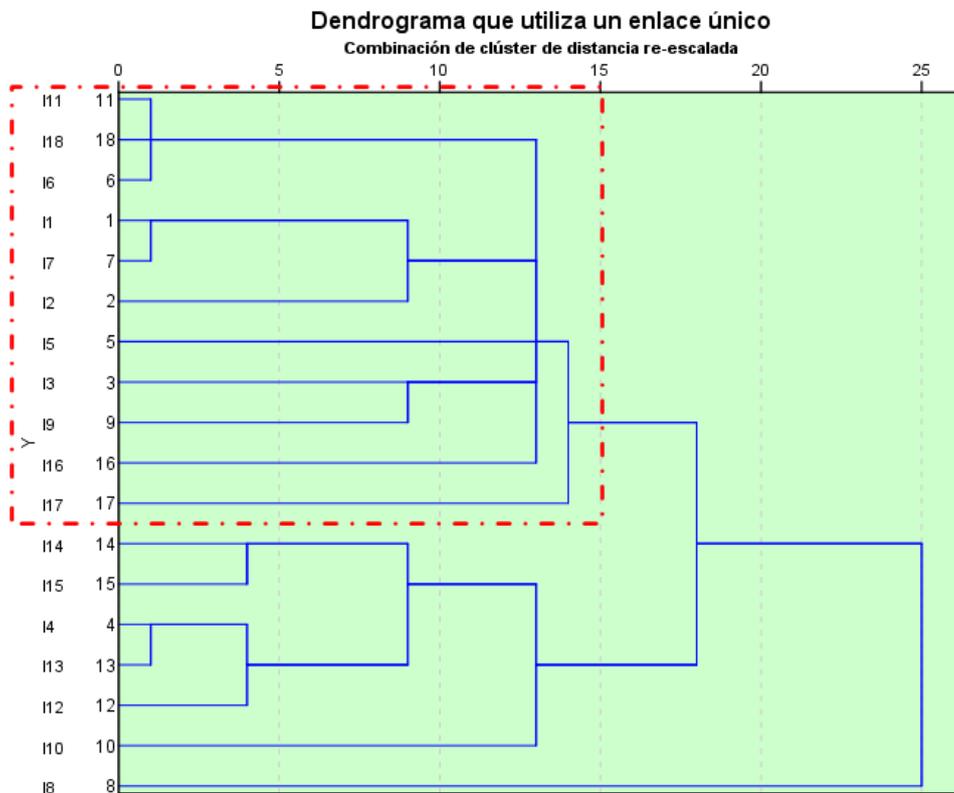


Figura 5. Dendrograma correspondiente a la agrupación de alumnos con nivel de aprovechamiento destacado.

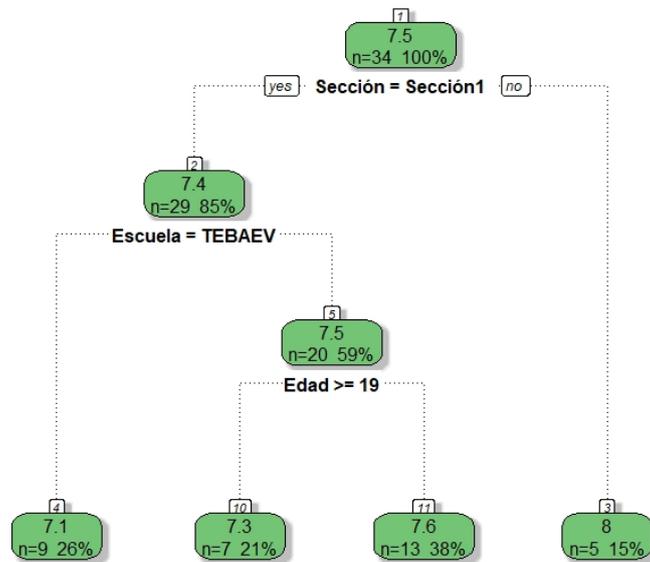
Una vez que se identifican los individuos del clúster más representativo se extraen de la base de datos para realizar un análisis minucioso y determinar las características generales de este grupo representativo de los estudiantes con nivel destacado, este grupo de estudiantes se observa en la Tabla 8, en donde se observan las características generales de este grupo, de la que se puede decir que son hombres pertenecientes a la sección del docente A, provienen de un bachillerato general, son pertenecientes a la región de Xalapa y tienen 18 años o más.

Tabla 8. Alumnos que pertenecen al grupo más representativo de estudiantes con nivel de aprovechamiento destacado.

Alumnos pertenecientes al clúster más representativo de nivel de desempeño destacado					
Nombre	Sexo	Sección	Escuela	Lugar	Edad
I1	2	1	2	2	18
I2	2	1	3	2	18
I3	2	1	2	3	20
I5	2	2	1	2	17
I6	2	1	1	2	19
I7	2	1	2	2	18
I9	2	1	3	3	20
I11	2	2	1	2	19
I16	2	2	1	3	19
I17	2	2	1	2	21
I18	2	2	1	2	19
	Hombre	Docente A	Bachillerato General	Regional	18 años 19 años

Desempeño regular

En este nivel de desempeño, únicamente se encuentran aquellos alumnos que obtuvieron un promedio final en un rango de 6 y 8 de calificación, en esta porción de la base de datos se cuenta con 42 alumnos que cumplen esta característica, en la Figura 6 se muestra el diagrama obtenido al aplicar la técnica de la inteligencia artificial (árboles de decisión) a la población perteneciente a este nivel de desempeño, en donde se puede observar que la agrupación más significativa de este grupo de estudiantes es la que pertenece a la sección 1, que no proviene de un bachillerato tipo TEBAEV, y que tiene 19 años o más.



Rattle 2021-mar.-12 13:12:27 52229

Figura 6. Árbol de decisión de la base datos con nivel de desempeño regular.

De igual manera se presenta a continuación los resultados obtenidos al realizar el análisis clúster, técnica propia de la estadística clásica a este mismo grupo de alumnos pertenecientes al nivel de desempeño regular. En las Figuras 7 y 8 se muestran el diagrama de témpanos donde se puede visualizar de manera inmediata tres agrupaciones principales que la estadística clásica realiza en base a la técnica clúster y el dendograma en donde se puede visualizar como la técnica ya mencionada fue realizando las agrupaciones de los individuos, marcando en este el grupo de alumnos más representativo de este nivel de aprovechamiento en un rectángulo rojo,

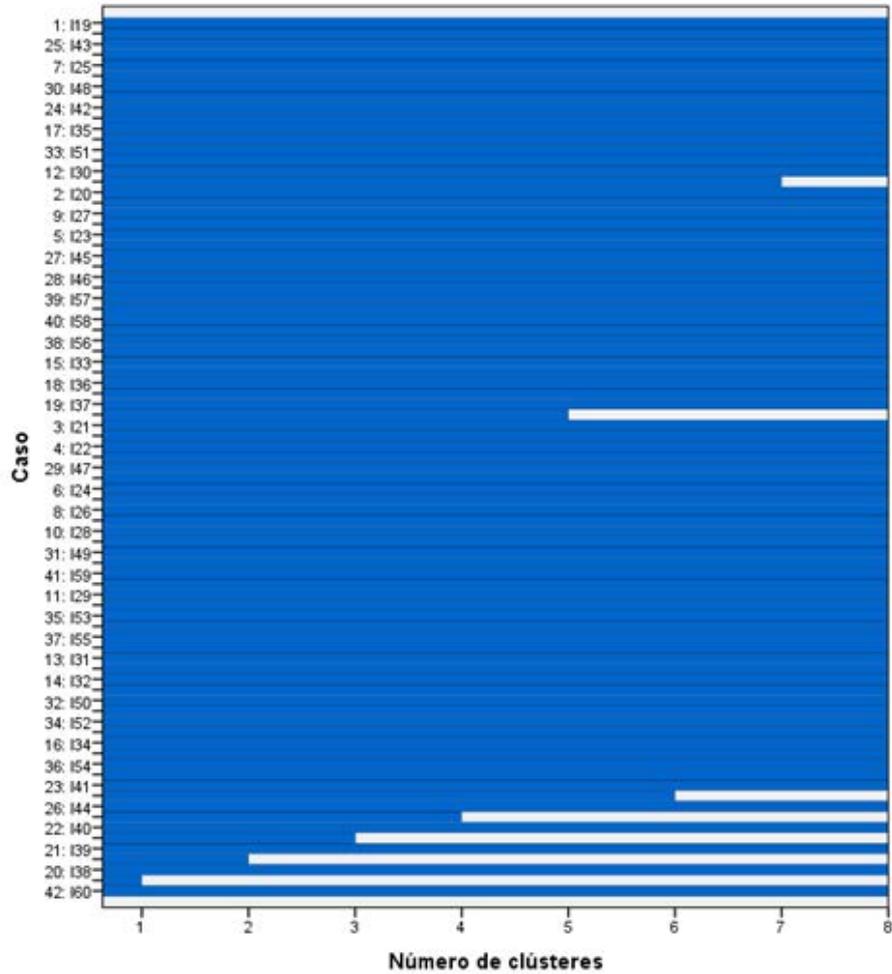


Figura 7. Diagrama de témpanos para el grupo con nivel de desempeño regular.

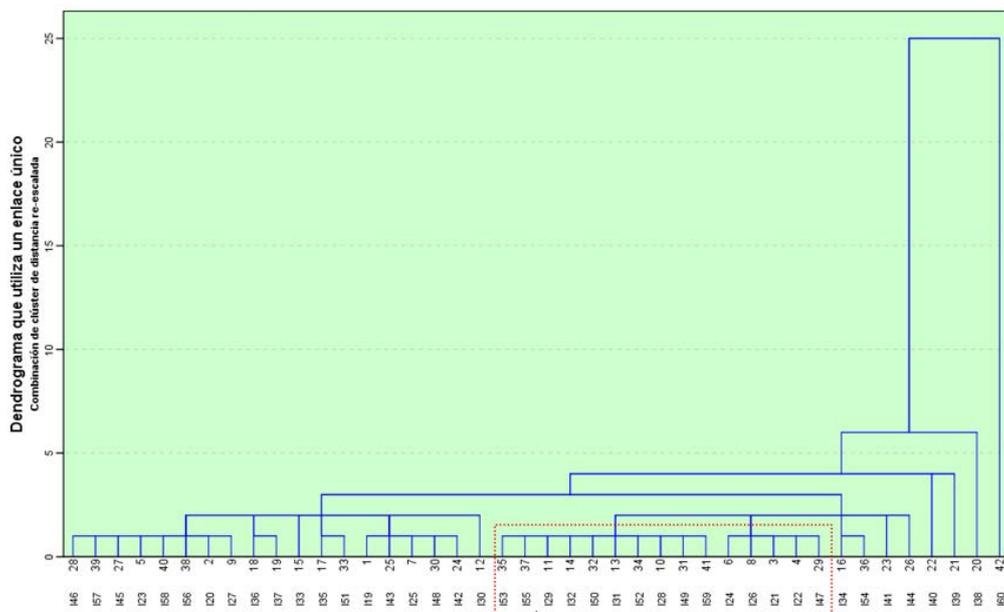


Figura 8. Dendrograma correspondiente a la agrupación de alumnos con nivel de aprovechamiento regular.

Al identificar a este grupo de individuos que pertenecen a este nivel de desempeño se extraen de la base de datos para poder realizar un análisis exhaustivo de sus características cualitativas para poder establecer el perfil del estudiante de álgebra lineal con altas posibilidades de obtener un nivel de desempeño regular. En la tabla 9 se muestran las características de cada uno de estos alumnos.

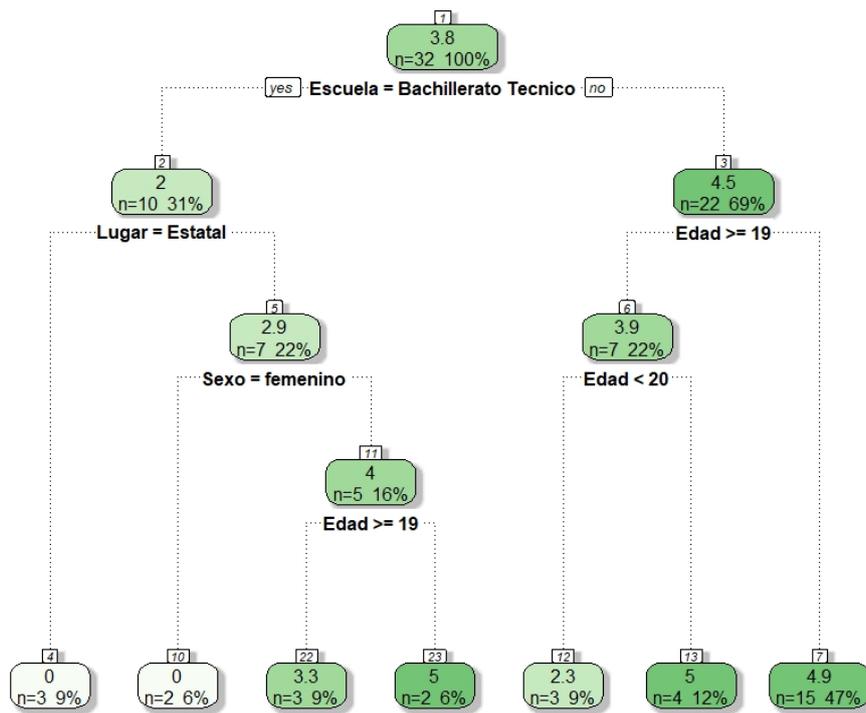
Tabla 9. Alumnos de desempeño regular.

Nombre	Sexo	Sección	Escuela	Lugar	Edad
I21	2	1	1	2	18
I22	2	1	1	2	18
I24	2	1	3	2	17
I26	2	1	3	2	20
I28	2	2	2	2	18
I29	2	2	1	2	19
I31	2	2	1	2	18
I32	2	2	1	2	18
I34	2	2	1	3	19
I41	2	1	1	3	18
I47	2	1	2	2	18
I49	2	2	2	2	21
I50	2	2	1	2	18
I52	2	2	1	2	17
I53	2	2	1	2	19
I54	2	2	1	3	19
I55	2	2	1	2	19
	Hombre	Docente A	Bachillerato General	Regional	18 años

Desempeño malo

En este nivel de desempeño se ubican aquellos alumnos que obtuvieron una calificación final igual o menor a 5, al igual que en los otros niveles de desempeño se realizaron los mismos dos análisis clasificatorios para establecer las principales características de este grupo de la población.

En la Figura 9 se muestra el árbol de decisión obtenido bajo la técnica de inteligencia artificial conocida bajo el mismo nombre, de acuerdo con esta técnica las características representativas de estos alumnos es que no egresan de un bachillerato técnico y tienen 19 años o más.



Rattle 2021-mar.-12 16:32:36 52229

Figura 9. Árbol de decisión correspondiente al grupo de alumnos con nivel de aprovechamiento malo.

Al desarrollar el análisis clúster para este mismo grupo de la población se obtuvo el diagrama de témpanos y el dendograma correspondientes, los cuales son mostrados en las Figuras 10 y 11 respectivamente, en donde se puede observar la formación de los grupos de alumnos con características similares.

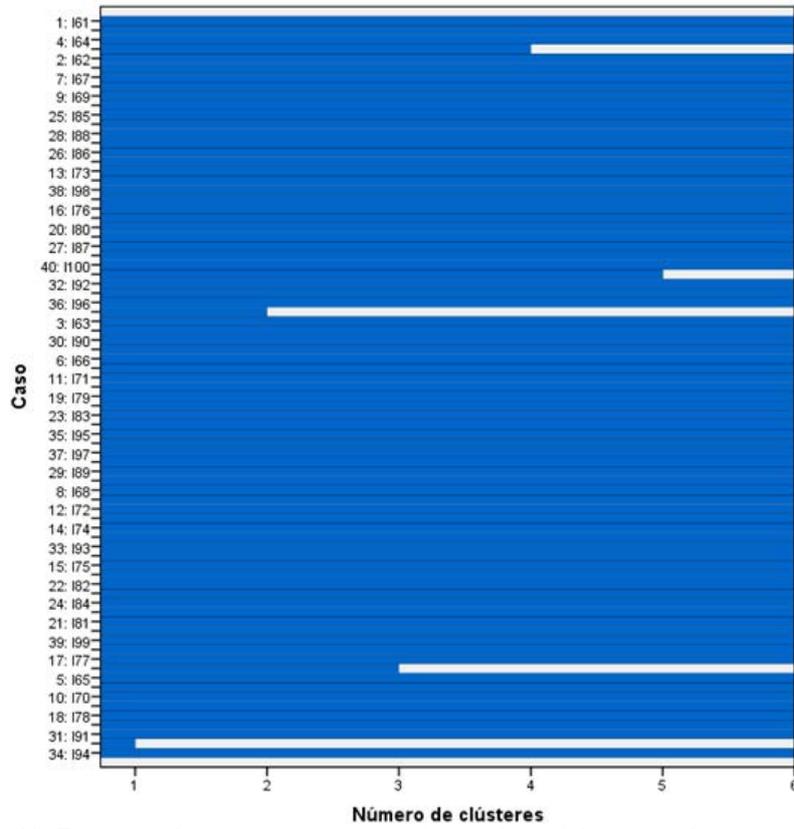


Figura 10. Diagrama de témpanos correspondiente al nivel de aprovechamiento malo.

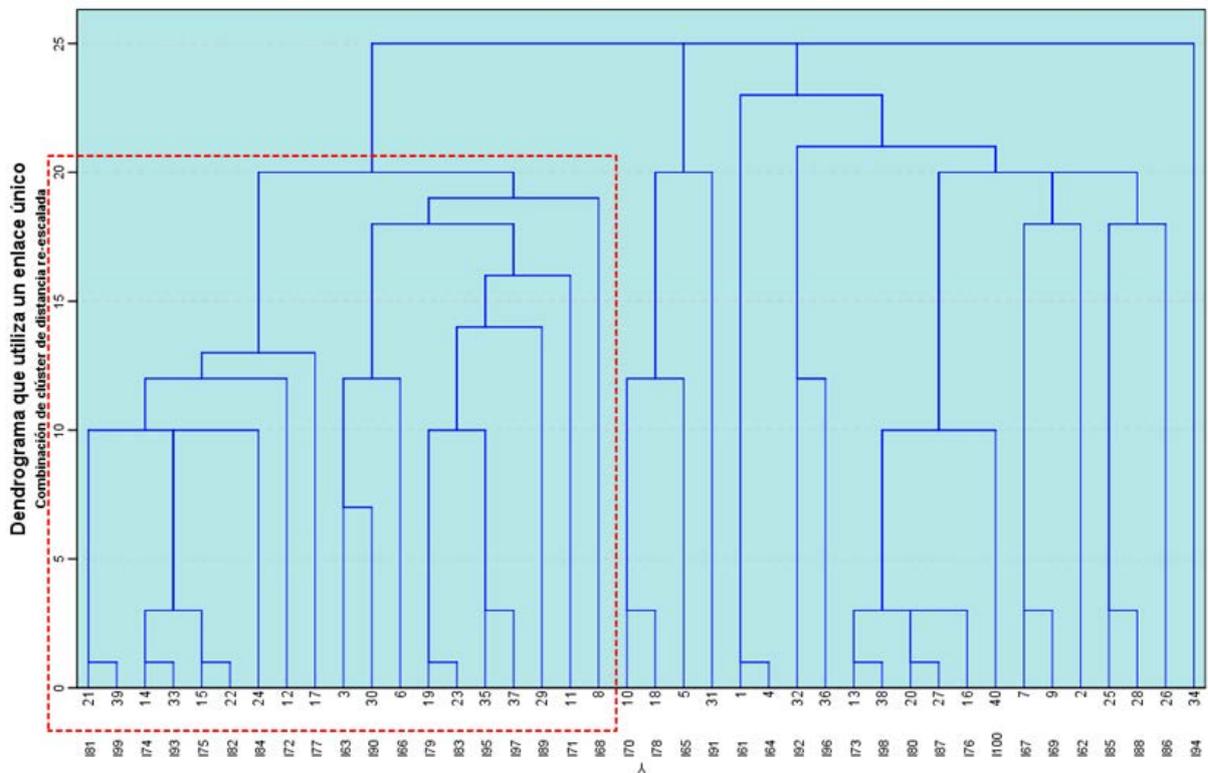


Figura 11. Dendrograma correspondiente a la agrupación de alumnos con nivel de aprovechamiento malo.

Al identificar a los alumnos más representativo de este nivel de aprovechamiento se realizó el mismo análisis minucioso que en los otros niveles de aprovechamiento para poder conocer de manera puntual cuáles son sus características que definen al alumno que tiene altas posibilidades de obtener un nivel de aprovechamiento malo en la materia de álgebra lineal 1 de la facultad de Estadística en la Universidad Veracruzana. En la tabla 10 se pueden observar las características de este grupo de alumnos representativos de este nivel de aprovechamiento, del cual se puede decir que son alumnos hombres, que toman clase con el docente B, provienen de un bachillerato técnico, que son de la localidad de Xalapa y tienen 18 años.

Tabla 10. Alumnos que pertenecen al nivel de desempeño regular.

Alumnos pertenecientes al clúster más representativo de nivel de desempeño regular					
Nombre	Sexo	Sección	Escuela	Lugar	Edad
I63	2	1	1	2	18
I66	2	2	1	3	18
I68	2	2	2	2	20
I71	2	3	1	4	18
I72	2	3	2	1	21
I74	2	3	2	1	19
I75	2	3	2	1	18
I77	2	3	3	1	22
I79	2	3	3	3	18
I81	2	3	1	1	18
I82	2	3	2	1	18
I83	2	3	3	3	18
I84	2	3	3	1	18
I89	2	2	3	3	19
I90	2	1	1	3	18
I93	2	3	2	1	19
I95	2	3	2	3	18
I97	2	3	2	3	19
I99	2	3	1	1	18
	Hombre	Docente B	Bachillerato Técnico	Local	18 años

Conclusión

A manera de disertación y tomando como base los resultados obtenidos en el análisis estadístico al calcular el coeficiente de correlación de Pearson y los coeficientes de regresión de las distintas variables involucradas en la investigación, se logró determinar que, ya sea promoviendo o no, el uso del teléfono inteligente, el impacto que causa este, para potencializar el aprovechamiento académico de manera positiva es prácticamente nulo.

Reforzando los hallazgos de Arias, Lugo y Morales (2016) quienes aseguran que la aplicación de los teléfonos inteligentes no se está contextualizando en el área académica, se encontró que el uso de aplicaciones educativo-matemático aumentó al final del semestre en los alumnos que estuvieron bajo el tratamiento experimental, aunque esto no impactó en su rendimiento final. Sin embargo, contrario a Oliva (2014), no podemos considerar al dispositivo inteligente como un enemigo del aprovechamiento académico, más bien como una herramienta tecnológica para la cual es necesario generar directrices que guíen de manera adecuada su implementación. Apoyados en los hallazgos de Basantes, Naranjo, Gallegos y Benítez (2017), quienes aseguran que el TI estimula la exploración fuera y dentro del área académica, queda claro que el estudiante refleja cierto grado de aceptación al incorporar el teléfono inteligente en las actividades académicas de la clase de álgebra lineal y lo adopta como herramienta de investigación, sin embargo, es primordial introducir a los teléfonos inteligentes en el contexto educativo de una manera regulada, para esto es necesario impulsar políticas educativas que incluyan al teléfono inteligente, apoyando a Figueroa (2016) es necesario la adaptación de los planes de trabajo para adherir el uso del dispositivo inteligente en las actividades académicas, puesto que al contemplarlos en la planeación curricular puede aumentar las competencias de los estudiantes y de esta manera el dispositivo inteligente será punta de lanza en logro de los aprendizajes esperados que el estudiante del primer curso de álgebra lineal de la Universidad Veracruzana necesita.

En lo que respecta a las técnicas de clasificación se lograron crear perfiles predictores de los niveles de aprovechamiento a los que los alumnos pueden aspirar en la clase de álgebra lineal en base a ciertas características cualitativas con las que el alumno inicia el curso, esto se traduce en un recurso poderoso al docente que imparte esta materia en la Universidad Veracruzana, para que pueda adecuar su planeación de trabajo en base a las características del grupo con el que inicia un nuevo curso.

La aplicación de estas dos técnicas clasificadoras, análisis clúster y arboles de decisión, de distinta naturaleza a la misma base de datos y con el mismo fin, permitió realizar un análisis a profundidad de las particularidades de cada una de ellas y así poder establecer ventajas y desventajas que representan al emplearlas para la construcción de predictores. De acuerdo con lo anterior y en base a los procesos realizados durante la aplicación de estos dos algoritmos, podemos decir que la Inteligencia Artificial facilita el proceso al clasificar la información, permitiéndole al usuario la manipulación de esta técnica sin tener conocimientos profundos en arboles de decisión, sin embargo, esta facilidad también está acompañada de cierto grado de imprecisión si la comparamos con la técnica de análisis clúster, esto da sustento a Solarte y Ocampo (2009) quienes establecen que las técnicas de categorización que utiliza clasificadores basados en patrones permiten catalogar el objeto o sujeto con una calidad competitiva con un nivel superior, logrando así un modelo más legible y de fácil interpretación a diferencia de la técnica de los árboles de decisión pues su eficiencia está estrechamente relacionada con la calidad de los patrones que utilizemos.

El análisis clúster, en comparación con la técnica de la inteligencia artificial, representa un mayor grado de dificultad, por lo que exige conocimientos estadísticos más especializados, lo que respalda lo planteado por Caridad (1997) quien establece que la inteligencia artificial trata de automatizar los procesos complejos de la estadística clásica minimizando los posibles errores y los tiempos que la aplicación de una técnica estadística pura demanda, sin embargo,

las dificultades que pueda presentar una técnica de la estadística clásica como es el análisis clúster, se justifica con la precisión y profundidad de análisis que se puede alcanzar al analizar los resultados que se obtienen al aplicar esta técnica, puesto que se tiene acceso pleno a todas las características de los individuos que la técnica agrupa en clúster, entendiendo por clúster el grupo de individuos que comparte ciertas características y es agrupado en base a ellas por la técnica del mismo nombre.

De acuerdo con los resultados obtenidos de estas dos técnicas clasificadoras se pueden afirmar tres cosas respecto al estudiante de álgebra lineal aplicada : 1) para aprovechamiento académico destacado las características que lo distinguen son sexo masculino, de la región de Xalapa (lugares circunvecinos), proviene de un bachillerato general y toma clase con el docente A; 2) de aprovechamiento académico regular reúne las características tales como sexo masculino, de la localidad de Xalapa, proviene de un bachillerato general y toma clase con el docente A, y 3) con aprovechamiento académico malo de sexo masculino, con radicación en Xalapa, Veracruz, proviene de un bachillerato técnico y toma clase con el docente B.

Referencias

- Arias. V.S., Lugo C. J. S. y Morales T. J. A. (2016). Uso de los dispositivos móviles en el aula de clase por parte de los estudiantes de grado 11 del colegio ciudadela cuba de la ciudad de Pereira (investigación cualitativa). Universidad Tecnológica de Pereira.
- Barrios G.B. (2020). Relación entre uso del teléfono inteligente y rendimiento escolar en estudiantes universitarios de álgebra lineal I. Universidad Veracruzana.
- Basantes A. B., Naranjo M. E., Gallegos M. C. y Benítez N. M (2017). Los Dispositivos Móviles en el Proceso de Aprendizaje de la Facultad de Educación Ciencia y Tecnología de la Universidad Técnica del Norte de Ecuador. *Formación Universitaria*. 10(2), 79-88.
- Beltrán M. B. (2016). Minería de datos [Archivo PDF]. Benemérita Universidad Autónoma de Puebla. <http://bbeltran.cs.buap.mx/NotasMD.pdf>
- Caridad O. J.M. (1997) Inteligencia artificial y estadística aplicada: algunas líneas modernas de desarrollo. Real academia de Córdoba, de ciencias, bellas letras y nobles artes.
- Cisneros H. L. y Robles R. S. (2017). ¿Para qué utilizan el celular en el aula, los estudiantes universitarios? Universidad de Guadalajara.
- Figuroa P.C.S. (2016). El uso del smartphone como herramienta para la búsqueda de información en los estudiantes de pregrado de educación de una universidad de Lima Metropolitana. *Educación*, 49, 29-44.
- Freedman D., Pisani R., Purves R., y Adhikari A. (1993). Estadística (Segunda ed.). Barcelona: Antoni Bosch.
- Oliva H. A. (2014). El uso de teléfonos móviles en el sistema educativo público de El Salvador: ¿Recurso didáctico o distractor pedagógico? *Realidad y reflexión*, 40, 60-75.
- Sánchez, A. (2019). Definición de Educación. Obtenido de <https://conceptodefinicion.de/educacion/>

- Silva C. A. C. y Martinez D. D.G. (2017). Influencia del Smartphone en los procesos de aprendizaje y enseñanza. *Suma de negocios*, 8, 11-18.
- Solarte M. G. R y Ocampo S. C. A. (2009). Técnicas de clasificación y análisis de representación del conocimiento para problemas de diagnóstico. *Scientia et Technica*, 42, 177-182.
- Tan P.N, Steinbach M. y Kumar V. (2006). Introduction to data mining. Boston: Pearson Education, Inc.
- Zamudio R. R. (2013). Descomposición de clases con diversas técnicas para mejorar la clasificación del Naive Bayes. [Tesis inédita de maestría]. Universidad Veracruzana.

Comportamiento del intervalo de confianza y de cobertura para la proporción en la distribución multinomial

Emmanuel Morales García¹, Julia Aurora Montano Rivas², Ernesto Pedro Menéndez Acuña³

¹Facultad de Estadística e Informática. Universidad Veracruzana. emmorales@uv.com

²Facultad de Estadística e Informática. Universidad Veracruzana. julmontano@uv.mx

³Facultad de Matemáticas. Universidad Veracruzana. emenendez@uv.com

Resumen

En esta investigación se desarrolló y analizó el comportamiento de intervalos de confianza y de cobertura para el parámetro proporción de una distribución multinomial (generalización multivariada de la binomial) con la finalidad de conocer que enfoque tiene mejor presentación de resultados. Dada la naturaleza cuantitativa de la investigación, se desarrolló el método de Sison-Glaz para la obtención de los intervalos de confianza, dicho método es empleado para este tipo de distribución y específicamente para la estimación de la proporción. En el caso bayesiano se utilizó como distribución a priori a la Dirichlet que es la generalización multivariada de la beta y se utiliza concretamente como la conjugada de la multinomial, posteriormente se implementó el método computacional cadenas de Márkov vía Monte Carlo (MCMC), para generar la simulación y obtención de la distribución a posteriori. Los datos procesados para realizar estos métodos fueron obtenidos de las elecciones 2016 para diputados del Estado de Veracruz, en donde se obtuvieron nueve clasificaciones para el tipo de nulidad en las boletas electorales, con esto se probará si la estimación del parámetro cae dentro del intervalo identificado para cada enfoque. Se concluye, que el método bayesiano da mejor resultado a pesar de que el frecuentista tiene buen enfoque, pero los intervalos son más amplios y existen casos donde el parámetro está afuera de los límites.

Palabras Clave: Distribución multinomial, intervalo de cobertura, intervalo de confianza, MCMC, distribución Dirichlet.

Introducción

En estadística se conocen dos tipos de intervalos, el de confianza y el de cobertura; los cuales permiten verificar la presencia de la variabilidad que se presenta en la mensuración de la población de estudio.

Los intervalos de confianza se construyen utilizando el teorema del límite central, sin embargo, existen otros métodos que apoyan a la búsqueda de estos valores, como lo son la máxima verosimilitud, el método de la F, intervalo de credibilidad por el enfoque Bayesiano y por la parte computacional el Bootstrap, entre otros.

En esta investigación se realiza el procedimiento comparativo del enfoque Bayesiano y el frecuentista para la construcción de intervalos de credibilidad y confianza de la distribución multinomial, también se implementa el proceso computacional para calcular los intervalos y la estimación de los parámetros de proporción de la distribución mencionada. El procedimiento se ejemplifica mediante el uso correcto del software estadístico R Project versión 4.0.1.

Antecedentes

Tortora (1978), publica una nota sobre la metodología para estimar el tamaño de muestra necesaria para la obtención de la confianza de manera precisa, sobre los parámetros de una población multinomial.

Esta propuesta calcula el intervalo de confianza simultáneamente mediante el procedimiento de Goodman, de acuerdo con los resultados obtenidos realiza una comparación y verifica el compartimiento con una población con distribución Binomial. En esta investigación también se aborda como Cochran propone un enfoque que modifica el tamaño de la muestra en entorno a la estimación del intervalo en una distribución multinomial.

Sison-Glaz (1995), propone el método para lograr obtener intervalos de confianza simultáneos para los parámetros de proporción de una distribución multinomial, de acuerdo con la literatura consultada, estos nuevos procedimientos son más precisos, debido a que se evalúan por medio del volumen de las regiones de confianza de probabilidad, de la cobertura nominal y de la probabilidad de cobertura que se obtiene de manera final. Por otra parte, estos autores refieren a su método como eficiente, de solución apta en tiempo y precisión para implementarlo a problemas que abarquen este tipo de datos.

En 2007, Wang, desarrolla una investigación en la cual escribe sobre los coeficientes de confianza exactos e intervalos de confianza para proporciones multinomiales, el objetivo principal es desarrollar un proceso para calcular los coeficientes de confianza exactos para los intervalos de confianza en una proporción multinomial, para cualquier tamaño de muestra fijo. Con esta metodología los coeficientes pueden derivarse oportunamente y así encontrar el punto donde la probabilidad es mínima.

Para el 2010, González, realiza su tesis de Magister en estadística, en la Universidad de Colombia, dicho trabajo se titula Comparación de intervalos de confianza para la distribución multinomial, de manera general realiza una comparativa de diversas metodologías para mostrar la eficiencia de ambas, teniendo como principal al teorema del límite central.

Alvarado (2014), defiende su tesis doctoral en la Universidad de Autónoma Barcelona, el trabajo muestra los aportes metodológicos en la estimación de tamaño de muestras en estudios poblacionales de prevalencia. El trabajo se centra en seis enfoques estadísticos para estimar tamaños de muestras para poblaciones multinomiales. Otro aspecto importante es que de acuerdo con la literatura los procedimientos para estimar muestras en el área de la salud son muy controversial. Realizan una estructura teoría de cómo trabaja la distribución multinomial y posteriormente realizan simulaciones con la finalidad de ver el comportamiento de estas estimaciones.

González, Vélez y Correa (2015), publican un artículo, pero ahora considerando 13 intervalos de confianza para realizar una comparación para los parámetros de la distribución multinomial, entre los métodos utilizados son, Teorema del límite central, método Bootstrap, intervalos de credibilidad, F, verosimilitud, entre otros. Según la aplicación que realizaron mediante simulación, los métodos que presentan mejor solución son la distribución F, verosimilitud, Quesenberry y Hurst.

Marco conceptual

Distribución Multinomial ($\sim M$)

Es una distribución discreta multivariante y representación contigua de la \sim binomial: Se efectúa cuando los ensayos tienen más de dos resultados posibles, a su vez las probabilidades son las mismas para cada ensayo, es importante mencionar que cada ensayo es independiente. Las variables aleatorias X_1, X_2, \dots, X_k poseen una distribución multinomial (González,2010; Alvarado,2014). Su función de probabilidad está dado como:

$$f(x_1, \dots, x_k; p_1, \dots, p_k, n) = \frac{\Gamma(n + 1)}{\prod_{i=1}^k \Gamma(x_i + 1)} \prod_{i=1}^k p_i^{x_i}$$

Las k categorías que se consideren de la población de unidades divididas deben ser mutuamente excluyentes y exhaustivas (Tortora,1978).

Definición. Sea $X = (X_1, X_2, \dots, X_k)$, una v.a. k -dimensional, entonces se considera que tiene una distribución multinomial con parámetros n, p_1, p_2, \dots, p_k , dado lo anterior se denota de la siguiente manera $X \sim M(n; p_1, p_2, \dots, p_k)$.

Distribución Dirichlet: ***Dir***(α)

Es una distribución continua multivariante, con parámetros $\alpha_1, \dots, \alpha_k$ perteneciente al conjunto de números reales positivos. Es una generalización multivariante de la distribución Beta. El modelo que se introduce para presenciar a las variables debe tomar valores en el intervalo de cero y uno, cuya suma es igual a la unidad, la Dirichlet observa la variación de las proporciones (Lin, 2016).

La densidad de probabilidad, con vector de parámetros α , se denota como:

$$Dir(\pi | \alpha_1, \dots, \alpha_m) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \pi_i^{\alpha_i - 1}$$

Inferencia estadística Bayesiana

El enfoque bayesiano, aprueba la interpretación subjetiva de la probabilidad como una medida condicional de la duda y que se puede corregir vía la evidencia muestral, mediante el teorema de Bayes, con esto se logra generar la información primordial del parámetro y se realiza la inferencia de éste, mediante la teoría de decisión (Box y Tiao, 1973; Olmos, 2016).

Regla de Bayes

El teorema de Bayes se describe como la actualización de las probabilidades de las hipótesis cuando se les proporciona evidencia. Trabaja mediante la probabilidad condicional, se utiliza para resolver problemas que involucra actualizaciones de las creencias. (Box y Tiao, 1973; Canavos,1988).

Razón de verosimilitud

Si $L(\theta)$ es la función de verosimilitud, se define la función de verosimilitud como:

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$$

El conjunto de valores para θ las cuales $R(\theta) \geq P$ es llamado intervalo de 100% de verosimilitud para θ (González,2010).

Distribución a priori

Cuantifica la inquietud que se conoce acerca de este antes de considerar la información muestral, a su vez se representan en una distribución de probabilidad (Olmos, 2016; Correa y Barrera, 2018). Para el caso de esta investigación se utilizará como distribución a priori a la Dirichlet.

Distribución a priori conjugada

$P(\theta)$ y $P(\theta|x)$ son distribuciones de probabilidad sobre θ . La primera solo incorpora información a priori y la segunda realiza una actualización de la información con la evidencia muestral que se obtenga. Ambas distribuciones son de la misma familia, es decir una familia conjugada (Olmos, 2016; Correa y Barrera, 2018).

Distribución predictiva

Es una herramienta que los usuarios tienen para determinar la calidad de la distribución a priori, además de proporcionar información, es útil para determinar los procedimientos de análisis posteriores (Bernardo,2003; Olmos,2016).

Se realiza una predicción de las observaciones futuras del fenómeno aleatorio, donde se considera la distribución marginal, esto es $P(x)$,

$$P(x) = \int P(x|\theta) P(\theta) d\theta$$

ó

$$P(x) = \sum P(x|\theta) P(\theta)$$

Esta expresión es el denominador del Teorema de Bayes y se origina al marginalizar la distribución de probabilidad conjunta:

$$P(x, \theta) = P(x|\theta) P(\theta)$$

En este sentido $P(x)$ se le conoce también, como distribución predictiva a priori o inicial, representa el conocimiento de valores actuales de x basado únicamente en la información que contiene $P(\theta)$.

Intervalo de confianza

Los intervalos de confianza se utilizan para una media, proporción o razón. Para θ_0 que se le conoce como el conglomerado de valores teniendo en cuenta que el nivel de significancia es mayor o igual que α , se le conoce como intervalo de confianza para θ al $100(1 - \alpha)\%$ (González, 2010; Cepeda et.al, 2008).

Intervalo de credibilidad

La estimación de un intervalo de cobertura (Bayesiano) es igual que la estimación clásica, es decir $100(1 - \alpha)\%$, la única diferencia que para este caso se emplea una distribución a posteriori. Estos a su vez se eligen mediante el 100% p que tiene los valores centrados entre los puntos finales (Curran,2008). El intervalo $a < \theta < b$ se le conoce como intervalo Bayesiano $100(1 - \alpha)\%$. Para θ sí cumple lo siguiente:

$$\int_{-\infty}^a \pi(\theta|x)d\theta = \int_b^{\infty} \pi(\theta|x)d\theta = \frac{\alpha}{2}$$

La estimación de los intervalos de cobertura se desarrolla por medio de una evaluación simple de las distribuciones a posteriori de los parámetros. Entonces si $\theta \in \Theta$ es una cantidad desconocida, $C \in \Theta$ es una región de $(1 - \alpha)100\%$ de credibilidad para θ si $P(\theta \in C|x) \geq 1 - \alpha$. En esta situación, $1 - \alpha$ es llamado el nivel de credibilidad (González, 2010; Cepeda et.al,2008).

Metodología

Método de obtención de los intervalos en ambos enfoques

Método Sison-Glaz

Este método se basa en dos principales vertientes, el primero es una aproximación para multinomiales probabilidades y se implementa el procedimiento de Levin (1981). La segunda opción utiliza una estructura sobre dependencia negativa inherente de la multinomial y la relación que tiene con la probabilidad, introducida por Glaz y Johnson (1984).

Método Bayesiano

Para el desarrollo y obtención de los intervalos de credibilidad, se utilizó la distribución Dirichlet como a priori, obteniendo finalmente el modelo Dirichlet- Multinomial.

Modelo Dirichlet- Multinomial

Es un modelo que puede observarse como la extensión del proceso binomial, utilizando como previa una distribución beta (Álvarez, Armero y Forte, 2017). A continuación, el proceso Dirichlet-Multinomial, se entiende de la siguiente manera:

$$P_1, \dots, P_k \sim Dir(\alpha_1, \dots, \alpha_k)$$

$$Y_1, \dots, Y_k \sim Mult(P_1, \dots, P_k)$$

Distribución a posterior

Modelo Dirichlet-Multinomial

$$f(\theta|D) \propto f(\theta, D)$$

$$= f(P_1, \dots, P_K | \alpha_1, \dots, \alpha_K) \prod_{y_i \in D} f(y_i | P_1, \dots, P_K)$$

$$\propto \prod_{j=1}^K P_j^{\alpha_j - 1} \prod_{y_i \in D} \prod_{j=1}^K P_j^{y_i^{(j)}}$$

$$= \prod_{j=1}^K P_j^{\alpha_j - 1 + \sum_{i \in D_j} y_i^{(j)}}$$

Donde: $f(\theta|D) = \text{Dir}(\alpha'_1, \dots, \alpha'_k)$

Distribución predictiva

$$\begin{aligned} f(y|D) &= \int f(y|\theta) f(\theta|D) d\theta \\ &= \int f(y|P_1, \dots, P_K) f(P_1, \dots, P_K|D) dS_K \\ &= \int \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \prod_{j=1}^K P_j^{y^{(j)}} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \prod_{j=1}^K P_j^{\alpha'_j - 1} dS_K \\ &= \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \int \prod_{j=1}^K P_j dS_K \\ &= \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \frac{\prod_{j=1}^K \Gamma(y^{(j)} + \alpha'_j)}{\Gamma(n + \sum_{j=1}^K \alpha'_j)} \end{aligned}$$

Valor esperado

$$E[n_i] = E[E[n_i|\theta_i]] = N E[\theta_i] = \frac{N\alpha_i}{\sum_{j=1}^J \alpha_j}$$

La máxima a posteriori (MAP) o moda a posteriori

$$\hat{H} = \underset{H}{\operatorname{argmax}} P(H|D) = \underset{H}{\operatorname{argmax}} P(D|H)P(H)$$

Para distribuciones H Dirichlet con parámetros α el MAP es:

$$\hat{\theta}_j = \frac{\alpha_j - 1}{\sum_{j'=1}^m (\alpha'_{j'} - 1)}$$

El MAP o moda posterior para una Dir $(N+\alpha)$ estimación de θ , es:

$$\hat{\theta}_j = \frac{N_j + \alpha_j - 1}{n + \sum_{j'=1}^m (\alpha'_{j'} - 1)}$$

Donde $\alpha = 1$; entonces el valor de $\hat{\theta}_j = \frac{N_j}{n}$ es la máxima verosimilitud estimada (MLE) para θ .

Para mostrar el procedimiento anterior se realizó una simulación usando el método Cadenas de Márkov Monte Carlo (MCMC), y se calculó distribución a posteriori, para obtener los nuevos resultados del modelo propuesto.

Cómputo Bayesiano

Los Métodos de Monte Carlo Vía Cadena de Márkov (MCMC), son métodos de simulación para generar muestras de las distribuciones a posteriores y la estimación de cantidades de interés a posteriori. En los MCMC se simulan valores sucesivamente de una densidad propuesta, que no tiene que ser necesariamente parecida a la densidad a posteriori. Cada valor generado depende solo del anterior valor simulado, de ahí la noción de cadena de Márkov (Forster et. al 1999).

Resultados

Aplicación con datos reales

Con la finalidad de mostrar el procesamiento de los métodos propuestos se empleó una base de datos sobre los resultados electorales de diputados locales en el Estado de Veracruz, 2016. El problema que se originó fue la clasificación de votos nulos que se presentaron durante el conteo de las boletas. A continuación, se muestra la clasificación de dichos votos:

Tabla 1. Clasificación de votos nulos.

Boleta depositada en la urna sin marca alguna
Marcas en toda la boleta que denotan cancelación o anulación
Múltiples marcas que no denotan preferencia alguna por las opciones en la boleta
Expresiones que denotan rechazo
Expresiones que no representan ningún sentido de elección
Marcas en 2 o más recuadros de candidatos y/o partidos no coaligados
Marca/s en recuadro de un partido político o coalición que además contiene expresiones que generan contradicción o incertidumbre sobre la preferencia marcada.
Marcas y/o expresiones en el recuadro de candidato no registrado que carece de elementos para otorgar el voto a cnr
Voto válido anulado por funcionarios de casilla

De acuerdo con las variables de la Tabla 1, se procedió a realizar los métodos empleando el software estadístico R Project, teniendo en cuenta que los datos provienen de una distribución multinomial, primero se trabajó con el enfoque frecuentista y posteriormente con el Bayesiano. A continuación, se muestran las tablas 2 y 3 en donde podemos observar los resultados de los procedimientos de ambos enfoques.

Intervalos de confianza: Para la proporción de una ~ Multinomial “Método Sison-Glaz”

Clasificación de votos nulos/ Distrito	CV1		CV2		CV3		CV4		CV5		CV6		CV7		CV8		CV9	
	<i>LI</i>	<i>LS</i>																
01 PÁNUCO	0.153	0.192	0.275	0.313	0.046	0.085	0.000	0.031	0.004	0.043	0.300	0.339	0.024	0.063	0.000	0.034	0.032	0.071
	0.172		0.294		0.065		0.012		0.023		0.319		0.044		0.015		0.051	
02 TANTOYUCAN	0.196	0.229	0.248	0.282	0.062	0.095	0.025	0.059	0.005	0.039	0.307	0.341	0.007	0.040	0.000	0.030	0.000	0.030
	0.212		0.265		0.078		0.042		0.022		0.324		0.023		0.016		0.013	
04 ÁLAMO TEMAPACHE	0.175	0.206	0.198	0.229	0.083	0.114	0.006	0.037	0.007	0.038	0.306	0.337	0.039	0.070	0.009	0.039	0.034	0.065
	0.191		0.214		0.099		0.021		0.022		0.321		0.054		0.024		0.049	
05 POZA RICA	0.031	0.068	0.291	0.329	0.096	0.133	0.023	0.060	0.019	0.056	0.019	0.056	0.000	0.036	0.002	0.040	0.023	0.060
	0.049		0.310		0.114		0.041		0.037		0.365		0.017		0.021		0.041	
06 PAPANTLA	0.107	0.138	0.153	0.184	0.037	0.068	0.000	0.029	0.000	0.028	0.526	0.557	0.013	0.044	0.000	0.027	0.030	0.061
	0.122		0.168		0.052		0.014		0.012		0.542		0.028		0.012		0.046	
12 COATEPEC	0.074	0.105	0.239	0.270	0.098	0.129	0.012	0.043	0.001	0.032	0.421	0.452	0.000	0.028	0.006	0.038	0.009	0.041
	0.090		0.254		0.113		0.027		0.016		0.436		0.012		0.022		0.025	
13 EMILIANO ZAPATA	0.065	0.103	0.385	0.424	0.008	0.047	0.037	0.075	0.000	0.038	0.309	0.348	0.000	0.033	0.001	0.039	0.025	0.064
	0.084		0.405		0.027		0.056		0.019		0.328		0.013		0.020		0.044	
14 VERACRUZ	0.013	0.048	0.320	0.355	0.160	0.194	0.032	0.064	0.036	0.070	0.243	0.277	0.000	0.030	0.000	0.033	0.043	0.078
	0.030		0.337		0.177		0.049		0.053		0.260		0.013		0.016		0.060	
15 VERACRUZ	0.023	0.055	0.367	0.399	0.151	0.184	0.052	0.084	0.030	0.062	0.180	0.213	0.001	0.033	0.017	0.049	0.029	0.062
	0.039		0.383		0.167		0.068		0.046		0.197		0.017		0.033		0.046	
16 BOCA DEL RIO	0.043	0.078	0.406	0.441	0.000	0.028	0.060	0.095	0.000	0.019	0.342	0.375	0.000	0.020	0.004	0.039	0.024	0.059
	0.060		0.422		0.011		0.077		0.016		0.358		0.002		0.022		0.041	
17 MEDELLIN	0.075	0.111	0.254	0.289	0.000	0.025	0.024	0.060	0.001	0.019	0.472	0.508	0.000	0.020	0.006	0.042	0.046	0.081
	0.093		0.271		0.010		0.042		0.001		0.490		0.002		0.024		0.063	
18 HUATUSCO	0.091	0.124	0.278	0.310	0.000	0.028	0.021	0.053	0.000	0.019	0.455	0.487	0.000	0.018	0.006	0.039	0.031	0.064
	0.107		0.294		0.011		0.037		0.002		0.471		0.002		0.023		0.047	
19 CORDOBA	0.036	0.067	0.351	0.382	0.059	0.090	0.054	0.085	0.016	0.046	0.309	0.340	0.001	0.031	0.0274	0.058	0.005	0.036
	0.051		0.367		0.074		0.069		0.031		0.324		0.016		0.042		0.020	
20 ORIZABA	0.037	0.072	0.533	0.567	0.035	0.070	0.023	0.058	0.022	0.057	0.197	0.232	0.000	0.033	0.000	0.032	0.000	0.033
	0.054		0.550		0.052		0.041		0.039		0.215		0.015		0.014		0.015	

APLICACIONES A LA METODOLOGÍA ESTADÍSTICA EN DIFERENTES DISCIPLINAS

21 CAMERINO Z. MENDOZA	0.069	0.100	0.290	0.324	0.046	0.077	0.028	0.037	0.000	0.029	0.533	0.386	0.000	0.020	0.000	0.029	0.085	0.116
	0.084		0.305		0.061		0.042		0.013		0.371		0.006		0.013		0.100	
22 ZONGOLICA	0.162	0.189	0.141	0.168	0.049	0.076	0.000	0.018	0.002	0.029	0.493	0.520	0.036	0.063	0.001	0.028	0.000	0.027
	0.175		0.154		0.063		0.004		0.015		0.507		0.050		0.014		0.014	
23 COSAMALOAPAN	0.081	0.114	0.230	0.262	0.013	0.045	0.020	0.053	0.000	0.027	0.463	0.496	0.000	0.028	0.005	0.037	0.049	0.082
	0.098		0.246		0.029		0.037		0.010		0.479		0.012		0.021		0.065	
24 SANTIAGO TUXTLA	0.100	0.134	0.240	0.273	0.020	0.054	0.000	0.025	0.004	0.037	0.376	0.410	0.009	0.043	0.000	0.028	0.111	0.145
	0.117		0.256		0.037		0.008		0.020		0.393		0.026		0.011		0.128	
25 SAN ANDRES TUXTLA	0.087	0.112	0.127	0.151	0.004	0.029	0.004	0.029	0.008	0.031	0.633	0.658	0.000	0.021	0.000	0.023	0.027	0.052
	0.100		0.139		0.016		0.016		0.020		0.646		0.008		0.011		0.039	
27 ACAYUCAN	0.100	0.144	0.249	0.292	0.094	0.137	0.000	0.040	0.008	0.051	0.353	0.396	0.004	0.043	0.000	0.036	0.011	0.054
	0.122		0.270		0.115		0.018		0.029		0.374		0.021		0.014		0.032	
28 MINATITLAN	0.073	0.110	0.298	0.325	0.085	0.122	0.009	0.046	0.010	0.046	0.346	0.382	0.000	0.036	0.000	0.036	0.022	0.095
	0.091		0.307		0.103		0.027		0.028		0.364		0.017		0.018		0.040	
29 COATZACOALCOS	0.046	0.039	0.299	0.334	0.104	0.140	0.012	0.047	0.027	0.063	0.366	0.401	0.000	0.029	0.009	0.045	0.023	0.058
	0.022		0.316		0.122		0.029		0.045		0.384		0.011		0.027		0.040	
30 COATZACOALCOS	0.029	0.066	0.232	0.269	0.085	0.122	0.016	0.052	0.021	0.057	0.401	0.437	0.004	0.040	0.000	0.033	0.047	0.084
	0.048		0.250		0.103		0.034		0.039		0.419		0.022		0.015		0.065	

Aplicación 2
Tabla 3. Método Bayesiano.

Intervalos de credibilidad: Para la proporción de una ~ Multinomial “Bayesiano”

Clasificación de votos nulos/ Distrito	CV1		CV2		CV3		CV4		CV5		CV6		CV7		CV8		CV9	
	LI	LS																
01 PÁNUCO	0.158	0.186	0.279	0.311	0.056	0.074	0.008	0.016	0.018	0.029	0.301	0.336	0.036	0.051	0.011	0.020	0.044	0.059
	0.172		0.294		0.065		0.012		0.023		0.319		0.044		0.015		0.051	
02 TANTOYUCAN	0.199	0.225	0.251	0.278	0.071	0.088	0.036	0.048	0.017	0.027	0.308	0.339	0.019	0.029	0.013	0.021	0.010	0.016
	0.212		0.265		0.078		0.042		0.022		0.324		0.023		0.016		0.013	
04 ÁLAMO TEMAPACHE	0.180	0.202	0.201	0.224	0.091	0.108	0.017	0.026	0.019	0.027	0.307	0.335	0.048	0.062	0.020	0.029	0.043	0.056
	0.191		0.214		0.099		0.021		0.022		0.321		0.054		0.024		0.049	
05 POZA RICA	0.042	0.056	0.292	0.324	0.103	0.124	0.034	0.048	0.031	0.044	0.347	0.381	0.013	0.022	0.016	0.027	0.035	0.048

	0.049	0.310	0.114	0.041	0.037	0.365	0.017	0.021	0.041								
06 PAPANTLA	0.112 0.132	0.155 0.177	0.046 0.059	0.010 0.018	0.009 0.016	0.524 0.554	0.023 0.033	0.009 0.015	0.039 0.052								
	0.122	0.168	0.052	0.014	0.012	0.542	0.028	0.012	0.046								
12 COATEPEC	0.081 0.098	0.241 0.267	0.104 0.123	0.023 0.032	0.013 0.020	0.422 0.449	0.008 0.015	0.018 0.026	0.020 0.029								
	0.090	0.254	0.113	0.027	0.016	0.436	0.012	0.022	0.025								
13 EMILIANO ZAPATA	0.072 0.093	0.384 0.419	0.022 0.033	0.048 0.065	0.014 0.024	0.309 0.343	0.009 0.018	0.015 0.025	0.037 0.052								
	0.084	0.405	0.027	0.056	0.019	0.328	0.013	0.020	0.044								
14 VERACRUZ	0.025 0.036	0.322 0.353	0.165 0.190	0.042 0.057	0.046 0.061	0.247 0.276	0.009 0.017	0.012 0.067	0.052 0.067								
	0.030	0.337	0.177	0.049	0.053	0.260	0.013	0.016	0.060								
15 VERACRUZ	0.033 0.045	0.367 0.398	0.155 0.179	0.061 0.076	0.040 0.053	0.183 0.207	0.013 0.021	0.028 0.039	0.040 0.053								
	0.039	0.383	0.167	0.068	0.046	0.197	0.017	0.033	0.046								
16 BOCA DEL RIO	0.053 0.068	0.406 0.437	0.008 0.015	0.068 0.086	0.000 0.003	0.342 0.374	0.001 0.004	0.017 0.026	0.035 0.048								
	0.060	0.422	0.011	0.077	0.016	0.358	0.002	0.022	0.041								
17 MEDELLIN	0.082 0.102	0.256 0.285	0.007 0.014	0.035 0.049	0.000 0.003	0.472 0.505	0.000 0.004	0.019 0.029	0.055 0.072								
	0.093	0.271	0.010	0.042	0.001	0.490	0.002	0.024	0.063								
18 HUATUSCO	0.098 0.117	0.279 0.307	0.088 0.015	0.032 0.043	0.001 0.004	0.455 0.486	0.001 0.004	0.018 0.028	0.041 0.054								
	0.107	0.294	0.011	0.037	0.002	0.471	0.002	0.023	0.047								
19 CORDOBA	0.045 0.058	0.353 0.381	0.066 0.081	0.062 0.077	0.027 0.037	0.310 0.337	0.013 0.020	0.037 0.049	0.017 0.024								
	0.051	0.367	0.074	0.069	0.031	0.324	0.016	0.042	0.020								
20 ORIZABA	0.047 0.062	0.531 0.567	0.045 0.060	0.034 0.048	0.033 0.048	0.200 0.228	0.011 0.020	0.011 0.019	0.011 0.020								
	0.054	0.550	0.052	0.041	0.039	0.215	0.015	0.014	0.015								
21 CAMERINO Z. MENDOZA	0.076 0.093	0.292 0.319	0.054 0.068	0.036 0.048	0.010 0.017	0.355 0.384	0.004 0.008	0.010 0.017	0.091 0.108								
	0.084	0.305	0.061	0.042	0.013	0.371	0.006	0.013	0.100								
22 ZONGOLICA	0.164 0.185	0.145 0.164	0.057 0.069	0.003 0.006	0.012 0.019	0.492 0.520	0.045 0.052	0.011 0.017	0.011 0.017								
	0.175	0.154	0.063	0.004	0.015	0.507	0.050	0.014	0.014								
23 COSAMALOAPAN	0.089 0.107	0.232 0.258	0.024 0.034	0.030 0.042	0.007 0.014	0.463 0.493	0.009 0.015	0.016 0.025	0.058 0.073								
	0.098	0.246	0.029	0.037	0.010	0.479	0.012	0.021	0.065								
24 SANTIAGO TUXTLA	0.107 0.128	0.242 0.271	0.032 0.043	0.005 0.011	0.016 0.026	0.375 0.406	0.021 0.031	0.007 0.014	0.117 0.138								
	0.117	0.256	0.037	0.008	0.020	0.393	0.026	0.011	0.128								
25 SAN ANDRES TUXTLA	0.092 0.107	0.129 0.147	0.014 0.020	0.013 0.020	0.017 0.024	0.632 0.657	0.006 0.011	0.008 0.014	0.034 0.045								
	0.100	0.139	0.016	0.016	0.020	0.646	0.008	0.011	0.039								
27 ACAYUCAN	0.109 0.135	0.252 0.287	0.105 0.129	0.013 0.024	0.022 0.037	0.353 0.392	0.016 0.028	0.010 0.020	0.024 0.039								
	0.122	0.270	0.115	0.018	0.029	0.374	0.021	0.014	0.032								
28 MINATITLAN	0.082 0.101	0.291 0.325	0.092 0.114	0.022 0.033	0.022 0.034	0.346 0.380	0.013 0.022	0.013 0.022	0.034 0.048								
	0.091	0.307	0.103	0.027	0.028	0.364	0.017	0.018	0.040								
29 COATZACOALCOS	0.017 0.027	0.301 0.330	0.112 0.133	0.024 0.035	0.037 0.052	0.367 0.399	0.008 0.015	0.022 0.033	0.034 0.047								
	0.022	0.316	0.122	0.029	0.045	0.384	0.011	0.027	0.040								
30 COATZACOALCOS	0.041 0.056	0.235 0.266	0.091 0.113	0.028 0.040	0.033 0.046	0.402 0.438	0.017 0.028	0.011 0.020	0.057 0.074								
	0.048	0.250	0.103	0.034	0.039	0.419	0.022	0.015	0.065								

En la evaluación de las tablas mencionadas anteriormente se puede observar las siguientes, características notables de ambos métodos, véase la Tabla 4.

Tabla 4. Características: Métodos para obtener los intervalos.

Características	Sison & Glaz	Bayesiano
Mayor amplitud (Li-Ls)	●	
Menor amplitud (Li-Ls)		●
Resultados mejor presentados		●
Parámetro contenido		●
Valores Ls y Li similares	●	
Procedimiento en el software	●	●

Como podemos observar la tabla 4 es un resumen general de lo que se observó en las Tablas 2 y 3, teniendo en cuenta que el método Sison y Glaz (1995) presentan características más débiles a comparación del método bayesiano, para esta aplicación con datos reales, se comprueba que el método más funcional para este tipo de estudio es el Bayesiano, sin embargo, no se puede concluir que el método frecuentista no fue funcional para este ejemplo, debido a que presentan resultados aceptables. Ambos métodos son procedimientos que se adaptan de manera eficaz, precisa y rápida en el software empleado.

Para esta investigación fue relevante mostrar los códigos en el software R, como parte de los resultados, esto con la finalidad de que los usuarios que tengan problemas adaptados a estos métodos estadísticos puedan comprender el procedimiento implementado en la parte computacional, a continuación, se muestra el proceso y describe cada sección:

Código en R Project

```
#INTERVALO DE CONFIANZA "método sisonglaz"#
##Paqueterías y librerías##
install.packages("EMT")
install.packages("DescTools")
library (EMT)
library (DescTools)
##Declarar las variables y sus categorías##
VOTOSCN=c("C1V", "C2V", "C3V", "C4V", "C5V", "C6V", "C7V", "C8V", "C9V")
##Colocar los valores esperados y observados##
observados = c (151,788,326,108,123,1318,71,49,207)
esperados = c (1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9)
#Obtener el IC, Colocar el nivel de significancia y el método#
MCI=MultinomCI(observados,conf.level=0.95,method="sisonglaz")
MCI
```

Para el desarrollo del método Sison-Glaz se emplearon dos paquetes con sus respectivas librerías, ya que son necesarias para este procedimiento:

EMT: Prueba multinomial exacta: Prueba de bondad de ajuste para datos discretos multivariados (Menzel,2015).

DescTools: Herramientas para estadística descriptiva, en este paquete se encuentra la función para obtener el IC mediante Sizon -Glaz para la proporción de una distribución multinomial (Signorell, *et.al*, 2021).

```
##INTERVALO DE CREDIBILIDAD##
#Paqueterías y librería#
install.packages("Bolstad2")
library(Bolstad2)
install.packages("MCMCpack")
library(MCMCpack)
require(coda)
#Procedimiento para encontrar la distribución a posteriori#
posterior<-
MCMultinomdirichlet(c(151,788,326,108,123,1318,71,49,207),
c(1,1,1,1,1,1,1,1,1), mc=1000)
colnames(posterior)<-
c("VN1","VN2","VN3","VN4","VN5","VN6","VN7","VN8","VN9")
posterior
##Resumen general: puedes ver los IC y otras estadísticas
descriptivas#
summary(posterior)
#Despliega solamente los IC#
BIC<-HPDinterval(posterior, prob = 0.95)
BIC
```

Ahora se describen las paqueterías que fueron requeridas para el proceso de obtener los intervalos de credibilidad.

Bolstads: Funciones de Bolstads, en esta paquetería se encuentra el desarrollo para encontrar los intervalos de credibilidad (Curran,2015).

MCMCpack: Cadena de Márkov-Monte Carlo paquete, en esta función se encuentra la instrucción para obtener y encontrar la distribución a posteriori (Martín *et.al*, 2021).

Coda: Diagnóstico vía *MCMC*: contiene otra alternativa para obtener IC bayesiano (Plummer *et.al*, 2020).

Conclusiones

Se desarrollaron los métodos para obtener intervalos de confianza y cobertura en una distribución multinomial. El método Bayesiano fue el más preciso para esta aplicación, debido a que la amplitud de los intervalos de cobertura es menor y se considera que la estimación es más precisa. El método Sison y Glaz (Intervalo de confianza), es funcional para la aplicación, a diferencia del Bayesiano aquí la amplitud fue mayor, lo que indica menor seguridad del valor de la estimación. Se presentaron intervalos similares en ambos enfoques (en dos situaciones). Ambos enfoques son muy accesibles al manejarlos en el software R. El código en R está adaptado para que los usuarios que lo requieran solo puedan reemplazar por sus datos y así obtener esos intervalos de confianza que se requiera.

Referencias

- Álvarez, D. Armero, C. y Forte A. (2017). What Does Objective Mean in a Dirichlet-multinomial process. *International statistical review*. Valencia, Spain.
- Alvarado S. (2014). Aportes metodológicos en la estimación de tamaños de muestra en estudios poblacionales de prevalencia. (Tesis). Doctorado en Ciencia. Universidad Autónoma de Barcelona. Facultad de Medicina.
- Box, G. y Tiao, G. (1973). *Bayesian Inference In Statistical Analysis*. Massachusetts/ Addison-Wesley Publishing Company.
- Bernardo, J. (2003). Bayesian Statistics. Probability and Statistics (R. Viertl, ed) of the encyclopedia of life support systems (EOLSS). Oxford, UK.
- Canavos, G., (1988). *Probabilidad y estadística: aplicaciones y métodos*. México D.F., McGraw Hill.
- Cepeda, E., Aguilar, W., Cervantes, V., Corrales, M., Díaz., Rodríguez, D. (2008). Intervalos de confianza de credibilidad para una proporción. *Revista Colombiana de Estadística*.
- Correa J. y Barrera, C. (2018). *Introducción a la estadística Bayesiana*. Medellín, Colombia: Instituto Tecnológico Metropolitana.
- Curran, J. (2015). Packages Bolstad2. Versión 1.0-28. R function and data sets for the book.
- Curran, J. (2008). Introducción a intervalos creíbles bayesianos para el error en muestras y en los perfiles de ADN. *Revista de probabilidad y riesgo de ley*. Estados Unidos Americanos.
- Forster, J. Mcdonald, J. y Smith, P. (1999). Markov chain Monte carlo exact inference for binomial and multinomial logistic regression models. *Statistics and Computing*. UK.
- Glaz, J. y Johnson, B. (1984). Probability for Multivariate Distribution with Dependence Structures', *Journal of the American Statistical Association* 79, 411–436.
- González D. (2010). Comparación de intervalos de confianza para la distribución multinomial. (Tesis). Magister en Estadística. Universidad Nacional de Colombia, Facultad de Ciencias.
- González D. Correa, J., Vélez, J. (2015). Comparación de 13 intervalos de confianza para los parámetros de la distribución multinomial. *Revista Facultad de ciencias (Universidad Nacional de Colombia)*. Sede Medellín.
- Levin, B. (1981). A Representation for Multinomial Cumulative Distribution Functions, *The Annals of Statistics* 9(429), 1123–1126.
- Lin, J. (2016). On The Dirichlet Distribution. A report submitted to the Department of Mathematics and Statistics. (Tesis Maestría). Queen's University Kingston, Ontario, Canada.
- Martín, D. *et al.* (2021). Packages MCMCpack. Versión 1.5.0. R function and data sets for the book.
- Menzel, U. (2015). Packages EMT. Versión 1.1. R function and data sets for the book.
- Olmos, Z., R. (2016). *Paradigmas de la inferencia estadística: Un enfoque Bayesiano*. (Trabajo práctico educativo). Especialización en métodos estadístico. Universidad Veracruzana. Facultad de Estadística e Informática.
- Plummer, M. *et al.* (2020). Packages Coda. Versión 0.19.4. R function and data sets for the book.
- R Equipo Central (2020). *R: Un lenguaje y un entorno para la informática estadística*. Fundación R para la Computación Estadística, Vienna, Austria.

- Singorell, A. *et al.* (2021). Packages DescTools. Versión 0.99.41. R function and data sets for the book.
- Sison, C. y Glaz, J. (1995). Simultaneous Confidence Intervals and Sample Size Determination for Multinomial Proportions, *Journal of the American Statistical Association*, 90:429, 366-369.
- Tortora, R. (1978). Nota sobre la estimación del tamaño de la muestra para poblaciones multinomiales. *Revista Americana de estadística*. Estados Unidos Americanos.
- Wang, H. (2007). coeficientes de confianza exactos e intervalos de confianza para proporciones multinomiales. *Revista de estadística multivariante*. Estados Unidos Americanos.

Anexos

Distribuciones a priori y posteriori

Azul = a priori

Verde = a posteriori

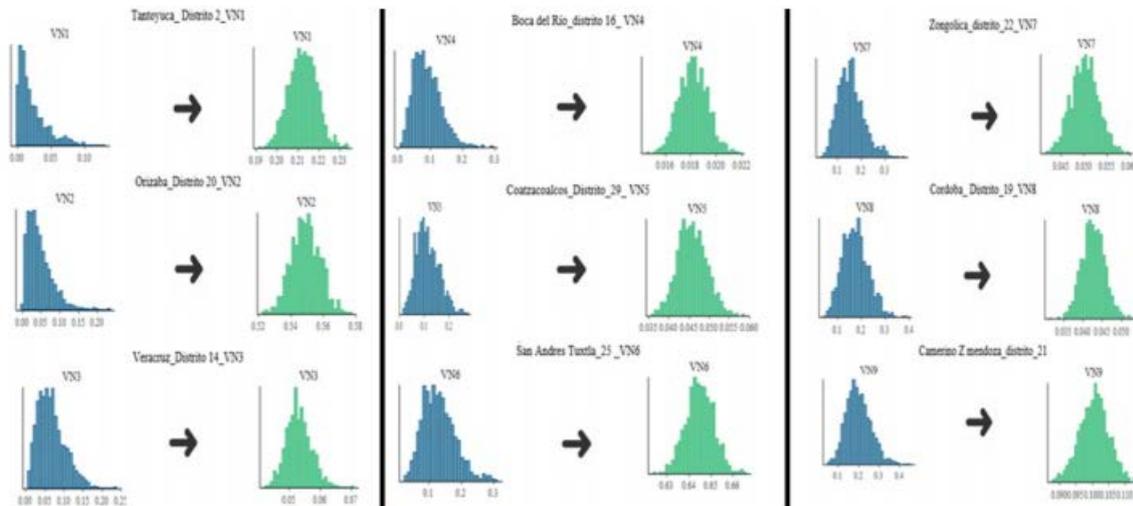


Figura 1. Representación de las distribuciones empleadas.

Influencia de la actividad física en sobrevivientes de cáncer de mama y colorrectal

Alarcón Rojas Carlos Adrián¹, Menéndez Acuña Ernesto Pedro², Montano Rivas Julia Aurora³,
Álvarez Bañuelos María Teresa⁴.

¹ Facultad de Estadística e Informática. Universidad Veracruzana. cadalro24@gmail.com

² Facultad de Matemáticas. Universidad Veracruzana. emenedez@uv.mx

³ Facultad de Estadística e Informática. Universidad Veracruzana. julmontano@uv.mx

⁴ Instituto de Salud Pública. Universidad Veracruzana. talvarez@uv.mx

Resumen

El cáncer de mama y el colorrectal se han constituido a nivel mundial, nacional y local en un problema sanitario prioritario y creciente. Siendo las neoplasias más frecuentes y una de las primeras causas de muerte en la población. Conocida por ser una enfermedad de comportamiento heterogéneo, multicausal y de gran variabilidad geográfica en el comportamiento epidemiológico debido al efecto combinado de los diferentes factores de riesgo que influyen en los factores pronósticos. El objetivo de este trabajo es conocer la relación entre los factores pronósticos del cáncer de mama y del colorrectal asociados con los determinantes sociodemográficos que influyen en la actividad física de los supervivientes de dichas neoplasias en el Estado de Veracruz, mediante la identificación de perfiles de comportamiento en los que se pueda incidir. La aplicación del análisis de correspondencia múltiple como herramienta para observar patrones entre la población estudiada permitió conocer los diferentes perfiles que integran a los individuos, con la finalidad de identificar puntos donde sea posible el desarrollo de estrategias que modifiquen dichas conductas.

Palabras clave: Actividad física, Análisis de Correspondencia Múltiple, Cáncer colorrectal, Cáncer de mama, Sedentarismo.

Introducción

De acuerdo con las proyecciones basadas en GLOBOCAN 2018, se predice un aumento de 19.3 millones de casos nuevos de cáncer para el año 2025. Más de las dos terceras partes de las defunciones por cáncer se producen en países de ingresos bajos y medianos, donde sobresale el cáncer de mama y el colorrectal, como una de las principales causas de muerte en mujeres y hombres (Bray et al., 2018a; Siegel et al., 2016).

En el mundo, 2.1 millones de mujeres fueron diagnosticadas con cáncer de mama (CM); la incidencia de este tipo de cáncer ha aumentado desde 2008 a 2018 en más de un 20%, mientras que la mortalidad incrementó en un 14% (Bray et al., 2018a).

Además, se diagnosticaron 1.8 millones de nuevos casos de cáncer colorrectal (CCR), con un estimado de 800,000 muertes en 2018. Considerando así a esta patología el tercer cáncer con mayor incidencia y el segundo de mayor mortalidad (Bray et al., 2018b).

El perfil epidemiológico de estas problemáticas de salud se encuentra en constante cambio. Se estima que del 50 a 60% de todos los casos de cáncer de mama, en México, se detectan

en etapas avanzadas y solo el 22% de las mujeres reportan tener una mamografía en el último año (Reynoso-Noverón & Torres-Domínguez, 2018). Mientras tanto, el cáncer colorrectal ha registrado un incremento en la tasa de mortalidad en el periodo de 1980 a 2008, pasando de 0.9 a 3.1 por 100 000 habitantes, respectivamente.

Por otro lado, la Encuesta Nacional de Salud y Nutrición de Medio Camino, en 2016, reporta que la prevalencia nacional de sobrepeso y obesidad ha aumentado 9.5%, en comparación con 2012 (Shamah-Levy et al., 2018). La Organización para la Cooperación y el Desarrollo Económico (OCDE) informó este año que, una de cada cuatro personas es obesa y un 72.6% de la población mexicana tiene sobrepeso, como resultado, el impacto en la economía es evidente: las enfermedades asociadas con el sobrepeso representan el 8.4% del gasto en salud y muestra una reducción en la expectativa de vida entre todos los países analizados (OECD, 2019).

Por otra parte, estudios han asociado un menor riesgo a desarrollar cáncer cuando se realiza ejercicio intenso. Del mismo modo, el realizar actividad física durante al menos 150 minutos por semana reduce los niveles de obesidad (Jurado et al., 2015).

Por este motivo, la Sociedad Americana de Oncología Clínica ha establecido una iniciativa polifacética para hacer frente a la carga compartida entre la obesidad y pacientes con cáncer, considerando el incremento del riesgo en la mortalidad a causa del sobrepeso y la obesidad en varias enfermedades, incluyendo el CM y CCR (Ligibel et al., 2014). La relación entre el índice de masa corporal (IMC) y las neoplasias, muestra que la obesidad se ha asociado con una peor supervivencia.(Chan et al., 2014) Sin embargo, se han encontrado conclusiones contradictorias, modificadas por la población de estudio (Amadou et al., 2014; Sedjo et al., 2014).

En relación con la actividad física, se ha reportado evidencia que apoya su realización como un factor que influye en el desarrollo de CCR (Lauby-Secretan et al., 2016). Por el contrario, el riesgo de CM y su modificación por los niveles de actividad han mostrado resultados menos concluyentes (Dieli-Conwright et al., 2020).

Como se ha mencionado anteriormente, en el desarrollo del CCR y CM intervienen diversos factores que a su vez se relacionan, por ello, su etiología es complicada; dichos elementos se dividen en ambientales y genéticos (Bray & Piñeros, 2016). Algunos componentes ambientales incluyen el tipo, cantidad y calidad de la dieta, estilos de vida y uso de terapia de restitución hormonal. Por otro lado, las causas genéticas contemplan a la predisposición hereditaria (Jochems et al., 2017; Zorrón Cheng Tao Pu et al., 2020).

Existe controversia entre la influencia de diversos factores pronósticos en la supervivencia, principalmente modificados por el tipo de población estudiada, tales como: la edad, padecer obesidad y obesidad mórbida al diagnóstico, factores socioeconómicos, entre otros (Alarcón Rojas et al., 2019; Wang et al., 2019).

Metodología

Se tomaron los pacientes con cáncer colorrectal y sin ningún tipo de cáncer del estudio de Ramírez-Díaz (2017), y del estudio de Álvarez-Bañuelos et al. (2020) los que padecen cáncer de mama; por lo tanto, los datos de interés corresponden al periodo 2017 - 2018.

Se realizaron estadísticas descriptivas, tales como: tablas de frecuencias y gráficos con la finalidad de comparar las proporciones de acuerdo con los factores sociodemográficos, estilos de vida y comorbilidades según el tipo de cáncer en supervivientes en la población de Veracruz.

Se aplicó estadística inferencial para comparar las proporciones al 95% de confianza de los niveles de actividad física de acuerdo con la población estudiada.

Se buscaron patrones de comportamiento aplicando el análisis de correspondencia múltiple entre los tipos de cáncer de acuerdo con su actividad física y los factores sociodemográficos, estilos de vida y comorbilidades.

Los datos fueron analizados en los paquetes estadísticos SPSS versión 24 y R versión 4.0.2.

Resultados

Se obtuvo una cohorte inicial de 535 pacientes, siendo el 45.05% de CM, el 18.32% de CCR y 36.63% de población sin cáncer (PSC). La edad media del diagnóstico fue de 51 ± 11 años para cáncer de mama y 56 ± 14 años para cáncer colorrectal y población sin cáncer. En ambos tipos de cáncer y la población sin cáncer, la edad mínima fue muy cercana (21 y 22 años), mientras que los pacientes con mayor edad (86 años) tienen diagnóstico de CCR o son población sin cáncer.

La zona de procedencia rural fue más frecuente en las pacientes con cáncer de mama, observando que las zonas vecinas a los puntos más densamente poblados presentaron frecuencias mayores (Figura 1). Por otro lado, los pacientes con cáncer de colon y sin cáncer provinieron en su mayoría de zonas urbanas, 62.2% y 74.5% respectivamente (Tabla 1).

De acuerdo con la realización de actividad física vigorosa, destaca que solo el 25.2% la llevan a cabo al menos 10 minutos diarios y quienes realizan mayor actividad son las pacientes con diagnóstico de CM (9.35%), correspondiendo al doble que CCR. Además, sobresale que en los pacientes sin cáncer la realización de actividad vigorosa no muestra un contraste elevado.

Respecto al sexo, la población con diagnóstico de cáncer de mama se encuentra compuesta en su totalidad por mujeres, mientras que los pacientes con cáncer colorrectal y sin cáncer comparten proporciones en hombres y mujeres, 59.2% y 40.8%, respectivamente (Tabla 1).



Fuente: Elaboración propia

Figura 1. Mapa de frecuencias de la procedencia de las pacientes con diagnóstico de cáncer de mama en el Estado de Veracruz.

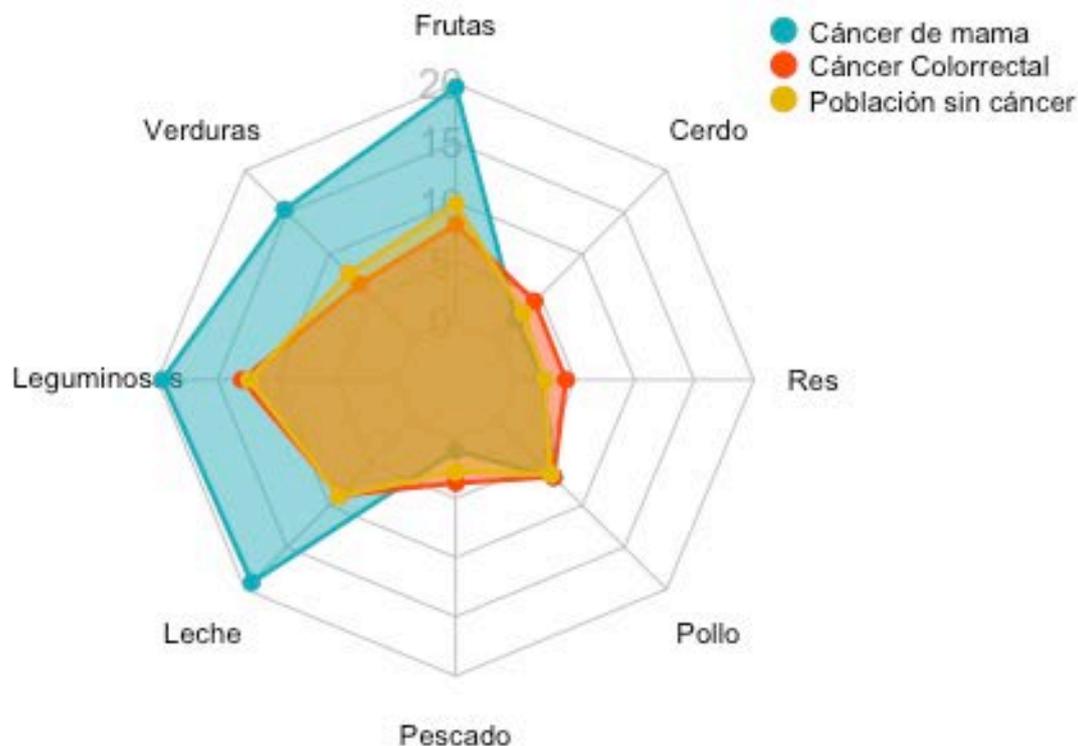
En relación con el nivel de escolaridad, en las tres poblaciones estudiadas se distribuye mayormente en educación básica o menor (CM 93.3% vs CCR/PSC 87.8%). Mientras tanto, el estudio de nivel socioeconómico ubica a las pacientes con cáncer de mama en nivel bajo (98.7%) y a los pacientes de cáncer colorrectal en alto (87.8%), siendo los pacientes sin cáncer quienes tuvieron los niveles más altos (88.3%) (Tabla 1). Además, el estado nutricional por índice de masa corporal muestra que las mujeres con CM presentan mayores niveles de sobrepeso, obesidad u obesidad mórbida, mientras que la población de CCR tiene una tendencia al bajo peso o Normopeso. Por otro lado, los pacientes sin cáncer tienen mayor tendencia al Normopeso. Así mismo, el desarrollo de patologías asociadas como diabetes e hipertensión tiene un comportamiento similar en las tres poblaciones, 15.8% vs 13.3% vs 15.3% y 19.5% vs 22.4% vs 18.9%, respectivamente (Tabla 1).

Tabla 1. Características sociodemográficas, estilos de vida y patológicas por tipo de cáncer de 535 pacientes atendidos en el Centro Estatal de Cancerología y Centro de Alta Especialidad.

	Cáncer de mama		Cáncer colorrectal		Población sin cáncer		<i>p</i>
	No.	%	No.	%	No.	%	
Sexo							
Masculino	0	0.0	58	59.2	116	59.2	<0.001
Femenino	241	100	40	40.8	80	40.8	
Nivel Socioeconómico							
Bajo	238	98.7	12	12.2	23	11.7	<0.001
Alto	3	1.2	86	87.8	173	88.3	
Lugar de procedencia							
Rural	151	62.7	37	37.8	50	25.5	<0.001
Urbano	90	37.7	61	62.2	146	74.5	
Escolaridad							
Analfabeta	23	9.5	27	27.6	39	19.9	<0.001
Básica	202	83.8	59	60.2	133	67.9	
Media o superior	16	6.6	12	12.2	24	12.2	
IMC							
Bajo peso	0	0.0	33	33.7	26	13.3	< 0.001
Normopeso	98	40.7	49	50.0	107	54.6	
Sobrepeso	93	38.6	13	13.3	51	26	
Obesidad	50	20.7	3	3.1	12	6.1	
Diabetes Mellitus Tipo 2							
No	203	84.2	85	86.7	166	84.7	0.841
Si	38	15.8	13	13.3	30	15.3	
Hipertensión arterial							
No	194	80.5	76	77.6	159	81.1	0.759
Si	47	19.5	22	22.4	37	18.9	
Consumo de tabaco							
No	222	92.1	61	62.2	116	59.2	<0.001
Si	19	7.9	37	37.8	80	40.8	
Consumo de alcohol							
No	204	84.6	32	37.7	81	41.3	<0.001
Si	37	15.4	66	66.3	115	58.7	

Fuente: Elaboración propia

Al comparar el consumo de porciones promedio por semana por tipo de alimento, se observa que las pacientes con cáncer de mama tienen mayor tendencia a consumir frutas y verduras, mientras que los pacientes con cáncer de colon y la población sin cáncer presentan comportamiento alimenticio similar, con un alto consumo de proteínas animales (Figura 2).



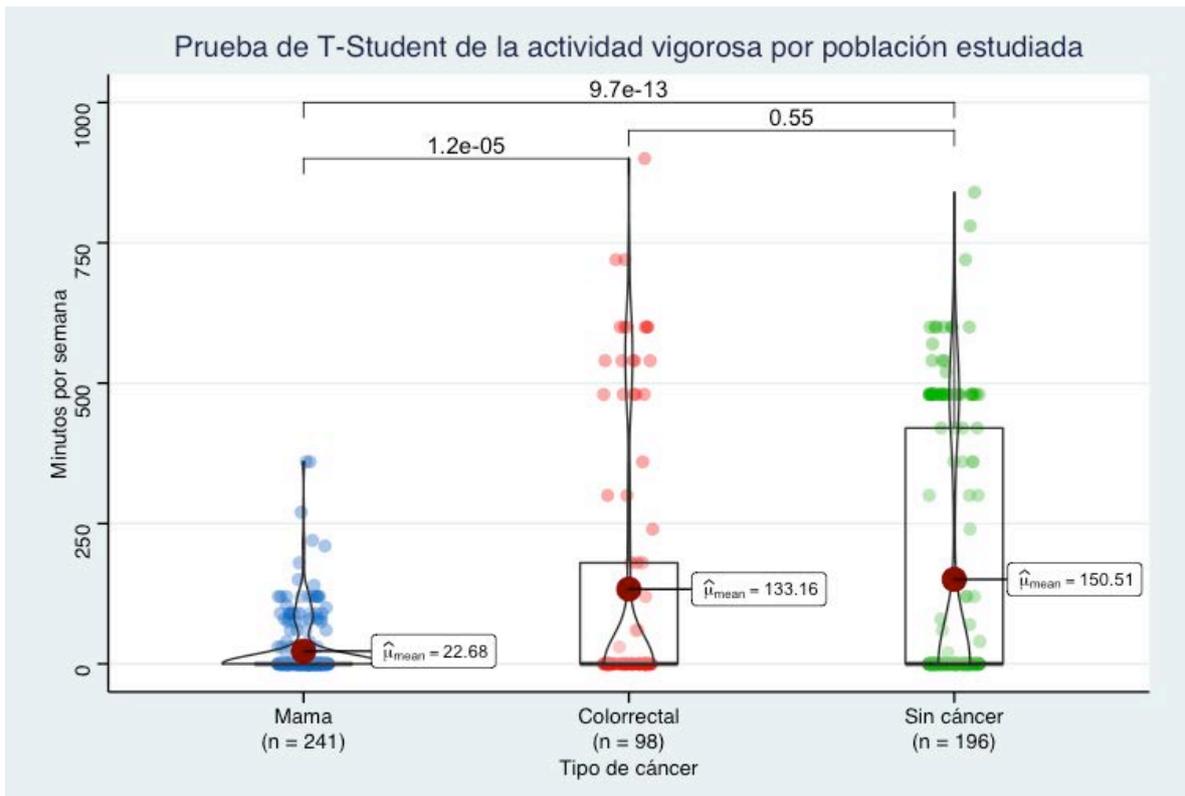
Fuente: Elaboración propia

Figura 2. Diagrama de radar del consumo promedio de porciones alimenticias en la población con cáncer de mama, colorrectal y sin cáncer.

Mediante la prueba de Kolmogorov-Smirnov se realizó la comprobación del supuesto de normalidad para el tiempo de actividad física con intensidad vigorosa y moderada, rechazando que los datos se distribuyan de forma normal en los diferentes tiempos analizados ($p < 0.001$).

En relación con la homocedasticidad, se comprobó el supuesto de homogeneidad de varianzas mediante la prueba de Levene, de acuerdo con el tiempo de actividad física para las tres poblaciones, rechaza la hipótesis de que los datos presenten homogeneidad de varianzas ($p < 0.001$).

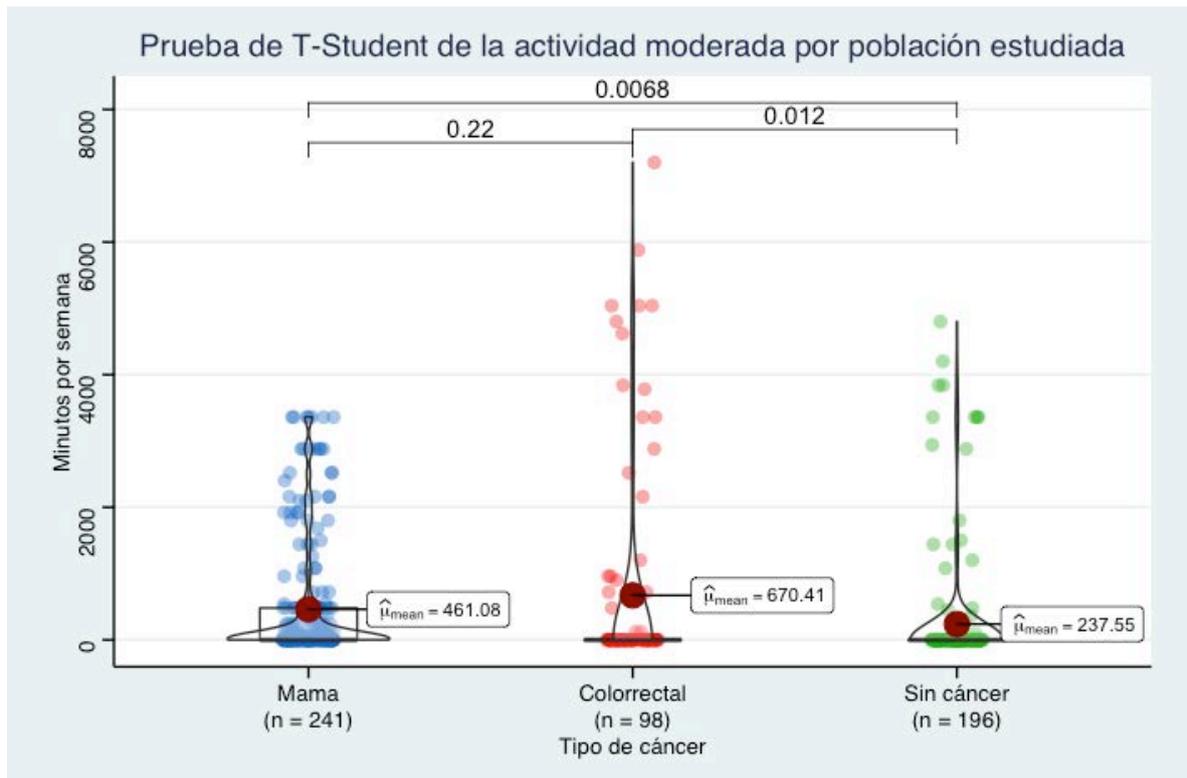
Con la información obtenida en la comprobación de supuestos, se consideró el uso de métodos no paramétricos para la comparación de la actividad física. Pese a esto, tomando en cuenta el tamaño de la muestra y el teorema central del límite, se decidió utilizar una prueba T de student para la comparación de medias entre los grupos.



Fuente: Elaboración propia

Figura 3. Gráficos de violines y cajas con prueba de T-Student entre grupos de acuerdo con el tiempo de actividad física vigorosa por semana.

Al comparar el tiempo de actividad física de intensidad vigorosa se rechazó igualdad de medias entre los tipos de cáncer y el cáncer de mama con la población sin cáncer, observándose una mayor cantidad de pacientes con cáncer de colorrectal que realizaban tiempos más elevados, respecto con los pacientes con cáncer de mama. La población sin cáncer no tuvo suficiente evidencia estadística para rechazar la igualdad de medias entre los pacientes con cáncer colorrectal (Figura 3).

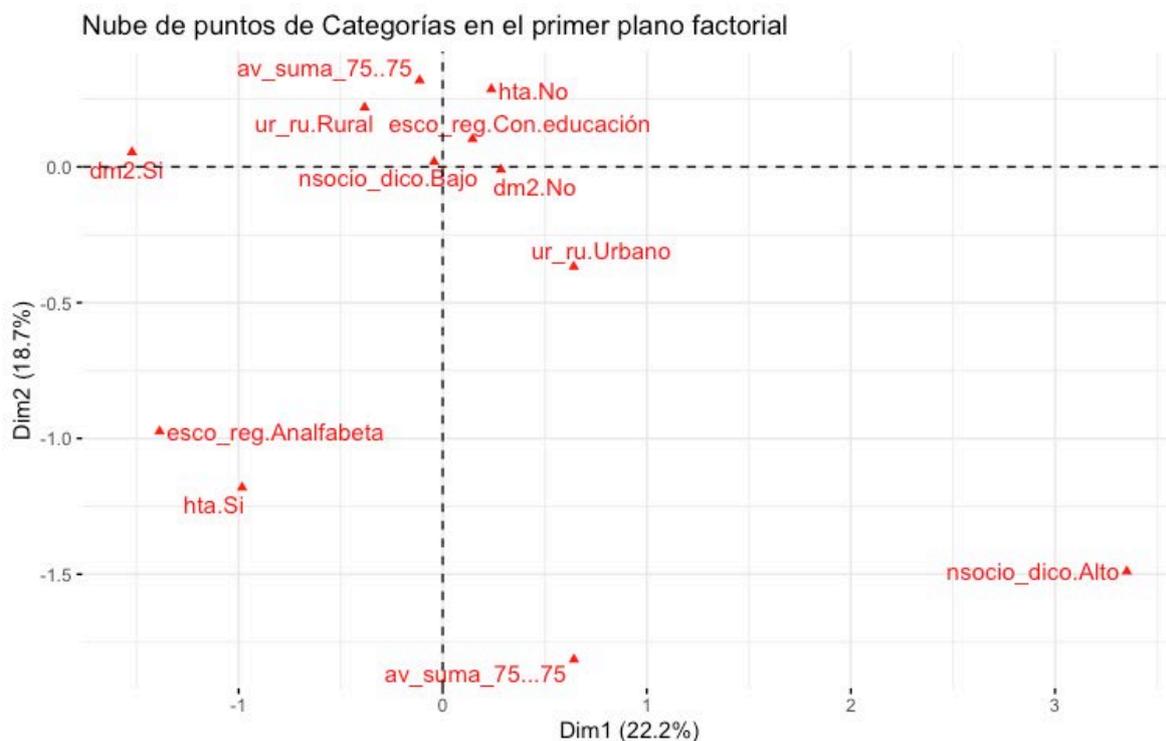


Fuente: Elaboración propia

Figura 4. Gráficos de violines y cajas con prueba de T-Student entre grupos de acuerdo con el tiempo de actividad física moderada por semana.

En cuanto a la actividad física moderada, la población con cáncer de mama mostró una mayor distribución. Por otro lado, los pacientes con cáncer de colon presentaron tiempos mayores. Pese a esto, no existe diferencia significativa entre los tiempos de actividad moderada de los dos cánceres, pero sí muestran una diferencia al compararse con la población sin cáncer. (Figura 4)

El análisis de correspondencia múltiple se realizó para las tres poblaciones de acuerdo con el cumplimiento o no del tiempo recomendado por semana de actividad física según intensidad (75 minutos por semana de actividad vigorosa y 150 minutos por semana de actividad moderada), se introdujeron las variables sociodemográficas de escolaridad, zona de procedencia, nivel socioeconómico y patologías asociadas.

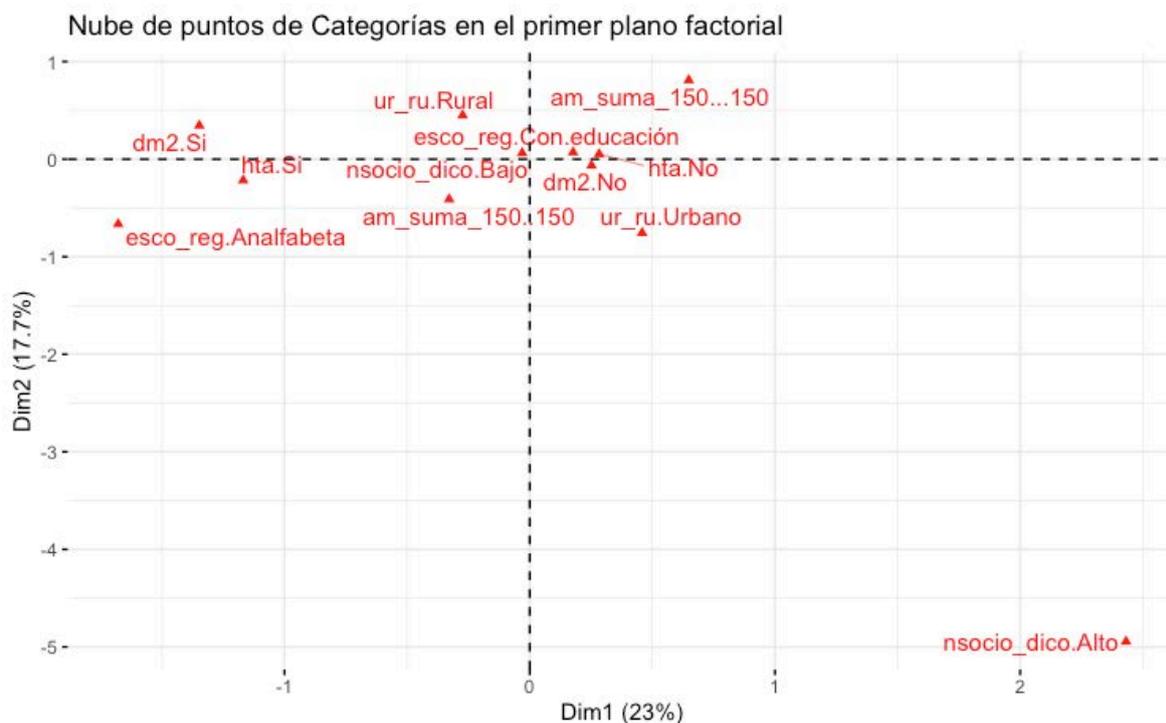


Fuente: Elaboración propia

Figura 5. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes con cáncer de mama de acuerdo con el tiempo de actividad física vigorosa.

Las pacientes con cáncer de mama que realizan menos del tiempo recomendado de actividad física vigorosa por semana tuvieron una mayor tendencia a ser de estratos socioeconómicos bajos, proceder de zonas rurales y no padecer diabetes mellitus tipo 2 o hipertensión (Figura 5).

Por otro lado, las pacientes con cáncer de mama quienes siguen la recomendación de tiempo de actividad con intensidad moderada fueron principalmente aquellas con algún nivel educativo, sin comorbilidades y de procedencia urbana, observándose que, el nivel socioeconómico alto no tuvo gran interacción posiblemente por la distribución de la población (Figura 6).

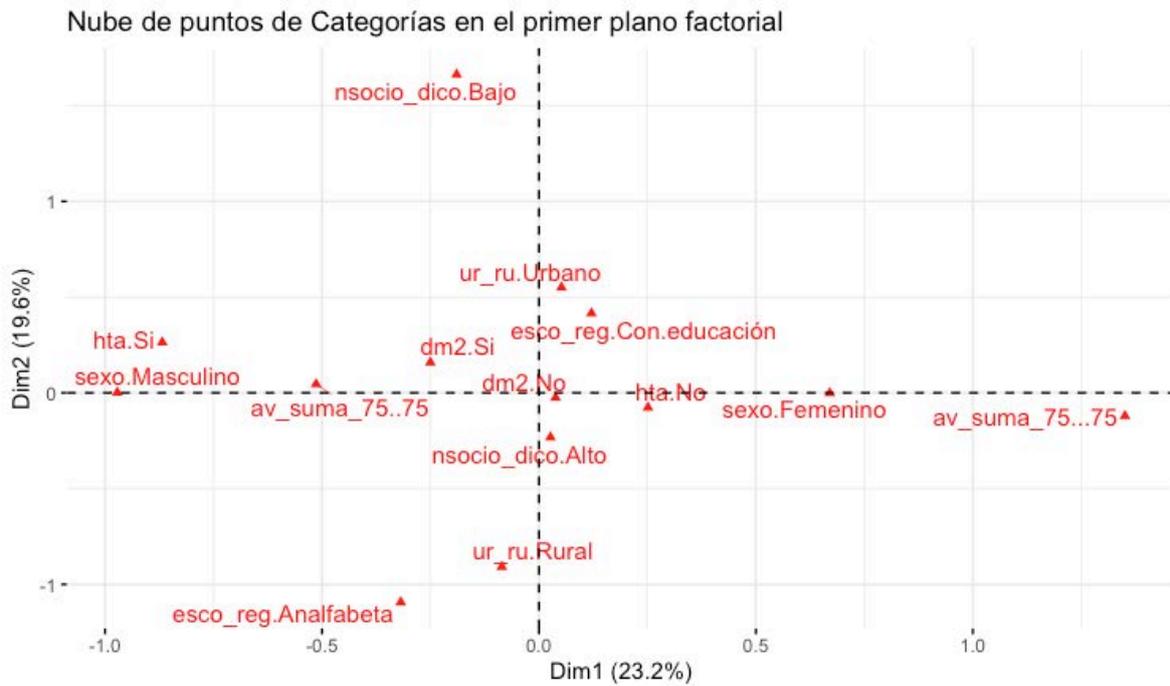


Fuente: Elaboración propia

Figura 6. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes con cáncer de mama de acuerdo con el tiempo de actividad física moderada.

Los pacientes con cáncer colorrectal de acuerdo con la actividad física de intensidad vigorosa mostraron un patrón de comportamiento más relacionado con el sexo, donde las mujeres fueron quienes realizaron más del tiempo recomendado de actividad vigorosa por semana pertenecientes a estratos socioeconómicos altos, procedentes de zonas urbanas, con algún grado educativo y no padecer diabetes mellitus tipo 2 o hipertensión (Figura 7).

En contraste, los pacientes con cáncer colorrectal, quienes siguen la recomendación de tiempo de actividad con intensidad moderada, fueron principalmente hombres con analfabetismo, con tendencia a padecer alguna comorbilidad y de procedencia rural, mayormente de nivel socioeconómico alto (Figura 8).



Fuente: Elaboración propia

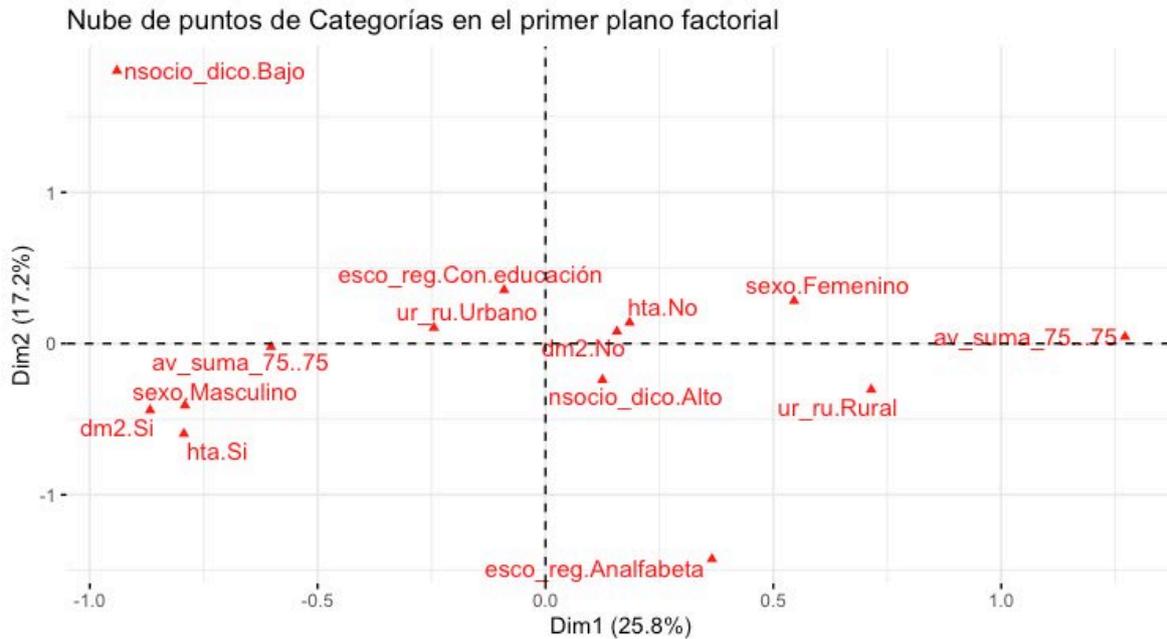
Figura 7. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes con cáncer colorrectal de acuerdo con el tiempo de actividad física vigorosa.



Fuente: Elaboración propia

Figura 8. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes con cáncer colorrectal de acuerdo con el tiempo de actividad física moderada.

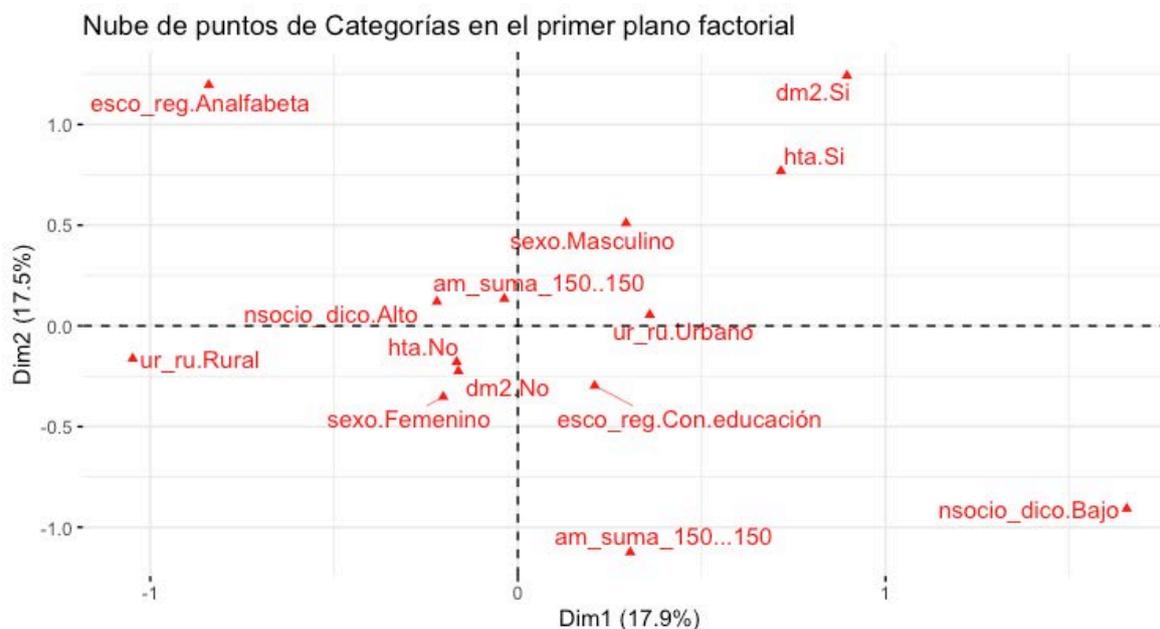
La población sin cáncer muestra un comportamiento influenciado por el sexo, observando que quienes realizan menos del tiempo recomendado de actividad física vigorosa por semana fueron principalmente hombres, además tuvieron una mayor tendencia a ser de estratos socioeconómicos bajos, procedentes de zonas urbanas y con padecimientos crónicos como diabetes mellitus tipo 2 o hipertensión (Figura 9).



Fuente: Elaboración propia

Figura 9. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes sin cáncer de acuerdo con el tiempo de actividad física vigorosa.

Sin embargo, las mujeres del grupo que no padece cáncer con un nivel socioeconómico alto, sin patologías crónicas, provenientes de zonas rurales, con tendencia al analfabetismo fueron principalmente aquellas que realizaron menos del tiempo por semana recomendado de actividad física con intensidad moderada (Figura 10).



Fuente: Elaboración propia

Figura 10. Análisis de Correspondencia Múltiple de las características sociodemográficas y patológicas de pacientes sin cáncer de acuerdo con el tiempo de actividad física moderada.

Conclusiones

Como resultado del análisis realizado, se observa una modificación en los patrones de actividad física de acuerdo con la población estudiada. Además, estos patrones tienden a verse alterados por el nivel socioeconómico y el sexo de los pacientes.

El deterioro físico de los pacientes con cáncer de colorrectal corresponde a la mayor presencia de índice de masa corporal con bajo peso, mientras que las pacientes con cáncer de mama tuvieron mayor tendencia a el exceso de peso y, en el caso de la población sin cáncer, se observó una tendencia al normopeso.

Las porciones de alimentos consumidos son diferentes según el tipo de cáncer, destacando que, tanto los pacientes con cáncer colorrectal y la población sin cáncer mostraron un comportamiento similar. Pese a esto, el consumo de proteínas animales, un factor de riesgo conocido para el desarrollo de cáncer de colon se encuentra muy presente en los pacientes con esta neoplasia, manteniendo dietas ricas en este grupo alimenticio.

El tiempo de actividad física vigorosa muestra una diferencia, principalmente, entre los pacientes con cáncer de mama y el resto de la población estudiada, se observa que los pacientes con cáncer de colon son quienes realizan actividades laborales de mayor intensidad, al igual que la población sin cáncer.

En relación con la actividad física moderada, no existió una diferencia entre los tipos de cáncer, pero si se observan comportamientos diferentes al compararse con la población sin cáncer, quienes realizan en promedio menos actividad de intensidad media.

El uso del análisis de correspondencia múltiple permitió realizar la búsqueda de tipologías o patrones de los individuos en relación con las variables sociodemográficas y patológicas, identificando tendencias en el comportamiento.

Las pacientes con cáncer de mama tuvieron una mayor tendencia a no realizar el tiempo recomendado de actividad física moderada e intensa a medida que su escolaridad y nivel socioeconómico eran menores, al igual que provenir de zonas rurales y padecer diabetes o hipertensión.

El comportamiento de los pacientes con cáncer colorrectal y sin cáncer tuvo una mayor tendencia a verse modificado de acuerdo con el sexo, siendo los hombres quienes realizaban menos de lo indicado para la actividad vigorosa y más para la actividad moderada, además de padecer una patología crónica.

Referencias

- Alarcón Rojas, C. A., Alvarez-Bañuelos, M. T., Morales-Romero, J., Suárez-Díaz, H., Hernández-Fonseca, J. C., & Contreras-Alarcón, G. (2019). Breast Cancer: Metastasis, Molecular Subtypes, and Overweight and Obesity in Veracruz, Mexico. *Clinical Breast Cancer*, *19*(1), e166–e171. <https://doi.org/10.1016/j.clbc.2018.08.003>
- Amadou, A., Torres Mejia, G., Fagherazzi, G., Ortega, C., Angeles-Llerenas, A., Chajes, V., Biessy, C., Sighoko, D., Hainaut, P., & Romieu, I. (2014). Anthropometry, silhouette trajectory, and risk of breast cancer in Mexican women. *American Journal of Preventive Medicine*, *46*(3 SUPPL. 1). <https://doi.org/10.1016/j.amepre.2013.10.024>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018a). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. <https://doi.org/10.3322/caac.21492>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018b). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. <https://doi.org/10.3322/caac.21492>
- Bray, F., & Piñeros, M. (2016). Cancer patterns, trends and projections in latin america and the caribbean: A global context. *Salud Publica de Mexico*, *58*(2), 104–117. <https://doi.org/10.21149/spm.v58i2.7779>
- Chan, D. S. M., Vieira, A. R., Aune, D., Bandera, E. V., Greenwood, D. C., McTiernan, A., Navarro Rosenblatt, D., Thune, I., Vieira, R., & Norat, T. (2014). Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, *25*(10), 1901–1914. <https://doi.org/10.1093/annonc/mdu042>
- Dieli-Conwright, C. M., Fox, F. S., Tripathy, D., Sami, N., Van Fleet, J., Buchanan, T. A., Spicer, D., Lee, K., Mortimer, J., Bernstein, L., Demark-Wahnefried, W., & Courneya, K. S. (2020). Hispanic ethnicity as a moderator of the effects of aerobic and resistance exercise on physical fitness and quality-of-life in breast cancer survivors. *Journal of Cancer Survivorship*. <https://doi.org/10.1007/s11764-020-00918-3>

- Jochems, S. H. J., Van Osch, F. H. M., Bryan, R. T., Wesselius, A., Van Schooten, F. J., Cheng, K. K., & Zeegers, M. P. (2017). Impact of dietary patterns and the main food groups on mortality and recurrence in cancer survivors: A systematic review of current epidemiological literature. *BMJ Open*, 8(2), e014530. <https://doi.org/10.1136/bmjopen-2016-014530>
- Jurado, D., Bravo, L. M., Cerón, C., Guerrero-R, N., & Yépez-Ch, M. C. (2015). Hábitos de vida y cáncer colorrectal: Un estudio de casos y controles en una población de ingresos medios y bajos. *Universidad y Salud*, 17(1), 7–17.
- Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., & Straif, K. (2016). Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMsrl606602>
- Ligibel, J. A., Alfano, C. M., Courneya, K. S., Demark-Wahnefried, W., Burger, R. A., Chlebowski, R. T., Fabian, C. J., Gucalp, A., Hershman, D. L., Hudson, M. M., Jones, L. W., Kakarala, M., Ness, K. K., Merrill, J. K., Wollins, D. S., & Hudis, C. A. (2014). American Society of Clinical Oncology position statement on obesity and cancer. *Journal of Clinical Oncology*, 32(31), 3568–3574. <https://doi.org/10.1200/JCO.2014.58.4680>
- OECD. (2019). *The Heavy Burden of Obesity*. <https://doi.org/10.1787/2c4847ef-es>
- Reynoso-Noverón, N., & Torres-Domínguez, J. A. (2018). Epidemiología del cáncer en México: carga global y proyecciones 2000-2020. *Revista Latinoamericana de Medicina Conductual*, 8(1), 9–15. <http://revistas.unam.mx/index.php/rlmc/article/view/65111>
- Sedjo, R. L., Byers, T., Ganz, P. A., Colditz, G. A., Demark-Wahnefried, W., Wolin, K. Y., Azrad, M., & Rock, C. L. (2014). Weight gain prior to entry into a weight-loss intervention study among overweight and obese breast cancer survivors. *Journal of Cancer Survivorship*, 8(3), 410–418. <https://doi.org/10.1007/s11764-014-0351-9>
- Shamah-Levy, T., Cuevas-Nasu, L., Gaona-Pineda, E. B., Gómez-Acosta, L. M., Morales-Rúan, M. D. C., Hernández-Ávila, M., & Rivera-Dommarco, J. Á. (2018). Sobrepeso y obesidad en niños y adolescentes en México, actualización de la Encuesta Nacional de Salud y Nutrición de Medio Camino 2016. *Salud Pública de México*, 60(3, may-jun), 244. <https://doi.org/10.21149/8815>
- Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J Clin*, 66(1), 7–30. <https://doi.org/10.3322/caac.21332> <https://doi.org/10.3322/caac.21332>
- Wang, R., Zhu, Y., Liu, X., Liao, X., He, J., & Niu, L. (2019). The Clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer. *BMC Cancer*, 19(1), 1091. <https://doi.org/10.1186/s12885-019-6311-z>
- Zorron Cheng Tao Pu, L., Rana, K., Singh, G., Nakamura, M., Yamamura, T., Koay, D. S. C., Ovenden, A., Edwards, S., Ruskiewicz, A., Hirooka, Y., Fujishiro, M., Burt, A. D., & Singh, R. (2020). Different factors are associated with conventional adenoma and serrated colorectal neoplasia. *Nagoya Journal of Medical Science*, 82(2), 335–343. <https://doi.org/10.18999/nagjms.82.2.335>

Desgaste profesional y compromiso laboral en médicos residentes en el estado de Veracruz

Teobaldo Ortiz-Ramírez¹; Claudio Rafael López-Castro²; Julia Aurora Montano-Rivas², Carlos Adrián Alarcón-Rojas¹

¹Facultad de Estadística e Informática. Universidad Veracruzana. a_ortiz@outlook.com

²Facultad de Estadística e Informática. Universidad Veracruzana. ccastro@uv.mx

²Facultad de Estadística e Informática. Universidad Veracruzana. julmontano@uv.mx

²Facultad de Estadística e Informática. Universidad Veracruzana. cadalro24@gmail.com

Resumen

El burnout o desgaste profesional se ha descrito como un síndrome tridimensional, el cual afecta al recurso humano en sus actividades laborales cotidianas, en contraste a este síndrome, se encuentra el compromiso laboral, el cual se caracteriza por el entusiasmo hacia el trabajo. Esta investigación tuvo como objetivo evaluar el compromiso laboral y desgaste profesional de los médicos residentes con sede en hospitales del estado de Veracruz. El estudio que se presenta es de tipo observacional, transversal y descriptivo. Se realizó un muestreo probabilístico, teniendo a 218 médicos residentes que se encontraban inscritos en algún programa de especialidad médica de la Universidad Veracruzana. Se utilizaron los instrumentos de Maslach Burnout Inventory (MBI) y el Utrecht Work Engagement Scale (UWES-9) previamente validados. Los datos fueron analizados en los paquetes estadísticos SPSS versión 26 y R versión 4.0.2. Se aplicaron las técnicas estadísticas de chi-cuadrado y correspondencia múltiple donde se identificaron asociaciones entre las variables vigor, agotamiento emocional y realización personal con variables demográficas, así como niveles de compromiso laboral, en términos generales, de tipo medio e indicios de desgaste profesional en más de una dimensión. Los médicos residentes con estado civil de casado fueron los que presentaron mayores niveles de compromiso laboral, mientras que el sexo femenino tuvo niveles más altos en las dimensiones de desgaste profesional. Los datos obtenidos muestran un panorama general de cómo se encuentran los médicos residentes acerca del compromiso laboral y desgaste profesional, por lo que, realizar investigaciones más detalladas serán de utilidad para identificar las áreas de mejora para incrementar los niveles positivos y disminuir aquellos de desgaste profesional.

Palabras clave: burnout, work engagement, médicos residentes, compromiso, desgaste.

Introducción

Hoy en día, la investigación relacionada al bienestar del recurso humano ha cobrado un gran interés para las organizaciones, debido que se ha identificado que contar con colaboradores saludables física y mentalmente impacta en el desempeño y productividad de sus actividades.

Las organizaciones de salud han mejorado las condiciones laborales identificando las situaciones de estrés y desgaste laboral e implementando estrategias para conseguir calidad de vida en el trabajo, con lo cual se conseguirá un compromiso laboral que se traduce en una mayor satisfacción laboral (García Rodríguez *et al.*, 2015).

El compromiso laboral o engagement se define como “un estado mental positivo de realización personal relacionado con el trabajo, que se caracteriza por vigor, dedicación y absorción” (Salanova *et al.*, 2000 p. 119). Estas características se hacen presentes cada que el personal realiza sus actividades cotidianas. Mientras que el burnout o desgaste profesional se define como un “síndrome el cual está determinado por un conflicto interpersonal crónico durante la ejecución de sus actividades laborales” (Terrones-Rodríguez *et al.*, 2016).

Por otra parte, organismos nacionales como la Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud (CIFRHS) en su filosofía institucional tiene como objetivo propiciar cambios innovadores en apoyo a la mejora de la calidad de la educación en salud y por ende de la atención médica, se rige mediante sus principios de productividad, calidad, alto desempeño y bienestar, entre otros. (Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud (CIFRHS), 2017).

Mientras que a nivel internacional la Organización Panamericana de la Salud, en su Agenda de Salud Sostenible 2018-2030, considera al fortalecimiento de la gestión y el desarrollo de los recursos humanos para la salud con competencias que apoyen el abordaje integral (Organización Mundial de la Salud (OMS), 2019).

A partir de lo antes mencionado, se puede evidenciar la importancia que representan los recursos humanos para la salud, debido a que en ellos se confía la atención médica de calidad como resultado final de una cadena de valor que inicia desde el bienestar físico y mental de los médicos residentes que se encuentran en adiestramiento.

Se ha encontrado que los profesionales de la salud con un alto nivel de compromiso laboral dedican siempre el máximo esfuerzo en brindar la mejor atención posible y perfeccionar las estrategias para la calidad de la atención, incluso en situaciones de dificultad (Fernández Merino & Yáñez Gallardo, 2014).

Además, los empleados altamente comprometidos se caracterizan por tener mayor creatividad y desempeño, dando como resultado la satisfacción del cliente, es por ello que es deseable que las organizaciones públicas y privadas cuenten con empleados comprometidos (Bakker *et al.*, 2014).

La investigación en recursos humanos en salud está centrada mayormente en las complicaciones que experimentan los profesionales de la salud durante su formación. Por lo que, realizar investigación relacionada con el compromiso laboral y desgaste profesional de los médicos residentes aporta información relevante sobre las condiciones laborales, en especial de cómo afectan su desempeño, en las organizaciones de salud.

Marco teórico y antecedentes

En este apartado se definen las generalidades sobre los médicos residentes y los términos relacionados a las residencias médicas, así como la definición de compromiso y desgaste profesional que fueron la temática de estudio.

Generalidades de los médicos residentes

De acuerdo con la Ley Federal del Trabajo, en su capítulo XVI, se entiende que un médico residente es un profesional de la salud con estudios de licenciatura en medicina, que se encuentra en periodo de adiestramiento en alguna especialidad en una Unidad Médica Receptora de Residentes, que cuenta con el aval y reconocimiento de una Institución de Educación Superior (H. Congreso de la Unión, 2021). Las sedes o Unidades Médicas Receptoras deben contar con la infraestructura, equipamiento y personal para que el médico residente pueda realizar parcial o totalmente el programa académico de alguno de los siguientes grupos de especialidades: quirúrgicas, clínicas, salud pública y atención primaria o auxiliares, y de diagnóstico.

Las sedes se clasifican en tres niveles de atención, entendiéndose como la forma ordenada y estratificada de organizar sus recursos para satisfacer las requerimientos de la población (García Salabarría, 2006). A continuación, se describen brevemente los tres niveles de atención médica (Balderas, 2009).

- El primer nivel de atención médica representa el acceso al sistema de salud, es el contacto inicial entre la población y el personal de salud. Ejemplo: consultorios, clínicas, centros de salud y unidades de medicina familiar.
- Segundo nivel de atención se caracteriza por la orientación al diagnóstico y tratamiento oportuno, así como la limitación al daño. Ejemplo: hospitales regionales o generales de zona que cuentan con los servicios de medicina interna, cirugía, ginecología-obstetricia y pediatría.
- En el tercer nivel de atención, se tratan problemas de salud que requieren alta especialización, equipo sofisticado o servicios de rehabilitación avanzada. Ejemplo: Hospitales especializados en cancerología.

Desgaste profesional

El término burnout surge en la década de los 70's cuando el psiquiatra Herbert Freudenberg, observó que en un periodo de uno a tres años la mayoría de los sujetos que trabajaban en un clínica de Nueva York, experimentaba una pérdida progresiva de energía, desmotivación y ausencia de interés por el trabajo, al límite de llegar al agotamiento, aunado con varios síntomas como lo son la ansiedad y la depresión (Carlin & Garcés de los Fayos Ruiz, 2010).

En 1976, la psicóloga Cristina Maslach, alude al término de burnout, como una condición a la que están predispuestos con mayor frecuencia los trabajadores de los servicios de salud, ya que ellos brindan atención médica de manera directa, la cual suele ser intensa y prolongada (Reyes García et al., 2014).

Entonces, el síndrome de desgaste profesional o burnout se define como una sensación de fracaso y una existencia agotada que es resultado del exceso, esto por los requerimientos de energía, recursos personales o fuerza de voluntad del trabajador (Freudenberger, 1974).

En el ámbito de los trabajadores de la salud, Maslach y Jackson (año), fueron las primeras en conceptualizar al burnout para este sector y lo definen como un “síndrome de estrés crónico que se manifiesta en aquellas profesiones de servicios caracterizadas por una atención intensa y prolongada a personas que están en una necesidad o dependencia” (Maslach & Jackson, 1981a).

Por su parte Brill (citado en López Carballeira, 2017) define al burnout como:

Estado disfuncional y disfórico relacionado con el trabajo, en donde el sujeto no padece otra alteración psicopatológica mayor que (a) ha desarrollado su puesto de trabajo a un desempeño y unos niveles afectivos adecuados a ese mismo contexto laboral; (b) no recuperará los niveles previos sin ayuda externa o una nueva disposición ambiental (p. 22).

Dimensiones Burnout

Para Maslach y Jackson (1981), en la definición propuesta de burnout consideran que este síndrome consta de tres dimensiones o etapas por las que atraviesan los sujetos: 1) El agotamiento emocional, 2) la despersonalización y 3) la relacionada a la realización personal. A continuación, se describen estas dimensiones con mayor detalle.

La dimensión de cansancio emocional, hace referencia a la sensación de cansancio, fatiga y vacío que experimentan los sujetos, esto a causa de los altos requerimientos y sobrecarga de trabajo (Maslach & Jackson, 1981b). Por otra parte, esta dimensión se caracteriza por el hastío a causa de la incesante atención de los prestadores de servicios con sus usuarios (Montoya Zuluaga & Moreno Moreno, 2012).

En cuanto a la dimensión de despersonalización o cinismo, esta expone actitudes con poca sensibilidad, apáticas y desarrollo de acciones cínicas hacia los usuarios, así como la aparición de actitudes neutras y negativas hacia el entorno de trabajo (Montoya Zuluaga & Moreno Moreno, 2012). Esta dimensión surge como respuesta al exceso de agotamiento emocional, inicialmente como una manera de autoprotección. Sin embargo, si los sujetos continúan trabajando en exceso estos comenzaran a distanciarse y a reducir su productividad, dado que medida que se desarrolla el cinismo, los sujetos pasan de hacer su mejor esfuerzo a solo hacer el mínimo (Maslach, 2009).

Finalmente, la dimensión de realización personal se vincula con la predisposición que tienen los sujetos a evaluarse de forma negativa, así como la sensación de fracaso, aumento de inseguridad, incompetencia y falta de logros – productividad. La sensación que experimentan de reducción de eficacia es intensificado por la escasez de recursos en el trabajo, así mismo por la privación de apoyo social y la falta oportunidades para el desarrollo profesional (Maslach, 2009).

Compromiso Laboral

El compromiso laboral o engagement se define como un estado mental positivo, placentero que se relaciona con la actividad laboral, el cual esta caracterizado por los aspectos de vigor, dedicación y absorción. Además, el engagement hace referencia a un estado afectivo – cognitivo que es persistente e influyente, y no solo un estado específico y momentáneo. (Schaufeli *et al.*, 2002).

Dimensiones del engagement

- El vigor, está determinado por elevados niveles de energía y resistencia mental, así como el deseo de dedicar el máximo esfuerzo durante la realización del trabajo.
- Dedicación, manifiesta el alto nivel de implicación en el trabajo, aunado con la exteriorización de un sentimiento de importancia, entusiasmo, inspiración, orgullo y desafío por el trabajo.
- Y finalmente, la absorción, sucede cuando se está completamente absorto en el trabajo, es decir, que el sujeto experimenta la sensación de que el tiempo es insuficiente y resulta difícil separarse de lo que se está haciendo, esto debido al alto nivel de concentración y disfrute experimentado.

Problemática de los recursos humanos para la salud

En las últimas décadas, las investigaciones sobre la participación laboral se han incrementado, principalmente las relacionadas al compromiso laboral. Esto debido a que el capital humano es un recurso estratégico para el buen desempeño de cualquier organización y en los sistemas de salud no son la excepción, dado que el compromiso del personal es fundamental para su buen funcionamiento, pero a pesar de que en las organizaciones de salud se ha hecho un cambio para incrementar el compromiso, aún sigue presentándose la situación contraria, el desgaste profesional.

Los residentes médicos, son reconocidos como parte esencial en la prestación de servicios, aun cuando todavía son personal en formación. Sin embargo, se ha evidenciado que las residencias médicas se vinculan con depresión y burnout, estrés, exceso de trabajo, fatiga e incluso el uso de drogas (Gaona Valle *et al.*, 2018).

La calidad de vida del médico residente se asocia con jornadas extenuantes, así como altos niveles de estrés. En un estudio, se estimó una media de 70.1 ± 21.2 horas semanales de

trabajo para 12 especialidades médicas. Se reportó que las cargas horarias de trabajo generalmente aumentaban, además debían realizar su trabajo con rapidez y esmero. Pese a esto el sistema institucional no garantizaba un horario de trabajo adecuado, por lo que, el tiempo para realizar su trabajo no era suficiente y esto les impedía tener tiempo personal (Cázares-Ramírez *et al.*, 2015).

Por otro lado, en la determinación de niveles de estrés, se identificó que 80.6% de los residentes padecían niveles moderado a severo y percibieron exceso de supervisión, carga de trabajo excesiva, poca libertad para realizarlo y gran número de pacientes atendidos por día (Palacios-Nava & Paz-Román, 2014). Se han identificado los factores anteriormente mencionados como modificadores del compromiso laboral e indicios de desgaste profesional en los médicos residentes.

Se conoce que el personal de salud que se mantiene saludable y entusiasmado por realizar sus actividades de trabajo, resiste las situaciones de estrés de una manera más positiva (Rabin *et al.*, 2005) y su desempeño laboral se ve incrementado, así como la calidad que ofrecen a los usuarios de los servicios (Hernandez-Vargas *et al.*, 2014).

Metodología

El diseño de este estudio fue de alcance observacional, ya que la obtención de los datos fue mediante un instrumento en línea, por lo que no hubo intervención directa con los médicos residentes. Así mismo, fue transversal dado que la de información se recolectó en un único momento y de acuerdo con el nivel de profundidad, fue descriptivo.

La población de estudio estuvo integrada por 1171 médicos distribuidos en los 5 campus universitarios de la Universidad Veracruzana que estuvieran cursando una especialidad médica. La obtención de la muestra fue probabilística, obtenida por la fórmula general para poblaciones finitas, con un nivel de confianza de 95% y margen de error de 0.06, obteniendo una muestra de 217 médicos residentes. La fórmula utilizada para el cálculo se muestra a continuación.

$$n = \frac{Z^2 * N * p * q}{e^2 * (N - 1) + Z^2 * q * p}$$

Donde:

Z = Nivel de confianza

p = Porcentaje de población que tiene el atributo deseado

q = Porcentaje de población que no tiene el atributo deseado

N = Tamaño de la población

e = Error de estimación máximo aceptado

n = Tamaño de la muestra

Se aplicaron 218 cuestionarios en el periodo comprendido entre el 12 de febrero y el 15 de marzo de 2021. El cuestionario estuvo integrado por un consentimiento informado y 39 preguntas, de las cuales, 9 fueron de tipo sociodemográficas y de identificación, 9 de compromiso laboral, 22 de desgaste profesional y una pregunta abierta. La distribución del cuestionario fue en formato digital mediante correo electrónico a través de la Coordinación de Especialidades Médicas de la Universidad Veracruzana.

Por otra parte, para dar respuesta al objetivo de esta investigación se emplearon análisis estadísticos del tipo univariado, bivariado, multivariado y de confiabilidad. El análisis de tipo univariado consiste en estudiar cada una de las variables de interés por separado. Una de las técnicas más usadas es la distribución de frecuencias, mientras que los análisis bivariados estudian la relación entre dos variables y los análisis multivariantes, son una serie de técnicas que analizan simultáneamente los datos de más de dos variables.

En este estudio, se utilizó la técnica univariada de distribución de frecuencias para analizar todas las variables de estudio; así mismo, se utilizó la técnica de chi-cuadrado para determinar si había dependencia o independencia entre variables, la cual es de tipo bivariada; y de las técnicas multivariadas se utilizó el análisis de correspondencias múltiples, el cual es un método de interdependencia para datos cualitativos. Los análisis y gráficos de las técnicas antes mencionadas se realizaron en el programa R versión 4.0.2.

El instrumento utilizado para evaluar el compromiso laboral fue el *Utrecht Work Engagement Scale* en su versión corta (UWES-9) y para el desgaste profesional se aplicó el *Maslach Burnout Inventory* (MBI) con una escala Likert de 6 puntos. Las opciones de respuesta para ambos instrumentos fueron: “nunca”, “casi nunca”, “algunas veces”, “regularmente”, “bastantes veces” y “siempre”.

El UWES-9 considera que el compromiso laboral está compuesto por tres dimensiones: 1) vigor, hace referencia a los altos niveles de energía y voluntad invertidos en el trabajo a pesar de las dificultades, 2) dedicación, se refiere a estar fuertemente comprometido en el trabajo y 3) absorción, implica un estado placentero de concentración (Salanova *et al.*, 2000).

Mientras que el MBI, se encuentra conformado por tres subescalas de: 1) Agotamiento o cansancio emocional, la cual evalúa la vivencia de estar exhausto emocionalmente por los requerimientos del trabajo, 2) despersonalización, indica el grado en que cada sujeto reconoce actitudes de frialdad y distanciamiento, y 3) la subescala de realización personal evalúa los sentimientos de autoeficiencia y realización personal en el trabajo (Kulakova *et al.*, 2017).

En la Tabla 1, se muestran los niveles de evaluación para ambos instrumentos, así como los puntos de corte para categorizarlos.

Tabla 2. Puntos de corte para determinar el nivel de compromiso y desgaste profesional.

UWES-9		Niveles	MBI	
Dimensión	Punto de		Punto de corte	Dimensión
Vigor	> 95	Muy Alto	-	Agotamiento
Dedicación	> 75 - < 95	Alto	> 66	Despersonalización
Absorción	> 25 - < 75	Medio	> 33 - < 66	Realización personal
	> 5 - < 25	Bajo	< 33	
	< 5	Muy Bajo	-	

Fuente: Elaboración propia basado en (Schaufeli et al., 2002) y (Maslach & Jackson, 1981a).

Por otro lado, para confirmar la fiabilidad de los instrumentos se realizó un Alfa de Cronbach, en el que se obtuvo de manera global un 0.945 para el UWES-9 y para el MBI un 0.869, en la Tabla 2, se muestra el Alpha de Cronbach obtenido para cada una de las dimensiones.

Tabla 3. Fiabilidad de los instrumentos utilizados.

Indicador	No de ítems	Alfa de cronbach
Vigor	3	0.901
Dedicación	3	0.939
Absorción	3	0.898
UWES-9	9	0.945
Agotamiento emocional	9	0.927
Realización personal	8	0.898
Despersonalización	5	0.755
MBI	22	0.869

Fuente: Elaboración propia

El puntaje obtenido para el MBI se considera “muy bueno”, mientras que para el UWES-9 obtuvo una puntuación alta que podría considerarse como “redundante”, sin embargo, se decidió aplicarlo debido a que no era posible hacer una reducción de dimensiones y principalmente para mantener el modelo tridimensional propuesto.

Resultados

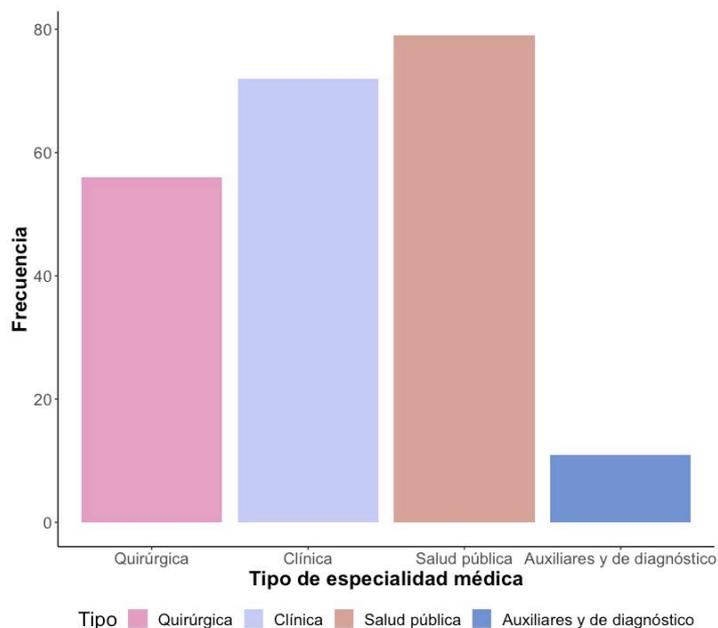
La muestra final estuvo integrada por 218 médicos residentes, la edad promedio fue de 31 ± 4 años, el mínimo de edad fue de 25 años y el máximo fue de 51 años. En la Tabla 3, se muestran los resultados para las variables demográficas y de identificación.

Tabla 4. Frecuencias y porcentajes de las variables sociodemográficas y de identificación (n=218).

	Variable	Frecuencia	%
Sexo	Femenino	104	47.7
	Masculino	114	52.3
Estado civil	Soltero	152	69.7
	Unión libre	26	11.9
	Casado	39	17.9
	Divorciado	1	0.5
Campus	Coatzacoalcos - Minatitlán	7	3.2
	Orizaba - Córdoba	36	16.5
	Poza Rica - Tuxpan	16	7.3
	Veracruz	96	44.0
	Xalapa	63	28.9
Años de residencia	1 ^{er} año	58	26.6
	2 ^{do} año	69	31.7
	3 ^{er} año	71	32.6
	4 ^{to} año	20	9.2
Tipo de guardia	ABC	26	11.9
	ABCD	184	84.4
	ABCDE	2	0.9
	No aplica	6	2.8
Servicios de salud	IMSS	179	82.1
	ISSSTE	5	2.3
	SSA	34	15.6
Nivel de Atención	1 ^{er} nivel	76	34.9
	2 ^{do} nivel	45	20.6
	3 ^{er} nivel	97	44.5
	Total	218	100.0

Fuente: Elaboración propia

En la Figura 1, se observa que el tipo de especialidad que estuvo mayormente representada en este estudio, que fue la del tipo de salud pública con un 36.2% y atención primaria con 36.2% de médicos residentes, seguido de las especialidades de tipo clínico con 33% y las quirúrgicas con un 25.7%, las especialidades que tuvieron menor participación fueron las del tipo de auxiliares y de diagnóstico con tan solo un 5% del total de médicos residentes.



Fuente: Elaboración propia

Figura 11. Frecuencias del tipo de especialidad médica.

Por otra parte, en la Figura 2, se muestran las regiones de procedencia, donde se observa que la mayoría de los médicos residentes provienen de las regiones sureste y centro del país acumulando más del 78% de los médicos que se encuentran cursando una especialidad médica ofertada por la Universidad Veracruzana.



Fuente: Elaboración propia

Figura 12. Regiones de procedencia de los médicos residentes inscritos en la Universidad Veracruzana.

En cuanto al análisis descriptivo para compromiso laboral, se observa en la Tabla 4 que en general los médicos residentes poseen un nivel de medio a alto, dado que el 73.4% se encuentra entre estas categorías, está proporción también coincide con la dimensión de vigor, dedicación y absorción concentran al 80.8% y 67.4% respectivamente.

Tabla 5. Frecuencias de los niveles de compromiso laboral por dimensión.

Dimensión	Nivel									
	Muy		Bajo		Medio		Alto		Muy	
	f	%	f	%	f	%	f	%	f	%
Vigor	7.2	3.2	39	17.9	97	44.5	63	28.9	12	5.5
Dedicación	10	4.6	32	14.7	117	53.7	59	27.1	-	-
Absorción	9	4.1	43	19.7	85	39	62	28.4	19	8.7
Engagement	10	4.6	37	17	102	46.8	58	26.6	11	5

Fuente: Elaboración propia

Por otra parte, para describir el desgaste profesional, se obtuvo para la dimensión de agotamiento emocional un nivel bajo con 35.8 % del total de médicos residentes, mientras que para la variable de despersonalización se acumuló el 71.5% en las categorías de bajo y medio, y para la variable de realización personal en un nivel medio, como se muestra en la Tabla 5.

Tabla 6. Frecuencias de los niveles de desgaste profesional.

Dimensión	Nivel					
	Bajo		Medio		Alto	
	f	%	f	%	f	%
Agotamiento emocional	78	35.8	66	30.3	74	33.9
Despersonalización	77	35.3	79	36.2	63	28.4
Realización personal	76	34.9	85	39	57	26.1

Fuente: Elaboración propia

Una vez definidos los niveles para las dimensiones de compromiso laboral y desgaste profesional, se realizó un análisis estadístico de chi-cuadrado (χ^2) para determinar la asociación con las variables sociodemográficas, en las que solo se encontró significancia entre la variable sexo, tipo de especialidad y estado civil, en la Tabla 6 se muestran los valores obtenidos.

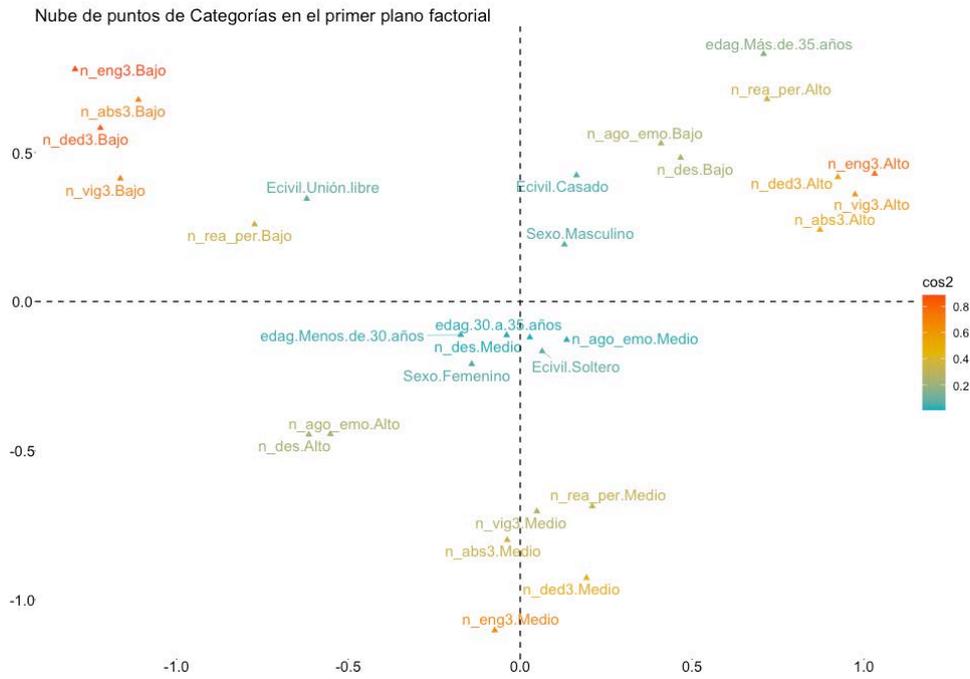
Tabla 7. Prueba X2 para compromiso laboral y variables sociodemográficas.

Variable	Dimensión	χ^2	Grados de libertad	Significancia
Sexo	Vigor	12.17	4	0.016
	Dedicación	11.35	3	0.009
Tipo de especialidad	Dedicación	83.88	57	0.011
Sexo	Agotamiento emocional	6.45	2	0.039
	Realización personal	8.72	2	0.012
Estado civil	Agotamiento emocional	12.39	4	0.014

Fuente: Elaboración propia

En la Figura 3, se muestra el mapa de correspondencia múltiple, que forma parte los análisis multivariados, en los que se observan las variables de compromiso laboral y desgaste profesional, junto con las variables de sexo, estado civil y la edad, que para este análisis se agruparon. En el primer cuadrante se agrupan los médicos residentes con estado civil de casado, los cuales no presenta indicios de desgaste profesional y alto niveles para compromiso laboral.

Mientras que, en el segundo cuadrante, se encuentran aquellos con estado civil de unión libre, que presentan un nivel de compromiso laboral bajo, así como en sus tres dimensiones y tienen indicios de desgaste profesional en la dimensión de realización personal.



Fuente: Elaboración propia

Figura 13. Mapa de correspondencia múltiple para la variable compromiso laboral y desgaste profesional.

Los resultados anteriormente expuestos fueron obtenidos mediante la aplicación de métodos estadísticos, los cuales nos permitieron manipular datos cualitativos y cuantitativos. Estos métodos están presentes desde el establecimiento de la muestra, la depuración de los datos incompletos, así como la presentación y análisis analítico y gráfico. En este trabajo fueron de gran utilidad, puesto que se pudieron analizar varias variables cualitativas en un solo análisis y su representación gráfica brinda una completa y fácil interpretación. Los métodos estadísticos son imprescindibles al momento de realizar cualquier investigación, ya que para cada tipo de dato y objetivo tiene una técnica adecuada para tratar y dar una respuesta confiable.

Conclusiones

El compromiso laboral y desgaste profesional juegan un papel muy importante en las organizaciones, debido a que afectan directamente a los recursos humanos y, a su vez, este ejerce fuerte influencia en todo ámbito organizacional, puesto que del recurso humano depende la toma de decisiones, así como la atención que se brinda. Los recursos humanos para la salud, en este caso los médicos residentes, son una parte estratégica para las organizaciones de este ámbito, aun cuando se encuentran en adiestramiento.

A partir de los resultados obtenidos, se identificó que en su mayoría los médicos residentes tienen un nivel medio de compromiso laboral y sola una quinta parte se encuentra con niveles de bajo a muy bajo, por lo que aún existe margen de mejora para incrementar estos niveles. Con respecto al desgaste profesional, se identificó que una tercera parte de los médicos residentes presenta un nivel bajo de realización personal y alto para agotamiento emocional, lo cual es un indicador de desgaste profesional. En términos generales, más de la tercera parte de médicos residentes no presentó indicios de desgaste profesional en ninguna de las dimensiones.

Por otro lado, con base en las asociaciones, se encontró que el vigor, la dedicación, agotamiento emocional y realización personal dependen de la variable sexo. Mientras que el estado civil y el tipo de especialidad médica influyen en las dimensiones de agotamiento emocional y dedicación respectivamente.

Finalmente, a partir del mapa de correspondencias múltiples se identificó que los médicos residentes de sexo masculino, mayores de 35 años con estatus de casado, presentan altos niveles de compromiso, en contraparte, los médicos con estatus de unión libre presentan afectación en la dimensión de realización personal y compromiso laboral bajo. Además, el sexo femenino presenta una correspondencia con niveles de tipo alto para las dimensiones de agotamiento emocional y despersonalización.

Esta investigación es un referente sobre el panorama general en el que se encuentran el compromiso laboral y desgaste profesional en los médicos residentes en hospitales del estado de Veracruz, y con base en los resultados obtenidos se expone que hay área de oportunidad para mejorar e implementar estrategias que permitan incrementar los niveles positivos y disminuir los negativos, con la finalidad de garantizar un ámbito de formación agradable para los recursos humanos en salud.

Finalmente, se propone realizar un muestreo estratificado, con la finalidad de que cada campus se encuentre representado y, así poder, describir de manera más objetiva los resultados. Así mismo, realizar entrevistas a profundidad, con la intención de conocer aquellos factores que pudieran estar afectando los niveles de compromiso e incrementado los de desgaste profesional, a fin de obtener más información para la propuesta de un plan de mejora.

Referencias

- Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and Work Engagement: The JD–R Approach. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 389–411. <https://doi.org/10.1146/annurev-orgpsych-031413-091235>
- Balderas, M. de la luz. (2009). *Administración de los servicios de enfermería* (6ª ED.). Mc GrawHill.
- Carlin, M., & Garcés de los Fayos Ruiz, E. J. (2010). El síndrome de burnout: Evolución histórica desde el contexto laboral al ámbito deportivo. *Anales de Psicología*, 26(1), 169–180. <http://www.redalyc.org/articulo.oa?id=16713758020>
- Cázares-Ramírez, S. E., Prieto-Miranda, C. A., Jiménez-Bernardino, G., Vera-Haro, M. de J., & Esparza-Pérez, R. I. (2015). Jornadas laborales y sus repercusiones en médicos residentes en un hospital de segundo nivel. *Medicina Interna de México*, 31(6), 669–679.
- Comisión Interinstitucional para la Formación de Recursos Humanos para la Salud (CIFRHS). (2017). *Acerca de la CIFRHS*. Visión CIFRHS. <http://www.cifrhs.salud.gob.mx>
- Fernández Merino, C., & Yáñez Gallardo, R. (2014). Describiendo el engagement en profesionales de enfermería de atención primaria de salud. *Ciencia y Enfermería*, 20(3), 131–140. <https://doi.org/10.4067/S0717-95532014000300012>
- Freudenberger, H. J. (1974). Staff Burn-Out. *Journal of Social Issues*, 30(1), 159–165. <https://doi.org/10.1111/j.1540-4560.1974.tb00706.x>
- Gaona Valle, L. S., Consuelo Estrada, J. R., Tomp Osnaya, E. P., & Rizo Amézquita, J. N. (2018). La importancia de la salud en los médicos residentes y su relación con la calidad y seguridad del paciente. *BOLETÍN CONAMED - OPS*, 3(18), 15.18.
- García Rodríguez, J. J., Labajos Manzanares, M. T., & Fernández Luque, F. (2015). Students in graduate programs in nursing and commitment to studies. *Enfermería Global*, 14(38).
- García Salabarría, J. (2006). Hospital Psiquiátrico de La Habana ¿Sobrevivirán los niveles de atención a la revolución de la salud pública cubana? In *Rev Cubana Salud Pública* (Vol. 32, Issue 1).
- H. Congreso de la Unión. (2021). Ley Federal del Trabajo. *Diario Oficial de La Federación, DOF 01-05-2019*, 325.
- Hernandez-Vargas, C. I., Llorens-Gumbau, S., & Rodríguez-Sánchez, A. M. (2014). Empleados saludables y calidad de servicio en el sector sanitario: un estudio de caso. *Anales de Psicología*, 30(1). <https://doi.org/10.6018/analesps.30.1.143631>
- Kulakova, O., Moreno Jiménez, B., Garrosa, E., Sánchez Hernández, M. O., & Aragón, A. (2017). Universalidad del constructo del Maslach Burnout Inventory en un contexto latinoamericano. *Acta de Investigación Psicológica*, 7(2), 2679–2690. <https://doi.org/10.1016/j.aippr.2017.05.001>
- López Carballeira, A. (2017). El Burnout. In *El síndrome de Burnout: Antecedentes y consecuentes organizacionales en el ámbito de la sanidad pública gallega* (1st ed., pp. 3–56). Universidad de Vigo.
- Maslach, C. (2009). Comprendiendo el Burnout. *Ciencia y Trabajo*, 32(2), 37–43.
- Maslach, C., & Jackson, S. E. (1981a). *The Maslach Burnout Inventory Manual*. https://www.researchgate.net/publication/277816643_The_Maslach_Burnout_Inventory_Manual/citations
- Maslach, C., & Jackson, S. E. (1981b). The measurement of experienced burnout. *Journal of Organizational Behavior*, 2(2), 99–113. <https://doi.org/10.1002/job.4030020205>
- Montoya Zuluaga, P. A., & Moreno Moreno, S. (2012). Relación entre síndrome de burnout , estrategias de afrontamiento y engagement. *Psicología Del Caribe*, 29(1), 204–227.
- Organización Mundial de la Salud (OMS). (2019). 57.º CONSEJO DIRECTIVO 71.ª SESIÓN DEL COMITÉ REGIONAL DE LA OMS PARA LAS AMÉRICAS. https://www.paho.org/hq/index.php?option=com_docman&view=download&alias=50291-cd57-od359-s-plan-estrategico-ops&category_slug=cd57-es&Itemid=270&lang=es
- Palacios-Nava, M. E., & Paz-Román, M. del P. (2014). Condiciones de trabajo, estrés y

- manifestaciones psicosomáticas en médicos de hospitales de la ciudad de México. *Medicina y Seguridad Del Trabajo*, 60(235), 322–334. <https://doi.org/10.4321/S0465-546X2014000200005>
- Rabin, S., Matalon, A., Maoz, B., & Shiber, A. (2005). Keeping doctors healthy: A salutogenic perspective. *Families, Systems, & Health*, 23(1), 94–102. <https://doi.org/10.1037/1091-7527.23.1.94>
- Reyes García, F. A., Guerra Vargas, M., Ortega Reyes, A. O., Robles Acosta, C., & Hernández Gress, E. S. (2014). Revisión teórica de los elementos de impacto del síndrome de burn-out en el clima organizacional. *Boletín Científico de Ciencias Básicas e Ingenierías*, 2.
- Salanova, M., Schaufeli, W. B., Llorens, S., Peiró, J. M., & Grau, R. (2000). Desde el “burnout” al “engagement”: ¿una nueva perspectiva? *Revista de Psicología Del Trabajo y de Las Organizaciones*, 16, 117–134.
- Schaufeli, W. B., Salanova, M., González-romá, V., & Bakker, A. B. (2002). The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *Journal of Happiness Studies*, 3(1), 71–92. <https://doi.org/10.1023/A:1015630930326>
- Terrones-Rodríguez, J. F., Cisneros-Pérez, V., & Arreola-Rocha, J. de J. (2016). Síndrome de burnout en médicos residentes del Hospital General de Durango, México. *Revista Medica Instituto Mexicano Del Seguro Social*, 2, 242–248.

Análisis estadístico de los factores de riesgo asociados con la actividad sexual en adolescentes mexicanos entre 2012 y 2018

Areli Keila Acatitla Dávila¹, Claudio Rafael Castro López², Lilliam Marie Pinzón³

¹Facultad de Estomatología. Benemérita Universidad Autónoma de Puebla. areli.keila@gmail.com

²Facultad de Estadística e Informática, Universidad Veracruzana. ccastro@uv.mx

³Facultad de Odontología. Universidad de Utah. lilliam.pinzon@hsc.utah.edu

Resumen

La adolescencia es una etapa fundamental en la vida y desarrollo de un individuo, por lo que determinar los factores de riesgo asociados al inicio de la actividad sexual en adolescentes de 10-19 años de México, es un factor clave para prevenir e implementar estrategias de promoción a la salud que protejan a la población adolescente ante cualquier desenlace no deseado, al mismo tiempo se buscó reportar los hallazgos más importantes sobre las características de la población adolescente en México, cuyo perfil se asocia a los factores de riesgo que presentan los adolescentes en función al inicio de la actividad sexual mediante el uso diverso de las diferentes técnicas estadísticas, desde las más tradicionales, como son las tablas de contingencias hasta las más actuales, como la técnica de segmentación TAID-LCA, cada una de estas se complementaron una a la otra para fortalecer y concluir los hallazgos encontrados en la búsqueda de la asociación entre el inicio sexual y los factores de riesgo de esta población. Se encontró que existe necesidad en fortalecer los programas de educación sexual en los adolescentes ya que al comparar el año 2012 y 2018 no hay evidencia para declarar que los conocimientos y los factores de riesgo que influyen al inicio de la actividad sexual han mejorado, por lo que se resalta la importancia del mejoramiento ante el tema de las Infecciones de Transmisión Sexual (ITS) para la promoción y prevención de la salud de los adolescentes.

Palabras clave: Factores de riesgo; Adolescentes; Anticonceptivos; Inicio de actividad sexual; Conducta sexual; ITS; Métodos anticonceptivos, Análisis estadístico TAID-LCA.

Introducción

La adolescencia es un periodo en el que el sujeto busca el reconocimiento y autoexploración de sí mismo, de esta manera busca comprender los cambios físicos, psicológicos y emocionales que esta etapa conlleva (Feldman J, 2002; Blythe M, 2000), este periodo tan importante en el desarrollo de una persona parte de tres facetas significativas: la primera de ellas es la etapa temprana, la cual está comprendida de los 10 a los 14 años, la segunda etapa es la media, dicha etapa abarca de los 15 a los 17 años y, por último, la etapa tardía que va de los 18 a los 20 años (Blythe M, 2000). Este periodo tan vital en el desarrollo de una persona contiene transformaciones activas y dinámicas en donde las peculiaridades de los jóvenes y la examinación de sí mismos es lo más común y frecuente, por lo que se origina el acrecentamiento de la sexualidad (Feldman J, 2002).

El comportamiento sexual se asocia a componentes ambientales e individuales que involucran la parte biopsicosocial del individuo, prácticamente estos dos factores tienen una

relevancia de suma importancia; en cuanto a los componentes ambientales podemos destacar precedentes en torno a las relaciones personales y sociales dentro y fuera del hogar, así como, factores culturales que llevan al conocimiento o desconocimiento de enfermedades de transmisión sexual, uso de métodos anticonceptivos, adicciones, entre otros; al hablar de los factores individuales podemos destacar la sintomatología depresiva del individuo, es decir, estado de ánimo del adolescente, el comportamiento en relación al ámbito social en el que se desenvuelve y aquellos relacionados a los trastornos de las conductas alimentarias. (Grunbaum J, 2002; Conard L, 2003; Cohen D, 2002)

“Se consideran factores de riesgo aquellos que incrementan la posibilidad de un desenlace no deseado (acto sexual) y factores de protección aquellos que la reducen” (Feldman J, 2002).

A través del tiempo se ha observado que la actividad sexual de un adolescente cada vez comienza en momentos más tempranos de su vida y mientras el coito sexual ocurra de manera más precoz, los riesgos a presentar cualquier enfermedad de transmisión sexual se multiplican al aumentar el número de parejas sexuales que el individuo puede tener a lo largo de su vida (Cutié JR, 2005). La actividad sexual de un adolescente inicia aproximadamente a los 16 años de edad, según las estadísticas de algunos estudios (Marchetti R, 1996) y, en algunos otros el promedio del inicio de la actividad sexual oscila alrededor de los 13 hasta los 15 años de edad (UNICEF, 2015), acompañado de una importante presencia en personas jóvenes con Infecciones de Transmisión Sexual (ITS) y más particularmente, del Virus de Inmunodeficiencia Humana (VIH), además de un aumento en el índice de embarazos en adolescentes, este último argumento se ve asociado al carente conocimiento en cuanto al manejo de los métodos anticonceptivos existentes en la actualidad o simplemente por tabúes y creencias en cuanto a ellos (Juárez, 2005).

Diversos estudios exponen que una cantidad significativa de adolescentes mexicanos están en riesgo de tener embarazos no deseados, además de estar expuestos a contraer alguna ITS al estar comprometiendo su integridad física (Caballero-Hoyos R, 2001; Gayet C, 2003).

Publicaciones recientes presentan resultados en donde se observa que los jóvenes inician su actividad sexual en edades prematuras. Se aprecia que adolescentes entre 15 y 19 años de edad ya han aprobado tener relaciones sexuales alguna vez en su vida, lo que representa una tercera parte de jóvenes en esta situación; en cuanto a las mujeres los estudios reportan que antes de los 12 años ya han tenido su primera relación sexual (19%), en referencia a los hombres, la situación empieza en edades comprendidas entre los 12 y 14 años de edad (99%) (Instituto Mexicano de la Juventud., 2006). “Ante estos hechos, las autoridades de salud en México tienen como prioridad la prevención de conductas sexuales de riesgo en poblaciones que se desarrollan en medios vulnerables” (Secretaría de Salud., 2002).

A pesar del aumento en cuanto a las publicaciones acerca del uso de métodos anticonceptivos, la información que se encuentra acerca de estos y el conocimiento de las prácticas sexuales y las conductas que tienen los adolescentes aún se muestran con múltiples mitos y tabúes, sociales y culturales que obstaculizan la información certera. (Menkes, 2003).

Los resultados muestran un efecto positivo de esas intervenciones en el conocimiento sobre VIH/SIDA y formas de protección; sin embargo, la práctica de uso del condón o anticonceptivos no se ha modificado. (Martínez-Donate AP, 2004) (Walker D, 2006)

Se precisan modelos de comportamiento educativos y conductuales basados en estudios científicos cuyas investigaciones han sido aprobados para favorecer la eventualidad de las prácticas no saludables y de esta manera minimizar la incidencia y prevalencia de las infecciones de transmisión sexual (ITS). Por lo que, la finalidad del presente capítulo de investigación fue determinar los factores de riesgo asociados al inicio de la actividad sexual en adolescentes de 10-19 años, con base a los datos de la Encuesta Nacional de Salud y Nutrición (ENSANUT) del año 2012 y 2018 y reportar los hallazgos sobre las características de la población adolescente en México, cuyo perfil se asocia a los factores de riesgo que presentan los adolescentes en función al inicio de la actividad sexual mediante el uso de técnicas estadísticas que fortalecen los hallazgos de la misma.

El uso de las diversas técnicas estadísticas disponibles en la actualidad para la asociación de variables como son: tablas de contingencia, prueba de independencia de ji-cuadrada, análisis de correspondencia, análisis de segmentación CHAID y TAID-LCA fueron implementadas en el presente trabajo con el objetivo de aprovechar cada una de ellas y demostrar el uso y ventajas de estas para la obtención de la información y cumplimiento de los objetivos, se describen de forma breve la aplicación de cada una de las técnicas mencionadas anteriormente, con el fin de aportar en la selección de estrategias estadísticas para datos cualitativos en cuanto a asociación de variables.

Metodología

Tipo de estudio

El diseño del estudio correspondió a un estudio observacional, analítico, transversal y retrospectivo.

Instrumento

El presente trabajo de investigación analizó los datos obtenidos del Cuestionario de Salud de adolescentes de 10-19 años realizado por la Encuesta Nacional de Salud y Nutrición (ENSANUT) con las correspondientes bases de datos de dicho instrumento de los años 2012 y 2018 propuesto por el Instituto Nacional de Salud Pública (INSP) y el Instituto Nacional de Estadística y Geografía (INEGI).

Esta encuesta se aplicó a nivel nacional, contiene 111 preguntas, las cuales están agrupadas en ocho secciones: I. Factores de riesgo (15). Aplican a hombres o mujeres de 10 a 19 años. II. Infecciones de transmisión sexual y salud reproductiva (44). Aplican a hombre o mujer de 12 a 19 años (20); solo mujeres de 12 a 19 que tuvieron relaciones sexuales (5); solo mujeres con último hijo nacido vivo y sigue vivo (18); solo mujeres de 12 a 19 años (1). III. Vacunación (16). Hombre o mujer de 10 a 19 años. IV. Conductas alimentarias (1). Hombre o mujer de 10 a 19 años. V. Sintomatología depresiva (4). Hombre o mujer de 10 a 19 años.

VI. Accidentes (8). Hombre o mujer de 10 a 19 años. VII. Ataque y violencia (21). Hombre o mujer de 10 a 19 años. VIII. Disciplina (2). Adolescentes de 10 a 14 años.

Variables

Para fines de la investigación, solo se analizaron las dos primeras secciones del cuestionario y dentro de estas preguntas, correspondientes a las secciones I y II, se seleccionaron las variables de interés que corresponden y delimitan la investigación.

Las variables del estudio fueron seleccionadas de acuerdo a la literatura, mismas que cumplían con los objetivos en función a los factores de riesgo que se relacionan al inicio de la actividad sexual, tales como, la variable del inicio de actividad sexual, la variable factor alcohol, factor tabaco, uso de métodos anticonceptivos, conocimientos de ITS, así como, entidad y sexo, cabe mencionar que las variables fueron cualitativas en su mayoría nominales y ordinales.

Análisis estadístico

Las bases de datos del cuestionario de la ENSANUT 2012 y 2018 fueron recodificadas para fines didácticos, análisis y mejor entendimiento del investigador. Se codificaron las respuestas para agrupar por grupo de edad de acuerdo a la etapa de la adolescencia a la que pertenecen y al estado de procedencia, las bases de datos se analizaron con los paquetes estadísticos Rstudio, SPSS statistics v23 y TAID-LCA para el correspondiente análisis en función a su objetivo.

Para el cumplimiento de los objetivos anteriormente mencionados, se estructuró una serie de análisis estadísticos en función a sus objetivos, como primera etapa se realizaron análisis exploratorios para cada una de las variables, posteriormente, se describió a las mismas mediante análisis bivariado para datos categóricos mediante el uso de tablas de contingencia, seguidamente se aplicó la prueba de hipótesis de ji-cuadrada (Carreón,2016;Mendivelso y Rodríguez, 2018) como un primer acercamiento y comprobar la asociación entre las variables, es decir, el inicio de la actividad sexual con respecto a las variables consideradas factor de riesgo, la variable factor alcohol, factor tabaco, uso de métodos anticonceptivos y conocimientos de ITS.

Una vez obtenido el resultado de la prueba de independencia de ji-cuadrada se procedió con el análisis multivariado, usando análisis de correspondencia simple, para observar de manera gráfica la asociación entre los factores de riesgo y el inicio de la actividad sexual, consecutivamente, como análisis confirmatorio se realizó la técnica estadística de segmentación mediante el algoritmo CHAID (Chi Squared Automatic Interaction Detector). El algoritmo CHAID (Kass, 1980) se propone para una variable respuesta de tipo cualitativo y predictores cualitativos y se basa en el test ji-cuadrado para contrastar las independencias en las diferentes etapas en las que lleva a cabo el algoritmo (Ver Figura 1).



Figura 1. Etapas del algoritmo de CHAID

Por último, se procedió a segmentar la población de adolescentes mexicanos para la creación de perfiles, mediante el análisis de segmentación no simétrico usando el algoritmo TAID (*Detección Automática de Interacciones basada en el Índice Tau*) con el fin de obtener árboles ternarios, la aportación fundamental de dicho algoritmo es que este, considera varias variables respuestas (dependientes) en lugar de tener sólo una como en el algoritmo CHAID. El algoritmo TAID busca grupos semejantes entre las variables y de esa manera hallar las asociaciones entre ellas para crear clases latentes de manera exhaustiva y excluyente. (Castro y Galindo, 2011) (Ver Figura 2).

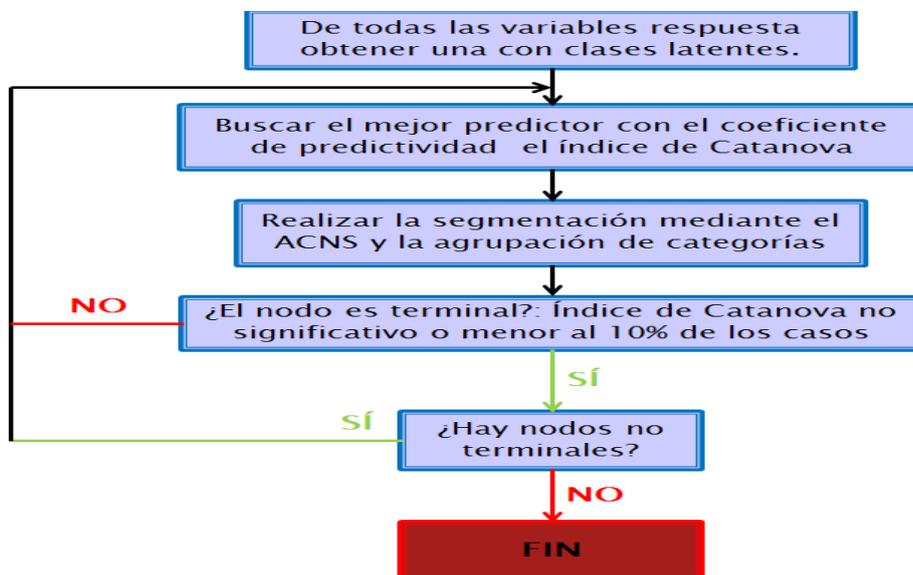


Figura 2. Etapas del algoritmo TAID

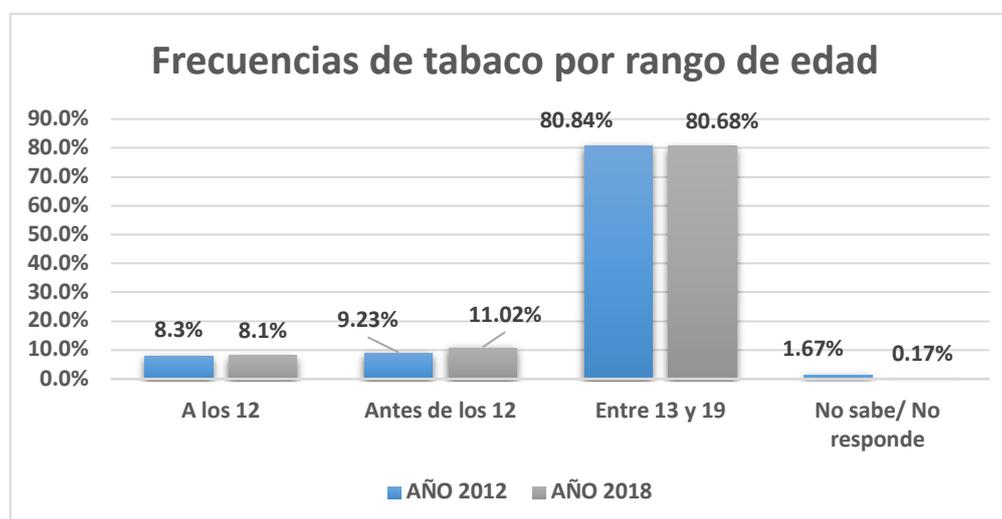
Las dos bases de datos se analizaron con el mismo análisis multivariante y se compararon los resultados para los dos años y, de esa forma, conocer el comportamiento de los factores de riesgo y los perfiles en los adolescentes mexicanos.

Resultados

Se realizaron las tablas de contingencia en relación con los factores de riesgo de cada variable, con respecto a los años 2012 y 2018, como análisis exploratorio; para fines didácticos se muestran los resultados de las variables factor de riesgo de tabaco y alcohol con el inicio de actividad sexual, así como, conocimientos con respecto a las ITS y conocimiento en cuanto al uso de métodos anticonceptivos.

La base de datos del año 2012 contiene 21509 observaciones, por lo que, para el análisis de la variable factor de riesgo de tabaco se clasificó en función a la respuesta edad, es decir, la edad en la que fumaron por primera vez, el análisis exploratorio arrojó que solo el 17.78 % del total de adolescentes refirió haber fumado por primera vez, sin embargo, entre los 13 y 19 años de edad se encuentra el 80.84 % de la población que fuma. En cuanto a la base de datos 2018, esta contó con una muestra de 4767 adolescentes.

Se observó que de los 4767 adolescentes presentados en la base de datos de la ENSANUT 2018, solo el 12.37% del total de la población ya han fumado por primera vez (Figura 3).



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Figura 3. Frecuencias de factor tabaco para el año 2012 y 2018.

Los resultados anteriores muestran el patrón de comportamiento de los adolescentes entre el año 2012 y 2018, en donde se puede observar que del año 2012 a 2018 la incidencia en cuanto a factor tabaco ha disminuido, (17.78% y 12.37%) respectivamente.

En cuanto a la variable alcohol, para el año 2012, se observa que el 60.27% de los adolescentes mexicanos nunca han tomado una gota de alcohol y los que ya han probado esta sustancia también se encuentran, en su mayoría, entre los 13 y los 19 años de edad. En cuanto a esta misma variable, para el año 2018, se encontró que el 24.94% de los adolescentes ha tomado alguna vez en su vida, porcentaje que se encuentra disminuido comparado con el año 2012 en donde se tiene que el 39.56 % de la población adolescente ha tomado alguna vez, el

comportamiento de este factor de riesgo en referencia a estos años se ve disminuido para el año 2018 (Figura 4).

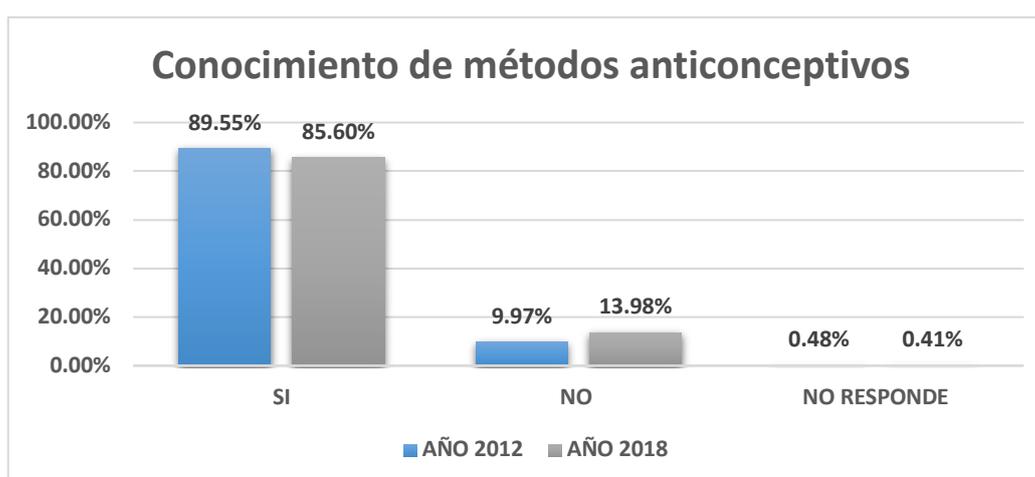


Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Figura 4. Frecuencias de factor alcohol para el año 2012 y 2018.

Los resultados encontrados para la variable factor conocimiento de métodos anticonceptivos mostraron asociación con el conocimiento de los mismos y el inicio de la actividad sexual.

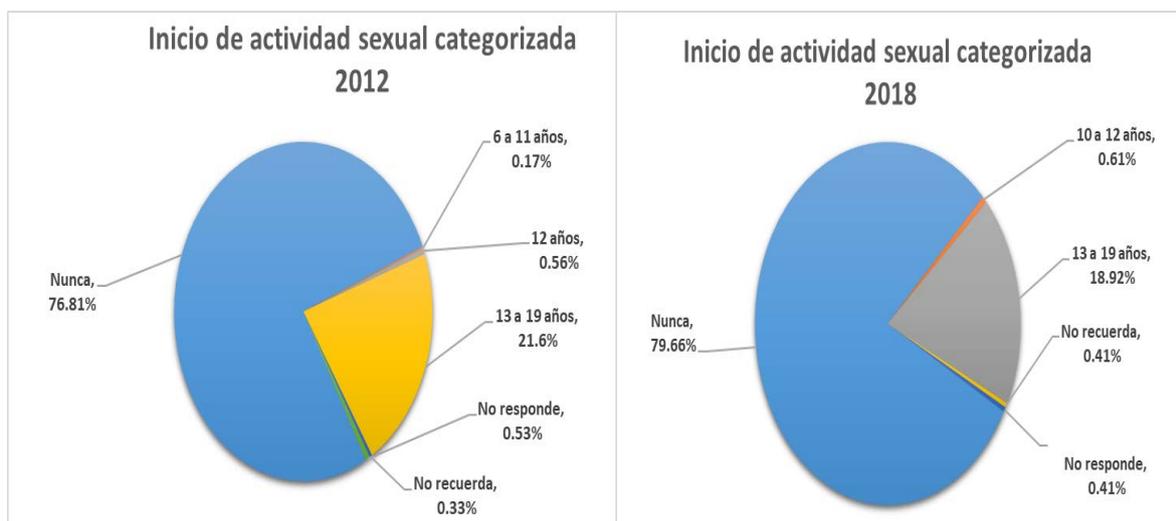
Se encontró que el 89.55 % de la población adolescente de 12 a 19 años, tienen conocimiento acerca del uso de métodos anticonceptivos. Los hallazgos encontrados para el conocimiento de métodos anticonceptivos mostraron que el 85.60% de los adolescentes cuentan con conocimientos acerca de los métodos anticonceptivos, porcentaje que se encuentra disminuido con respecto al año 2012 (89.55%). Con estos hallazgos se observa que al igual que en los anteriores factores ya descritos, el comportamiento de estos entre el año 2012 y 2018 no varía de manera significativa (Figura 5).



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018

Figura 5. Frecuencias de conocimiento de uso de métodos anticonceptivos para el año 2012 y 2018.

De igual manera se analizaron las frecuencias de los adolescentes que iniciaron su actividad sexual en función a la edad de los mismos años (2012 y 2018), se observa que para el 2012 que la mayoría de los adolescentes que han iniciado su actividad sexual lo hicieron entre los 12 y 19 años de edad. En cuanto al inicio de actividad sexual de los adolescentes para el año 2018, se observó que el 60.26% de los adolescentes iniciaron su actividad sexual, dicho comportamiento subió muy poco comparado con el del año 2012 (59.18%). Los patrones de comportamiento no se modifican de manera muy notoria entre un año y otro. (Figura 6)



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018

Figura 6. Frecuencias de edad de inicio de actividad sexual para el año 2012 y 2018.

Con base en lo anterior, se clasificaron las edades de inicio de actividad sexual por etapa de adolescencia según la literatura, y con ello, se describe la asociación de las variables factor de riesgo tabaco e inicio de la actividad sexual para ambos años. Las variables estudiadas son categóricas nominales, por lo que fue necesario describir la tabla de contingencia de dichas variables (tabaco, alcohol, conocimiento de métodos anticonceptivos y conocimiento de métodos anticonceptivos con el inicio de la actividad sexual).

En referente a la variable tabaco e inicio de ac sexual para el año 2012, se observa que gran parte de los adolescentes que consumieron tabaco por primera vez, en edades comprendidas entre los 13 y 19 años, también iniciaron su actividad sexual en una etapa media de su adolescencia, es decir, entre los 15 y 17 años de edad (Véase Tabla 1). La tabla de contingencia para el año 2018 muestra que la mayoría de los adolescentes inicio su actividad sexual en una etapa media de su adolescencia (15-17 años) y comenzó a fumar entre los 13 y 19 años de edad (Ver Tabla 2).

Tabla 1. Tabla de contingencia entre tabaco e inicio de la actividad sexual (2012)

Tabla 2. Tabla de contingencia entre tabaco e inicio de la actividad sexual (2018)

Edad de inicio de consumo de tabaco.	Etapa de inicio de actividad sexual				Edad de inicio de consumo de tabaco.	Etapa de inicio de actividad sexual			
	Etapa media**	Etapa tardía***	Etapa temprana*	Nunca		Etapa media**	Etapa tardía***	Etapa temprana*	Nunca
12 años	65	3	60	186	12 años	13	2	7	26
Antes de los 12 años	41	4	47	202	Antes de los 12 años	7	0	13	34
Entre 13 y 19 años	1169	204	344	1331	Entre 13 y 19 años	180	31	44	213

*Etapa temprana: 10-13 años

**Etapa media: 14-16 años

***Etapa tardía: 17-19 años

Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Una vez obtenidas las tablas de contingencias para ambos años se llevó a cabo la prueba Ji-cuadrada para probar las siguientes hipótesis:

H_0 : No existe asociación entre las variables

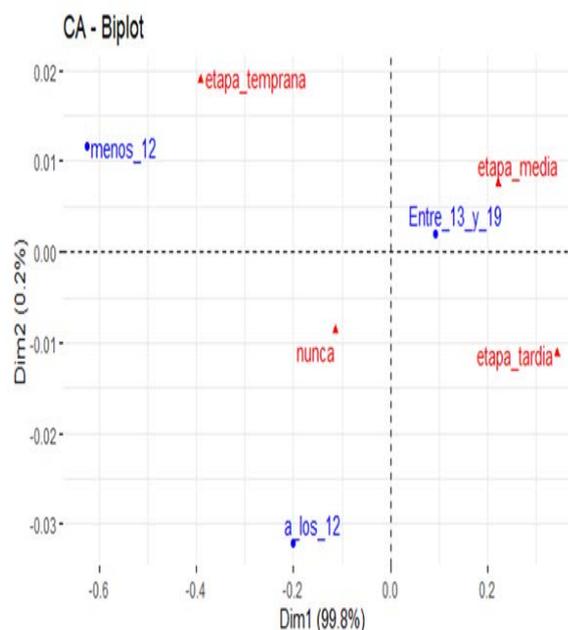
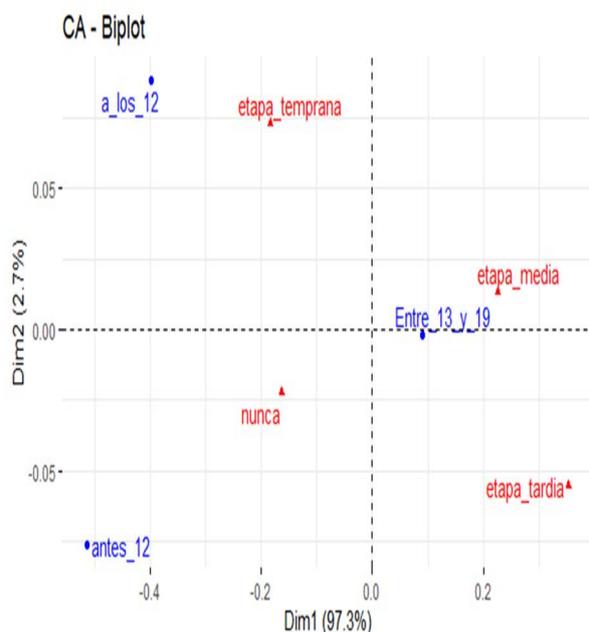
H_1 : Existe asociación entre las variables

Considerando un nivel de significancia del 0.05, se encontró evidencia estadística suficiente para rechazar la hipótesis nula (ji-cuadrada= 156.84, $gl = 6$, $p\text{-valor} < 2.2e-16$); por lo que se determina que existe asociación entre la variable factor tabaco e inicio de actividad sexual para el año 2012 y para 2018 se encontró que también existe asociación entre las variables. (ji-cuadrada = 27.152, $gl = 6$, $p\text{-valor} = 0.00013567$).

Para entender mejor la asociación entre las variables estudiadas se realizó un análisis de correspondencia.

El análisis de correspondencia proporciona una representación visual de la asociación de dos variables categóricas mediante la clasificación de acuerdo con los atributos que poseen los objetos (Hair, et al. 1999). Entonces, el análisis de correspondencia nos permitió entender las asociaciones de las categorías de nuestras variables de interés, factor tabaco e inicio de la actividad sexual. A continuación, se muestra el gráfico obtenido del análisis de correspondencia, el gráfico expone la siguiente situación: adolescentes que tuvieron su primera experiencia con el tabaco entre los 13 y 19 años de edad fueron proclives a tener su primera relación sexual en las últimas dos etapas de su adolescencia (etapa media o tardía) y para el 2018 el análisis de correspondencia expone la siguiente situación: adolescentes que tuvieron su primera experiencia con el tabaco entre los 13 y 19 años de edad fueron proclives a tener su primera relación sexual en las últimas dos etapas de su adolescencia (etapa media

o tardía), mientras que aquellos adolescentes que tuvieron su primera experiencia con el tabaco a edades inferiores a los 12 años de edad fueron proclives a experimentar su primera relación sexual en la etapa temprana de la adolescencia (Figura 7 y Figura 8).



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Figura 7. Mapa perceptual relación entre inicio de actividad sexual e inicio de consumo de tabaco. (2012)

Figura 8. Mapa perceptual relación entre inicio de actividad sexual e inicio de consumo de tabaco. (2018)

La tabla de contingencia de la variable alcohol e inicio de la actividad sexual del año 2012 (Tabla 3), mostró que los adolescentes entre los 13 y los 19 años que ya han ingerido alcohol, también iniciaron su actividad sexual en una etapa media de su adolescencia (15-17 años). En cuanto a la tabla de contingencia de la variable ingesta de alcohol e inicio de actividad sexual (2018) mostró que el 25.76% de los adolescentes entre 13 y 19 años ya han ingerido bebidas alcohólicas e iniciado su actividad sexual (Tabla 4).

Tabla 3. Tabla de contingencia entre factor de alcohol e inicio de la actividad sexual (2012).

Edad de inicio de ingesta de alcohol	Inicio de actividad sexual				Total	Total %
	Etapa media**	Etapa tardía***	Etapa temprana*	Nunca		
Antes de los 12 años	64	10	71	557	1032	6.29
12 años	76	8	69	887	710	4.33
entre 13 y 19 años	1797	385	462	3483	6127	37.36
Nunca ha tomado	473	97	177	7785	8532	52.02
Total	2410	500	779	12712	16401	
Total %	14.69	3.05	4.75	77.51		

Tabla 4. Tabla de contingencia entre factor de alcohol e inicio de la actividad sexual (2018).

Edad de inicio de ingesta de alcohol	Inicio de actividad sexual				Total	Total %
	Etapa media**	Etapa tardía***	Etapa temprana*	Nunca		
Antes de los 12 años	6	2	13	87	110	3.06
12 años	18	1	9	65	93	2.58
entre 13 y 19 años	302	66	66	493	927	25.76
Nunca ha tomado	147	28	48	2245	2468	68.59
Total	475	97	136	2890	3398	
Total %	13.2	2.7	3.78	80.32		

*Etapa temprana: 10 a 13 años.

**Etapa media: 14 a 16 años.

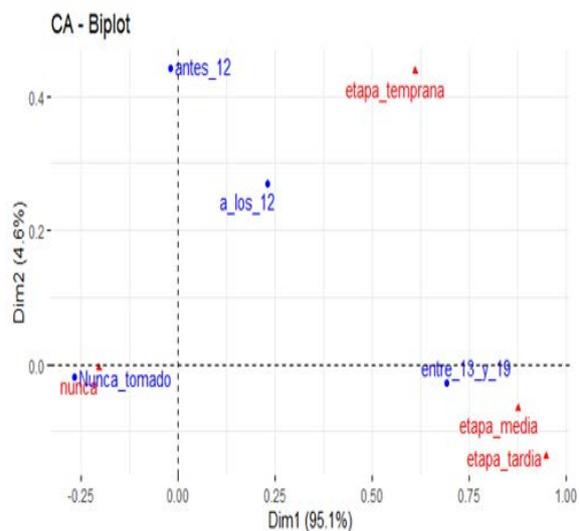
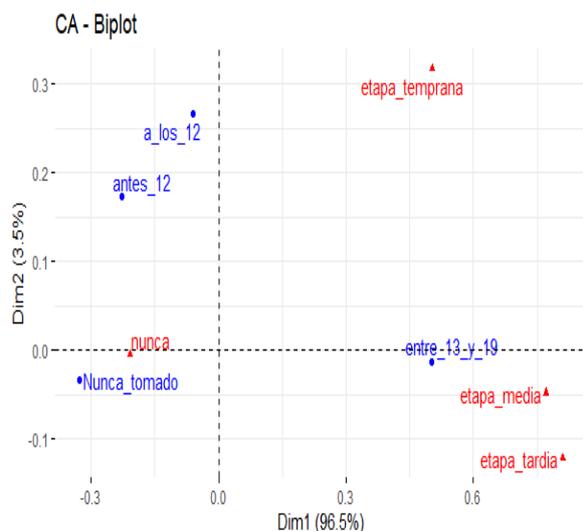
***Etapa tardía: 17 a 19 años.

Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

La prueba Ji-cuadrada para el 2012 resultó con 2604, con 9 grados libertad y un p valor $< 2.2e-16$. Considerando un nivel de significancia del 0.05 se encontró evidencia estadística suficiente para rechazar la hipótesis nula; existe asociación entre la variable alcohol e inicio de actividad sexual. Para el 2018 también se encontró que existe asociación entre las variables (ji-cuadrada = 655.99, gl = 9, p-valor $< 2.2e-16$).

El análisis de correspondencia para la variable alcohol expone la siguiente situación para el año 2012: adolescentes que consumieron alcohol por primera vez entre los entre 13 y 19 años fueron proclives a tener su primera experiencia sexual en las últimas dos etapas de su adolescencia; mientras que aquellos que no tomaron ninguna gota de alcohol en ninguna etapa de su adolescencia no fueron proclives a experimentar su primera relación sexual en ninguna etapa de su adolescencia. Para el año 2018, el análisis de correspondencia expone la siguiente situación: adolescentes que consumieron alcohol por primera vez entre los entre 13 y 19 años de edad fueron proclives a tener su primera experiencia sexual en las últimas dos etapas de su adolescencia; aquellos que no tomaron ninguna gota de alcohol no fueron proclives a experimentar su primera relación sexual en ninguna etapa de su adolescencia, mientras que aquellos adolescentes que experimentaron su primera experiencia con el alcohol

a los 12 años fueron proclives a tener su primera relación sexual en la etapa temprana de su adolescencia (Figura 9 y Figura 10).



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Figura 9. Mapa perceptual relación entre inicio de actividad sexual e inicio de consumo de alcohol 2012.

Figura 10. Mapa perceptual relación entre inicio de actividad sexual e inicio de consumo de alcohol 2018.

En referencia a los conocimientos de métodos anticonceptivos que poseen los adolescentes con relación al inicio de su actividad sexual, la tabla de contingencia del año 2012 mostró que de los adolescentes que aún no inician su actividad sexual la mayoría cuenta con conocimiento acerca de los métodos anticonceptivos, además de mostrar que en la mayoría de las etapas de inicio de actividad sexual casi todos tienen conocimientos acerca de métodos anticonceptivos. La tabla de contingencia del año 2018 determinó que solo un 14.02% de los adolescentes no cuentan con conocimientos suficientes, de los cuales la mayoría de los adolescentes que se encuentran en desconocimiento son de edades de entre 15 y 17 años (etapa media) (Tabla 5 y Tabla 6).

Tabla 5. Tabla de contingencia de uso de métodos anticonceptivos e inicio de actividad sexual (2012).

Conocimiento	Inicio de actividad sexual				Total	Total %
	Etapa tardía	Etapa media	Etapa temprana	Nunca		
Si	2326	477	742	11172	14717	89.6
No	87	24	38	1484	1633	9.94
NS/NR	1	1	0	74	76	0.46
Total	2414	502	780	12730	16426	
Total %	14.7	3.06	4.75	77.5		

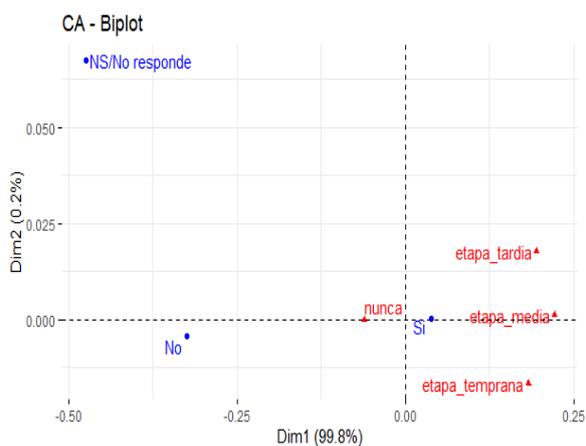
Tabla 6. Tabla de contingencia de uso de métodos anticonceptivos e inicio de actividad sexual (2018).

Conocimiento	Inicio de actividad sexual				Total	Total %
	Etapa tardía	Etapa media	Etapa temprana	Nunca		
Si	88	453	126	2418	3085	85.62
No	9	22	11	463	505	14.02
NS/NR	0	0	0	13	13	0.36
Total	97	475	137	2894	3603	
Total %	2.69	13.18	38	80.32		

Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

La prueba ji- cuadrada para estas variables en 2012 mostró ji-cuadrada= 2604, gl = 9, p-valor < 2.2e-16 y para el 2018 la prueba arrojo lo siguiente: ji-cuadrada = 53.861, grados libertad = 6, p-valor= 7.87e-10 se consideró un nivel de significancia del 0.05 y en ambos años se encontró evidencia estadística suficiente para rechazar la hipótesis nula; es decir, existe asociación entre las variables conocimiento de métodos.

El gráfico del análisis de correspondencia (2012) para esta variable muestra que los adolescentes que cuentan con conocimientos en métodos anticonceptivos fueron proclives a tener su primera experiencia sexual en las últimas dos etapas de su adolescencia; así como, en etapas tempranas, además de mostrar que los adolescentes quienes aún no inician su actividad sexual también cuentan con conocimientos de métodos anticonceptivos; y en el año 2018 se encontró que los adolescentes con conocimiento en métodos anticonceptivos tienen su primera experiencia sexual en cualquier etapa de su adolescencia; así como, en etapa temprana de su adolescencia (Figura 11 y Figura 12).



Fuente: Elaboración propia utilizando Rstudio con datos de la ENSANUT 2012 y 2018.

Figura 11. Mapa perceptual relación entre inicio de actividad sexual y conocimiento de métodos anticonceptivos. (2012)

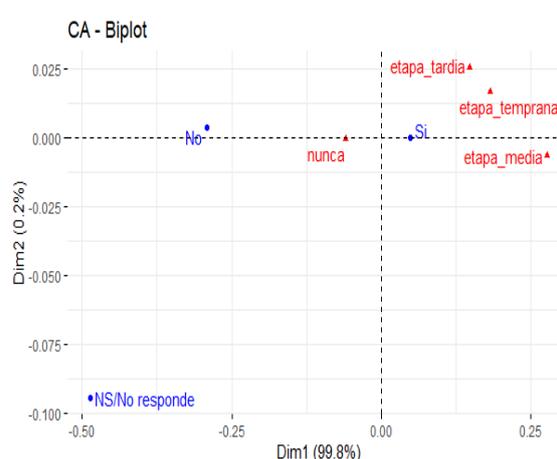
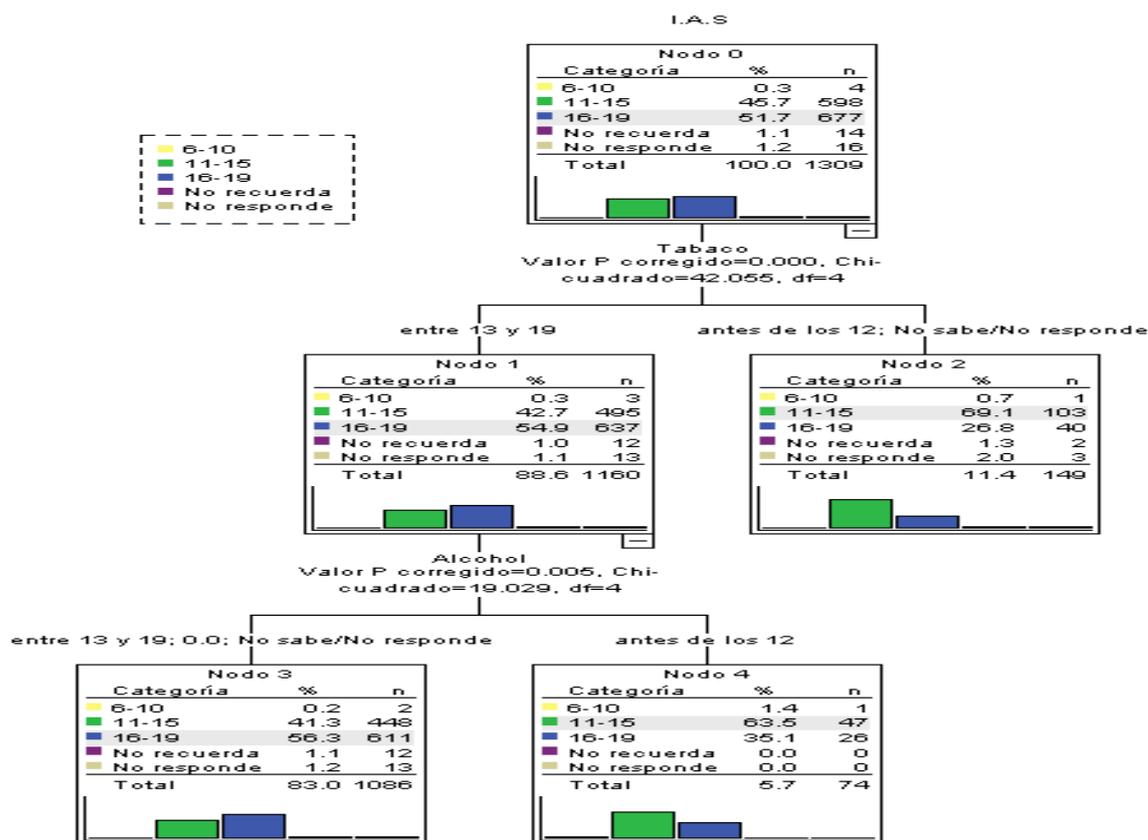


Figura 12. Mapa perceptual relación entre inicio de actividad sexual y conocimiento de métodos anticonceptivos. (2018)

Un análisis exploratorio de mayor profundidad se realizó mediante la técnica de segmentación con el uso del algoritmo CHAID, se partió de cuatro variables predictoras para comparar con los resultados anteriores, (factor alcohol, factor tabaco, conocimiento de métodos anticonceptivos y conocimiento de ITS) y la variable respuesta de interés, es decir, el inicio de actividad sexual. El análisis para el año 2012 arrojó 5 nodos de los cuales 3 son terminales, que son los que caracterizan a la variable respuesta, por lo que encontramos tres perfiles de los adolescentes en cuanto al inicio de la actividad sexual.

El algoritmo mostró que el 51.7% de los adolescentes de entre 16 a 19 años han iniciado su actividad sexual, el 45.7% inicio su actividad sexual entre los 11 y 15 años, 0.3% inicio entre los 6 y 10 años y el 2.3% no sabe o no respondió. Posteriormente, el algoritmo seleccionó el mejor predictor, es decir, la variable factor tabaco, en este caso, vemos que esta variable presenta una agrupación óptima de sus categorías, se observa que el predictor es significativo para cada una de sus categorías (ji-cuadrada 42.055, gl=4, p-valor 0.0000), se observa que los adolescentes que fumaron por primera vez entre los 13 y 19 años iniciaron su actividad sexual en la etapa tardía de su adolescencia (54.9%) y los jóvenes que fumaron por primera antes de los 12, iniciaron su actividad sexual en la etapa temprana y media de su adolescencia (69.8%). Para la categoría de 13 a 19 años el algoritmo realiza otro cruce y para ello escoge como mejor predictor la variable factor alcohol. En este caso, se observa que el algoritmo realiza una agrupación similar a la anterior, los dos nodos obtenidos son significativos (ji-cuadrada 19.029, gl=4, p-valor 0.005) y, por tanto, terminales.

Con los resultados anteriores se concluye que el 63.5% de los adolescentes han tomado por primera vez antes de los 12 años y consumido tabaco por primera vez entre los 13 y 19 años de edad e iniciado su actividad sexual en su etapa tardía de la adolescencia, mientras que del resto de adolescentes se encontró que empezaron a ingerir bebidas alcohólicas y consumir tabaco entre los 13 y 19 años, sin embargo, el 69.1% que empezó a fumar antes de los 12 años inicio su actividad sexual en su etapa tardía (Figura 13)



Fuente: Elaboración propia utilizando SPSS statistics v23 con datos de la ENSANUT 2012.

Figura 13. Árbol final para el inicio de la actividad sexual.

Como se logró observar, el análisis de segmentación arrojó resultados similares que, en los análisis anteriores, por lo que, se omite el análisis con esta técnica para la base 2018 y se procede a la siguiente técnica de segmentación para la creación de perfiles de los adolescentes en México.

Se buscó crear grupos para la construcción de perfiles más precisos, mediante el aumento de variables respuesta, por lo que fue necesario usar el algoritmo TAID, es decir, se buscó hacerlo por entidad (zona del país a la que pertenecen) y sexo, como se percibió en el algoritmo CHAID, las variables respuestas no pueden ser usadas de manera conjunta, de modo que se tenía que realizar tantos análisis individuales como variables respuestas tenemos (uno por cada una) para explicarlos, así como se realizó con el análisis de correspondencia simple. Así que, esta fase consistió en un análisis de clases latentes con todas las variables respuestas, como ya hemos dicho con el fin de obtener una única variable respuesta con varias categorías (clase latente).

La estimación se realizó para un mínimo de 2 clases y un máximo de 5, sin embargo, se selecciona el modelo para 2 clases para los dos años de acuerdo con los resultados obtenidos para la bondad de ajuste de los 2 modelos (Ver Tabla 7).

Tabla 7. Tabla de bondad de ajuste para 2 clases del año 2012 y 2018

Tabla de bondad de ajuste 2012					
Clases	BIC	AIC	Chi²	G²	log-Verosimilitud
2	7988.78	7900.77	21.82516	24.1918	-3933.385
Tabla de bondad de ajuste 2018					
Clases	BIC	AIC	Chi²	G²	log-Verosimilitud
2	1660.215	1605.693	11.9363	13.32145	-787.8466
BIC: criterio bayesiano de información; AIC: criterio de información Akaike					

Fuente: Elaboración propia utilizando TAID-LCA con datos de la ENSANUT 2012 y 2018.

Con base en los modelos elegidos, se describe cada uno para su correspondiente año, a partir de las proporciones (probabilidades a priori) y las probabilidades condicionales (de las categorías de cada variable dada la clase a la que pertenece el adolescente) de cada una de las bases de datos.

Respecto al año 2012, se observa que el mayor número de casos este contenido en la clase 1 (59.2%), mientras que en la clase 2 se encuentra el 40.79% de los casos, en cuanto al año 2018 se observa que al igual que en el año 2012, el mayor número contenido de los casos se encuentra en la clase 1 (53.59%), en ambos años se observa que en la clase 1, los jóvenes que tienen mayor probabilidad de pertenecer esta clase son hombres pertenecientes a la zona Sur del país e iniciaron su actividad sexual entre los 11 y 15 años de edad. Con referencia a la clase 2, del año 2012, se encuentra que los jóvenes con mayor probabilidad de pertenecer a este grupo son mujeres pertenecientes a la zona centro del país que iniciaron su actividad sexual entre los 16 y 19 años de edad, al igual que los casos del 2018 con la única diferencia de que los hombres tienen mayor probabilidad de pertenencia a este grupo (Ver Tabla 8).

Tabla 8. Modelos de 2 clases según el inicio de la actividad por sexo y la entidad años 2012 y 2018

Año 2012				Año 2018			
Variable	Respuesta	Clase 1	Clase 2	Variable	Respuesta	Clase 1	Clase 2
Proporciones		0.5921	0.4079	Proporciones		0.5359	0.4641
Sexo	Hombre	0.7418	0.4454	Sexo	Hombre	0.8121	0.6859
	Mujer	0.2582	0.5546		Mujer	0.1879	0.3141
Entidad	Norte	0.1167	0.3175	Entidad	Norte	0.1907	0.1107
	Norte/occidente	0.2007	0.0626		Norte/occidente	0.0548	0.1214
	Centro/norte	0.2419	0.0197		Centro/norte	0.1378	0.0948
	Centro	0.1327	0.5265		Centro	0.1049	0.6716
	Sur	0.3078	0.0737		Sur	0.512	0.0016
Inicio de actividad sexual	6 a 10 años	0.0051	-10.131	Inicio de actividad sexual	6 a 10 años	NA	NA
	11 a 15 años	0.4979	0.3972		11 a 15 años	0.5831	0.3116
	16 a 19 años	0.4854	0.5633		16 a 19 años	0.4168	0.6191
	NS/NR	0.0115	0.0395		NS/NR	-13.821	0.0693

Norte: Baja California, Sonora, Chihuahua, Coahuila, Nuevo León, Tamaulipas.
 Nore/Occidente: Baja California Sur, Sinaloa, Nayarit, Durango, Zacatecas.
 Centro/Norte: Jalisco, Aguascalientes, Colima, Michoacán, San Luis Potosí.
 Centro: Guanajuato, Querétaro, Hidalgo, Estado de México, Distrito Federal, Tlaxcala, Puebla.
 Sur: Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán, Quintana Roo.

Fuente: Elaboración propia utilizando TAID-LCA con datos de la ENSANUT 2012 y 2018.

Una vez obtenidas las probabilidades de las clases, se construyen los árboles ternarios para cada año, conforme a las variables independientes o predictoras, se encuentran las características que contienen al mayor número de casos de acuerdo con los nodos que arroja el algoritmo, es decir, adolescentes que consumieron por primera vez alcohol o tabaco, así como, aquellos que cuentan con conocimientos acerca del uso de métodos anticonceptivos e infecciones de transmisión sexual.

El Figura 14, que se presenta a continuación, muestra el árbol de segmentación resultante para ese año, en él se describen las características de cada uno de los grupos o nodos pertenecientes con mayor probabilidad al grupo de adolescentes que iniciaron su actividad sexual.

La segmentación TAID toma como mejor predictor la variable conocimiento de infecciones de transmisión sexual (ITS), en donde se observa que se divide el porcentaje de los adolescentes que saben y no saben acerca de las ITS.

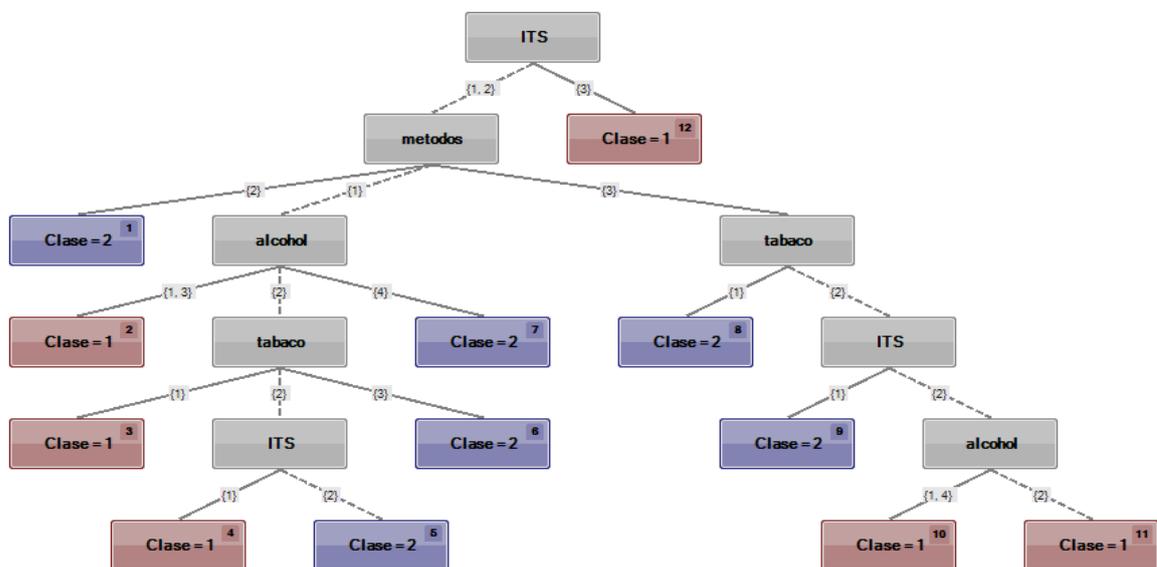
Nodo 1: Adolescentes mujeres pertenecientes principalmente a los estados de Guanajuato, Guerrero Hidalgo, Edo. De México, D.F., Morelos, Tlaxcala y Puebla que iniciaron su actividad sexual entre los 16 y 19 años y no tenían conocimiento acerca del uso de métodos anticonceptivos, sin embargo, se divide el porcentaje en cuanto al conocimiento de ITS, es decir, el 50.69% si contaban con los conocimientos necesarios mientras que el 49.31% no los tenía.

Nodo 3: Adolescentes hombres pertenecientes a los estados de Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán y Quintana Roo e iniciaron su actividad sexual entre los

11 y 15 años, también consumieron alcohol y tabaco antes de los 12 años y casi la mitad de ellos no cuenta con conocimientos en cuanto a las ITS.

Nodo 4: Adolescentes hombres pertenecientes a los estados de Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán y Quintana Roo e iniciaron su actividad sexual entre los 11 y 15 años, no sabían del uso de métodos anticonceptivos e ingirieron alcohol y tabaco entre los 13 y 19 años.

Nodo 7: Adolescentes mujeres pertenecientes principalmente a los estados de Guanajuato, Guerrero Hidalgo, Edo. De México, D.F., Morelos, Tlaxcala y Puebla que iniciaron su actividad sexual entre los 16 y 19 años, nunca han ingerido alcohol, si tienen conocimientos de métodos anticonceptivos, pero casi la mitad de los casos no cuentan con conocimientos acerca de las infecciones de transmisión sexual.



Fuente: Elaboración propia utilizando TAID-LCA con datos de la ENSANUT 2012.

Figura 14. Árbol de segmentación para el inicio de la actividad sexual 2012.

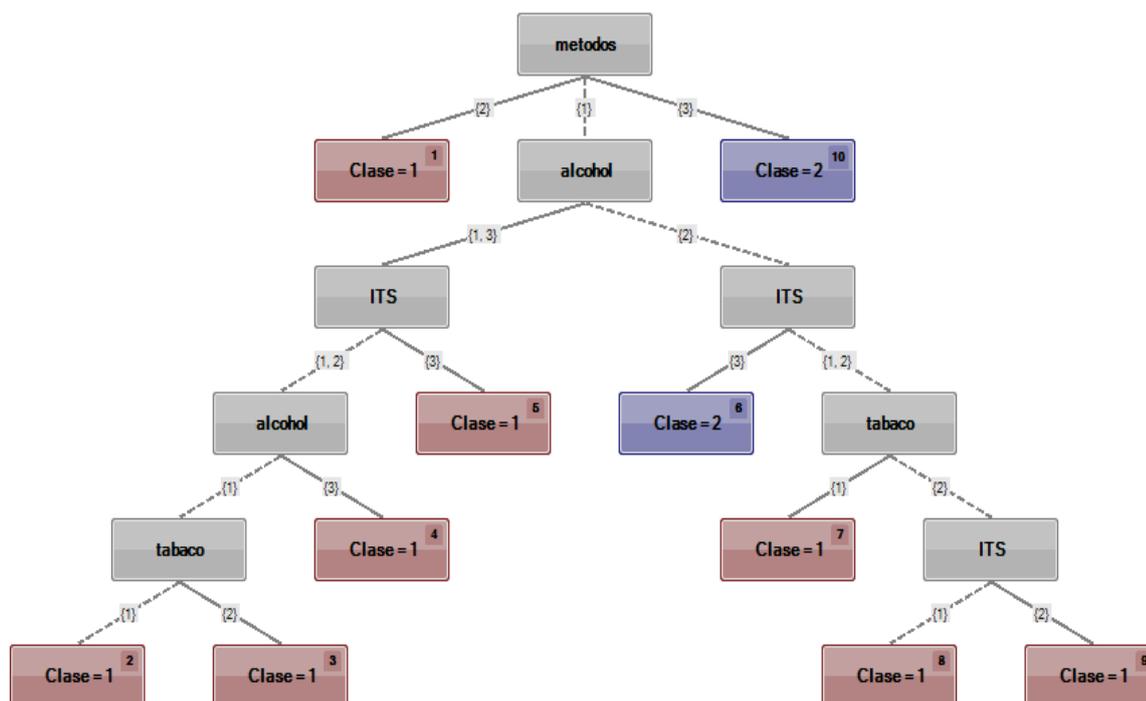
Para el año 2018, la Figura 15 presenta el árbol de segmentación resultante en que se describen las características de cada uno de los grupos o nodos pertenecientes, con mayor probabilidad, al grupo de adolescentes que iniciaron su actividad sexual.

En este caso, la segmentación TAID tomó como mejor predictor la variable conocimiento de uso de métodos anticonceptivos.

Nodo 2: Adolescentes hombres pertenecientes a los estados de Guerrero, Oaxaca, Chiapas, Veracruz, Tabasco, Campeche, Yucatán y Quintana Roo e iniciaron su actividad sexual entre los 11 y 15 años que ingirieron por primera vez alcohol y tabaco antes de los 12 años, sin embargo, el 42% de ellos no tenía conocimientos suficientes acerca de las infecciones de transmisión sexual.

Nodo 8: Adolescentes hombres pertenecientes a los estados del sur del país e iniciaron su actividad sexual entre los 11 y 15 años, si tenían conocimientos suficientes acerca de las ITS y de los métodos anticonceptivos y fumaron y tomaron por primera vez entre los 13 y 19 años.

Nodo 9: Adolescentes hombres pertenecientes a los estados del sur del país e iniciaron su actividad sexual entre los 11 y 15 años, no tenían conocimientos acerca de las ITS, pero *sí* en cuanto al uso de métodos anticonceptivos y fumaron y tomaron por primera vez entre los 13 y 19 años.



Fuente: Elaboración propia utilizando TAID-LCA con datos de la ENSANUT 2018.

Figura 15. Árbol de segmentación para el inicio de la actividad sexual 2018

Conclusión

El inicio de la actividad sexual se asocia a la presión de grupo ejercida y sentida por los adolescentes en el entorno social, es decir, por los amigos, quienes construyen espacios para socializar información y narrar las experiencias sexuales previas, lo cual influye para aumentar la curiosidad, misma que incide en un mayor deseo a tener una práctica sexual a temprana edad; datos que se hacen notar al encontrar que los adolescentes, quienes nunca han consumido alcohol y nunca han consumido tabaco, tampoco han iniciado su actividad sexual.

Los hallazgos en función a los comportamientos de los adolescentes entre el año 2012 y 2018 no se mostraron notorios por lo que se concluye que los adolescentes mantienen, de manera similar, los comportamientos y conocimientos acerca de los factores de riesgo y los protectores, encontrando la necesidad de fortalecer el conocimiento de los mismos, es de suma importancia para la promoción y prevención de la salud en el adolescente.

Se demostró que existe una necesidad primordial en los estados del centro y sur del país en los que no cuentan con conocimientos suficientes en cuanto a las infecciones de transmisión sexual, tema esencial en el desarrollo de los adolescentes, ya que el conocimiento del mismo es un factor de protección hacia el adolescente quien está iniciando su actividad sexual en una etapa temprana y media de su adolescencia.

De la misma manera, se concluye que un caso esencial en el tema son las mujeres adolescentes, a quienes se encontró que no tienen conocimientos suficientes en cuanto al uso de métodos anticonceptivos, los programas nacionales de protección y prevención son de suma importancia para la prevención de infecciones de transmisión sexual, así como, embarazos no deseados en edades tempranas.

El uso de las diversas técnicas estadísticas descritas en el presente capítulo, permitió aportar mayor información en cuanto a la búsqueda de los factores asociados al inicio de la actividad sexual en los adolescentes del país; se describen los métodos estadísticos esenciales e importantes para estudios posteriores como ayuda en la elección de estrategias estadísticas en futuros trabajos, cuyo objetivo sea asociar variables cualitativas; se sugiere el uso de análisis de segmentación TAID-LCA ya que permite enfocar un principio asimétrico de las variables al permitir trabajar con diversas variables respuesta simultáneamente, mientras que el algoritmo CHAID no enfoca un principio asimétrico de las variables de estudio y aplica tablas colapsadas.

Referencias

- Blythe M, R. S. (2000). Female adolescent sexuality: promoting healthy sexual development. *Obstet Gynecol Clin North Am*, 27: 125-41.
- Caballero-Hoyos R, V.-S. A. (2001). El estrato socioeconómico como factor predictor del uso constante de condón en adolescentes. *Rev. Salud Pub*, 35:531-538.
- Carrión F. (2016) Aplicaciones estadísticas en la relación de variables: Autoconcepto y grado de madurez analizado en estudiantes. *Revista EAC*,5(1),8-13. ISSN: 1390-9029
- Castro, C., y Galindo, P. (2011). Colapsabilidad en tablas de contingencia multivariantes: Usando modelos log-lineales, modelos gráficos y algoritmos de segmentación, se tiene información de variables relevantes. Editorial académica española.
- Cohen D, F. T. (2002). When and where do youths have sex? The potential role of adult supervision. *Pediatrics*, 110(6): e66.
- Conard L, B. M. (2003). Sexual function, sexual abuse and sexually transmitted diseases in adolescence. *Best Pract Res Clin Obstet Gynecol*, 17: 103-16.
- Consejo Nacional para la Prevención y Control de SIDA (CONASIDA). (2006). México: El SIDA en México, categorías de transmisión. México, DF: Secretaría de Salud.
- Cutié, J. L. (2005). Primera Relación en Adolescentes Cubanos. *Revista Chilena OBSTET GINECOL*. 70) 2: 83-86.
- D, K. (2001). Emerging Answers: Research findings on programs to reduce teen pregnancy. National Campaign to Prevent Teen Pregnancy, 13-72.
- Feldman J, M. A. (2002). Adolescent sexuality and sexual behavior. *Curr Opin Obstet Gynecol*, 14: 489-93.

- Gayet C, J. F. (2003). Uso del condón entre adolescentes mexicanos para la prevención de las infecciones de transmisión sexual. *Salud Publica Mex*, 45 supl 5: S632-S640.
- Grunbaum J, K. L. (2002). Youth Risk Behavior Surveillance, 2001. United States MMWR. 51(SS-4): 1-64.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis multivariante*. Madrid: Prentice Hall Iberia.
- Instituto Mexicano de la Juventud. (2006). Encuesta Nacional de Juventud 2005, resultados preliminares. México, DF: Secretaria de Educación Pública.
- Instituto Nacional de Estadística, Geografía e Informática. (2006.). México: Porcentaje de nacimientos registrados de madres adolescentes (menores de 20 años), 1990-2005. Obtenido de INEGI.
- Juárez, F. &. (2005). Salud sexual y reproductiva de los adolescentes en México: Un nuevo marco de análisis para la evaluación y diseño de políticas. *Papeles de la Población. Papeles de población*, 11(45), 177-219. Recuperado el 01 de 12 de 2020, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S140574252005000300008&lng=es&tlng=es.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 127-199.
- Marchetti R, R. G. (1996). Los adolescentes y el SIDA: La importancia de la educación para la prevención. *SIDA-FTS*, 2:61-9.
- Martínez-Donate AP, H. M. (2004). Evaluation of two schools-based HIV prevention interventions in the border city of Tijuana, México. *JSR*, 41:267-278.
- Mendivelso F. & Rodríguez M. (2018). Prueba de chi cuadrado de independencia aplicadas a tablas 2xN. *Rev. Medica Sanitas*, 21(2),92-95. Doi:10.26852/01234250.6
- Menkes, C. &. (2003). Sexualidad y embarazo adolescente en México. *Papeles de población*, 9(35). Recuperado el 27 de noviembre de 2020, de http://www.scielo.org.mx/scielo.php?script=sci_isoref&pid=S140574252003000100011&lng=es&tlng=es
- Meza, J. (2015). Salud sexual y reproductiva en adolescentes y jóvenes en México en el siglo XXI. *Uaricha.*, 12(29), 129-14. Recuperado el 03 de diciembre de 2020, de https://www.revistauaricha.umich.mx/ojs_uaricha/index.php/urp/article/view/21/20
- Ojeda, N. ((2013).). Cambios y continuidades de la unión libre en Tijuana, México: el caso de las jóvenes en Tijuana. *Perfiles latinoamericanos*, 21(42), 193-224. Recuperado el 01 de 12 de 2020, de http://www.scielo.org.mx/scielo.php?script=sci_art
- P., G. (2014.). Las conductas sexuales de riesgo de los adolescentes españoles. *Revista de Seapa.*, 2 (2): 38-61.
- Secretaría de Salud. (2002). Programa de Acción VIH/SIDA e Infecciones de Transmisión Sexual. México, DF: Secretaria de Salud.
- Sieving R, O. J. (2002). Adolescent sexual behavior and sexual health. *Pediatr Rev*, 23: 407-16.
- UNICEF. (2015). Una aproximación a la situación de adolescentes y jóvenes en América. Una aproximación a la situación de adolescentes y jóvenes en América, (págs. 6-51). México, D.F.
- Villaseñor-Sierra A, C.-H. R.-S.-P. (2003). Conocimiento objetivo y subjetivo sobre el VIH/SIDA como predictor del uso de condón en adolescentes. *Salud Publica Mex*, 45 supl 1: S73-S80.
- Walker D, G. J. (2006). HIV prevention in Mexican schools: Prospective randomized evaluation of intervention. *BMJ*, 332:1189-1194.

Predicción de casos y mortalidad por diabetes, cambios alimentarios para la reducción del impacto

Keila Raquel Medina Aguilar¹, María Yesenia Zavaleta Sánchez²

¹Facultad de Estadística e Informática, Universidad Veracruzana. keilarmedina@gmail.com

²Facultad de Estadística e Informática, Universidad Veracruzana. yzavaleta@uv.mx

Resumen

La diabetes mellitus (DM) es una enfermedad crónica resultante de la incapacidad del páncreas para producir insulina o del cuerpo para utilizarla adecuadamente (OMS, 2020). Durante 1980, la tasa de defunciones ocasionadas por diabetes en México fue de 21.8 por cada 100,000 habitantes y 86.0 en 2016 (SSA/DGE, 2010; 2018). Un importante factor preventivo de esta enfermedad es llevar una alimentación adecuada con alimentos ricos en fibra, bajos en grasas, azúcares y calorías (NIDDK, 2016; Clínica Mayo, 2019). El objetivo de este estudio es crear modelos estadísticos para predecir tasas futuras de muerte e incidencia de diabetes en México y conocer si existen cambios alimenticios que puedan regular dichas tasas. La metodología de estudio consistió en la identificación de tasas de mortalidad e incidencia de diabetes (tipos 1 y 2) y calorías consumidas por alimentos en México entre los años 2000 y 2017, clasificación de los alimentos por grupos de acuerdo con el Sistema Mexicano de Alimentos Equivalentes, descripción y exploración de las variables, análisis de correlación entre las tasas y los grupos alimentarios, análisis de componentes principales (ACP) para los grupos alimenticios y modelación estadística de predicción futura de las tasas y modulación de estas mediante cambios en el consumo de alimentos. Los análisis estadísticos fueron realizados con el software R Project v.4.0.3. Se encontró que, por cada kilocaloría consumida/día de alimentos ricos en grasas y azúcares, la tasa de mortalidad incrementa 13.4/100,000 hab. Y por cada kilocaloría consumida/día de alimentos fuentes de grasas y azúcares, la tasa de incidencia de diabetes aumenta 21.9/100,000 hab., no así para el consumo de fuentes de fibra. De mantener los mismos patrones alimenticios de los últimos años, la tasa de mortalidad por diabetes en México sería de 97.70/100,000 habitantes para el año 2037. Los modelos creados detectaron relación entre el incremento en las tasas de incidencia y mortalidad por diabetes en México y la ingesta de alimentos ricos en grasas y azúcares. Una reducción en el consumo de alimentos altos en grasas y azúcares disminuirían las tasas futuras de casos y muertes por diabetes en México.

Palabras clave: Diabetes, dieta, mortalidad, incidencia, predicción, regresión múltiple.

Introducción

La diabetes mellitus (DM) es un padecimiento de tipo crónico, que resulta de la incapacidad del páncreas para producir suficiente insulina (tipos 1 y 2) o del cuerpo para utilizarla adecuadamente (tipo 2) (OMS, 2020). Debido a que los síntomas pueden presentarse varios años después del inicio de la DM, su diagnóstico no es inmediato. Esto último facilita la aparición de complicaciones y que sean difíciles de prevenir. Además, genera consecuencias asociadas que empeoran la salud y la calidad de vida del paciente (Organización Mundial de la Salud [OMS], 2020). Aunque la DM2 es más común en adultos y el tipo 1 en niños o jóvenes, en años recientes, se ha observado la presencia de DM2 también en menores de edad (OMS, 2020).

En el año 2015, se registraron alrededor de 1.6 millones de muertes asociadas a diabetes en el mundo, de las cuales el 43% fueron en individuos de 70 años o menos (OMS, 2016; 2020). En el 2014, 422 millones de adultos (4.7%) tenían diabetes; cuatriplicándose desde 1980 cuando la cifra era de 108 millones (8.5%). La OMS estima que entre el 24 y 62% de las personas que viven con diabetes tipo 2 desconocen su enfermedad debido a que aún no han sido diagnosticadas (OMS, 2016).

En México, durante el 2016, la diabetes fue la segunda causa de muerte (15.4% de las defunciones). La prevalencia de esta enfermedad aumentó del 9.7 al 10.3% en mujeres entre el 2012 y 2016, y se mantuvo alrededor del 8.5% en hombres entre esos mismos años. Durante 1980, la tasa de defunciones ocasionadas por diabetes mellitus fue de 21.8 por cada 100,000 habitantes y durante el 2016, 86.0 por cada 100,000 habitantes (Federación Mexicana de Diabetes [FMD], 2016; Dirección General de Epidemiología de la Secretaría de Salud [SSA/DGE], 2010; 2018).

Los factores de riesgo para desarrollar diabetes mellitus se han clasificado en genéticos y derivados del estilo de vida. Entre los genéticos se encuentran mutaciones y polimorfismos en los genes PPAR γ , KCNJ11/ABCC8, y HNF4A, así como antecedentes familiares de diabetes y algunas etnias (Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales [NIDDK por sus siglas en inglés], 2016, Udler, 2019). Algunos factores de riesgo derivados del estilo de vida son tener obesidad o sobrepeso, acantosis nigricans, mujeres que hayan presentado diabetes gestacional o con hijos con un peso al nacer \geq 4Kg, sedentarismo y alimentación alta en azúcares simples (NIDDK, 2016; Clínica Mayo, 2019).

Para prevenir la presencia de la diabetes tipo 2 es relevante disminuir el peso corporal entre un 5 y un 10% en personas con sobrepeso y obesidad, realizar actividad física moderada entre 30 y 60 minutos o actividad física intensa de tipo aeróbica 15 a 30 minutos por lo menos cuatro días a la semana, así como llevar una alimentación adecuada con alimentos ricos en fibra, bajos en grasas, azúcares y calorías (NIDDK, 2016; Clínica Mayo, 2019).

Marco contextual

En este sentido, llevar una alimentación baja en carbohidratos simples, grasas y sodio, es uno de los hábitos más relevantes en la prevención y el tratamiento de la diabetes. En poblaciones europeas se han implementado estrategias en la prevención de esta enfermedad, encontrando reducciones del riesgo para diabetes entre el 20 y el 70% gracias a la alimentación adecuada por sí sola o acompañada de actividad física (Hernández Ruiz de Eguilaz *et al.*, 2016). Además, el consumo de frutos secos redujo el riesgo a presentar diabetes en 18% y 40% por el consumo de aceite de oliva (Hernández Ruiz de Eguilaz *et al.*, 2016). Así mismo, la ingesta de productos cárnicos derivados de res y alimentos procesados (como galletas, pasteles, chocolate, dulces, fruta enlatada y leche evaporada, pan blanco, papas fritas) se han asociado con mayor riesgo de incidencia atribuible a diabetes en población canadiense (RIA = 1.38 IC 95% 1.02-1.86) (Reeds *et al.*, 2016).

También, el consumo de 1 a 3 raciones de 30g al día de cereales enteros disminuye significativamente el riesgo relativo de DM2 (RR 0.75 a 0.78), 1 a 7 raciones al día de vegetales y frutas (80g cada una) es factor protector significativo para DM2 (RR 0.90 a 0.96). El consumo de 1 a 5 raciones de 200g al día de lácteos también es un factor protector (RR 0.94 a 0.97). Mientras que, el consumo de huevos, granos refinados, carnes rojas o procesadas, dulces y bebidas

azucaradas son factores de riesgo significativos. La ingesta de nueces, legumbres y pescados no mostraron asociaciones con DM2 (Schwingshackl *et al.*, 2017).

En una población de indios asiáticos, se encontró que el consumo de almendras mostró una disminución significativa en los niveles medios de hemoglobina glucosilada en pacientes con DM2 ($dm = -0.3$ IC 95% -0.5 a -0.01) luego de 24 semanas de consumo (Gulati, Misra & Pandey, 2017). En otro estudio realizado en población del norte de Suecia se encontró que el llevar una dieta paleolítica (consumo de carne magra, pescado, mariscos, huevos, verduras, frutas, bayas y nueces) durante 12 semanas, aumentó significativamente la sensibilidad a insulina en un 45% con o sin ejercicio. Además, los niveles de hemoglobina glucosilada disminuyeron en un 0.9% (IC 95% -1.2 a -0.6) en los pacientes con diabetes que sólo llevaron dieta paleolítica y en un 1.1% (IC 95% -1.7 a -0.7) en pacientes que acompañaron la dieta paleolítica con una rutina de ejercicio (Otten *et al.*, 2017).

En un análisis sistemático del estudio de carga global de enfermedad (*Global Burden of Disease Study*) realizado en 2017, se encontró que México tuvo el ratio más alto de muertes por diabetes tipo 2 relacionadas con la dieta (35 [28–44] muertes por cada 100,000 habitantes) y 1605 (1231–2034) casos de discapacidad ajustada por años de vida por cada 100,000 habitantes relacionados también con la dieta. México se encuentra entre los países que se caracterizan por una ingesta baja de granos enteros, frutas, vegetales, nueces y semillas; ingesta de alimentos con bajo aporte de fibra y ácidos grasos omega 3, consumo elevado de sodio, grasas trans, azúcares y carnes rojas (Afshin *et al.*, 2019).

Durante el año 2000, la población mexicana invirtió el 5% de sus ingresos destinados a la alimentación en consumo de frutas, 11.4% en verduras, 16.9% en cereales, 23% en carnes y 6.9% en alimentos procesados de todo tipo. Para el año 2018, los mexicanos aumentaron su consumo de verduras y de cereales, con un gasto del 11.7% y 17.8% respectivamente. Sin embargo, disminuyeron su consumo de frutas a un 4.6% e incrementaron la ingesta de alimentos procesados a un 12% del gasto total en alimentos (Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria [CEDRSSA], 2019).

Asimismo, un estudio determinó que entre 1961 y 2013, las tasas de mortalidad de DM2 incrementaron significativamente y que la ingesta calórica aumentó en promedio 647.9 kcal por persona al día; también, hubo correlación entre la ingesta de azúcar y la elevación de la tasa de defunciones por diabetes. Además, se observó mayor consumo de azúcares, alimentos y grasas de origen animal y aceites vegetales, así como disminución en la ingesta de cereales y legumbres (Soto Estrada, Moreno Altamirano, García-García, Ochoa Moreno & Silberman, 2018).

Aunque la causa de la Diabetes tipo 1 (DM1) es desconocida, diferentes investigaciones han identificado relación entre la dieta y su incidencia, mortalidad y aparición prematura de complicaciones. Por ejemplo, la ingesta elevada de leche de vaca y carnes rojas, así como deficiente en cereales y ácidos grasos poliinsaturados han aumentado el riesgo de autoinmunidad en los islotes pancreáticos, riesgo de complicaciones y empeoramiento del estado de salud del paciente con DM1 (Norris, 2010; Rewers & Ludvigsson, 2016). Igualmente, en población danesa se compararon los efectos de una dieta baja en carbohidratos (<100 g carbohidratos/día) y una dieta alta en carbohidratos (>250 g carbohidratos/día) en pacientes con DM1, observando menor variación

glucémica (32,7 vs. 37,5%, $p < 0,013$) en los pacientes que llevaron la dieta baja en carbohidratos (Schmidt *et al.*, 2019).

Para el estudio predictivo en enfermedades como la diabetes ya se han tomado como referencia diferentes factores de riesgo y parámetros que evalúan el buen o mal control de dicha patología, mediante distintas herramientas estadísticas. Por una parte, en un análisis sistemático de la carga global de enfermedades se evaluaron factores de riesgo dietéticos, acompañados de otras variables como edad, sexo, país y año de registro utilizando un modelo de regresión del proceso espaciotemporal gaussiano (Afshin *et al.*, 2019).

En otro estudio se buscaron asociaciones entre la dieta y la incidencia de DM2. Se realizó un análisis factorial para identificar patrones alimentarios y una regresión logística multivariante para detectar asociaciones entre dichas variables. En dicho análisis, se realizaron ajustes por factores sociodemográficos y factores de confusión del estilo de vida como edad, sexo, circunferencias de cintura y niveles sanguíneos de interleucina 6 (IL-6) y adiponectina (Reeds *et al.*, 2016).

Por otra parte, se han desarrollado modelos predictivos más específicos que se han adaptado de mejor manera a las variables contempladas en esos estudios. Por ejemplo, en un estudio de proyección de prevalencia de diabetes para adultos mayores en México, se utilizó el modelo de ancianos futuro (FEM, por sus siglas en inglés). Este modelo es una simulación económica y geográfica creada originalmente para predecir costos en salud de adultos mayores, considerando variables indicadoras del estado de salud actual y exposición a factores de riesgo de diabetes. Se realizó un muestreo con reemplazo cada dos años considerando la edad, sexo, escolaridad, estado civil, Índice de Masa Corporal (IMC), hipertensión arterial (HTA), tabaquismo, actividad física, si el individuo residía en una localidad rural o urbana y si contaba o no con seguro de salud. Así como 4 escenarios, con aplicación o no de intervenciones y diferentes porcentajes de reducción de la incidencia de diabetes dos años posteriores al estudio para la predicción de la tasa de prevalencia (González González, Tysinger, Goldman & Wong, 2017).

Asimismo, una investigación predictiva del riesgo de diabetes con mediciones de cambio en los niveles de glucosa en ayunas y HbA1c (hemoglobina glucosilada), tomando en cuenta el IMC, edad, número de visitas al médico, sexo, tabaquismo y etnia. Se crearon dos modelos predictivos, el primero se trató de un modelo estático de riesgos proporcionales de Cox y el segundo un modelo dinámico también llamado modelo histórico de riesgos proporcionales de Cox. Este último fue nombrado así debido a que el modelo se actualiza en un momento histórico determinado a uno, dos o tres años partiendo del riesgo inicial y reflejando los cambios observados en los biomarcadores. Dicho modelo da como resultado un riesgo acumulado al momento t (Parast, Mathews & Friedberg, 2019).

Recientemente han surgido nuevos métodos llamados modelos de clasificación, que compiten con los métodos estadísticos convencionales, adaptando dichos análisis a los datos de los investigadores para lograr un mejor ajuste. Por ejemplo, una investigación de 2018 realizó una predicción en la aparición de complicaciones de diabetes (a 3, 5 y 7 años después del inicio del estudio) mediante el empleo de modelos de clasificación tales como regresión logística (LR), Bayes ingenuo (NB), máquinas de soporte vectorial (SVM) y bosque aleatorio (RF). Para estos modelos se midieron la sensibilidad, especificidad, precisión, valor predictivo positivo (VPP), valor predictivo negativo (VPN), área bajo el ROC (AUC) y coeficiente de correlación de Matthews (MCC) y se incluyeron

variables como sexo, edad, tiempo transcurrido desde el diagnóstico de diabetes, IMC, hemoglobina glucosilada, HTA y hábitos de tabaquismo (Dagliati A. *et al.*, 2018).

Sin embargo, estos estudios contemplaron variables intrínsecas del individuo, que son difíciles de obtener en datos poblacionales, pues sería costoso y bastante tardado recabarlos.

Justificación

Se estima que el gasto total de diabetes al año, en México, es de 224,350 millones de pesos (Salcedo Álvarez, Rivas Herrera, González Caamaño, Zárate Grajales & Villalobos San Luis, 2016) y que un 87% de este gasto se usa para tratar complicaciones (SSA, 2018). Según la OMS, aproximadamente el 80% de las muertes por diabetes ocurren en países de medios y bajos ingresos. México ocupa el lugar número 6 de personas con diabetes en el mundo. Se estima que cada 6 segundos sucede una defunción causada por diabetes en el mundo y que la mitad de las personas a quienes ya les fue diagnosticada esta patología no son conscientes de su condición (FMD, 2020).

En 2015, aproximadamente 1.6 millones de fallecimientos en el mundo fueron por diabetes, el 43% ocurrieron en personas menores de 70 años (OMS, 2016; 2020). En el 2014, la prevalencia mundial fue de 422 millones de adultos; siendo 4 veces superior que en 1980 (108 millones) (OMS, 2016). En 2016, la diabetes ocupó el segundo lugar en causas de defunciones en nuestro país, responsable del 15.4% de las defunciones; incrementando de 14,626 en 1980 a 105,574 en 2016 (FMD, 2016). Además, se ha estimado que la prevalencia en el año 2050 de diabetes en México para la población de adultos de 50 años y más será del 34% (González González *et al.*, 2017) y 13.7 a 22.5% en adultos de más de 20 años (15 y 25 millones de adultos) (Meza *et al.*, 2015).

A pesar de las estrategias implementadas para reducir la mortalidad e incidencia de diabetes, en México y en el mundo, estas cifras siguen en aumento. Existen estudios sobre predicciones de casos futuros de prevalencia diabetes tipo 2 y su mortalidad, pero se enfocan sólo en adultos mayores. Sin embargo, en los últimos años se han dado diagnósticos de este padecimiento en adultos jóvenes, adolescentes y niños (Temneanu, Trandafir & Purcarea, 2016). Esto último hace que se requieran predicciones más realistas, que incluyan a la población general.

Por otro lado, no existen estudios que predigan incidencia y mortalidad de diabetes asociadas a la alimentación, sino únicamente de manera general. Además, la mayoría de los estudios de alimentación y diabetes se centran en el consumo de alimentos industrializados, pero pocos se basan en los grupos de alimentos naturales que son los que se consumen día con día por toda la población, ni se ha establecido con claridad cuáles son los que contribuyen mayoritariamente al desarrollo de la diabetes mellitus y los que podrían atenuar el riesgo. Se han estudiado dietas específicas como la mediterránea y la paleolítica (compuesta principalmente de frutos y semillas de origen africano) (Hernández Ruiz de Eguilaz *et al.*, 2016; Otten *et al.*, 2017) pero la dieta mexicana incluye una amplia variedad de alimentos que difieren en gran medida de los que se consumen en otras regiones del mundo.

Debido a lo anterior se planteó como preguntas de investigación si ¿habrá un incremento de casos y muertes por diabetes en un futuro próximo? y ¿se podrá regular ese impacto con cambios en la ingesta dietaria? Bajo la hipótesis de que existe una relación directa entre el incremento en las tasas de incidencia y mortalidad por diabetes en México y la ingesta dietaria, y los cambios positivos en esta tiene una relación directa con la reducción de casos y muertes por diabetes en México.

El objetivo del estudio fue crear un modelo estadístico que permita predecir el número de casos y muertes por diabetes en México (en un futuro cercano) y conocer si existen cambios en la dieta que puedan regular la morbilidad y mortalidad futura de diabetes en México.

Metodología

Se incluyeron bases de datos sobre la alimentación de la población mexicana entre los años 2000 a 2018, obtenidas de los balances de alimentos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés) (FAO, 2020). Así como de casos incidentes de diabetes mellitus 1 y 2 y defunciones por diabetes mellitus 1 y 2 (tasas por cada 100,000 habitantes) entre el mismo periodo de tiempo obtenidos de la Secretaría de Salud y del Instituto de Estadística y Geografía (INEGI) (SSA/DGE, 2018; INEGI, 2018).

Los balances de alimentos se filtraron para obtener las cantidades de consumo de cada alimento excluyendo las cantidades de producción y de otros usos no relacionados a la ingesta humana. Cada alimento fue clasificado en los grupos de alimentos de acuerdo con el Sistema Mexicano de Alimentos Equivalentes (SMAE). Los grupos de clasificación fueron verduras, frutas, cereales y tubérculos, leguminosas, alimentos de origen animal, leches, aceites y grasas, azúcares, alimentos libres de energía y bebidas alcohólicas (Pérez Lizaur, Palacios González, Castro Becerra & Flores Galicia, 2014). La base de datos final se compone de 19 casos y 13 variables. El software estadístico utilizado fue R Project v4.0.3.

Se realizaron análisis descriptivos, exploratorios y correlaciones de Pearson, se aplicó un análisis de componentes principales para reducir la cantidad de variables alimentarias y se diseñaron modelos de regresión lineal múltiple para estimar las tasas futuras y los cambios alimentarios necesarios para reducir el impacto futuro de la diabetes en México.

Resultados y discusión

Análisis descriptivo y correlaciones

En la Tabla 1 se pueden observar las medias y las desviaciones estándar de casos nuevos y muertes por diabetes del año 2000 al 2018 en México. Así como el consumo total de calorías al día per cápita, del cual sobresale la ingesta de cereales y tubérculos, azúcares, alimentos de origen animal y, grasas y aceites.

También, se realizó un análisis de correlación de Pearson para conocer qué variables alimentarias se correlacionan con las tasas de incidencia y mortalidad por diabetes entre los años 2000 y 2018. Por una parte, como se aprecia en la Tabla 2, variables como consumo calórico obtenido de las leguminosas y azúcares, parecen disminuir la tasa de muertes por diabetes, así como la ingesta de cereales y tubérculos con la tasa de incidencia. Lo hallado para la ingesta de leguminosas y, de cereales y tubérculos podría deberse al alto contenido de grasas protectoras y de fibra contenidas en estos grupos de alimentos (INFOODS/FAO, 2018; Bouchenak & Lamri-Senhadji, 2013). Sin embargo, el resultado observado para el consumo de azúcares es contrario a lo esperado puesto que este grupo de alimentos, en estudios previos, se han asociado con riesgo y empeoramiento de la diabetes mellitus (Reeds *et al.*, 2016; Schwingshackl *et al.*, 2017; Schmidt *et al.*, 2019).

Tabla 1. Medias y desviaciones estándar de las variables de estudio.

Variable	Media	Desviación estándar
Tasa de incidencia	380.5	39.3
Tasa de mortalidad	67.6	14.3
Verduras	94.9	8.7
Frutas	107.6	5.0
Cereales y tubérculos	1439.9	45.9
Leguminosas	109.1	9.1
Alimentos de origen animal	422.5	50.4
Leche	164.5	4.4
Grasas y aceites	366.7	27.9
Azúcares	467.9	13.2
Alimentos libres de energía	7.4	1.6
Bebidas alcohólicas	58.6	7.0
Total kilocalorías	3239.2	91.5

Consumo promedio medido en kilocalorías/persona/día. Tasas por cada 100,000 habitantes de diabetes mellitus tipos 1 y 2.

Fuente: elaboración propia.

En contraste con lo anterior, el aporte de calorías que provienen de alimentos de origen animal, leches, aceites y grasas, alimentos libres de energía, bebidas alcohólicas y verduras parecen incrementar las tasas de casos y defunciones por diabetes mellitus en México. Esto último coincide con lo reportado para riesgo de diabetes en otras investigaciones (Reeds *et al.*, 2016; Schwingshackl *et al.*, 2017; Norris, 2010; Rewers & Ludvigsson, 2016). Sin embargo, no existe concordancia con el grupo de “alimentos libres de energía” compuesto en su mayoría por especias y sazónadores de alimentos, ni con la ingesta de verduras (Schwingshackl *et al.*, 2017). Aunque actualmente no hay evidencia que apoye estas correlaciones, pudiera ser que el abuso en el consumo represente riesgos para la salud.

Tabla 2. Correlaciones entre las variables alimentarias y las tasas de incidencia y mortalidad.

Variable	Tasa de incidencia	Tasa de mortalidad
Verduras	0.081	0.508*
Frutas	-0.204	-0.111
Cereales y tubérculos	-0.478*	-0.146
Leguminosas	-0.313	-0.581**
Alimentos de origen animal	0.550*	0.951**
Leche	0.195	0.654**
Grasas y aceites	0.572*	0.922**
Azúcares	-0.400	-0.651**
Alimentos libres de energía	0.373	0.698**
Bebidas alcohólicas	0.377	0.759**

Consumo promedio en kilocalorías/persona/día. ** $p < 0.01$. * $p < 0.05$.

Fuente: elaboración propia.

Posteriormente, se estimaron las medias y las desviaciones estándar de calorías totales y de aquellas provenientes de los macronutrientes per cápita al día, así como los porcentajes de consumo de

macronutrientes con relación al consumo calórico total. Estas estimaciones se compararon con las ingestas calóricas recomendadas por macronutrientes en el Codex Alimentarius con referencia en una dieta estándar de 2000 kilocalorías (Codex Alimentarius, 2012) y con los porcentajes de consumo de macronutrientes recomendadas para el tratamiento de la diabetes mellitus en la Norma Oficial Mexicana (NOM) 015-SSA2-2018 (Diario Oficial de la Federación [DOF], 2018). Los resultados de esta comparación se muestran en la Tabla 3.

Analizando el porcentaje de lípidos consumido, se observa que se encuentra dentro de los porcentajes que recomienda la NOM-015; por su parte, el porcentaje de proteínas se observa bajo para lo recomendado y el porcentaje de hidratos de carbono es un poco elevado. Sin embargo, al analizar las calorías de lípidos comparada con la dieta estándar recomendada por la FAO en el Codex Alimentarius, se puede ver que la ingesta es mucho más elevada, algo similar ocurre con los carbohidratos.

Tabla 3. Comparación de la ingesta observada y la ingesta recomendada de macronutrientes y calorías.

Variable	Ingesta media observada ± desviación estándar	Ingesta recomendada
Porcentaje de lípidos	26.47 ± 1.02	25 – 35%
Porcentaje de proteínas	10.91 ± 0.07	15 – 20%
Porcentaje de carbohidratos	62.63 ± 1.06	50 – 60%
Kcals de lípidos	857.84 ± 51.18	500 – 700 Kcals
Kcals de proteínas	353.26 ± 10.59	300 – 400 Kcals
Kcals de carbohidratos	2028.11 ± 47.58	1000 – 1200 Kcals
Kcals totales	3239.21 ± 91.47	2000 Kcals

Kcals: kilocalorías. Los porcentajes de ingesta recomendada se basan en la NOM-015 -SSA2-2018: para la prevención, detección, diagnóstico, tratamiento y control de la Diabetes Mellitus. La ingesta calórica recomendada está basada en los lineamientos de la FDA y el Codex Alimentarius de la FAO-OMS.

Fuente: elaboración propia.

A pesar de que la ingesta de calorías provenientes de las proteínas, el consumo calórico total promedio es mayor que lo recomendado; por lo cual, aunque el porcentaje de lípidos esté dentro del rango de la norma oficial, no es el mejor, dado que sigue siendo mayor. La población mexicana tuvo un excedente aproximado de 1240 kilocalorías al día, que puede estar afectando y agravando la mortalidad e incidencia por diabetes.

Análisis de Componentes Principales (ACP)

Posteriormente, se realizó un análisis de componentes principales (con la finalidad de reducir el número de variables alimentarias), este análisis arrojó la presencia de 3 componentes con autovalores mayores que 1 es 3. Sin embargo, debido a que en el tercer componente únicamente se agrupaba una variable, se decidió hacer el ACP con solamente 2 componentes (Tabla 4) que explican el 73% de la varianza. Además, se aplicó una rotación varimax para una mejor comprensión de los componentes. Este análisis obtuvo un valor de Medida Kaiser-Meyer-Olkin de adecuación de muestreo (KMO) de 0.768, es decir, fue un análisis aceptable; y una prueba de esfericidad de Bartlett de 142.18 (45 gl, $p < 0.001$) que indica que es factible aplicar este análisis. En la Tabla 4 se aprecian las variables que integran a cada componente resaltadas en negritas, ya que cada una tiene un peso distinto entre un componente y otro. También se observa que algunas cambian de componente al aplicar la rotación varimax. Esta rotación permitió una mejor

interpretación de los componentes, los cuales se nombraron como “fuentes de grasas y azúcares” (componente 1) y “fuentes de fibra” (componente 2). Este análisis se realizó con el fin de ocuparse en los análisis de regresión lineal múltiple, en caso de que los modelos de las variables alimentarias originales no cumplieran con los supuestos de la regresión múltiple.

Tabla 4. Componentes principales sin rotar vs. Rotación varimax.

Variable	Componentes sin rotar		Componentes rotados	
	Componente 1	Componente 2	Componente 1	Componente 2
Verduras	.789	.493	.573	.733
Frutas	.247	.755	-.026	.794
Cereales y tubérculos	.162	.708	-.090	.721
Leguminosas	-.647	.370	-.734	.126
Alimentos de origen animal	.847	-.308	.901	.000
Leches	.843	.187	.728	.464
Grasa y aceites	.932	-.226	.953	.106
Azúcares	-.651	.451	-.766	.201
Libres de energía	.890	.192	.770	.485
Bebidas alcohólicas	.885	-.016	.838	.287

Fuente: elaboración propia.

Modelo para la tasa de mortalidad

Lo siguiente fue probar modelos de regresión lineal múltiple, tanto para la tasa de mortalidad como para la tasa de incidencia, mediante la fórmula:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_i + \varepsilon$$

Donde

y: variable dependiente

β_0 : intercepto

β_j : coeficientes del modelo

x_i : variables independientes

ε : error del modelo (Montgomery D. C., Peck E. A. & Vining G. G., 2006)

En el caso de la tasa de mortalidad, se generó un modelo de regresión múltiple que evaluara la tasa de mortalidad respecto a los componentes principales rotados (“fuentes de grasas y azúcares” y “fuentes de fibra”). El modelo de mortalidad con componentes principales dio un error estándar residual de 5.292, un R^2 de 0.879 y un R^2 ajustado de 0.864, es decir, los dos componentes son capaces de explicar el 86% de la mortalidad por diabetes. Solamente el componente uno, que agrupó alimentos ricos en grasas y azúcares, fue significativo (Tabla 5); por lo que, fue posible utilizar el modelo ya que se cumplieron todos los supuestos (Tabla 6).

Tabla 5. Coeficientes del modelo para la mortalidad.

	Coefficiente	Error estándar	t	p
Intercepto	67.608	1.214	55.69	<0.001
Componente 1	13.401	1.247	10.74	<0.001
Componente 2	-0.844	1.247	-0.68	0.508

Componente 1: Fuentes de grasas y azúcares. Componente 2: Fuentes de fibra.

Fuente: elaboración propia.

El modelo final para la tasa de mortalidad fue conformado de la siguiente manera:

$$\hat{y} = 67.608 + (13.401)componente1 + (-0.844)componente2$$

Donde

\hat{y} : Tasa de mortalidad promedio por diabetes en México.

Componente1: Consumo de kilocalorías en alimentos ricos en grasas y azúcares.

Componente2: Consumo de kilocalorías en alimentos ricos en fibra.

Tabla 6. Supuestos, valores influyentes y de apalancamiento del modelo para la mortalidad.

Supuesto	Prueba	Estadístico	p
Normalidad	Shapiro-Wilk	0.967	0.719
Homocedasticidad	Breusch-Pagan estudentizado	2.575	0.276
Independencia	Durbin-Watson	1.987	0.309
No colinealidad	VIF	1(CP1) 1(CP2)	-----
Influencia	Distancia de cooks	Todos <1	-----
Outliers	Valores de hat	0.25	-----

Fuente: elaboración propia.

De acuerdo con estos resultados, se puede decir que, por cada kilocaloría consumida de alimentos ricos en grasas y azúcares, la tasa de mortalidad incrementa en 13.4 por cada 100,000 habitantes y este incremento es estadísticamente significativo. Por el contrario, por cada kilocaloría ingerida de alimentos ricos en fibra, la tasa de mortalidad disminuye en 0.84 por cada 100,000 habitantes; además de que este último hallazgo no mostró ser significativo, tampoco resulta clínicamente relevante al ser una reducción tan pequeña. Actualmente, no existen estudios similares con el que se puedan comparar los resultados que se muestran en la tabla 5, pero otros estudios muestran que la ingesta de alimentos fuentes de grasas y azúcares tienen una relación directa con la mortalidad por diabetes mellitus, tal como se observa en este análisis (Norris, 2010; Rewers & Ludvigsson, 2016; Schmidt *et al.*, 2019; Reeds *et al.*, 2016; Schwingshackl *et al.*, 2017).

Posteriormente, se generaron las predicciones para los siguientes años (Tabla 7) con sus intervalos de confianza al 95%. Encontrando que, de mantener los mismos patrones alimenticios de los últimos años, la tasa de mortalidad por diabetes mellitus (1 y 2) en México para el año 2037 sería de 97.70 por cada 100,000 habitantes, llegando hasta 104.18/100,000 habs. en el peor de los casos.

Tabla 7. Análisis predictivo para la tasa futura de mortalidad por diabetes mellitus en México.

Año	Mortalidad*	IC 95%
2019	50.82	46.43 a 55.21
2020	50.68	45.58 a 55.78
2021	53.42	49.17 a 57.66
2022	51.82	46.53 a 57.12
2023	53.66	49.25 a 58.06
2024	69.45	66.84 a 72.05
2025	60.01	57.01 a 63.00
2026	66.62	64.04 a 69.20
2027	64.58	61.81 a 67.36
2028	64.22	60.53 a 67.91
2029	64.88	59.65 a 70.11
2030	77.37	71.82 a 82.93
2031	67.61	63.05 a 72.16
2032	68.85	65.34 a 72.37
2033	68.10	65.49 a 70.71
2034	79.85	75.85 a 83.85
2035	87.82	82.40 a 93.24
2036	87.10	80.97 a 93.24
2037	97.70	91.23 a 104.18

*Tasa por cada 100,000 habitantes. IC 95%: intervalo de confianza al 95%.

Fuente: elaboración propia.

Aunque la tasa de mortalidad entre los años 2000-2018 mostró valores entre 44.88 y 101.26 por cada 100,000 habitantes, las estimaciones puntuales de la predicción van de 50.68 a 97.70 y sus intervalos también fueron cercanos a los datos originales. De ese modo, esta estimación indica que, para el año 2020, la tasa de mortalidad por DM en México debía llegar a 50.68 (IC95% 45.58 a 55.78) y a 77.37/100,000 hab. para el 2030 (71.82 a 82.93 en el mejor y peor de los casos respectivamente). Sin embargo, estas cifras son inferiores a lo reportado en otros estudios en los que las predicciones van de 120 a 124.5/100,000 hab. para los mismos años (Reyes Mendoza L., 2016, Fuentes M. L. & Arellano S., 2018). Además, otro estudio enfocado en la región del Estado de México estimó que, para el año 2030, la tasa de mortalidad por diabetes será entre 88.9 y 606.54/100,000 hab. únicamente para dicho estado (Reza Curiel Y. B., 2018). Todo lo anterior podría deberse a que en este estudio fue difícil conocer si la mortalidad seguirá aumentando en función de un sólo factor de riesgo como lo es la alimentación, cuando la DM y sus complicaciones son multifactoriales (NIDDK, 2016). Así, los estudios citados evaluaron diversos tipos de variables en los análisis predictivos o se trató de un análisis de series de tiempo que sólo contempla como variables el año y la tasa anual.

Modelo para la tasa de incidencia

Para la tasa de incidencia se generó un modelo de regresión múltiple que evaluara la tasa de incidencia respecto a los componentes principales rotados (“fuentes de grasas y azúcares” y “fuentes de fibra”). El modelo de incidencia y los componentes principales dieron un error estándar residual de 31.59, un R^2 de 0.42 y un R^2 ajustado de 0.35 ($p < 0.05$); es decir, los dos componentes alimentarios son capaces de explicar el 35% de la incidencia de diabetes. Al igual que en el modelo

para la mortalidad, el componente 1 que agrupó alimentos ricos en grasas y azúcares fue el único significativo (Tabla 8).

Tabla 8. Coeficientes del modelo para la incidencia.

	Coefficiente	Error estándar	t	p
Intercepto	380.514	7.25	52.50	<0.001
Componente 1	21.940	7.45	2.95	<0.01
Componente 2	-13.131	7.45	-1.76	0.096

Fuente: elaboración propia.

Tabla 9. Supuestos, valores influyentes y de apalancamiento del modelo para la incidencia.

Supuesto	Prueba	Estadístico	p
Normalidad	Shapiro-Wilk		0.98
Homocedasticidad	Breusch-Pagan estudentizado		5.80
Independencia	Durbin-Watson		1.05
No colinealidad	VIF	1(CP1) 1(CP2)	-----
Influencia	Distancia de cooks	Todos <1	-----
Outliers	Valores de hat	0.15	-----

Fuente: elaboración propia.

En cambio, no se cumplió con el supuesto de independencia (Tabla 9) de acuerdo con lo obtenido por el estadístico de Durbin-Watson y su valor p asociado. Aunque este y el supuesto de multicolinealidad no son necesarios de verificar, dado que el modelo se realizó con los componentes principales como variables independientes y este análisis genera variables independientes e incorrelacionadas entre sí (Montgomery D. C. *et al.*, 2006).

El modelo final para la tasa de incidencia fue formulado de la siguiente manera:

$$\hat{y} = 380.51 + (21.94)componente1 + (-13.13)componente2$$

Donde

\hat{y} : Tasa de incidencia promedio por diabetes en México.

Componente1: Consumo de kilocalorías en alimentos ricos en grasas y azúcares.

Componente2: Consumo de kilocalorías en alimentos ricos en fibra.

Dados los resultados del modelo, se tiene que, por cada kilocaloría de alimentos fuentes de grasas y azúcares que una persona consume al día, la tasa de incidencia de diabetes aumenta en 21.9 por cada 100,000 habitantes y que, por cada kilocaloría ingerida de alimentos ricos en fibra, la tasa de incidencia disminuye en 13.1 por cada 100,000 habitantes, aunque este último dato no fue significativo a un valor de α de 0.05. Como ya se mencionó en el modelo de la mortalidad, no existen estudios similares que contemplen variables alimentarias para predecir tasas de mortalidad e incidencia de DM, pero el consumo de alimentos ricos en grasas y azúcares ya se han asociado con el riesgo de diabetes mellitus (Norris, 2010; Rewers & Ludvigsson, 2016; Schmidt *et al.*, 2019; Reeds *et al.*, 2016; Schwingshackl *et al.*, 2017).

Pese a que en este estudio la ingesta de alimentos ricos en fibra no mostró disminuir las tasas de mortalidad e incidencia por diabetes de manera significativa, existen otros estudios en los que el

consumo de dichos alimentos fue un factor protector (Hernández Ruiz de Eguilaz *et al.*, 2016; Gulati, Misra & Pandey, 2017; Otten *et al.*, 2017). De forma similar a este hallazgo, una investigación previa no encontró relación entre la ingesta de alimentos fuentes de fibra con riesgo a DM (Schwingshackl *et al.*, 2017).

Puesto que el modelo para la tasa de incidencia por diabetes no obtuvo un buen ajuste, es decir, las variables alimentarias explican muy poco la tasa de incidencia (35%), no se generaron predicciones futuras. Esto coincide con la literatura para la diabetes mellitus tipo 1 dado que la aparición de esta enfermedad se ha asociado más a otros factores de riesgo y causas desconocidas que a la alimentación (NIDDK, 2016). A pesar de que, para las tasas también se incluyeron las incidencias y defunciones del tipo 2 de diabetes mellitus, el estar presente el tipo 1 pudo haber influenciado para que, en el modelo de incidencia, la alimentación no explicara un porcentaje mayor de los casos de diabetes en México. Por el contrario, en el modelo de la mortalidad, la alimentación explicó un porcentaje mayor de la tasa de defunciones por DM para los tipos 1 y 2. Aunque, como ya se mencionó, en la aparición del tipo 1 no es tan relevante la dieta como para el tipo 2, el llevar una mala alimentación sí se ha asociado con mayor riesgo de complicaciones y mortalidad en pacientes que viven con DM1 (Norris, 2010; Rewers & Ludvigsson, 2016; Schmidt *et al.*, 2019).

Conclusiones

De acuerdo con el análisis de correlación, el consumo de unos u otros grupos de alimentos sí ha influenciado los patrones de tasas de incidencia y mortalidad por diabetes mellitus en México en los últimos años. También, existe una relación directa entre el incremento en las tasas de incidencia y mortalidad por diabetes en México y la ingesta de alimentos ricos en grasas y azúcares. Por el contrario, no se observó influencia del consumo de fuentes de fibra sobre las tasas de mortalidad e incidencia de DM.

Además, la implementación de cambios positivos en la ingesta dietaria de los mexicanos, como la reducción en el consumo de alimentos altos en grasas y azúcares (leguminosas, alimentos de origen animal, leches, grasas, azúcares, libres de energía, bebidas alcohólicas), disminuirían significativamente las tasas futuras de casos y muertes por diabetes en México. Finalmente, se logró crear un modelo estadístico de regresión lineal múltiple para la tasa de mortalidad que permitió realizar una predicción hasta el año 2037.

Referencias

- Afshin A., Sur P.J., Fay K.A., Cornaby L., Ferrara G., Salama J.S., *et al.* (2019). Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* (London, England), 393(10184), 1958–1972. [https://doi.org/10.1016/S0140-6736\(19\)30041-8](https://doi.org/10.1016/S0140-6736(19)30041-8)
- Bouchenak M. & Lamri Senhadji M. (2013). Nutritional quality of legumes, and their role in cardiometabolic risk prevention: a review. *Journal of medicinal food*, 16(3), 185–198. <https://doi.org/10.1089/jmf.2011.0238>
- Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria [CEDRSSA]. (2020, 10 de noviembre). Evolución de la producción y el consumo de alimentos en México (2000-2018). http://www.cedrssa.gob.mx/post_n-evolucionin-n_de_la_n-produccinin-n_y_el_n-consumo-n_de_alimentos_en_mn-xico_-2000-2018.htm

- Comité del Codex sobre nutrición y alimentos para regímenes especiales. (2020, 20 de diciembre). Programa conjunto FAO/OMS sobre normas alimentarias. http://www.fao.org/tempref/codex/Meetings/CCNFSDU/ccnfsdu34/nf34_05s.pdf
- Dagliati A., Marini S., Sacchi L., Cogni G., Teliti M., Tibollo V., De Cata P., Chiovato L., & Bellazzi R. (2018). Machine Learning Methods to Predict Diabetes Complications. *Journal of diabetes science and technology*, 12(2), 295–302. <https://doi.org/10.1177/1932296817706375>
- Federación Mexicana de Diabetes [FMD]. (2020a, 4 de abril). Estadísticas en México. <http://fmdiabetes.org/estadisticas-en-mexico/>
- Federación Mexicana de Diabetes [FMD]. (2020b, 4 de abril). Diabetes en México. <http://fmdiabetes.org/diabetes-en-mexico/>
- Food and Agriculture Organization of the United Nations [FAO]. (2020, 4 de abril). FAOSTAT. Food Balance. <http://www.fao.org/faostat/en/#data>
- Fuentes M. L., Arellano S. Panorama de la diabetes en México. (2020, 16 de abril). México Social. <https://www.mexicosocial.org/panorama-de-la-diabetes-en-mexico/>
- González González C., Tysinger B., Goldman D. P. y Wong, R. (2017). Proyección de la prevalencia de diabetes en mexicanos de 50 años y más: el modelo del futuro anciano-México (FEM-México). *BMJ abierto*, 7(10),e017330. <https://doi.org/10.1136/bmjopen-2017-017330>
- Gulati, S., Misra, A., & Pandey, R. M. (2017). Effect of Almond Supplementation on Glycemia and Cardiovascular Risk Factors in Asian Indians in North India with Type 2 Diabetes Mellitus: A 24-Week Study. *Metabolic syndrome and related disorders*, 15(2), 98–105. <https://doi.org/10.1089/met.2016.0066>
- Hernández Ruiz de Eguilaz, M., Batlle, M. A., Martínez de Morentin, B., San-Cristóbal, R., Pérez-Díez, S., Navas-Carretero, S., & Martínez, J. A. (2016). Cambios alimentarios y de estilo de vida como estrategia en la prevención del síndrome metabólico y la diabetes mellitus tipo 2: hitos y perspectivas. *Anales del Sistema Sanitario de Navarra*, 39(2), 269-289. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272016000200009&lng=es&tlng=es.
- Instituto Nacional de Estadística y Geografía [INEGI]. (2020, 2 de mayo). Mortalidad general. <https://www.inegi.org.mx/sistemas/olap/Proyectos/bd/continuas/mortalidad/MortalidadGeneral.asp>
- INFOODS/FAO. (2020, 4 de mayo). Food Composition Databases. <http://www.fao.org/infoods/infoods/tablas-y-bases-de-datos/bases-de-datos-faoinfoods-de-composicion-de-alimentos/es/>
- Mayo Clinic. (2020, 5 de abril). Diabetes de tipo 2. <https://www.mayoclinic.org/es-es/diseases-conditions/type-2-diabetes/symptoms-causes/syc-20351193>
- Montgomery D. C., Peck E. A. & Vining G. G. (2006). *Introducción al Análisis de Regresión Lineal*. México: CECSA.
- Meza R., Barrientos Gutiérrez T., Rojas Martínez R., Reynoso Noverón N., Palacio Mejía L. S., Lazcano Ponce E., & Hernández Ávila M. (2015). Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates. *Preventive medicine*, 81, 445–450.
- National Institute of Diabetes and Digestive and Kidney Diseases [NIDDK]. (2020, 2 de abril). Factores de riesgo para la diabetes tipo 2. <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/factores-riesgo-tipo-2>

- Norma Oficial Mexicana. Para la prevención, detección, diagnóstico, tratamiento y control de la Diabetes Mellitus. 2018, 3 de mayo. D.O.F. No. 015-SSA2-2018 https://www.dof.gob.mx/nota_detalle.php?codigo=5521405&fecha=03/05/2018
- Norris, J. M. (2010). Dieta para lactantes y niños y riesgo de diabetes tipo 1: avances y perspectivas recientes. *Curr Diab Rep* 10, 345–349. <https://doi.org/10.1007/s11892-010-0131-5>
- Organización Mundial de la Salud [OMS]. (2020, 5 de abril). INFORME MUNDIAL SOBRE LA DIABETES. <https://apps.who.int/iris/bitstream/handle/10665/254649/9789243565255-spa.pdf;jsessionid=BB77C7FE8F00FD090CF22969A3B6675E?sequence=1>
- Organización Mundial de la Salud [OMS]. (2020, 6 de abril). Diabetes. <https://www.who.int/es/news-room/fact-sheets/detail/diabetes>
- Otten, J., Stomby, A., Waling, M., Isaksson, A., Tellström, A., Lundin-Olsson, L., Brage, S., Ryberg, M., Svensson, M., & Olsson, T. (2017). Benefits of a Paleolithic diet with and without supervised exercise on fat mass, insulin sensitivity, and glycemic control: a randomized controlled trial in individuals with type 2 diabetes. *Diabetes/metabolism research and reviews*, 33(1), 10.1002/dmrr.2828. <https://doi.org/10.1002/dmrr.2828>
- Parast L., Mathews M. & Friedberg M. W. (2019). Dynamic risk prediction for diabetes using biomarker change measurements. *BMC medical research methodology*, 19(1), 175. <https://doi.org/10.1186/s12874-019-0812-y>
- Pérez Lizaur A. B., Palacios González B, Castro Becerra A. L. & Flores Galicia I. (2014). Sistema Mexicano de Alimentos Equivalentes. (4ª ed.). México D.F.: Ogali.
- Reeds, J., Mansuri, S., Mamakeesick, M., Harris, S. B., Zinman, B., Gittelsohn, J., Wolever, T. M., Connelly, P. W., & Hanley, A. (2016). Dietary Patterns and Type 2 Diabetes Mellitus in a First Nations Community. *Canadian journal of diabetes*, 40(4), 304–310. <https://doi.org/10.1016/j.jcjd.2016.05.001>
- Rewers M., Ludvigsson J. (2016). Environmental risk factors for type 1 diabetes. *The Lancet* 387(10035), 2340-2348. [https://doi.org/10.1016/S0140-6736\(16\)30507-4](https://doi.org/10.1016/S0140-6736(16)30507-4).
- Reyes Mendoza L. (2016). Análisis de la mortalidad en México de enfermedades crónico degenerativas no transmisibles relacionadas con la obesidad: una proyección al año 2020. [Tesis de Licenciatura, Universidad Autónoma del Estado de México]. <https://core.ac.uk/download/pdf/154794805.pdf>
- Reza Curiel B. Y. (2018). Patrones de distribución espacial de la Diabetes Mellitus en el Estado de México y su relación con factores socioeconómicos 1990 al 2015. [Tesis de Maestría, Universidad Autónoma del Estado de México]. <http://ri.uaemex.mx/bitstream/handle/20.500.11799/99449/MAEG%20Brenda%20Yazmin%20Reza%20Curiel.pdf?sequence=1&isAllowed=y>
- Salcedo Álvarez R. A., Rivas Herrera J. C., González Caamaño B. C., Zárate Grajales R. A. & Villalobos San Luis S. (2016). Gasto per cápita en salud, cobertura de enfermería y diabetes en México. *Revista CONAMED*, 21(3):120-125.
- Schmidt S., Christensen M. B., Serifovski N., Damm-Frydenberg C., Jensen J. B., Fløyl T., Størling J., Ranjan A., & Nørgaard K. (2019). Low versus high carbohydrate diet in type 1 diabetes: A 12-week randomized open-label crossover study. *Diabetes, obesity & metabolism*, 21(7), 1680–1688. <https://doi.org/10.1111/dom.13725>
- Schwingshackl, L., Hoffmann, G., Lampousi, A. M., Knüppel, S., Iqbal, K., Schwedhelm, C., Bechthold, A., Schlesinger, S., & Boeing, H. (2017). Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *European journal of epidemiology*, 32(5), 363–375. <https://doi.org/10.1007/s10654-017-0246-y>

- Secretaría de Salud/Dirección General de Epidemiología [SSA/DGE]. (2020a, 18 de abril). *Panorama Epidemiológico y Estadístico de la Mortalidad en México 2010*. https://www.gob.mx/cms/uploads/attachment/file/267597/Mortalidad_2010.pdf
- Secretaría de Salud [SSA]. (2020b, 18 de abril). *Diabetes: una enfermedad con complicaciones costosas que pueden evitarse*. <https://www.gob.mx/salud/articulos/diabetes-una-enfermedad-con-complicaciones-costosas-que-pueden-evitarse>
- Secretaría de Salud/Dirección General de Epidemiología [SSA/DGE]. (2020c, 19 de abril). *Principales causas de defunción por año*. http://sinaiscap.salud.gob.mx:8080/DGIS/GenerarTabla?titulo=Principales%20causas%20de%20defunci%F3n%20por%20a%F1o&href=/tablero/defunciones/defunciones_principalesCausas_porAnio/1998-2018_defunciones_principalesCausas_porAnio.xlsx&ruta=/tablero/defunciones/defunciones_principalesCausas_porAnio/1998-2018_defunciones_principalesCausas_porAnio.xlsx&hoja=2016&adicionalesTitulo=2016
- Soto Estrada, G., Moreno Altamirano, L., García-García, J. J., Ochoa Moreno, I., & Silberman, M. (2018). Trends in frequency of type 2 diabetes in Mexico and its relationship to dietary patterns and contextual factors. *Gaceta sanitaria*, 32(3), 283–290. <https://doi.org/10.1016/j.gaceta.2017.08.001>
- Temneanu, O. R., Trandafir, L. M., & Purcarea, M. R. (2016). Type 2 diabetes mellitus in children and adolescents: a relatively new clinical problem within pediatric practice. *Journal of medicine and life*, 9(3), 235–239.
- Udler, M. S. (2019). Type 2 Diabetes: Multiple Genes, Multiple Diseases. *Current Diabetes Reports*, 19(8), 1–9.

Análisis de componentes principales en altas dimensiones aplicado a la metabolómica de suelos del café con *Pontoscolex corethrurus*

Marycruz Álvarez-Jiménez¹, Jesús Hernández-Suárez², Sergio Francisco Juárez-Cerrillo³, Isabelle Barois-Boullard⁴

¹Facultad de Estadística e Informática, Universidad Veracruzana. mewry.alvarez@gmail.com

²Facultad de Estadística e Informática, Universidad Veracruzana. jeshernandez@uv.mx

³Facultad de Estadística e Informática, Universidad Veracruzana. sejuarez@uv.mx

⁴Red de Ecología Funcional, Instituto de Ecología A.C. isabelle.barois@inecol.mx

Resumen

El Análisis de Componentes Principales en Altas Dimensiones (ACPAD) es una técnica de reducción de dimensionalidad para matrices de datos con más variables que casos. Este tipo de matrices son comunes en las ciencias biológicas, por ejemplo, en la metabolómica. En este capítulo se presentan los elementos teóricos del ACPAD y se demuestra su aplicación en un análisis metabolómico de suelo y rizosfera de dos especies de café en presencia y ausencia de la lombriz de tierra, *Pontoscolex corethrurus*. La aplicación del ACPAD permitió realizar un análisis inicial con una representación simplificada de los datos metabolómicos. Se utilizó el programa MetaboAnalyst con el cual se observó que el uso de este método es viable para el análisis exploratorio y reducción de dimensionalidad del análisis químico en suelos; además se identificó que la composición química de los suelos varía entre sí con base en la función de su cercanía a las raíces, la especie de café, la presencia de lombriz y el método de identificación de las moléculas.

Palabras Clave: Datos metabolómicos, matrices de datos dispersas, matrices de datos gordas, MetaboAnalyst, suelo.

Introducción

En las ciencias biológicas es frecuente recolectar datos para más variables, que número de muestras (casos). Por ejemplo, en metabolómica, los espectrómetros de masas pueden medir miles de metabolitos en tan solo una muestra biológica de tejidos de plantas o incluso del suelo (Ringnér, 2008; Swenson et al. 2015). Sin embargo, la mayoría de las técnicas tradicionales del análisis multivariado requieren que el número de casos n sea mayor que el número de variables p , de otro modo estas técnicas no funcionan adecuadamente. La visualización en bajas dimensiones de este tipo de datos se dificulta y por lo tanto se requieren de procedimientos de análisis alternos.

El Análisis de Componentes Principales en Altas Dimensiones (ACPAD) es una técnica de aprendizaje no supervisado que se aplica para el análisis exploratorio de matrices de datos $X n \times p$ con $p \gg n$. Se suele denominar a estas matrices como matrices de datos *gordas*. En el ACPAD se determina el subespacio de baja dimensión que captura la mayor parte de la varianza mediante combinaciones lineales de las variables originales (Gray, 2017). Además, se calculan los eigenvectores de la matriz de covarianza de los datos, resumiendo de esta manera los datos con menos variables (Wainwright, 2019). Los casos se pueden graficar para evaluar visualmente similitudes y diferencias entre ellos, y así poder determinar grupos. Cada caso puede estar representado por relativamente pocos eigenvectores, en lugar de miles de variables.

En metabolómica se facilita la comprensión de las diferencias entre diversos grupos de muestras con base en la composición química de las muestras. Sin embargo, a pesar de los avances en las plataformas analíticas disponibles para el análisis de matrices gordas aún es un desafío, debido a la necesidad de herramientas para analizar datos en altas dimensiones con pocos casos que no realicen transformaciones a los datos que produzcan pérdida de información o variabilidad (Chong y Xia, 2020). Se han propuesto novedosas herramientas para procesar y analizar datos metabolómicos con mayor rapidez y facilidad (Vignoli et al. 2019); MetaboAnalyst es una de las alternativas más utilizada actualmente debido a su facilidad y capacidad para realizar análisis estadísticos metabolómicos robustos, visualizarlos e interpretarlos, además de ser gratuita.

En este trabajo, se presenta con rigor matemático el ACPAD y se muestra una aplicación consistente del análisis de la composición química de ocho tipos de suelo. Se ilustra cómo este análisis permite identificar y visualizar grupos de muestras que corresponden al tipo de suelo. De esta manera, se corrobora la viabilidad del ACPAD para el análisis de similitud química de suelos.

Metodología

Obtención de las muestras

El proyecto se desarrolló en el invernadero “Planta piloto de cultivo celular” localizado en el Edificio C del campus III de INECOL- Biomimic en Xalapa, Veracruz, donde se estableció un experimento bajo condiciones semicontroladas de temperatura (entre 11.45 y 26.8° C) y humedad relativa (entre 59.80 y 96.12%) durante 210 días para caracterizar el suelo y rizosfera de *Coffea arabica* y *C. canephora* en presencia y ausencia de la lombriz de tierra *Pontoscolex corethrurus*. Las plantas de café de la especie *C. arabica* provinieron de Agroindustrias Unidas de México (AMSA) y las plantas *C. canephora* se germinaron y crecieron en el invernadero del INECOL.

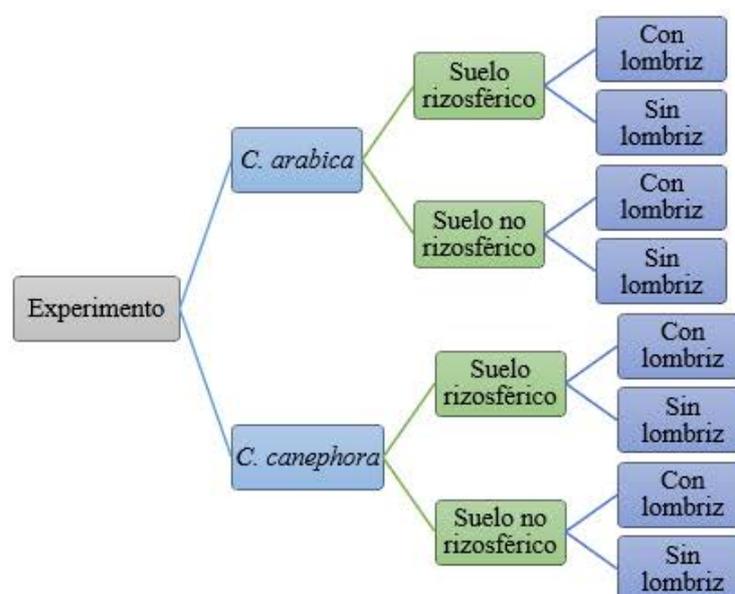


Figura 1. Diseño experimental 2^3 de *Coffea arabica* y *C. canephora* en suelo rizadosférico y no rizadosférico, con y sin lombriz.

Cada unidad experimental constó de una planta de seis meses de edad sembrada en macetas cilíndricas de 25 x 15 cm con suelo arcilloso pasteurizado y nueve perforaciones en el fondo del contenedor de 5 mm de diámetro. Las unidades experimentales con presencia de lombriz se inocularon con 10 individuos adultos con una biomasa total en promedio de 4.50 g de manera aleatoria. Además, se realizaron riegos de agua potable (600 ml) cada tercer día. Cada tratamiento contó con 10 réplicas biológicas, distribuidas completamente al azar en el invernadero. Las variables independientes consideradas fueron: especie de café (*C. arabica* y *C. canephora*), presencia de lombriz (presencia, ausencia), sitio de análisis (suelo rizosférico y no rizosférico, Figura 1) en el que se obtuvieron los componentes químicos del suelo según su relación masa carga y tiempo de retención.

Los análisis metabolómicos se llevaron a cabo en los Laboratorios de Química de Productos Naturales y Química Bioanalítica, de la Red de Estudios Moleculares Avanzados del Instituto de Ecología A.C. ubicados en el Edificio B del Campus 3 del INECOL, Clúster Científico y Tecnológico BioMimic®.

Procesamiento de las muestras

Para determinar si hay diferentes compuestos químicos en los suelos, se realizó el estudio metabolómico no dirigido o análisis contrastante de los perfiles metabólicos con un sistema de cromatografía de líquidos de ultra alta resolución (UPLC) Acquity Clase I (Waters™) acoplado a un analizador de espectrometría de masas de alta resolución de cuadrupolo y tiempo de vuelo (QTOF) modelo HDMI Synapt G2-Si (Waters™) utilizando como método de ionización una fuente de electrospray (ESI) en modo positivo (+) y negativo (-), fue a partir de lo anterior, que se obtuvo la base de datos metabolómicos.

ACPAD

Los datos espectrométricos obtenidos se procesaron a través de análisis comparativos discriminatorios con el ACPAD. Las variables en las muestras son los valores de las relaciones masa-carga (m/z) o masas exactas de los precursores, tiempos de retención (rt) y abundancia relativa de los diferentes perfiles químicos obtenidos en el programa MetaboAnalyst 4.0 y el paquete de R prcomp.

Matemáticamente, el PCA se expresa como el siguiente problema de eigenvalor:

$$\frac{1}{N-1} X'Xw = \lambda w \quad (1)$$

En donde:

X es una matriz de datos centrada en la media cuyas filas y columnas contienen datos de la muestra y valores de variables respectivamente,

N es el número de muestras.

λ y w denotan los autovalores y autovectores, respectivamente.

El producto Xw es la puntuación de CP (dada por $t = Xw$).

El coeficiente de correlación R entre t y la variable p está dada por:

$$R = \frac{\sqrt{\lambda_m} W_{m,p}}{\sigma_p}, \quad (2)$$

En donde:

- λ_m es el eigenvalor m ,
- σ_p es la desviación estándar de la variable p ,
- $W_{m,p}$ denota al componente m de la variable p .

Si cada variable se escala a tener varianza unitaria, es decir $\sigma_p = 1$, entonces el coeficiente de correlación R entre t y la variable p es proporcional a $W_{m,p}$.

$$R = \sqrt{\lambda_m} w_{m,p} \quad (3)$$

El vector w es el eigenvalor o el coeficiente de correlación R y siempre que cada variable tenga escala automática, el rango del vector propio w y el coeficiente de correlación R será consistente. Sin embargo, cuando la variable no es escalada automáticamente, el rango de w y R será diferente (Van den Berg et al., 2006).

Sea $S_+^{d \times d}$ el espacio de matrices positivas semidefinidas d -dimensionales, y denotemos a la esfera unitaria d -dimensional por $S^{d-1} = \{v \in R^d \mid \|v\|_2 = 1\}$. Consideremos a un vector aleatorio X de dimensión d , con vector de medias cero y matriz de covarianza $\Sigma \in S_+^{d \times d}$ denotamos con

$$\gamma_1(\Sigma) \geq \gamma_2(\Sigma) \geq \dots \geq \gamma_d(\Sigma) \geq 0 \quad (4)$$

A los eigenvalores ordenados de la matriz de covarianza Σ . El análisis de componentes principales busca la dirección $v \in S^{d-1}$ tal que la varianza de la variable aleatoria $\langle v, X \rangle$ sea máxima. Esta dirección se conoce como el primer componente principal. En términos analíticos tenemos

$$v^* = \arg \max_{v \in S^{d-1}} \text{var}(\langle v, X \rangle) = \arg \max_{v \in S^{d-1}} E[\langle v, X^2 \rangle] = \arg \max_{v \in S^{d-1}} \langle v, \Sigma v \rangle, \quad (5)$$

Por definición, el primer componente principal es el eigenvector de la matriz de covarianza Σ . De manera análoga, los r componentes principales se obtienen de la matriz ortogonal $V \in R^{d \times r}$, formada con las columnas ortogonales $\{v_1, \dots, v_r\}$, que maximizan la cantidad. Estos componentes principales son simplemente los r eigenvectores más importantes de la matriz de covarianza Σ .

$$E\|V^T X\|_2^2 = \sum_{j=1}^r E[\langle v_j, X \rangle^2] \quad (6)$$

Cuando no se conoce la matriz de covarianza, sino sólo una muestra finita $\{x_i\}_{i=1}^n$, la matriz de covarianzas se estima con $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$, haciendo el supuesto de que los datos tienen media 0. La versión empírica del ACPAD se basa en el remplazamiento de la covarianza desconocida Σ con esta versión empírica $\hat{\Sigma}$. Por ejemplo, el análogo empírico del primer componente principal (5) está dado por

$$\hat{v} = \arg \max_{v \in S^{d-1}} \langle v, \hat{\Sigma} v \rangle \quad (7)$$

El ACPAD es ideal para resumir las matrices en altas dimensiones en una cantidad menor de componentes. Sin embargo, cuando la dimensión es grande, $d \gg n$, no se pueden obtener más de n componentes principales porque el rango r de la matriz de covarianza S satisface $r \leq \min\{d, n\}$, (Koch, 2013). Desde un punto de vista estadístico se necesita saber en qué sentido el ACPAD proporciona una solución adecuada al problema de reducción de dimensionalidad.

Resultados y discusión

Del análisis metabolómico, se obtuvo una matriz gorda $A \in R^{n \times m}$ con $n \geq m$, con los valores individuales ordenados como:

$$\sigma_{\max}(A) = \sigma_1(A) \geq \sigma_2(A) \geq \dots \geq \sigma_m(A) = \sigma_{\min}(A) \geq 0$$

Se observaron 3,040 compuestos químicos (los cuales son las variables) y para cada variable se observaron 40 abundancias relativas de cada uno de los compuestos detectado por el analizador de espectrometría de masas de alta resolución de cuadrupolo y tiempo de vuelo (QTOF) para el suelo rizosférico y no rizosférico de *C. arabica* y *C. canephora*, tanto en presencia como en ausencia de *P. corethrurus*.

Se encontró que el tratamiento con mayor cantidad de variables o compuestos químicos con ESI+ fue el suelo no rizosférico de *C. canephora* con lombriz. Por otro lado, el tratamiento con menor cantidad de compuestos químicos fue el suelo de *C. arabica* no rizosférico sin lombriz (Figura 2).

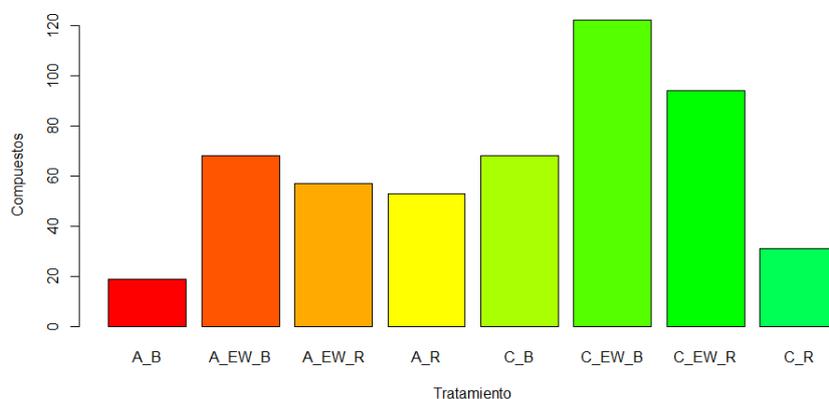


Figura 2. Número de compuestos químicos del suelo no rizosférico de *C. arabica* sin lombriz (A_B), suelo no rizosférico de *C. arabica* con lombriz (A_EW_B), suelo rizosférico de *C. arabica* con lombriz (A_EW_R), suelo rizosférico de *C. arabica* sin lombriz (A_R), suelo no rizosférico de *C. canephora* sin lombriz (C_B), suelo no rizosférico de *C. canephora* con lombriz (C_EW_B), suelo rizosférico de *C. canephora* con lombriz (C_EW_R) y suelo rizosférico de *C. canephora* sin lombriz (C_R) con ESI+.

Por otro lado, el tratamiento con mayor cantidad de compuestos químicos con ESI- fue el suelo rizosférico de *C. arabica* sin lombriz. Mientras que el tratamiento con menor cantidad de compuestos químicos fue el suelo no rizosférico de *C. arabica* con lombriz (Figura 3).

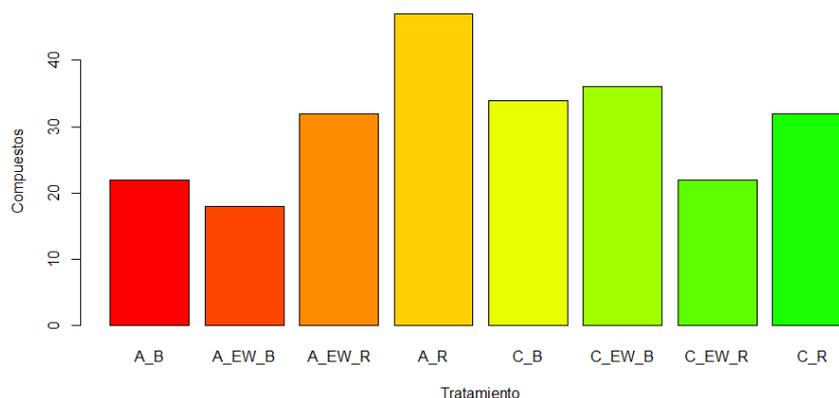


Figura 3. Número de compuestos químicos del suelo no rizosférico de *C. arabica* sin lombriz (A_B), suelo no rizosférico de *C. arabica* con lombriz (A_EW_B), suelo rizosférico de *C. arabica* con lombriz (A_EW_R), suelo rizosférico de *C. arabica* sin lombriz (A_R), suelo no rizosférico de *C. canephora* sin lombriz (C_B), suelo no rizosférico de *C. canephora* con lombriz (C_EW_B), suelo rizosférico de *C. canephora* con lombriz (C_EW_R) y suelo rizosférico de *C. canephora* sin lombriz (C_R) con ESI -.

Aunque no se realizaron procedimientos de prueba de hipótesis, los resultados dan pistas de que existe una mayor diversidad química en el suelo no rizosférico de *C. canephora* con lombriz en ESI+ y en el suelo rizosférico de *C. arabica* sin lombriz en ESI-. Por el contrario, existe una menor variabilidad química en el suelo no rizosférico de *C. arabica* tanto en ESI+ como ESI- con y sin lombriz, lo cual se explica debido a que los marcadores moleculares han demostrado que la diversidad genética de *C. arabica* es menor que la de *C. canephora* (Cubry et al., 2008; Spinoso-Castillo et al., 2020), la cual podría promover la producción de químicos tóxicos contra los herbívoros, explicándose de esta manera la resistencia de *C. canephora* al ataque de insectos en comparación con *C. arabica* (Granados-Sánchez, Ruíz-Puga, y Barrera-Escorcia, 2008).

Al procesar la matriz de datos gorda con Metabo Analyst, el ACPAD encontró que tanto para ESI+ (a) como ESI-(b) los dos primeros componentes explican más del 60% de la varianza (Figura 4).

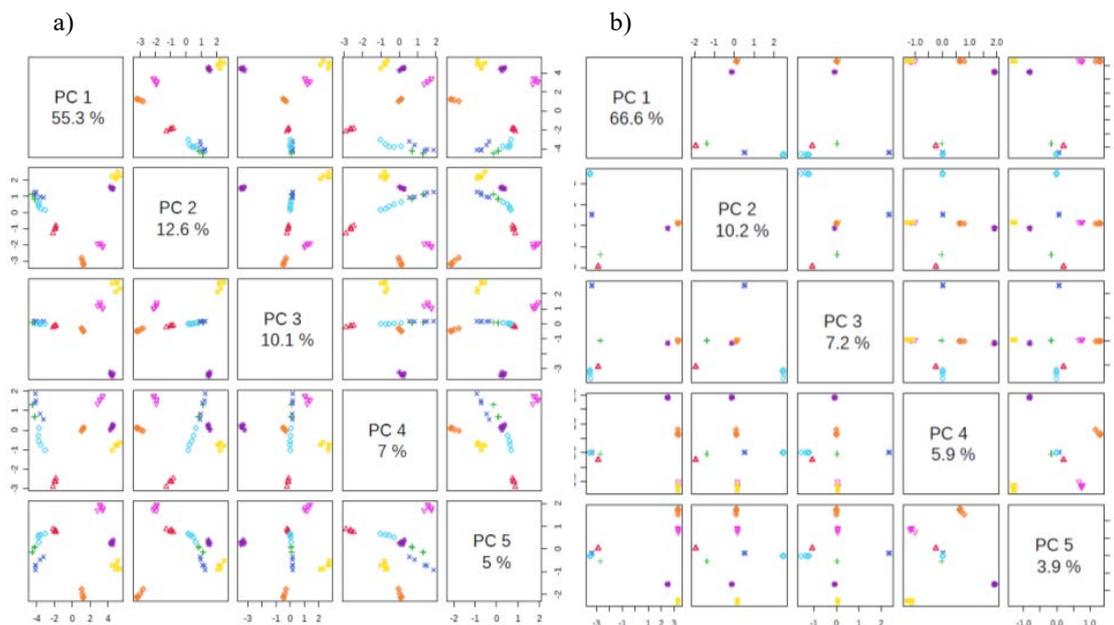


Figura 4. Gráficos de puntuaciones por pares entre las CP seleccionadas. La varianza explicada de cada CP se muestra en la celda diagonal correspondiente.

Complementando lo anterior, con el análisis se puede observar que la mayoría de los grupos están separados entre ellos (Figura 5 inciso a). Esto facilitó la visualización preliminar de diferencias entre ellos. Es decir, a través del ACPAD, pudo observarse que la composición química de los suelos varía entre sí en función de su cercanía a las raíces, con lo que se valida las hipótesis de Lavelle et al. (2016) y Pétriacq et al. (2017) quienes exponen que tanto la especie de planta como la cercanía con la raíz tienen efectos en la composición química de los suelos. Tal es el caso del suelo de *C. canephora* cuya rizósfera se diferencia del resto de suelo aun cuando están en contacto con la misma planta.

Otros grupos que se diferenciaron claramente a través del análisis metabólico por ESI+, véase la Figura 5 inciso a), fueron los suelos de *C. canephora* sin lombrices, tanto el rizosférico como el no rizosférico. Estos se mostraron separados entre sí, lo que implica que esta zona, la rizósfera, es un espacio en el que los componentes químicos difieren entre sí debido a la actividad biológica alrededor de la raíz (Kuzuyakov y Blagodatskaya, 2015). Esto no implica que las sustancias sean de alguna naturaleza en particular, más bien que hay una transformación química por efecto de la cercanía con la raíz, observable aun cuando no se encuentran las lombrices en los suelos de *C. canephora*.

El mencionado impacto radicular no fue tan evidente en *C. arabica*, pues las matrices de suelo con lombrices y el suelo rizosférico se encontraron muy cerca unas de otras (Figura 5 inciso a). Es destacable el hecho de que, por influencia de las lombrices, la composición química de un suelo no rizosférico puede llegar a semejarse a la del suelo rizosférico. Esto podría explicar los llamados efectos positivos observados en el crecimiento vegetal en presencia de lombrices mencionados en Van Groenigen et al. (2014) y Blouin et al. (2019).

De hecho, el único grupo que se separa es el correspondiente a *C. arabica* no rizosférico sin lombrices. Estos datos tienen la menor actividad biológica esperada por encontrarse lejos de la

zona radicular de la planta y no contar con lombrices para promover la actividad microbiana (Lavelle, Lattaud, Trigo y Barois, 1995; Puga-Freitas, Barot, Taconnat, Renou y Blouin, 2012; Blouin, 2018). Así, el efecto de las lombrices podría ser tan relevante como el de las plantas a través de sus raíces tal y como antes han sugerido otros autores (Domínguez, Aira y Gómez-Brandón, 2010; Van Groeningen, 2014).

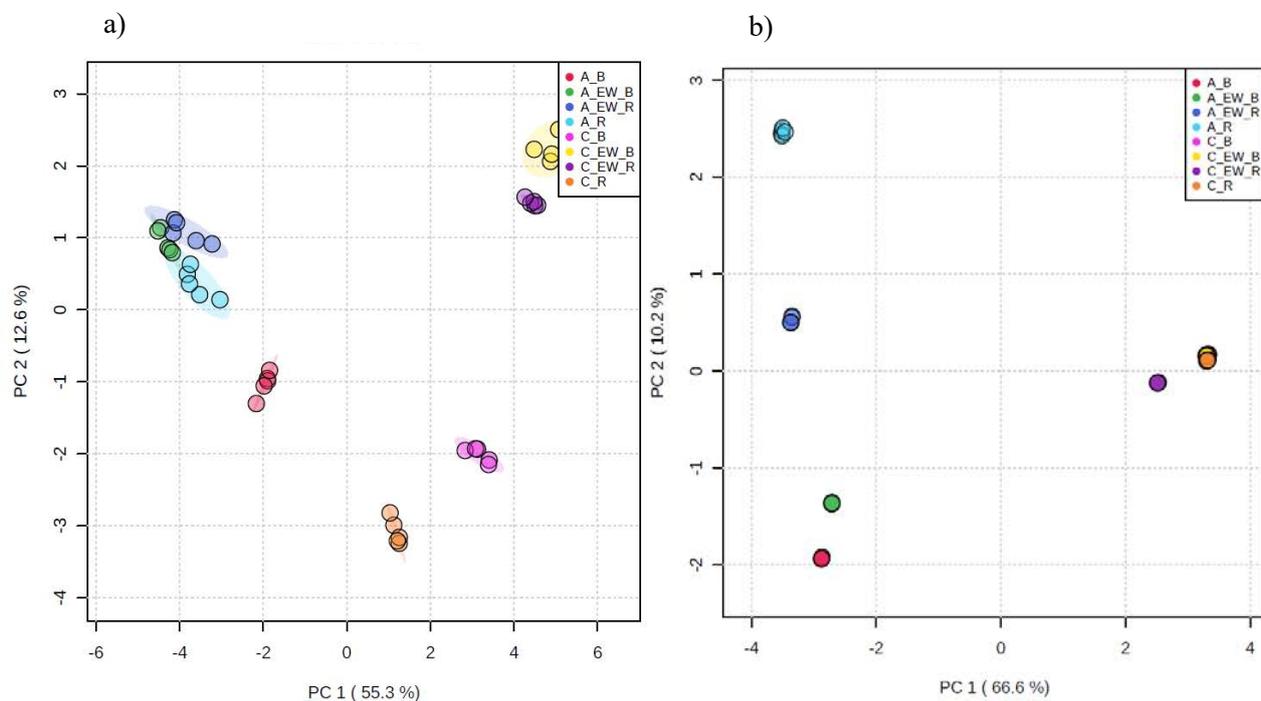


Figura 5. ACPAD del del suelo no rizosférico de *C. arabica* sin lombriz (A_B), suelo no rizosférico de *C. arabica* con lombriz (A_EW_B), suelo rizosférico de *C. arabica* con lombriz (A_EW_R), suelo rizosférico de *C. arabica* sin lombriz (A_R), suelo no rizosférico de *C. canephora* sin lombriz (C_B), suelo no rizosférico de *C. canephora* con lombriz (C_EW_B), suelo rizosférico de *C. canephora* con lombriz (C_EW_R) y suelo rizosférico de *C. canephora* sin lombriz (C_R). En la parte superior izquierda (a) se encuentran los resultados para el ESI + y en la parte superior derecha (b) del ESI -. Gráfica de puntuaciones de los dos primeros componentes principales. Las varianzas explicadas se muestran entre paréntesis con $n = 40$ por 3040 (picos / (mz / rt)), con cinco replica para cada tratamiento ($n = 5$). Las áreas sombreadas indican regiones de color de elipse de confianza del 95% según los puntos de datos para grupos individuales. Los círculos punteados indican la agrupación por factores en común y una buena separación entre el suelo rizosférico y no rizosférico de *C. arabica* y *C. canephora* con y sin *P. corethrurus*.

Por otro lado, los grupos expuestos a *C. canephora* con ESI- no se lograron visualizar separadamente (Figura 5 inciso b) debido a que son químicamente semejantes unos a otros. Nótese que estos grupos tienen una composición química parecida y fueron influenciados mayormente por la planta a través de su sistema radicular. Lo anterior se debe a que todos los grupos, con una excepción, se encuentran en el mismo espacio del ACPAD. Sólo el grupo de suelo rizosférico en presencia de lombriz se separa visiblemente debido a su composición. Esto se explica por

encontrarse en la región de mayor actividad biológica, la rizósfera, en presencia de lombrices de tierra como detonantes de actividad microbiana y de transformación química (Lavelle, 2002; Barret, Morrissey y O'gara, 2011).

Conclusiones

El ACPAD fue una herramienta que permitió identificar geoméricamente las semejanzas y diferencias de la composición química de las complejas matrices resultantes de análisis químico-biológico de suelos y su naturaleza molecular. Esto permite conjeturar sobre la relevancia de una variable independiente en la variabilidad química del suelo. Para nuestro ejemplo específico, a través del ACPAD, se identificó que tanto la especie de planta (*C. arabica* y *C. canephora*) como la presencia de lombriz (*P. corethrurus*) y la proximidad del suelo a la raíz, tienen una influencia en la naturaleza química de los suelos. Entre los factores, se sugiere que *C. arabica* y *C. canephora* tienen un mayor impacto en el tipo de compuestos químicos que se encuentran en el suelo, debido a que se agrupó un mayor número de muestras por especie de planta.

Agradecimientos

Marycruz Álvarez-Jiménez agradece a la Especialización en Métodos Estadísticos (EME) de la Facultad de Estadística e Informática de la Universidad Veracruzana por el respaldo recibido para realizar esta investigación. También, agradece al CONACyT por la beca otorgada durante el segundo semestre de la EME. Un especial agradecimiento al INECOL, por los espacios brindados para la realización de la parte experimental y los análisis químicos; Además al financiamiento recibido por parte del proyecto de ciencia básica CONACyT No. 257819 titulado “Diversidad bacteriana del suelo de la rizosfera del café antes, durante y después de su paso por el intestino de la lombriz de tierra *Pontoscolex corethrurus*”. Así como al Dr. José Antonio Guerrero Analco, al Dr. Juan Luis Monribot y a la Mtra. Laura Stefany Licona del laboratorio de Química de Productos Naturales por su valioso apoyo durante todo el trabajo de laboratorio y análisis químicos. Finalmente, al Mtro. Eduardo Hernández Hernández por sus valiosos aportes y correcciones al presente documento.

Referencias

- Barret, M., Morrissey, J. P. y O'gara, F. (2011). Functional genomics analysis of plant growth-promoting rhizobacterial traits involved in rhizosphere competence. *Biology fertility of soils*, 47 (7), 729-743.
- Blouin, M. (2018). Chemical communication: An evidence for co-evolution between plants and soil organisms. *Applied soil ecology*, 123, 409-415.
- Blouin, M., Barrere, J., Meyer N., Lartigue, S., Barot, S. y Mathieu, J. (2019). Vermicompost significantly affects plant growth. A meta-analysis. *Agronomy for Sustainable Development*, 39 (34), 1-15.
- Chong, J., y Xia, J. (2020). Using MetaboAnalyst 4.0 for metabolomics data analysis, interpretation, and integration with other omics data. In *Computational Methods and Data Analysis for Metabolomics* (pp. 337-360). Humana, New York, NY.
- Cubry, P., Musoli, P., Legnate, H., Pot, D., De Bellis, F., Poncet, V., Leroy, T. (2008). Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome*, 51 (1), 50-63.

- Domínguez, J., Aira, M. y Gómez-Brandón, M. (2010). Vermicomposting: earthworms enhance the work of microbes. En *Microbes at work*, 93-114, Berlin, Heidelberg: Springer.
- Granados-Sánchez, D., Ruíz-Puga, P. y Barrera-Escorcia, H. (2008). Ecología de la herbivoría. *Revista Chapingo. Serie ciencias forestales y del ambiente*, 14 (1), 51-63.
- Gray, V. (Ed.). (2017). *Principal component analysis: methods, applications and technology*. Nova Science Publishers, Incorporated.
- Koch, I. (2013). *Analysis of multivariate and high-dimensional data* (Vol. 32). Cambridge University Press.
- Kuzyakov, Y. y Blagodatskaya, E. (2015). Microbial hotspots and hot moments in soil: concept & review. *Soil Biology & Biochemistry*, 83, 184-199.
- Lavelle, P. (2002). Functional domains in soils. *Ecological research*, 17 (4), 441-450.
- Lavelle, P., Lattaud, C., Trigo, D. y Barois, I. (1995). Mutualism and biodiversity in soils. In *The significance and regulation of soil biodiversity* (pp. 23-33): Springer, Dordrecht.
- Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jiménez, J., McKey, D., Mathieu, J., Velásquez, E. y Zangerlé, A. (2016). Ecosystem engineers in a self-organized soil: a review of concepts and future research questions. *Soil Science*, 181(3/4), 91-109.
- Pétriacq, P., Williams, A., Cotton, A., McFarlane, A. E., Rolfe, S. A. y Ton, J. (2017). Metabolite profiling of non-sterile rhizosphere soil. *The Plant Journal*, 92 (1), 147-162.
- Puga-Freitas, R., Barot, S., Taconnat, L., Renou, J. P. y Blouin, M. (2012). Signal molecules mediate the impact of the earthworm *Aporrectodea caliginosa* on growth, development and defence of the plant *Arabidopsis thaliana*. *PLoS One*, 7 (12), e49504.
- Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology*, 26(3), 303-304.
- Spinoso-Castillo, J. L., Escamilla-Prado, E., Aguilar-Rincón, V. H., Ramos, V. M., de los Santos, G. G., Pérez-Rodríguez, P., y Corona-Torres, T. (2020). Genetic diversity of coffee (*Coffea* spp.) in Mexico evaluated by using DArTseq and SNP markers. *Genetic Resources and Crop Evolution*, 67(7), 1795-1806.
- Swenson, T. L., Jenkins, S., Bowen, B. P. y Northen, T. R. (2015). Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biology and Biochemistry*, 80, 189-198.
- Van den Berg, R. a., Hoefsloot, H. C. J., Westerhuis, J. a., Smilde, A. K. y van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7, 142.
- Van Groenigen, J. W., Lubbers, I. M., Vos, H. M., Brown, G. G., De Deyn, G. B. y Van Groenigen, K. J. (2014). Earthworms increase plant production: a meta-analysis. *Scientific reports*, 4 (1), 1-7.
- Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., Turano, P. y Luchinat, C. (2019). High-throughput metabolomics by 1D NMR. *Angewandte Chemie International Edition*, 58(4), 968-994.
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint* (Vol. 48). Cambridge University Press.

Validación de modelos de calidad de hábitat para el Chigüiro (*Hydrochoerus hydrochaeris*) en la Orinoquía Colombiana

Cyndy Nohemy Cardona Claros¹, Sergio Francisco Juárez Cerrillo²

¹ Facultad de Estadística e Informática, Universidad Veracruzana. cncardonac@gmail.com

² Facultad de Estadística e Informática, Universidad Veracruzana, sejuarez@uv.mx

Resumen

En la Orinoquía colombiana se ha reportado una disminución de la calidad del hábitat para el chigüiro (*Hydrochoerus hydrochaeris*) y se desconoce la relación del índice de idoneidad de hábitat (HSI) con las poblaciones de chigüiros. Para llenar este vacío, se conformaron cuatro bases de datos con las variables de hábitat a nanoescala y mesoescala, para las temporadas de lluvia y sequía, de 2015 y 2016, de la localidad de Miramar, Departamento del Casanare en Colombia. La variable respuesta es la presencia/ausencia del chigüiro y el número de individuos. Las dos bases de datos a nanoescala tienen 28 puntos de muestreo en los que se midieron variables geográficas y de vegetación, y cinco variables de hábitat incluida la calificación de HSI. Las bases de datos a mesoescala, tienen cada una, 239 registros correspondientes a las celdas de Unidades de Evaluación de Hábitat y ocho variables de hábitat. Con regresiones logísticas y regresiones Poisson inflado con cero, se evaluó el rendimiento del HSI y de las variables del hábitat. Se encontró que las variables del hábitat que determinan la presencia de los chigüiros a nanoescala son la cobertura de escape y la riqueza de alimento potencial, mientras que a mesoescala son la cobertura climática y la cobertura de escape principalmente, seguido de la riqueza de alimento potencial. Además, en ambas escalas en los dos períodos climáticos, la variable de calidad del modelo HSI determinó la presencia de la especie. En cuanto a las abundancias de chigüiros, a nanoescala la variable que mejor representa el número de individuos son la cobertura de escape y la climática en lluvias de 2015, y la riqueza de alimento potencial en sequía de 2016. Mientras que a mesoescala, la cobertura de escape es la variable en común que mejor representó el número de individuos en ambos períodos, específicamente en lluvias de 2015, el índice de dominancia del paisaje y la cobertura climática representan el número de individuos, y en sequía de 2016 son las siguientes variables: forraje de alimento potencial y el índice de yuxtaposición. La capacidad predictiva de los modelos ajustados se evaluó con indicadores de sensibilidad y probabilidades estimadas de presencia. Los resultados obtenidos permiten llevar a cabo acciones de manejo y aprovechamiento de la especie, así como de gestión del hábitat con confiabilidad y validez estadística.

Palabras Clave: Índice de idoneidad del hábitat, regresión logística, regresión Poisson inflado con cero.

Introducción

El chigüiro *Hydrochoerus hydrochaeris* es el roedor actual más grande del mundo y se distribuye en la zona tropical de América del Sur, desde Panamá hasta el norte de Argentina (Mones y Ojasti, 1973; Ojasti, 1990). Presenta una dieta herbívora con alta eficiencia digestiva (Forero et al., 2003; Corriale, 2010), compuesta principalmente de gramíneas, ciperáceas y plantas acuáticas (Ojasti, 1997; Forero, 1999). Requiere de asociaciones de interfase tierra-agua (Quintana, 1996), de manera que utiliza varios hábitats de forma diferencial (Ojasti, 1973; Herrera y Macdonalds, 1989; Soini y

Soini, 1992; Clavijo, 1993). Las coberturas vegetales que utiliza para el forraje y el refugio, tales como sabanas, pajonales, bosques de galería (Ojasti, 1973; Quintana y Rabinovich, 1993), suelen estar cercanas a cuerpos de agua. La disponibilidad de agua limita el movimiento y la dispersión de los chigüiros, ya que ésta incide en la termorregulación, la protección de depredadores (MacDonald, 1981; Ojasti, 1993) y, principalmente, en la reproducción de la especie para la cópula y la crianza.

Esta especie ha sido objeto de varias investigaciones en Sudamérica, en Colombia, particularmente en las sabanas inundables de la Orinoquía se desarrolló un índice de calidad del hábitat para el chigüiro en esta zona, el cual fue propuesto y desarrollado por Guzmán-Lenis et al. (2014a, 2014b) y fue nuevamente empleado por Cardona-Claros (2018) para evaluar los cambios espaciotemporales del hábitat de la especie. La evaluación de la calidad del hábitat permite identificar la capacidad de un hábitat para soportar una especie en particular, uno de los modelos más empleados para hacerlo es a través del índice de idoneidad del hábitat (HSI, Habitat Suitability Index, por sus siglas en inglés) (U.S. Fish and Wildlife Service, 1981). Este índice es una herramienta útil para representar la relación entre fauna silvestre y su hábitat, así como para orientar acciones de manejo (Litvaitis et al., 2004; Morrison et al., 2006; Yarrow, 2009).

En las investigaciones mencionadas en la Orinoquía colombiana, los modelos de HSI no fueron validados porque provenían de datos de campo. Cardona-Claros (2018) determinó que la calidad del hábitat disminuyó en casi una década, predominando áreas con calidades medias, y señala que éstas no estuvieron relacionadas con los datos poblacionales. Al respecto, son pocos los HSI que han sido verificados o validados mediante el uso de datos independientes. De manera que, Reiley et al. (2014), señalan que los HSI no validados pueden ser indicadores poco confiables de la calidad del hábitat. Por ello, el modelo de calidad de hábitat requiere de una validación estadística sólida que pruebe la precisión del modelo.

El desarrollo de métodos estadísticos para la validación es la tarea más difícil en el proceso de modelado, ya que no existe un estándar de oro contra el cual se puedan medir las predicciones (Fielding, 2002). La fase de validación es crucial para evaluar la precisión de las predicciones (Ottaviani et al., 2004). Para verificar la robustez de las predicciones, éstas deben compararse con un conjunto de datos independientes, es decir, uno que no se haya utilizado previamente en la construcción del modelo. El proceso de validación es necesario para determinar la confiabilidad y la utilidad de los modelos en cuanto a las medidas de conservación y el manejo de las especies (Brooks, 1997; Reiley et al., 2014). De manera que la validación implica probar el rendimiento de un modelo de idoneidad del hábitat contra datos cuantitativos independientes en espacio y tiempo, como es el caso de los datos poblacionales. Para ello, se utilizaron Modelos Lineales Generalizados (GLM), debido a la naturaleza de los datos, a los niveles de certidumbre, a la reproducibilidad y a las ventajas en el manejo de las variables (Jácome Flores, 2010).

Por lo tanto, en este estudio se validaron y compararon los modelos de HSI del chigüiro (*Hydrochoerus hydrochaeris*) a nanoescala y mesoescala en las sabanas inundables de la Orinoquía colombiana a partir de la validación estadística de las variables del hábitat y los datos poblacionales de la especie en dos períodos climáticos contrastantes.

Materiales y métodos

Se utilizó la base de datos de las variables de hábitat para el chigüiro proveniente de la tesis de maestría de Cardona-Claros (2018). Se empleó la base de datos que contiene información de las variables con las que se construyó el modelo de calidad de hábitat para la localidad de Miramar, en las sabanas inundables del Departamento de Casanare, en Colombia, en dos períodos climáticos (lluvia de 2015 y sequía de 2016) a dos escalas:

- 1) Nanoescala, correspondiente a 28 puntos de muestreo.
- 2) Mesoescala, correspondiente a las celdas de 16 hectáreas (400 x 400 m) para la totalidad del área de cada localidad. Se evaluaron un total de 478 filas y 14 variables en época de lluvias de 2015 (n=239) y en época de sequía de 2016 (n=239).

Se obtuvo la base de datos poblacionales y, a través del programa de QGIS, se realizaron las espacializaciones de los chigüiros registrados en la tesis de maestría de Tiboche-García (2016) en las celdas mencionadas anteriormente para lluvias de 2015 y sequía de 2016. A partir de lo anteriormente descrito, se construyó la base de datos de hábitat y poblaciones (presencias-ausencias y abundancias).

Validación y comparación de modelos HSI

Se realizaron modelos de regresión logística y modelos de regresión *Poisson inflado con cero* (PIC) con el programa R versión 4.0.5, se emplearon las paqueterías “gam”, “pROC”, “InformationValue”, “vcd”, “pscl” y “COUNT”. Las variables respuestas son la presencia del chigüiro y el número de individuos, respectivamente.

En los modelos de regresión logística se evaluaron los modelos de HSI (índice de idoneidad del hábitat) y las variables del hábitat de los chigüiros. En el segundo caso, la variable HSI no se incluyó dentro de la regresión múltiple, porque este índice contiene las variables del hábitat, además, en el análisis se seleccionaron las variables con el AIC (Criterio de información de Akaike) más bajo y estadísticamente significativas con p-valor menor a 0.05.

La capacidad predictiva de los modelos ajustados de la regresión logística se evaluó con indicadores de sensibilidad y probabilidades estimadas de presencia de chigüiros. En la evaluación del modelo se consideraron las siguientes medidas de ajuste: Devianza, la prueba de razón de verosimilitud y el criterio de información de Akaike (AIC). La devianza es una medida de los residuos generados por el modelo propuesto comparado con el modelo nulo (sin predictores). En general, si disminuye, el modelo es mejor., tiende a cero cuando el modelo predice bien. La prueba de razón de verosimilitud calcula la significancia de la diferencia de residuos entre el modelo de interés y el modelo nulo (en este caso con el modelo en blanco). El estadístico sigue una distribución chi-cuadrado con grados de libertad equivalentes a la diferencia de grados de libertad de los dos modelos. El AIC evalúa la calidad del modelo a través de la comparación con otros modelos penalizando por la inclusión de predictores, de manera que, a menor AIC mejor es el modelo.

Para determinar que la probabilidad estimada de que las predicciones y los resultados sean correspondientes se empleó el Índice de concordancia (ROC). Un ROC de 1 representa un modelo

que predice perfectamente la presencia de la especie mientras que un índice de concordancia de 0,5 es el equivalente a adivinar al azar. Para evaluar el rendimiento o la capacidad de clasificación del modelo se emplearon las medidas del área bajo la curva ROC (AUC) y el p-valor de la matriz de confusión de McNemar, en otras palabras, éstas pruebas indican qué tan bien funciona el modelo. Así que, se considera no discriminativo cuando el AUC= 0.5 y cuando se acerca a 1 aumenta la capacidad discriminativa, es decir que, el modelo predice con mucha exactitud la respuesta observada. Se tuvieron en cuenta los rangos de discriminación propuestos por Hosmer y Lemeshow (2000).

Los datos presentan exceso de ceros, por lo que usamos modelos de regresión Poisson inflado con cero.

Resultados

Modelo de Regresión logística

En la Figura 14 se muestran las probabilidades de la presencia de chigüiros a nanoescala y mesoescala en los dos periodos climáticos. A nanoescala, en lluvias de 2015 y en sequía de 2016, el valor de probabilidad de la presencia de chigüiros disminuye en la medida que los valores de HSI son mayores, lo mismo ocurre a mesoescala en lluvia de 2015. Mientras que, en sequía de 2016, el valor de probabilidad de registrar chigüiros aumenta al mejorar la calidad del hábitat.

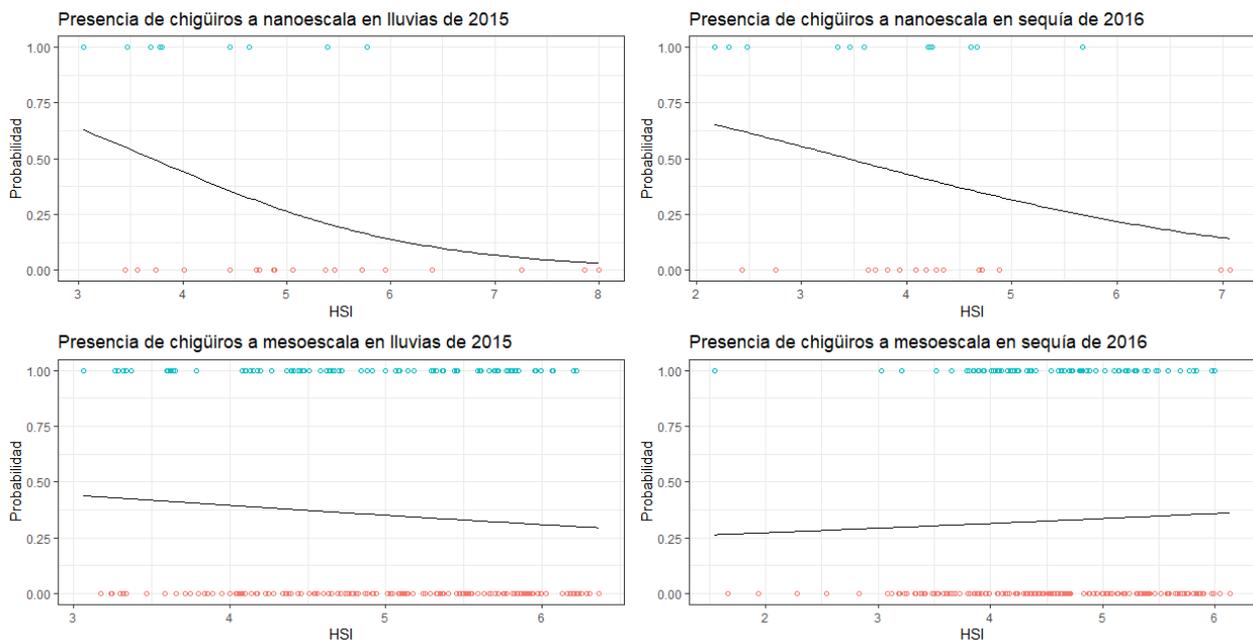


Figura 14. Probabilidad de la presencia de chigüiros según el HSI a nanoescala y mesoescala en dos periodos climáticos.

En la prueba de razón de verosimilitud del efecto de las variables comparando modelos, sólo a nanoescala, en lluvias de 2015, el modelo con HSI es significativo (

Anexo I). Por lo tanto, el predictor HSI ofrece un mejor ajuste a los datos. A nanoescala en sequía de 2016, y a mesoescala en lluvias 2015 y sequía 2016, la variable HSI no ofrece un mejor ajuste a los datos (

Anexo I). Por otra parte, los valores más pequeños del índice de razón de verosimilitud y de devianza muestran que el modelo con la variable HSI a mesoescala, se ajusta mejor a los datos de presencia-ausencia de los chigüiros (

Anexo I).

Las variables de hábitat a nanoescala que presentan el menor AIC y la mayor significancia estadística en lluvia de 2015, son la Riqueza de alimento potencial, la Cobertura de escape y el HSI (Tabla 8). El modelo multivariante con el AIC más bajo y con una mayor significancia estadística es el que está conformado por la Riqueza de alimento potencial y la Cobertura de escape (Mod4), y ofrece un mejor ajuste a los datos de la presencia de los chigüiros (

Anexo I). Aunque en sequía de 2016 ninguna variable contribuye de manera significativa a la presencia de la especie, el modelo 4 presenta los valores más bajos de devianza y de AIC (

Anexo I).

A mesoescala, las variables con menor AIC y las de mayor significancia estadística, en lluvias de 2015 son las variables de refugio (Cobertura climática y Cobertura de escape) y la Riqueza de alimento potencial (Tabla 8), y en sequía de 2016, son las variables de refugio y el Agua (Tabla 8). Los modelos multivariantes con las mejores medidas de ajuste en lluvias son: el modelo 5 que contiene las variables de refugio, y el modelo 6 que contiene la Riqueza de alimento potencial y la Cobertura climática (Anexo I). Mientras que, en sequía, es el modelo conformado por la Cobertura de escape y la disponibilidad de agua (Mod5) y el modelo con estas dos variables y la Cobertura climática (Mod7) (Anexo I).

Tabla 8. Variables asociadas a la presencia de los chigüiros a nanoescala y mesoescala en dos períodos climáticos.

Variables	Lluvia 2015				Sequía 2016				
	Df Deviance	AIC	LRT P	Pr(>Chi)	Df Deviance	AIC	LRT P	Pr(>Chi)	
Nanoescala									
Riqueza de alimento	28.82	32.82	6.3451	0.01177 *	38.235	42.235	0.00768	0.9302	
Forraje de alimento	31.553	35.553	3.6118	0.05737 .	38.239	42.239	0.00338	0.9536	
Cobertura de escape	29.442	33.442	5.7223	0.01675 *	37.265	41.265	0.97743	0.3228	
Cobertura climática	33.27	37.27	1.8951	0.16862	37.492	41.492	0.75094	0.3862	
HSI	30.997	34.997	4.168	0.04119 *	36.331	40.331	1.91162	0.1668	
Mesoescala									
Riqueza de alimento	305.23	309.23	5.891	0.0152183 *	298.2	302.2	3.6922	0.0546672 .	
Forraje de alimento	308.84	312.84	2.2738	0.1315807	298.24	302.24	3.652	0.0560004 .	
Cobertura de escape	299.67	303.67	11.45	0.0007149 ***	288.66	292.66	13.2278	0.0002758 ***	

Cobertura climática	294.48	298.48	16.6369	4.526e-05 ***	296.77	300.77	5.1175	0.0236863 *
Agua	311.06	315.06	0.0619	0.8035628	297.81	301.81	4.0841	0.0432885 *
Índice de yuxtaposición (IJ)	311.1	315.1	0.0145	0.9040358	301.52	305.52	0.3685	0.5437998
Índice de dominancia (IDO)	308.99	312.99	2.1234	0.1450618	299.36	303.36	2.5316	0.1115902
Longitud Bordes	310.87	314.87	0.2482	0.6183756	299.07	303.07	2.8221	0.0929743 .
HSI	309.77	313.77	1.3496	0.2453501	301.52	305.52	0.3699	0.5430805

Probabilidad de las predicciones del HSI

En la Figura 15 se muestran las probabilidades de predecir la presencia de chigüiros. A nanoescala, en los dos períodos climáticos y a mesoescala en lluvia de 2016, los valores más altos de probabilidad están asociados a valores bajos de HSI, de manera que, a mayores pérdidas de HSI habrá una mayor probabilidad de que la variable dependiente tome valor de uno. En otros términos, el valor de probabilidad disminuye en la medida que los valores de HSI son mayores. En sequía de 2016, se observa que el valor de probabilidad aumenta en la medida que aumenta el HSI.

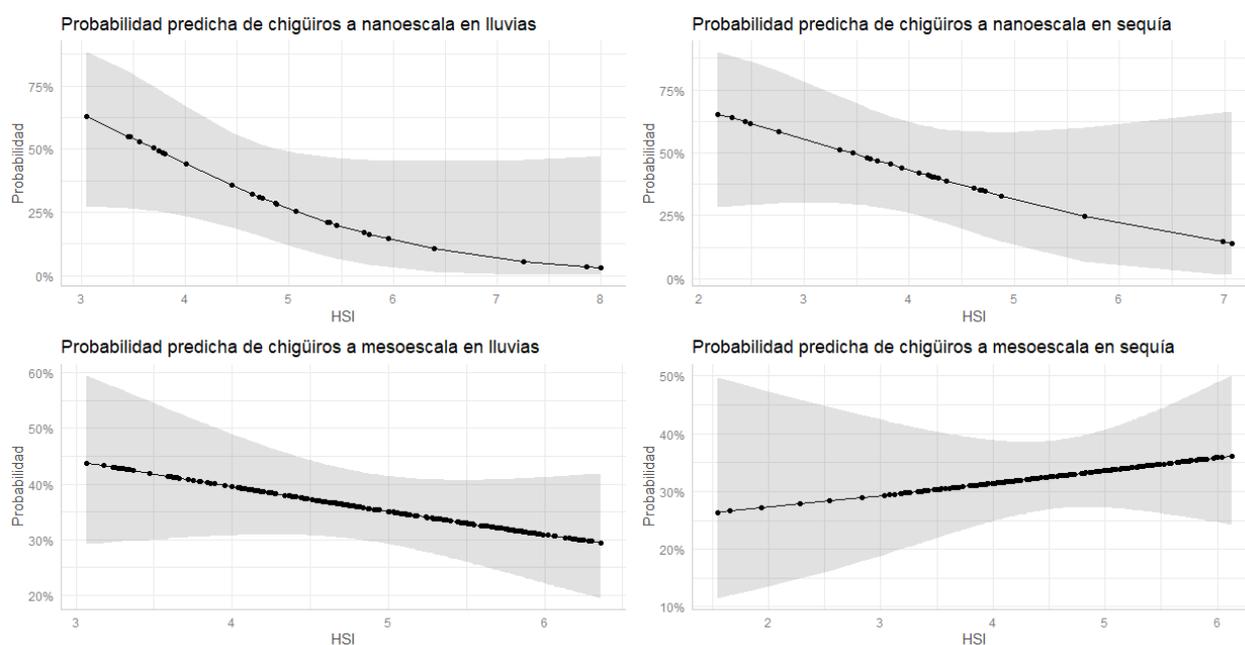


Figura 15. Probabilidad de predicción de la presencia de chigüiros según el HSI a nanoescala y mesoescala en dos períodos climáticos.

Calidad predictiva del modelo de HSI

En la Figura 16 se observan las curvas ROC para el modelo de HSI, en los dos períodos climáticos, a las dos escalas de evaluación. El índice de concordancia de los modelos de HSI como predictor a nanoescala indica que hace un buen trabajo al predecir correctamente la presencia de chigüiros en lluvia de 2015 de manera aceptable (AUC= 0.711), mas no en sequía de 2016 (AUC= 0.6354). A mesoescala, las predicciones están por debajo de un AUC de 0.6, por lo que, la discriminación no es aceptable.

Por otra parte, a nanoescala el modelo de HSI en lluvia de 2015 clasificaría correctamente el 71.43% de las observaciones y en sequía de 2016 el 67.86%. En mesoescala el modelo de HSI en lluvia de 2015 clasificaría correctamente el 65.27% de las observaciones y en sequía de 2016 el 44.77%.

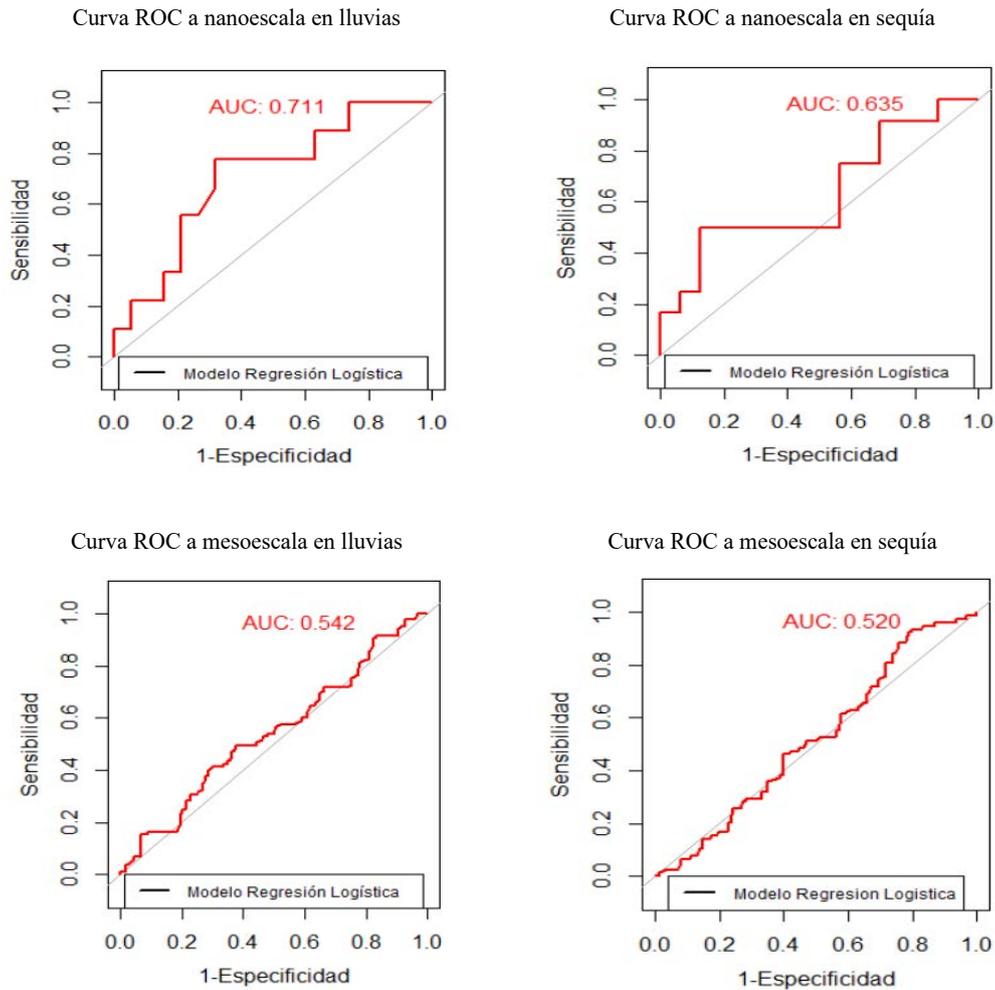


Figura 16. Curva de características operativas del receptor ROC para el modelo de regresión logístico.

Calidad predictiva del modelo de las variables de hábitat

A nanoescala, en lluvias de 2015, el índice de concordancia del modelo 4 (AUC= 0.8275) que tiene las variables la Riqueza de alimento potencial y Cobertura de escape como predictores, indica que hace un mejor trabajo al predecir correctamente los resultados que con el modelo univariante de HSI

(AUC= 0.711). La prueba de McNemar ($\chi^2= 1.5$, $p= 0.2207$) señala que estos dos modelos tienen clasificadores con proporciones similares de error en los conjuntos de prueba, por lo tanto, cualquier de los dos es igual de eficiente. Como se mencionó anteriormente, en el caso de la época de sequía de 2016 ninguno de los modelos multivariantes ofrece un mejor ajuste a los datos, por lo tanto, el modelo que tiene el HSI (AUC= 0.6354) es el mejor modelo que determina la presencia de chigüiros.

A mesoescala, en lluvias de 2015, el índice de concordancia del modelo 4 (AUC= 0.645) que tiene las variables la Riqueza de alimento potencial y Cobertura de escape como predictores, indica que hace un mejor trabajo al predecir correctamente los resultados que con el modelo 6 (Presencia ~ Riqueza de alimento potencial + Cobertura climática) (AUC= 0.6153). La prueba de McNemar ($\chi^2= 0.071429$, $p= 0.7893$) señala que estos dos modelos tienen clasificadores con proporciones similares de error en los conjuntos de prueba, por lo tanto, cualquier de los dos es igual de eficiente. En sequía de 2016, el índice de concordancia del modelo 7 que tiene las variables Cobertura de escape, Cobertura climática y Disponibilidad de agua como predictores indica que hace un mejor trabajo al predecir correctamente los resultados que el modelo 4 que contiene las variables del refugio. El resultado de la prueba de McNemar ($\chi^2= 4.9231$, $p= 0.0265$) es significativo y la hipótesis nula se rechaza, por lo tanto, ambos clasificadores no tienen proporciones similares de error en los conjuntos de prueba, de manera que el modelo 4 es el mejor modelo que determina la presencia de chigüiros.

Modelo de Regresión Poisson inflado con cero

En la Figura 17 se observa una alta concentración de valores en cero en la distribución del número de individuos de chigüiros a nanoescala y mesoescala en los dos períodos climáticos. A nanoescala, en lluvias de 2015 y sequía de 2016, se observan 19 (67.87%) y 16 (57.14%) valores en cero, respectivamente, de un total de 28 observaciones por cada época climática. A mesoescala, se tienen 154 (64.43%) y 161 (67.36%) registros de ceros en lluvias de 2015 y sequía de 2016, respectivamente, de un total de 239 celdas. Por lo tanto, para analizar este conjunto de datos es adecuado emplear el modelo de regresión *Poisson inflado con cero* (PIC).

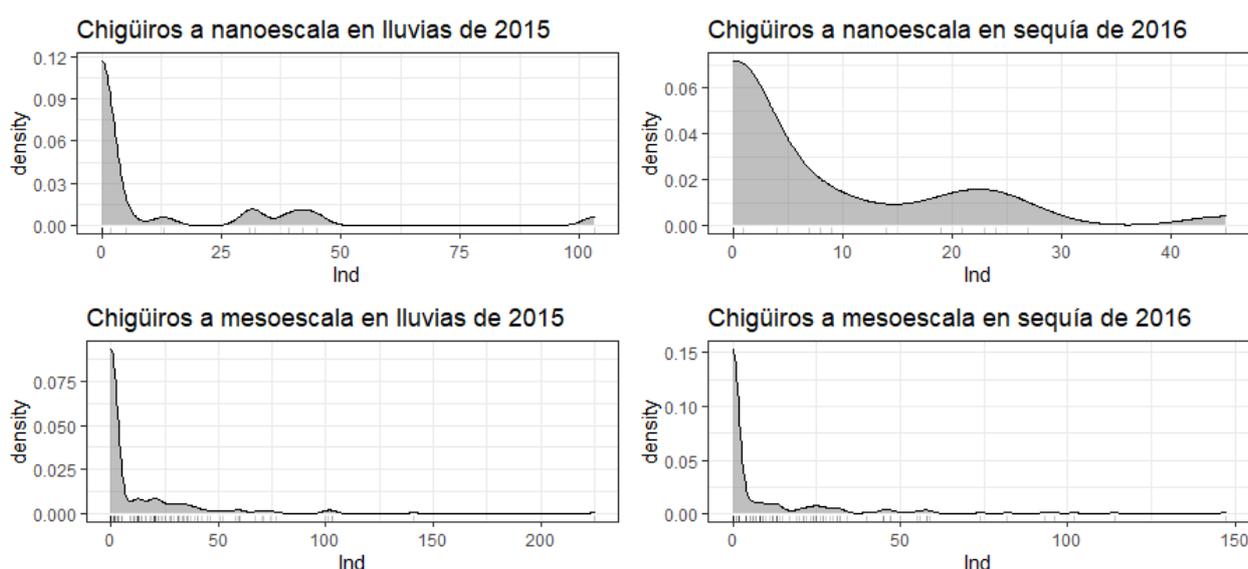


Figura 17. Distribución de chigüiros a nanoescala y mesoescala en dos períodos climáticos.

En la Tabla 9 y en la Tabla 10 se presentan los resultados de los modelos PIC para HSI a nanoescala y mesoescala en los dos períodos climáticos. A nanoescala, el efecto del HSI sobre el número de individuos de chigüiros es positiva y altamente significativa en lluvias de 2015 (Tabla 9). A mesoescala, la relación entre el HSI y el número de chigüiros es negativa y significativa al 10% en los dos períodos climáticos (Tabla 10).

Tabla 9. Modelo PIC estimado para el efecto del HSI sobre el número de chigüiros a nanoescala.

Período climático	VARIABLES	Componente de conteo	Componente cero inflado
Lluvia 2015	(Intercepto)	6.07333*** (0.31842)	-2.9327 (2.0748)
	HSI	0.62849*** (0.08154)	0.7907 (0.4568)
Sequía 2016	(Intercepto)	3.156*** (0.26365)	-1.7196 (1.5895)
	HSI	-0.08983 (0.06948)	0.4985 (0.3884)

El error estándar aparece entre paréntesis.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Tabla 10. Modelo PIC estimado para el efecto del HSI sobre el número de chigüiros a mesoescala.

Período climático	VARIABLES	Componente de conteo	Componente cero inflado
Lluvia 2015	(Intercepto)	4.06141*** (0.10613)	-0.3318 (0.8067)
	HSI	-0.11764*** (0.02204)	0.1894 (0.1633)
Sequía 2016	(Intercepto)	4.21857*** (0.12179)	1.1814 (0.7682)
	HSI	-0.19591*** (0.02679)	-0.1002 (0.1653)

El error estándar aparece entre paréntesis.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

En la Tabla 11 y en la Tabla 12 se presentan los resultados de los modelos PIC para las variables del hábitat a nanoescala y mesoescala en los dos períodos climáticos. Las variables del hábitat que tienen relación negativa y significativa al 10%, con el número de chigüiros a nanoescala en lluvias de 2015, son la Riqueza y el Forraje de alimento potencial y la Cobertura climática (Tabla 11). En sequía de 2016, las estimaciones de los coeficientes son significativas para todas las variables del

hábitat; las variables que tienen relación negativa y significativa al 10% son el Forraje de alimento potencial y la Cobertura de escape, y las variables con relación positiva y significativa son la Riqueza de alimento potencial y la Cobertura climática (Tabla 11).

Tabla 11. Modelo PIC estimado para el efecto de las variables del hábitat sobre el número de chigüiros a nanoescala.

Período climático	Variables	Componente de conteo	Componente cero inflado
Lluvia 2015	(Intercepto)	8.29864*** (0.58452)	3.9186 (4.2132)
	Riqueza de alimento potencial	-0.55085** (0.16809)	-1.2932 (0.9213)
	Forraje de alimento potencial	-0.52687*** (0.06002)	-0.2972 (0.5204)
	Cobertura de escape	0.04066 (0.03087)	0.3227 (0.2177)
	Cobertura climática	-0.26508*** (0.06348)	0.1275 (0.3548)
Sequía 2016	(Intercepto)	-2.95054** (1.13877)	-7.7083 (7.328)
	Riqueza de alimento potencial	3.12873*** (0.59622)	2.27509 (3.18119)
	Forraje de alimento potencial	-0.22597** (0.06881)	0.03989 (0.50755)
	Cobertura de escape	-0.32843*** (0.05765)	0.23422 (0.27032)
	Cobertura climática	0.57044*** (0.09981)	0.63928 (0.54173)

El error estándar aparece entre paréntesis.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A mesoescala, en lluvias de 2015, las variables del hábitat que tienen relación negativa y significativa al 10%, con el número de chigüiros son el Forraje de alimento potencial, la Cobertura de escape y la Longitud de bordes; las variables con relación positiva y significativa son la Riqueza de alimento potencial, la Cobertura climática y el Índice de yuxtaposición (Tabla 12). En sequía de 2016, el Forraje de alimento potencial, la Cobertura de escape, la Cobertura climática, la Disponibilidad de agua y la Longitud de bordes son las variables que tienen relación negativa y significativa al 10%; mientras que la Riqueza de alimento potencial y el Índice de dominancia son las variables con relación positiva y significativa (Tabla 12).

Tabla 12. Modelo PIC estimado para el efecto de las variables del hábitat sobre el número de chigüiros a mesoescala.

Período climático	VARIABLES	Componente de conteo	Componente cero inflado
Lluvia 2015	(Intercepto)	8.497507*** (0.325238)	-3.28678 (2.243968)
	Riqueza de alimento potencial	0.64526*** (0.130167)	-0.173257 (0.622395)
	Forraje de alimento potencial	-1.302998*** (0.146669)	0.117197 (0.492143)
	Cobertura de escape	-0.323208*** (0.06772)	0.534341 (0.369403)
	Cobertura climática	0.436762*** (0.061197)	0.310595 (0.316031)
	Disponibilidad de agua	0.002372 (0.005862)	-0.001186 (0.039697)
	Índice de yuxtaposición (IJ)	0.067487*** (0.011569)	-0.068029 (0.068776)
	Índice de dominancia (IDO)	0.047951 (0.051426)	0.560843 (0.3262)
	Longitud de bordes	-0.493836*** (0.050906)	-0.113948 (0.357549)
	Sequía 2016	(Intercepto)	4.02862*** (0.157448)
Riqueza de alimento potencial		0.92459*** (0.167326)	-0.7272 (1.04536)
Forraje de alimento potencial		-0.297011*** (0.068661)	0.28271 (0.39724)
Cobertura de escape		-0.167931*** (0.032297)	0.43385* (0.18791)
Cobertura climática		-0.377648*** (0.05434)	-0.03097 (0.24999)
Disponibilidad de agua		-0.046324*** (0.006404)	-0.05762 (0.03855)
Índice de yuxtaposición (IJ)		-0.009011 (0.014939)	0.0485 (0.07264)
Índice de dominancia (IDO)		0.403468*** (0.052084)	-0.0456 (0.32319)
Longitud de bordes		-0.345356*** (0.051608)	-0.04503 (0.31968)

El error estándar aparece entre paréntesis.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

En los modelos PIC estimados se observa que los coeficientes significativos para el HSI y para las variables del hábitat son para el componente de conteo mas no para el componente cero inflado, excepto para la Cobertura de escape a mesoescala en sequía de 2016, en donde es significativa en ambos componentes. Por lo tanto, la probabilidad del proceso que genera los ceros no aumenta conforme disminuye la media del proceso Poisson.

Las variables de refugio (Cobertura de escape y Cobertura climática) son las que mejor representan el número de chigüiros a nanoescala en lluvias de 2015 y en sequía de 2016 las Riqueza de alimento potencial. A mesoescala, las variables que mejor representan el número de chigüiros en lluvias de 2015 son el Índice de dominancia y las variables de refugio y en sequía de 2016 son la Cobertura de escape, el Forraje de alimento potencial y el Índice de yuxtaposición.

Conclusiones y discusión

En las regresiones que se realizaron se observa que no existe una relación directa entre el aumento de la calidad del hábitat para el chigüiro y el número de individuos ni tampoco con la presencia de la especie en las dos escalas y períodos climáticos evaluados, excepto a mesoescala en sequía de 2016. Esto difiere con lo reportado en Argentina por Corriale (2010).

El modelo de HSI para los chigüiros en la Orinoquía colombiana aplicado por Cardona-Claros (2018) es aceptable y, por lo tanto, es una herramienta útil para la implementación de planes y acciones de manejo y aprovechamiento de la especie en la zona, sin embargo, como todo modelo puede mejorarse para aumentar su precisión en pronósticos. Se podrían incluir algunas variables al modelo de idoneidad de hábitat como la distancia a las carreteras y a cuerpos de agua, que no fueron consideradas en dicho modelo y resultaron importantes en la idoneidad alta del hábitat del modelo de HSI propuesto por Schivo et al. (2015) en Argentina.

El modelo de HSI a nanoescala tiene una mejor precisión que a mesoescala. Aunque en ambos casos el modelo es aditivo, difieren la cantidad de variables; por ejemplo, a mesoescala se considera más variables que a nanoescala como la disponibilidad de cuerpos de agua, tres variables de métricas del paisaje y una variable como factor de caza. Además, las diferencias entre las predicciones y las observaciones en estas dos escalas de evaluación pueden deberse a las valoraciones de las Unidades de Evaluación de hábitat (UEH); ya que a nanoescala son más precisas (existe una mejor resolución espacial) al estar asociadas a puntos de muestreo específicos, mientras que a mesoescala hay características de las coberturas vegetales que sólo son definidas a partir de las imágenes satelitales disponibles para la zona en los dos períodos climáticos. Otras cuestiones de los casos mal clasificados podrían ser por una baja detectabilidad de los individuos en coberturas vegetales donde la visibilidad de los individuos se reduce notoriamente como en las sabanas con pajonales, sabanas con arbustos y en bosques de galería; también podrían considerarse cuestiones poblacionales que no han sido evaluadas en campo.

Agradecimientos

Al CONACYT por la beca otorgada durante el tiempo de la Especialización en Métodos Estadísticos. A la Maestra Arlensiu Tiboche García, al Doctor Hugo Fernando López Arévalo y al

Grupo de Conservación y Manejo de Vida Silvestre de la Universidad Nacional de Colombia por proporcionar los datos poblacionales de *Hydrochoerus hydrochaeris* en la Orinoquía Colombiana. A la Licenciada Rebeca Melo García, al Maestro José Luis Juárez Báez y a la Doctora Cecilia Cruz López por sus valiosos comentarios al presente documento.

Referencias

- Bejarano M., P., Rodríguez, M., y Vélez Á., N. (2014). Estructura y productividad de las poblaciones silvestres de chigüiro (*Hydrochoerus hydrochaeris*) en tres municipios del departamento de Casanare-Colombia. En H. López-Arévalo, P. Sánchez-Palomino, y O. Montenegro (Eds.), *El Chigüiro Hydrochoerus hydrochaeris en la Orinoquía colombiana: ecología, manejo sostenible y conservación*. Bogotá: Biblioteca José Jerónimo Triana No. 25. Instituto de Ciencias Naturales. Universidad Nacional de Colombia.
- Brooks, R. P. 1997. Improving habitat suitability index models. *Wildlife Society Bulletin*, 25, 163–167.
- Cardona-Claros, C.N. (2018). *Cambios espaciotemporales del hábitat del chigüiro (Hydrochoerus hydrochaeris) en sabanas inundables del Casanare* [Tesis de Maestría, Universidad Nacional de Colombia].
- Clavijo, J. J. (1993). Aspectos de la biología y el manejo de los chigüiros en Colombia. Pp. 21-65, en *Manejo de fauna silvestre acuática y actualización en zootecnia*. Memorias.
- Corriale, M. J. (2010). *Uso y selección de hábitat del carpincho (Hydrochoerus hydrochaeris) a distintas escalas espaciotemporales en los Esteros del Iberá, Corrientes, Argentina* [Tesis de Doctorado, Universidad de Buenos Aires].
- Emmons, L. y Feer, F. 1997. Neotropical rainforest mammals: A field guide. 2.ed. Chicago, University of Chicago Press.
- Fielding, A.H., 2002. What are appropriate characteristics of an accuracy measure? In: Scott, J.M., Heglund, P.J., Morrison, M.L., et al. (Eds.), *Predicting Species Occurrences. Issues of Accuracy and Scale*. Island Press, Washington, pp. 271–280.
- Forero, J. (1999). *Dieta alimentaria del chigüiro (Hydrochaeris hydrochaeris) en Caño Limón, Arauca, Colombia* [Tesis de Biología, Universidad de los Andes].
- Forero-Montaña, J., Betancur, J., y Cavelier, J. (2003). Dieta del capibara *Hydrochaeris hydrochaeris* (Rodentia: Hydrochaeridae) en Caño Limón, Arauca, Colombia. *Revista de Biología Tropical*, 51(2), 579–590.
- Guzmán-Lenis, A. R., A. A. Maldonado-Chaparro, H. F. López-Arévalo, P. Sánchez-Palomino, O. L. Montenegro y M. A. Torres. (2014a). Calidad de hábitat disponible para el chigüiro en las sabanas inundables de la Orinoquía: Propuesta metodológica. En *El chigüiro Hydrochoerus hydrochaeris en la Orinoquía Colombiana: Ecología, manejo sostenible y conservación*. (pp. 59-75). Bogotá: Universidad Nacional de Colombia.
- Guzmán-Lenis, A. R., A. A. Maldonado-Chaparro, H. F. López-Arévalo, P. Sánchez-Palomino, O. L. Montenegro y M. A. Torres. (2014b). Evaluación de la calidad del hábitat disponible para el chigüiro en el municipio de Paz de Aripuro, Casanare. En *El chigüiro Hydrochoerus hydrochaeris en la Orinoquía Colombiana: Ecología, manejo sostenible y conservación*. (pp. 77-101). Bogotá: Universidad Nacional de Colombia.
- Gysel, L. W. y L. J. Lyon. (1980). Análisis y evaluación de hábitat. En: *Manual de técnicas de gestión de vida silvestre*. (pp. 321-344). Estados Unidos de América: Wildlife Society, Maryland.
- Hall, L.S., Krausman, P.R., Morrison, M.L. (1997). The habitat concept and a plea for standard terminology. *Wildlife Society Bulletin*, 25(1), 173-182.

- Herrera, E. A., y Macdonald, D. W. (1989). Resource utilization and territoriality in group-living capybaras (*Hydrochoerus hydrochaeris*). *Journal of Animal Ecology*, 58(2), 667–679.
- Hosmer, D.W. y Lemeshow, S. (2000). Applied Logistic Regression. 2nd ed. John Wiley & Sons, Inc. Pp. 156-164.
- Litvaitis, J.A., K. Titus y E.M. Anderson. (1994). Measuring vertebrate use of terrestrial habitats and foods. Pages 254–271 in T.A. Bookhout [ED.], Research and management techniques for wildlife and habitats. The Wildlife Society, Bethesda, MD U.S.A.
- Mones, A. y Ojasti, J. (1986). *Hydrochoerus hydrochaeris*. *Mammalian Species*, 264, 1-7.
- Ojasti, J. (1973). *Estudio biológico del chigüire o capibara*. Caracas: Fondo Nacional de Investigaciones Agropecuaria. Editorial Sucre.
- Ojasti, J. (1990). Las comunidades de mamíferos en sabanas neotropicales, En: Sarmiento, G. (ed.), Las sabanas americanas. Aspectos de su biogeografía, ecología y utilización. CIELAT, pp. 259-293.
- Ojasti, J. (1997). Explotación humana de capibaras. Pp 280-301, En: Uso y Conservación de la Vida Silvestre Neotropical (Robinson JG y KH Redford, eds.). Editorial Fondo de la Cultura Económica, México D.F., México.
- Quintana, R.D. y Rabinovich, J. (1993). Assessment of capybara (*Hydrochaerus hydrochaeris*) populations in the wetlands of Corrientes Argentina Wetlands. *Ecology and Management*, 2(4), 223-230.
- Quintana, R.D. (1996). *Análisis y evaluación de la aptitud de hábitat del carpincho (Hydrochaeris hydrochaeris) en relación con la heterogeneidad del paisaje y las interacciones con ganado doméstico* [Tesis de doctorado, Universidad de Buenos Aires].
- Reiley, B. M., Bednarz, J. C., and Brown, J. D. (2014). A test of the Swainson's warbler habitat suitability index model. *Wildl. Soc. Bull.*, 38, 297–304.
- Schivo, F., Kandus, P., Bolkovic, M.L., Minottie, P.G., Trilla, G., Quintana, R.D. (2015). A habitat suitability model for capybara (*Hydrochoerus hydrochaeris*) at its core area in Argentina. *Tropical Conservation Science*, 8(1), 150–168.
- Soini, P. y M. Soini. (1992). Ecología del roncoso o capybara (*Hydrochaeris hydrochaeris*) en la Reserva Nacional Pacaya-Samiria, Perú. *Folia Amazónica*, 4(2), 119-133.
- Tiboche García, A. (2016). *Análisis de los cambios en el tamaño, densidad y estructura poblacional del chigüiro (Hydrochoerus hydrochaeris), como herramienta para el establecimiento de una unidad de manejo sostenible en Paz de Ariporo, Casanare* [Tesis de Maestría, Universidad Nacional de Colombia].
- U.S. Fish and Wildlife Service. (1981). Standards for the development of habitat suitability index models (103 EMS). U.S. Dep. Inter. Fish and Wildl. Serv., Div. Ecol. Serv., Washington, D.C. 162 pp.
- UNAL-Universidad Nacional de Colombia y CORPORINOQUIA. (2015). Estado actual de las poblaciones de Chigüiro (*Hydrochoerus hydrochaeris*) y su hábitat bajo escenario de intervención antrópica y cambio climático, en el departamento de Arauca y Casanare en jurisdicción de Corporinoquia. Convenio interadministrativo 200-15-15-022. Informe final. Bogotá.
- Van Horne, B. (1983). Density as a misleading indicator of habitat quality. *Journal of Wildlife Management*, 7, 893–901.
- Yarrow, G. (2009). Habitat Requirements of Wildlife: Food, Water, Cover and Space. *Forestry and Natural Resources. Fact Sheet*, (14), 1-5.

Anexo 1. Medidas de ajuste para los modelos de regresión logística a nanoescala y mesoescala en dos períodos climáticos.

Lluvia 2015				Sequía 2016			
Modelos	AIC	Verosimilitud Pr(>Chi)	Devianza	Modelos	AIC	Verosimilitud Pr(>Chi)	Devianza
Nanoescala							
Mod3: Presencia ~ HSI	34.997	0.04119*	0.119	Mod3: Presencia ~ HSI	40.331	0.167	0.050
Mod4: Presencia ~ RiquezaAlimento + CoberturaEscape	31.845	0.009469**	0.265	Mod4: Presencia ~ RiquezaAlimento + CoberturaEscape	43.256	0.610	0.026
Mod5: Presencia ~ RiquezaAlimento + ForrajeAlimento + CoberturaClimática + CoberturaEscape	35.263	0.04212*	0.282	Mod5: Presencia ~ RiquezaAlimento + ForrajeAlimento + CoberturaClimática + CoberturaEscape	45.312	0.569	0.077
Mesoescala							
Mod3: Presencia ~ HSI	313.77	0.2454	0.004	Mod3: Presencia ~ HSI	305.520	0.5431	0.001
Mod4: Presencia ~ RiquezaAlimento + CoberturaEscape	300.50	0.0002465***	0.053	Mod4: Presencia ~ CoberturaEscape + CoberturaClimática	294.6	0.001301**	0.044
Mod5: Presencia ~ CoberturaClimática + CoberturaEscape	299.87	0.0001796***	0.055	Mod5: Presencia ~ CoberturaEscape + DisponibilidadAgua	292.22	0.0003951***	0.052
Mod6: Presencia ~ RiquezaAlimento + CoberturaClimática	300.37	0.0002306***	0.054	Mod6: Presencia ~ CoberturaClimática + DisponibilidadAgua	300.23	0.02167*	0.025
Mod7: Presencia ~ CoberturaClimática + CoberturaEscape + RiquezaAlimento	301.06	0.0004285***	0.058	Mod7: Presencia ~ CoberturaEscape + CoberturaClimática + Agua	294.22	0.001322**	0.052

* p<0.05; ** p<0.01; *** p<0.001.

Inoculación con hongos *Micorrízicos Arbusculares* como alternativa para optimizar el aprovechamiento de maíces criollos sembrados en policultivo

Alejandra Ortiz Gasca¹, Zoylo Morales Romero², Wendy Sangabriel Conde³

¹Facultad de Estadística e Informática, Universidad Veracruzana, aleortizga@gmail.com

²Facultad de Estadística e Informática. Universidad Veracruzana. zmorales@uv.mx

³Facultad de Ciencias Agrícolas, Universidad Veracruzana. wsangabriel@gmail.com.

Resumen

La interacción entre los hongos micorrízicos y las plantas ha sido descrita como una estrategia positiva para incrementar el rendimiento del agroecosistema sin comprometer al ambiente. El maíz es un cultivo de gran importancia en México, se ha observado que su asociación con diversas especies vegetales incrementa el rendimiento de éste, en contraste a su monocultivo. Es por ello que en este trabajo se estimó el efecto de la inoculación con hongos micorrízicos arbusculares (HMA) en el cultivo de maíz (*Zea mays* L.) en asociación con calabaza larga (*Cucurbita pepo*), frijol de mata (*Phaseolus vulgaris* L.) y cempasúchil (*Tagetes erecta* L.). El experimento se estableció en la localidad de la Orduña, Coatepec, Veracruz, México, bajo un diseño factorial con tres factores: variedad de maíz a tres niveles (maíz blanco, maíz amarillo y maíz negro), inoculación y los cultivos asociados a tres niveles. Las variables consideradas fueron peso fresco de la planta y mazorca, peso seco de grano, rendimiento de maíz y biomasa, las cuales se analizaron mediante un análisis de varianza (ANOVA) con un nivel de significancia del 5% ($\alpha = 0.05$) para la comparación de medias. La interacción entre el cultivo asociado y la aplicación de un inóculo mostró diferencias significativas ($p < 0.5$) para el peso fresco de la planta, donde la aplicación del inóculo en asociación con calabaza incrementó los pesos de la planta. No se encontraron diferencias entre los factores evaluados, variedades de maíz, tratamiento inoculado y no inoculado, cultivo asociado (calabaza, frijol y cempasúchil) en cuanto al peso de la mazorca, peso seco de grano y peso fresco de la planta. Los rendimientos obtenidos no presentaron respuestas diferentes a los tratamientos aplicados.

Palabras Clave: hongos formadores de micorriza, milpa, policultivo, maíz criollo.

Introducción

El cultivo del maíz es de gran importancia económica para México y el segundo con mayores rendimientos obtenidos para el año 2019 y 2020 (SIAP, 2020). Durante estos años, el estado de Veracruz ocupó el séptimo y octavo lugar, respectivamente, en producción de grano de maíz, con un rendimiento medio de 2.22 ton^{ha-1} (SIAP, 2021). A pesar de ello los resultados obtenidos se consideran bajos o deficientes en relación con la superficie sembrada. Rubén Purroy Vásquez y colaboradores (2016) registraron que para la zona centro del Veracruz más de un 50% del cultivo de maíz se encuentra en lomerío (zonas con pendientes que van desde 20 a 100m con respecto al nivel base, susceptibles a la erosión y con suelos principalmente ácidos) y en cuanto al tipo de agricultura, en un 57% es tradicional.

Se habla de agricultura tradicional cuando los sistemas se establecen “con base en conocimientos empíricos de la cultura campesina, indígena e indo mestiza ha heredado por generaciones, los cuales se han alimentado con la observación, recolección y permanente reflexión” (Pérez et al., 2015), el agricultor hace uso de técnicas e insumos que se adaptan según las condiciones locales, este tipo de sistema ha perdurado debido al modelo de sostenibilidad que representa, puesto que promueve la diversidad biológica y prescinde de agroquímicos (Altieri y Toledo, 2011). Por otra parte, la agricultura convencional suele estar ligado a tecnologías promovidas desde la “Revolución Verde” las cuales suelen realizarse a costa de los ecosistemas promoviendo el monocultivo e insumos de alta demanda energética (Córdova y León, 2013).

El uso de variedades de semilla suele asociarse al tipo de agricultura que se emplea, en regiones con agricultura campesina tradicional de un 80 a 100% es usada la semilla criolla, a diferencia de la agricultura convencional donde es usada menos de un 50% (Cabera, Macías López, Ruiz, Ramírez y Alvarado, 2002). Se ha registrado que estas pequeñas y medianas unidades de producción agrícola cultivan un mayor número de variedades permitiendo una elevada recombinación genética y la conservación de la biodiversidad (Escobar Moreno, 2006).

No obstante, existe una preferencia a variedades nativas sobre aquellas mejoradas debido a la falta de adaptación y limitantes ecológicas del agroecosistema, así como su asociación a paquetes tecnológicos, alta demanda de insumos (Trejo, Gil, Sánchez, Carballo y López, 2004). A pesar de ello, la producción resultante no es suficiente para satisfacer la subsistencia de las comunidades y cubrir la demanda del mercado.

Los sistemas agrícolas tradicionales suelen intercalar diversos cultivos, como lo es la milpa, este sistema de policultivo combina la siembra de diversas especies como leguminosas, tomates, chiles, calabaza y maíz (Macías García, 2017). Esta combinación reproduce diversas interacciones y principio ecológicos debido a la diversidad vegetal y comunidad biológica que albergan (Eguiarte, Equihua Zamora y Espinosa Asuar, 2017). Por otra parte, se sabe que la interacción del maíz, calabaza y frijol genera un aumento en la producción del primero, en contraste con su monocultivo (Aguilar, Illsley, & Marielle, 2003).

A su vez, se ha verificado que la aplicación de hongos micorrízicos arbusculares (HMA) en maíz trae beneficios para la planta, mejorando su crecimiento y desarrollo (Xu et al., 2018). En esta simbiosis el hongo aporta hasta el 60% del total del fósforo (P) requerido, y la aplicación de inóculo puede reducir hasta un 33% de la aplicación de fertilizantes fosforados (Mora y Leblanc, 2012). Además, en suelos donde el (P) es limitante los HMA promueven la absorción de fosfato (Sawers et al., 2017). Sin embargo, las variedades e híbridos difieren en la respuesta al porcentaje de colonización y capacidad de absorción de fósforo, por ejemplo, en la región de los Tuxtlas se encontró diferentes respuestas a la colonización nativa, para distintas variedades de maíz, y la relación en la concentración de (P) así como diferentes asociaciones de especies de HMA (Sangabriel et al., 2014).

En relación con la problemática expuesta, el propósito del presente trabajo fue entender el efecto de la aplicación de un inóculo de HMA sobre diferentes variedades de maíz criollo, sembradas en policultivo y condiciones de campo.

Metodología

Localización del sitio experimental

El sistema de policultivo se estableció en la localidad de La Orduña, municipio de Coatepec, Veracruz, México en las coordenadas 19°27'21.4"N 96°56'03.8"O a una altitud de 1185.8077m, en una superficie de 476.7 m². El experimento se realizó en el mes de julio de 2020 correspondiente a la temporada otoño invierno, bajo temporal.

El suelo de la zona ha sido identificado como luvisol (INEGI,2007), caracterizado por tener un enriquecimiento de arcilla en el subsuelo, cuentan con una capacidad de humedad moderada y susceptibles a la erosión (INEGI, 2004).

La selección del sitio se realizó considerando características como la importancia de la actividad agrícola en la zona, la presencia del sistema milpa con diferentes tipos de maíces criollos y especies tradicionales que allí se cultivan.

Diseño estadístico

Este trabajo se planteó bajo un diseño experimental factorial con tres factores: 1) variedades de maíz con tres niveles (maíz blanco, maíz amarillo y maíz negro); 2) tratamiento con dos niveles (inoculado y no inoculado); 3) cultivo asociado con tres niveles (calabaza, frijol y cempasúchil).

Los tipos de maíces sembrados fueron; maíz amarillo, maíz blanco y maíz negro, con semillas de la zona, debido a que son las variedades mayormente cultivadas por los productores de la región. Para el desarrollo del sistema de policultivo se consideraron dos de las especies más importantes de la milpa, *calabaza (Cucurbita pepo)*, *frijol de mata (Phaseolus vulgaris L.)* y *cempasúchil (Tagetes erecta L.)* por ser una especie descrita como altamente micotrófica.

El biofertilizante implementado fue Rizofermic, un inoculante a base de un consorcio de hongos micorrízicos arbusculares propios de la región con un porcentaje de colonización radical de 85%.

Distribución y aplicación de tratamientos

Cada subparcela constó de 44 plantas de maíz sembradas a una distancia de 40cm entre planta x 50cm entre surco. Se descartaron las plantas del perímetro para eliminar el efecto borde y se realizó el muestreo de 16 plantas por subparcela. Las plantas se identificaron bajo claves para la toma de datos de un mismo individuo. Para aquellos tratamientos con inóculo, este se aplicó al momento de la siembra, 15 mg. por planta y posteriormente a los 15 días se llevó a cabo una segunda aplicación. Previo al establecimiento del cultivo se limpió el terreno a mano, con la ayuda de un azadón, lima y machetes; durante el desarrollo, se realizaron deshierbes de manera manual y no se aplicó insumos de síntesis química. El aporque se llevó a cabo utilizando azadones y manteniendo el ensayo libre de malezas.

Sistema de muestreo y variables analizadas

El muestreo de las plantas y mazorcas se realizó conforme a la metodología establecida por Centro Internacional de mejoramiento de Maíz y Trigo (CIMMYT, 2012), esta misma metodología se utilizó para la estimación del rendimiento de maíz y biomasa. Para evaluar el

rendimiento del maíz fueron consideradas las variables agronómicas de pesos frescos de mazorca, pesos secos de grano, pesos frescos de la planta.

En cuanto al porcentaje de colonización, el muestreo de raíces se realizó a los 120 días después de la siembra (DDS) correspondiente a la etapa de maduración. Se colectó 2g de raíces las cuales fueron procesadas para su tinción siguiendo la técnica propuesta por Phillips y Hayman (1970), la colonización micorrízica se cuantificó a través de la presencia de estructuras fúngicas utilizando la técnica de Giovannetti y Mosse (1989).

Análisis estadístico

Los datos se procesaron mediante un análisis de varianza (ANOVA), para aquellos datos que resultaron significativos se realizó la comprobación de los supuestos mediante las pruebas de Lilliefors para normalidad, Levene para homogeneidad de varianzas y Durbin-Watson independencia. Al no cumplirse el supuesto de normalidad, para la interacción de tratamiento y cultivo asociado en relación con el peso fresco de la mazorca se aplicó la prueba Kruskal-Wallis. Mediante el análisis de componentes principales, sin rotación, se crearon dos indicadores de aprovechamiento para grano y maíz, se relacionaron con el porcentaje de colonización a través del coeficiente de correlación de Pearson. El nivel de significancia considerado para todas las pruebas fue de 0.05. El análisis estadístico se llevó a través de los softwares estadísticos R i386 4.0.3 y MINITAB Release 14.1.

Resultados

Debido a cuestiones inherentes a la variedad de maíz negro, el cultivo llegó a la etapa de maduración vegetal y se presentó precocidad en la floración, fenómeno reportado con anterioridad para este tipo de variedad (Espinosa Trujillo, Mendoza Castillo y Castillo Gonzales, 2006). Sin embargo, no se desarrollaron las mazorcas suficientes para la toma de datos por lo cual fue descartados para el presente documento, evaluando así las variedades de maíz criollo blanco y amarillo.

Variables Agronómicas

Para las variables respuesta evaluadas, peso de la mazorca, peso seco del grano y peso fresco de la planta, no se encontraron efectos de primer orden significativos ($p > 0.05$).

La interacción de los tratamientos a los que se aplicó el inóculo y las variedades de maíz no presentó diferencias estadísticas significativas ($p > 0.05$) en las variables peso de la mazorca, peso seco del grano y peso fresco de la planta. Es decir, se observó que aquellos tratamientos inoculados con HMA en comparación con los tratamientos no inoculados presentaron valores medianos similares o menores (Figura 1). Se ha reportado, principalmente bajo condiciones de invernadero, que la aplicación de HMA tiene efectos positivos en la producción maíz, tales como incremento en la altura de las plantas, biomasa, área foliar y por ende en el rendimiento, con diferentes respuestas según la variedad (Anozie y Orluchukwu, 2018). Esto puede significar que la aplicación del inóculo bajo esa dosis y en condiciones de campo no favoreció la producción para estas variables, esto concuerda con Robledo et al. (2017) al no encontrar diferencias en la aplicación de biofertilizante a base de HMA en maíz. A su vez

con Días Franco, Salinas, Garza y Mayek (2008) al reportar algo similar en condiciones semiáridas.

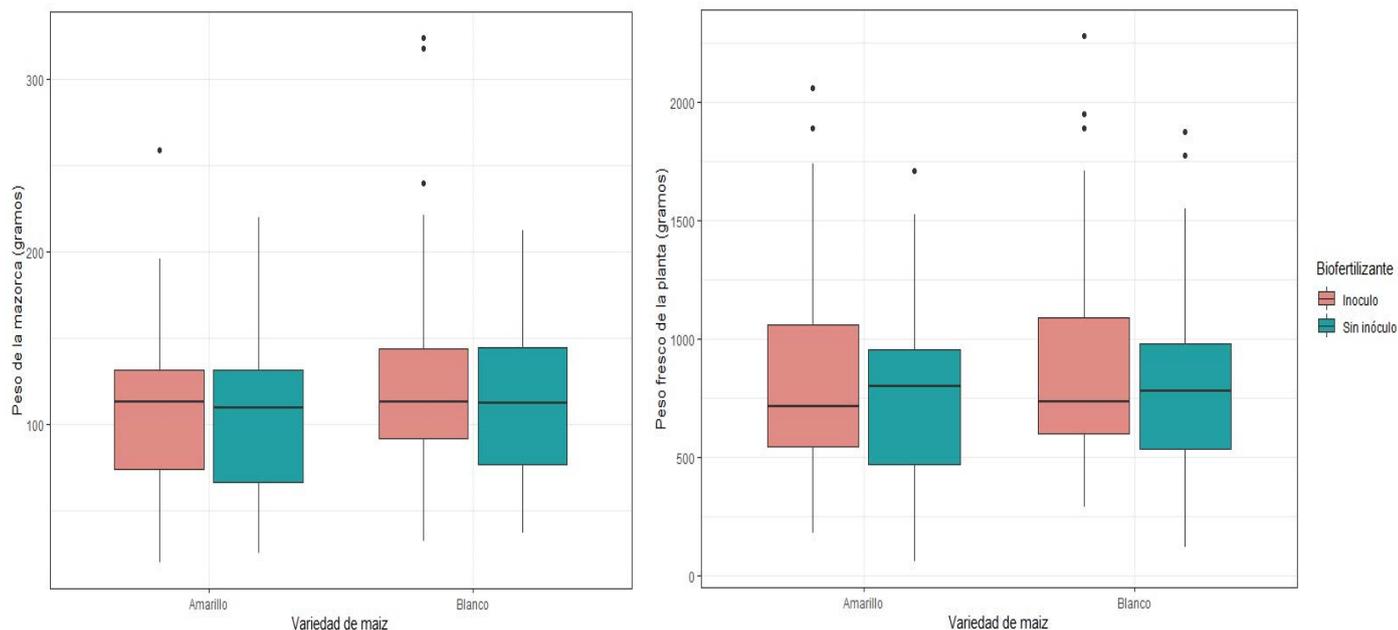


Figura 1. Pesos por variedad de maíz obtenidos, peso de la mazorca y peso fresco de la planta.

De igual manera, no se encontraron diferencias estadísticamente significativas ($p > 0.05$) entre las especies asociadas y las variedades de maíz. No obstante, los resultados obtenidos arrojaron que el maíz blanco obtuvo pesos mayores en asociación con calabaza, cempasúchil y frijol para las variables de peso de la mazorca, y peso fresco de la planta. Se ha registrado que la asociación del maíz con especies vegetales diferentes puede generar la modificación de la distribución de raíces en búsqueda de nutrientes del suelo, además en asociación con frijol el maíz incrementa la adquisición de fósforo (Zhang, 2020), esto podría explicar por qué la inoculación no resultó efectiva en esta asociación.

En cuanto al peso seco del grano, el maíz amarillo obtuvo los mayores pesos registrados cuando era asociado con cempasúchil y frijol (Figura 2). Son conocidos los efectos fungicidas, nematocidas e insecticidas del cempasúchil y como este puede influir en la reducción de problemas fitosanitarios de los cultivos aledaños (Gómez y Zavaleta, 2001), sin embargo, poco se sabe de los efectos que este pueda tener sobre las poblaciones de hongos micorrízicos nativos, ya que se ha observado respuestas favorables con anterioridad al inóculo aplicado.

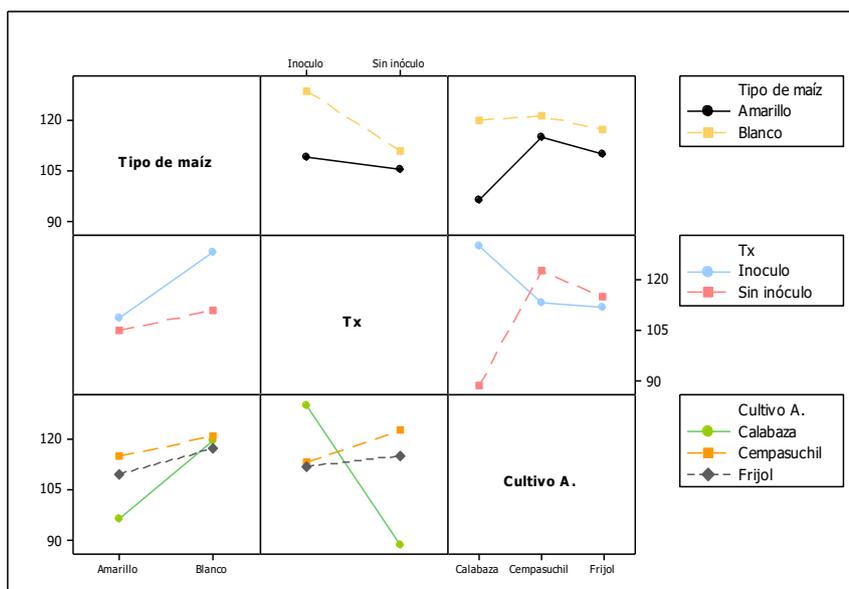


Figura 2. Gráfico de interacciones entre factores para el peso de la mazorca.

La aplicación de un inóculo y asociación de especies vegetales mostró diferencias significativas ($p < 0.05$), en cuanto al peso de la mazorca donde se observó que la aplicación de un inóculo incrementó los pesos para las variedades asociadas con calabaza, no así, en los casos de cempasúchil y frijol (Figura 2). Fueron comprobados los supuestos para el modelo, se utilizó el test de Lilliefors, y se encontró que el criterio de normalidad no se cumplió ($p < 0.05$), únicamente la homocedasticidad e independencia cumplieron los supuestos ($p > 0.05$). Es por ello que mediante la prueba Kruskal-Wallis se analizó la interacción entre los diferentes cultivos asociados y la aplicación del inóculo, donde se registró no haber diferencias significativas ($p=0.077$).

De igual manera para esta interacción se encontraron diferencias significativas ($p < 0.05$) para el peso fresco de la planta donde se logró observar una tendencia similar a lo obtenido en el peso de fresco de la mazorca. Es decir, cuando es aplicado el inóculo y el maíz se asocia con calabaza se registran mayores pesos que el tratamiento sin dicho inóculo, caso contrario al cempasúchil y frijol. Se sabe que la interacción maíz-calabaza es benéfica ya que esta última le proporciona la humedad necesaria a la planta (Aguilar, 2003), lo cual podría ser una condición favorable para el establecimiento de la simbiosis tras la inoculación. Finalmente, diferente a lo anterior, para el peso seco del grano no se encontraron diferencias estadísticas significativas ($p > 0.05$).

Porcentaje de colonización y dependencia micorrízica

Para la colonización micorrízica se encontró diferencias estadísticas significativas ($p < 0.05$) donde aquellos tratamientos con biofertilizante obtuvieron mayores porcentajes (Figura 3). La desigualdad de colonización fue notable para la variedad amarilla donde se dio una colonización de hasta 81.01% en el tratamiento con inóculo, mientras que la colonización menor se dio donde no se aplicaron los HMA (1.79%). La variación en respuesta a la

micorrización encontrada pudiera deberse a que las distintas variedades de maíz no se asociaron con el mismo ensamble de especies de HMA (Sangabriel Conde et al. 2015) o bien a que su respuesta al estar en simbiosis micorrízica es distinta (Oliveira et al. 2009). Por ejemplo, aun creciendo en condiciones similares, las variedades de maíz registraron diferentes porcentajes de colonización, posiblemente producto de las diferencias en las estrategias de colonización de las especies que colonizan a cada variedad en particular (Ortas, 2012).

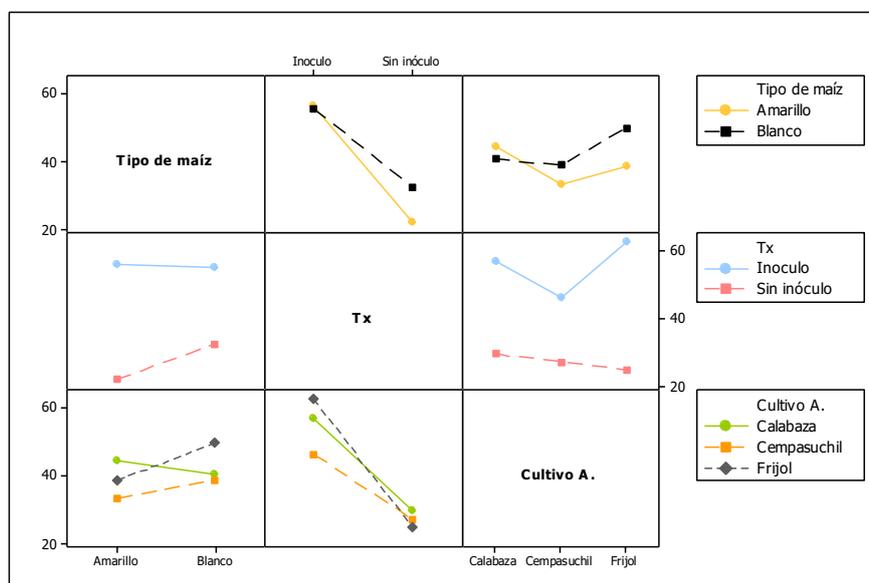


Figura 3. Gráfico de interacciones entre factores para el porcentaje de colonización

A su vez se encontró que la interacción entre los tres factores evaluados presentó diferencias estadísticas significativas ($p < 0.05$), donde la variedad de maíz blanco obtuvo mayores porcentajes de colonización micorrízica en asociación con frijol y la aplicación del inóculo, mientras que la variedad amarilla generó una respuesta favorable a la aplicación del inóculo cuando era asociado con cempasúchil. Las especies de HMA provenientes del inóculo fueron más eficientes para colonizar y facilitar los nutrientes a las plantas, se sabe que las especies de HMA difieren ampliamente en sus estrategias para colonizar las raíces y adquirir nutrientes (Thonar et al. 2011) y que estas estrategias pueden verse influenciadas por los cultivos asociados (Reyes, Lopez, Quiñones y Rincón, 2016).

Para evaluar la relación de los porcentajes de colonización se generaron dos índices del aprovechamiento del grano y otro de la planta, con las variables respuesta registradas. Este análisis dio un valor de 0.654 para la Medida Kaiser- Meyer-Olkin de adecuación de muestreo (KMO) que indicó un análisis regular, así como la prueba de esfericidad de Bartlett con un valor de $p < 0.001$ (chi- cuadrado 337.244 con 10 grados de libertad), lo cual indica que es pertinente realizar el modelo factorial, explicando un 86.88% (Tabla 1) de la variabilidad de los datos al tomar los dos factores obtenidos. El primer factor agrupó las variables de peso fresco de grano y mazorca, así como peso seco del grano (Tabla 2), en cuanto al segundo factor este agrupó los pesos secos y frescos de la planta. No se encontró correlación entre el porcentaje y el indicador de aprovechamiento de la planta, así mismo la correlación encontrada con el indicador de grano fue muy débil (0.203).

Tabla 1. Autovalores iniciales

Componente	Total	% de varianza	% acumulado
1	2.756	55.121	55.121
2	1.588	31.765	86.886
3	0.322	6.443	93.329
4	0.293	5.858	99.188
5	0.041	0.812	100

Tabla 2. Componentes generados

	Componente	
	1	2
Peso mazorca	0.862	-0.196
Peso fresco grano	0.941	-0.239
Peso seco grano	0.936	-0.221
Peso fresco planta	0.332	0.86
Peso seco planta	0.337	0.839

Rendimiento

En la Tabla 3 es posible observar que aun con la aplicación del inóculo se obtuvieron rendimientos de grano al 14% de humedad similares o cercanos y no se encontraron diferencias estadísticamente significativas ($p > 0.05$). Esta ausencia a la respuesta ante la aplicación del inóculo ha sido reportada con anterioridad (Sánchez de la cruz et al., 2008). Fue posible observar que el tratamiento de maíz amarillo en asociación con cempasúchil e inoculado fue aquel con mayor rendimiento estimado, sin embargo, el tratamiento de maíz blanco inoculado y sembrado con calabaza tiene un resultado similar a la misma variedad de maíz, pero sembrada con cempasúchil.

Se observó que en la mayoría de los casos la producción de biomasa fue más alta en las plantas inoculadas que en aquellas no inoculadas. El mayor peso de plantas se presentó en el tratamiento de maíz amarillo inoculado en asociación con cempasúchil, sin embargo, además de la posible ventaja dada por los HMA, la diferencia en peso y rendimientos con las demás variedades estudiadas puede ser atribuible a su genética (Marcos et al., 2016).

Tabla 3. Rendimientos estimados por tratamiento para las diferentes variedades de maíz.

Parcela	Cultivo asociado	Numero	Número	Índice de cosecha	Cociente grano: olote	Rendimiento seco al 14%	Rendimiento de biomasa (kg/ha)	Ton/ha
		de mazorcas buenas	de mazorcas malas					
BM	C	43	12	24.468379	1.310650888	786.6824912	2764.984739	0.7866825
BM	T	41	12	24.248578	1.488778055	583.7448301	2070.309219	0.5837448
BM	F	38	10	24.401616	3.12749004	779.4349392	2747.006741	0.7794349
BNM	C	42	10	32.816823	1.95709571	557.951254	1462.171008	0.5579513
BNM	T	38	4	33.261803	2.088948787	785.7372316	2031.561632	0.7857372
BNM	F	32	10	33.261803	2.088948787	654.3801125	5349.859281	0.6543801
AM	C	34	10	9.6179018	2.718894009	485.9884755	565.1028785	5052.957343
AM	T	40	6	15.804273	2.858085809	732.0411725	851.2106657	4631.919213
AM	F	36	7	10.23513	3.308035714	638.487536	742.4273675	6238.196663
ANM	C	36	6	11.474953	3	548.5540231	637.8535153	4780.446744
ANM	T	35	5	10.68125	2.992307692	652.2730592	758.4570455	6106.711187
ANM	F	32	3	9.6179018	2.36329588	512.2093305	595.5922447	5325.582865

BM= maíz blanco con inóculo, BNM= maíz blanco sin inóculo, AM= maíz amarillo con inóculo, ANM= maíz sin inóculo, C= calabaza, T= cempasúchil, F= frijol

La evidencia sugiere cierta especificidad funcional entre el consorcio de hongos presentes en el inoculante comercial Rizofermic y las plantas de maíz de las variedades estudiadas, y a su vez, una buena compatibilidad con el cultivo de calabaza, lo cual puede ser utilizado como estrategia para incrementar la producción de maíz criollo y generar un ahorro de insumos químicos.

Conclusiones

En este estudio no se encontraron diferencias estadísticas significativas entre los factores evaluados, variedades de maíz, tratamiento inoculado y no inoculado, cultivo asociado (calabaza, frijol y cempasúchil). Sin embargo, el uso de hongos micorrizicos arbusculares podría ser considerado como estrategia para optimizar la producción de grano de maíz blanco, mazorca y biomasa de esta variedad, cuando se asocie con calabaza y frijol. Por otro lado, cuando el maíz es asociado con cempasúchil la aplicación del inóculo no representa una alternativa para incrementar el rendimiento del maíz.

Agradecimientos

Agradecemos al CONACYT por el apoyo financiero para la realización de este trabajo. De igual manera al Laboratorio de Organismos benéficos de la Facultad de Ciencias Agrícolas por la disponibilidad del uso de este. Finalmente, a Ismael Gutiérrez Alarcón por su apoyo en la toma de datos de la presente investigación.

Referencias

- Aguilar, J., Illsley, C., y Marielle, C. (2003). Los sistemas agrícolas de maíz y sus procesos técnicos. *Sin maíz no hay país*, 83-122.
- Altieri, M. A., y Toledo, V. M. (2011). The agroecological revolution in Latin America: rescuing nature, ensuring food sovereignty and empowering peasants. *Journal of peasant studies*, 38(3), 587-612.
- Anozie, H. I., y Orluchukwu, J. A. (2018). Response of three varieties of maize (*Zea mays*) to arbuscular mycorrhizal fungi (*Gigaspora gigantea*) in the Humid Tropics. *Int. J. Agric. Earth Sci*, 4, 30-36.
- Cabrera, B., Macías López, A., Ruiz, R., Ramírez, M., y Alvarado, A. (2002). Uso de semilla criolla y caracteres de mazorca para la selección de semilla de maíz en México. *Revista Fitotec. México*, 25, 17.
- CIMMYT.2012. Manual de determinación de rendimiento. Centro Internacional de Mejoramiento de Maíz y Trigo, Mexico DF (Mexico).
- Díaz-Franco, A., J. Salinas G., I. Garza C. y N. Mayek P. 2008. Impacto de labranza e inoculación micorrízica arbuscular sobre la pudrición carbonosa y rendimiento de maíz z en condiciones semiáridas. *Rev. Fitotec. Mex.* 31: 257-263.
- Eguiarte, L. E., Equihua Zamora, C., y Espinosa Asuar, L. (2017). La milpa es un espejo de la diversidad biológica y cultural de México. *Revista Oikos*, 17, 7-9.
- Escobar Moreno, D. A. (2006). Valoración campesina de la diversidad del maíz estudio de caso de dos comunidades indígenas en Oaxaca, México: tesis doctoral. Universitat Autònoma de Barcelona.
- Espinosa Trujillo, E., del Carmen Mendoza Castillo, M., y Castillo González, F. (2006). Diversidad fenotípica entre poblaciones de maíz con diferentes grados de pigmentación. *Revista Fitotecnia Mexicana*, 29(2), 19-23.
- Giovannetti, M. y Mosse, B. (1980) 'An evaluation of techniques for measuring vesicular arbuscular mycorrhizal infection in roots', *New Phytologist*, vol. 84.
- Gómez-Rodríguez, O., y Zavaleta-Mejía, E. (2001). La asociación de cultivos una estrategia más para el manejo de enfermedades, en particular con *Tagetes* spp. *Rev. Mex. Fitopatol*, 19, 94-99.
- INEGI. (2004). Guía para la interpretación de la cartografía: edafología.
- Macías García. (2017). La milpa. *Revista Oikos*. 17, 5.
- Marcos, B., Martínez, A., López, G., López, C. y Arteaga, T. (2016). La biomasa de los sistemas productivos de maíz nativo (*Zea mays*) como alternativa a la captura de carbono. *Rev. Int. Contam. Ambie*, 32(3), 361-367
- Mora, A. N., y Leblanc, H. (2012) Evaluación del uso de micorrizas arbusculares para disminuir la aplicación de fertilizantes fosforados en el cultivo del maíz. *Tierra Tropical*, 8(2): 245-255.
- Oliveira C., Sá N., Gomes E., Marriel I., Scotti M., Guimarães C., Schaffert R. y Alves M. (2009) Assessment of the mycorrhizal community in the rhizosphere of maize (*Zea mays* L.) genotypes contrasting for phosphorus efficiency in the acid savannas of Brazil using denaturing gradient gel electrophoresis (DGGE). *Applied Soil Ecology* 41: 249-258.
- Ortas, I. (2012). The effect of mycorrhizal fungal inoculation on plant yield, nutrient uptake and inoculation effectiveness under long-term field conditions. *Field crops research*, 125, 35-48.
- Pérez Botho, B., Jiménez Velázquez, M. A., Sánchez Escudero, J., García Cué, J. L., y Muratalla Lúa, A. (2015). Agricultura tradicional en El Botho, Alto Mezquital, estado de Hidalgo. *Revista mexicana de ciencias agrícolas*, 6(6), 1215-1227.
- Philips, J. M. y Hayman, D. S. (1970) 'Improved procedures for clearing roots and staining parasitic and vesicular-arbuscular mycorrhizal fungi for rapid assessment of infection', *Transactions of the British mycological Society*, vol. 55.
- Purroy Vásquez, R., Gallardo López, F., Díaz Rivera, P., Ortega Jiménez, E., López Ortiz, S., y Torres Hernández, G. (2016). Flujo energético-económico como herramienta para tipificar

- agroecosistemas en el centro del estado de Veracruz, México. *Ecosistemas y recursos agropecuarios*, 3(7), 91-101.
- Reyes-Tena, A., López-Pérez, L., Quiñones-Aguilar, E. E., & Rincón-Enríquez, G. (2015). Evaluación de consorcios micorrícicos arbusculares en el crecimiento vegetal de plantas de maíz, chile y frijol. *Rev. Biol*, 17, 35-42.
- Robledo, M. T., Zavala, J. J. G., Lugo, H. J. A., Ortiz, R. L., Montiel, N. O. G., Macías, M. S., y Fernández, A. T. (2017). Biofertilización en híbridos de maíz androestériles y fértiles para los valles altos de México. *Revista Terra Latinoamericana*, 35(1), 65-72.
- Sánchez de la Cruz, R., A. Díaz-Franco, V. Pecina-Quintero, I. Garza-Cano y J. Loera-Gallardo. 2008. *Glomus intraradices* y *Azospirillum brasilense* en trigo bajo dos regímenes de humedad en el suelo. *Univ. Cienc.* 24: 239-245.
- Sangabriel C.W., Negrete Yankelevich, S., Maldonado Mendoza, I. E., y Trejo Aguilar, D. (2014). Native maize landraces from Los Tuxtlas, Mexico show varying mycorrhizal dependency for P uptake. *Biology and fertility of soils*, 50(2), 405-414.
- Sangabriel Conde W., Maldonado Mendoza I.E., Mancera López M.E., Cordero-Ramírez J.D., Trejo Aguilar D., y Negrete-Yankelevich S. 2015. Glomeromycota associated with Mexican native maize landraces in Los Tuxtlas, Mexico. *Applied Soil Ecology* 87: 63-71.
- Sawers, R. J., Ramírez Flores, M. R., Olalde Portugal, V., y Paszkowski, U. (2018). The impact of domestication and crop improvement on arbuscular mycorrhizal symbiosis in cereals: insights from genetics and genomics. *New Phytologist*, 220(4), 1135-1140.
- SIAP (Servicio de Información Agroalimentaria y Pesquera). 2020. Expectativas agroalimentarias 2021. Ciudad de México. https://www.gob.mx/cms/uploads/attachment/file/633847/Expectativas_2021_04_29_compressed.pdf
- SIAP (Servicio de Información Agroalimentaria y Pesquera). 2021. Resumen nacional por estado: Riego+Temporal, Año agrícola 2019-2020, Maíz grano. http://infosiap.siap.gob.mx:8080/agricola_siap_gobmx/ResumenProducto.do (Consulta: junio 10, 2021).
- Trejo, L., Gil, A., Sánchez, M., Carballo, A., y López, P. A. (2004). Producción de semilla mejorada por organizaciones de agricultores: caso productor de maíz Teocintle. *Revista Fitotecnia Mexicana*, 27(1), 93-100.
- Vargas, C. A. C., y Sicard, T. E. L. (2013). Resiliencia de sistemas agrícolas ecológicos y convencionales frente a la variabilidad climática en Anolaima (Cundinamarca-Colombia). *Agroecología*, 8(1), 21-32.
- Xu, H., Lu, Y., & Tong, S. (2018). Effects of arbuscular mycorrhizal fungi on photosynthesis and chlorophyll fluorescence of maize seedlings under salt stress. *Emirates Journal of Food and Agriculture*, 199-204.

Variables asociadas a la selección de hábitat de tres especies de tortugas Dulceacuícolas de la familia *Podocnemididae* presentes en la reserva natural Bojonawi (Vichada, Colombia), durante la estación seca

Keisy Daniela Parra Henao¹, Zoylo Morales Romero²

¹Facultad de Estadística e Informática, Universidad Veracruzana. kcparrah@gmail.com

²Facultad de Estadística e Informática, Universidad Veracruzana. zmorales@uv.mx

Resumen

Evalúamos el posible uso diferencial de tres macrohábitats (laguna, caño y río) por parte de los podocnemíidos que habitan en la Reserva Natural Bojonawi (Vichada, Colombia) durante la estación seca. Buscamos determinar cuáles características de los cuerpos de agua están asociadas a la selección de hábitat en tres especies de tortugas dulceacuícolas (*P. unifilis*, *P. vogli* y *P. dumerilianus*). Para esto, realizamos un análisis descriptivo, que permitió conocer la distribución y variabilidad de los datos; también se utilizaron pruebas de hipótesis para más de dos muestras para conocer si existían diferencias significativas de las variables ambientales y fisicoquímicas entre macrohábitats. Finalmente, implementamos gráficos Biplot, con la finalidad de identificar agrupaciones entre las variables que brindaran información sobre cuáles permitían diferenciar los macrohábitats de estudio, y así poder asociar la abundancia de cada especie de tortuga con las diferencias entre los sitios.

Palabras claves: Selección de hábitat, tortugas dulceacuícolas, gráficos Biplot, cuerpos de agua.

Introducción

Colombia, junto con Brasil, ocupan el primer puesto en cuanto a riqueza de especies de tortugas en Suramérica (Turtle Taxonomy Working Group 2017; Morales-Betancourt *et al.*, 2020). Sin embargo, el 33% de las especies se encuentran en alguna categoría de amenaza (Morales-Betancourt *et al.*, 2015). En la Orinoquia colombiana se encuentra el 48% de las tortugas continentales registradas para Colombia (Morales-Betancourt *et al.*, 2020). En la parte nororiental de la cuenca en el territorio colombiano, se encuentran algunas áreas protegidas como lo es la Reserva Natural Privada Bojonawi, ubicada en el departamento de Vichada. En ella se pueden encontrar de manera simpátrica varias especies de tortugas dulceacuícolas de la familia Podocnemididae con diferentes niveles de amenaza a nivel nacional, como son *Podocnemis unifilis* (terecay) catalogada como en peligro (EN), *Podocnemis vogli* (galápaga o sabanera) considerada como especie de preocupación menor (LC) y *Peltocephalus dumerilianus* (cabezón) en la categoría de datos insuficientes (DD, Morales-Betancourt *et al.*, 2015).

La familia Podocnemididae es una de las familias de tortugas dulceacuícolas con mayor número de especies amenazadas o en riesgo de extinción, pues constituyen un invaluable

recurso alimenticio y comercial para las comunidades aledañas a los grandes ríos de la Amazonia, Orinoquía, los valles del Magdalena y el Sinú en Colombia, al proveerles carne, huevos, aceite y otros productos de supervivencia (Rueda-Almonacid *et al.*, 2007). Sin embargo, en Colombia sólo hay información disponible sobre la demografía de *Podocnemis lewyana* y un poco sobre *Podocnemis unifilis*, para las cuales se tienen estudios principalmente sobre aspectos reproductivos y de alimentación debido a la mayor importancia social y económica de estas para las comunidades ribereñas del río Magdalena y de las cuencas de la Amazonía y Orinoquía, y son, además, en las que más se han enfocado la mayoría de los esfuerzos de conservación (Páez *et al.*, 2015).

La carencia de información sobre las dinámicas demográficas y ecológicas es uno de los principales factores que deben tenerse en cuenta en los programas de conservación y manejo de las especies de tortugas (Páez *et al.*, 2015). La falta de esta información en el diseño e implementación de planes de manejo puede llevarlos a una pérdida valiosa de tiempo, esfuerzo y dinero, ya que no se enfatiza en acciones que tendrían un impacto más efectivo para la persistencia de las poblaciones naturales de tortugas (Páez *et al.*, 2012). Las tortugas, por ser especies longevas, constituyen un reto particularmente difícil para los programas de conservación, porque requieren de programas de monitoreo a largo plazo, para documentar de forma confiable tendencias demográficas poblacionales que vayan más allá de posibles fluctuaciones naturales en respuesta a la variabilidad natural en las condiciones ambientales y/o en la dinámica misma de la población (Spencer y Jensen, 2010).

Para las grandes tortugas de río de la familia Podocnemididae presentes en Suramérica, se conoce que existe variabilidad en los tipos de hábitats usados por cada una de las especies, en donde al parecer la mayoría usan con diferente frecuencia hábitats de aguas negras, blancas y claras (Páez *et al.*, 2012; Rueda-Almonacid *et al.*, 2007). Por ejemplo, *Peltocephalus dumerilianus*, *Podocnemis erythrocephala* y *Podocnemis lewyana* parecen ser las especies más selectivas de la familia, las dos primeras prefieren aguas negras (Rueda-Almonacid *et al.*, 2007), y la última distribuyéndose únicamente en aguas turbias (Páez *et al.*, 2012; Ferrara *et al.*, 2017). Por su parte, *Podocnemis vogli* y *Podocnemis sextuberculata* pueden estar tanto en aguas blancas como en claras (Rueda-Almonacid *et al.*, 2007; Ferrara *et al.* 2017), y las dos especies más grandes de la familia *Podocnemis expansa* y *Podocnemis unifilis* parecen ser más generalistas y se encuentran en todo tipo de aguas (blancas, claras y negras; Páez *et al.*, 2012; Rueda-Almonacid *et al.*, 2007; Ferrara *et al.*, 2017).

El objetivo de este trabajo fue evaluar cuáles variables ambientales y fisicoquímicas están relacionadas con la selección de hábitat en tres especies de tortugas dulceacuícolas (*P. unifilis*, *P. vogli* y *P. dumerilianus*) presentes en la Reserva Natural Bojonawi (Vichada, Colombia) durante la estación seca. La RNB presenta un hábitat relativamente protegido, con poca presencia antrópica, lo que confiere a este estudio un gran valor no solo por su aporte al conocimiento demográfico y ecológico sobre estas especies en particular para Colombia,

sino porque sirve de referencia para estudios comparables en otros sitios con mayor nivel de intervención.

Metodología

Área de estudio

El estudio se realizó en la Reserva Natural Privada Bojonawi (RNB), ubicada 15 km al suroccidente del municipio de Puerto Carreño (6°05'52,79" N; 67°28'59,58" W) en el departamento de Vichada, Colombia. El muestreo se llevó a cabo durante la estación seca, entre enero y abril del año 2017. La RNB cuenta con un área de 4.680 ha y se encuentra localizada sobre el río Orinoco, limitando con caño Negro y la Reserva Natural de Agua Linda. Está ubicada a una altura de 51 m.s.n.m., con una temperatura promedio anual de 28°C y precipitación promedio anual cercana a los 2.176 mm (Fundación Omacha, 2012).

Métodos de muestreo

Métodos de muestreo

Para este estudio, entre los macro-hábitats presentes en la RNB se seleccionaron tres, los cuales de acuerdo con la clasificación de Lasso *et al.* (2016), se denominan: laguna de inundación (Laguna El Pañuelo), río (Orinoco) y caños (caño Charapa, afluente de la laguna y caño Tesoro, afluente del río Orinoco). En total, cada macro-hábitat fue muestreado durante 3 ciclos (cada ciclo constó de 5 días efectivos al mes), utilizando nueve trampas de doble embudo y seis trampas de un embudo, las cuales fueron ubicadas en la superficie del agua y se distribuyeron aproximadamente a 400 m de distancia entre ellas, en sitios adecuados para su instalación (poco profundos y sin mucha corriente). El número de trampas varió un poco entre los macro-hábitats (entre 14-16 trampas).

Las trampas fueron cebadas para atraer a los individuos utilizando pescado (obtenido de los mismos caños mediante pesca con flecha o arpones artesanales), maíz (*Zea mays*), ambos en descomposición, sandía (*Citrullus lanatus*), plátano (*Musa paradisiaca*) y hojas de chigo (*Campsiandra comosa*), con la finalidad de atraer las tres especies. Las trampas fueron revisadas cada 6 horas, tres veces al día para evitar el ahogamiento de los animales y la acumulación de material biológico diferente, los horarios de revisión fueron establecidos a las 9:00, 15:00 y 21:00 horas, y los cebos eran cambiados día por medio.

Adicionalmente, se implementaron otros dos métodos de muestreo: capturas con malla o red de pesca de ahorque (también llamada en la pesca artesanal como trasmallo) en jornadas diurnas y observación directa, y captura manual por medio de buceo a pulmón libre y braceo en jornadas diurnas y nocturnas. Estos métodos se utilizaron de acuerdo con las características de cada macro-hábitat. El método manual, consistía en la búsqueda directa de las tortugas en el agua, por recorridos de aproximadamente un kilómetro, observando desde

afuera del agua o utilizando careteo dentro del agua para buscarlos. Este método solo se implementó en los caños debido a las características de estos cuerpos de agua, al poseer aguas transparentes y poco profundas. Finalmente, el método de captura con trasmallo sólo pudo ser implementado en el río Orinoco y la laguna El Pañuelo, debido a que se requieren zonas amplias, profundas y sin vegetación o rocas para realizar los lances.

Cada individuo capturado fue medido, sexado y marcado; y posterior a los registros morfométricos, liberado en el mismo sitio de captura. Los individuos se pesaron con una pesola de 20 kg. Las variables morfométricas medidas en cm fueron: (a) longitud recta del caparazón (LRC), medida con calibrador desde la margen anterior del escudo nual (en este caso desde la cisura de los escudos marginales 1 hasta la margen posterior del caparazón en la unión entre los escudos marginales 12, (b) longitud máxima del plastrón (LMP), medida en la línea diagonal desde el margen anterior del plastrón, en la mitad del escudo intergular, hasta el margen posterior del plastrón, en el extremo del escudo anal derecho, (c) ancho del caparazón (AC) medido entre las cisuras de los marginales número 6 y 7 a cada lado del caparazón y (d) longitud pre-cloacal (LPC), medida desde el borde posterior del plastrón hasta la margen anterior de la cloaca (Rueda-Almonacid et al. 2007, Páez et al. 2015).

Para el marcaje de cada animal se realizaron muescas en las escamas marginales del caparazón utilizando la modificación del código de Cagle (1939) propuesto por Rueda-Almonacid *et al.* (2007), donde las escamas del caparazón tienen valores asignados que al ser sumados igualan el número del ID del individuo; las muescas se realizaron utilizando una sierra pequeña. Todos los individuos fueron clasificados según su LRC en tres categorías de tamaño como juveniles, sub-adultos y adultos (Rueda-Almonacid *et al.*, 2007). Los adultos y sub-adultos fueron sexados de acuerdo con el dimorfismo sexual descrito para cada especie, en tamaño y en cuanto a la longitud pre-cloacal (Rueda-Almonacid *et al.*, 2007).

Se realizó una caracterización ambiental de cada uno de los tres macro-hábitats con el fin de asociar diferencias entre las variables ambientales y las abundancias, proporciones sexuales y clases de tamaño observadas para cada sitio. Fueron evaluadas las siguientes variables: (a) temperatura del agua, (b) pH del agua, (c) concentración de oxígeno disuelto (mg/L), (d) sólidos disueltos totales (ppm) y (e) conductividad ($\mu\text{S}/\text{cm}$), medidas utilizando multiparámetro y oxímetro; (f) transparencia del agua, (g) ancho, (h) velocidad de la corriente y (i) cobertura vegetal. Estas variables fueron medidas en los puntos donde se ubicaron las trampas.

Análisis estadístico

Se obtuvo una base de datos de las variables fisicoquímicas, una base para las variables ambientales y una base para las características morfológicas de las tortugas. Para cada base de datos se realizó un análisis descriptivo, que permitió conocer la distribución y variabilidad de los datos. Para este análisis se usaron tablas de frecuencia y tablas de doble entrada,

gráficos de mosaico, gráficos de cajas y bigotes y se utilizaron estadísticas descriptivas como el promedio.

Con la finalidad de conocer si existen diferencias significativas de las variables ambientales y fisicoquímicas se utilizaron pruebas de hipótesis para más de dos muestras. Para realizar las pruebas de hipótesis, en primer lugar es básico saber si las muestras provienen de una distribución normal y, además, presentan igualdad de varianzas; para ello se comprobó el supuesto de homogeneidad de varianzas mediante la prueba de Levene y el supuesto de normalidad para las variables fisicoquímicas mediante la prueba de Shapiro-Wilk, ya que se trataba de una muestra pequeña y para las variables ambientales, se utilizó la prueba de Kolmogorov-Smirnov, debido a que se tienen más de 600 mediciones. Como los datos no cumplieron con los supuestos de normalidad y homocedasticidad ($p < 0.05$), se optó por utilizar métodos no paramétricos para las pruebas de hipótesis. Se utilizó la prueba de Kruskal-Wallis con un nivel de significancia del 5%, debido a que se cuenta con más de dos muestras para las comparaciones (tres macro-hábitats), adicionalmente se utilizó un análisis post-hoc (Prueba de rangos de Wilcoxon) para poder visualizar entre cuáles macro-hábitats existían diferencias significativas.

Adicionalmente, se realizaron gráficos Biplot, con la finalidad de reducir dimensionalidad y, a su vez, buscar agrupaciones entre las variables que puedan brindar información importante. También se utilizaron los gráficos Biplot para describir ambientalmente cada macrohábitat. Con la finalidad de homogeneizar la dimensión de las variables fisicoquímicas y ambientales, y la abundancia de las tortugas, se utilizó el método de remuestreo Bootstrap, este método permitió calcular las medias en cada grupo de variables, y así tener un nuevo conjunto de observaciones que corresponden a las medias de cada una de las variables, Con este método se busca reducir el sesgo debido al muestreo en los datos obtenidos y obtener muestras de manera aleatoria de la muestra original. Su principal uso es hallar una aproximación a la distribución de la variable analizada (Librero, 2015). Con lo anterior se construyó un HJ – Biplot de las medias de todas las bases de datos ambientales de los cuerpos de agua. Con el objetivo de concluir si existe relación entre las variables fisicoquímicas y ambientales, y la abundancia de las tortugas en los tres macrohábitats. Todos los análisis se llevaron a cabo con el software R.

Resultados

El número total de tortugas capturadas durante los tres meses de estudio fue de 216 individuos, de las cuales la especie *Podocnemis vogli* fue la más abundante capturando un total de 164 individuos, seguido de *Peltocephalus dumerilianus* la cual estuvo representada por 28 individuos y *Podocnemis unifilis* con 24 individuos capturados. *P. vogli* estuvo presente en los tres macrohábitats siendo más abundante en la Laguna (84 individuos) que en el Caño (52 individuos) y el Río (28 individuos). *P. dumerilianus* se encontró en su

mayoría en el Caño (27 individuos) y sólo un individuo en la Laguna. Por su parte, *P. unifilis* fue encontrada en la Laguna (20 individuos) y en el Río (4 individuos). Encontramos que el macrohábitat que presentó mayor número de capturas de tortugas fue la Laguna (105 individuos), seguida del Caño (19 individuos) y el Río (32 individuos) (Tabla 1).

Tabla 1. Frecuencia de absoluta de individuos capturados para cada especie y por cada macrohábitat.

Especie / Sitio	Caño	Laguna	Río	Total general
<i>Peltocephalus dumerilianus</i>	27	1	0	28
<i>Podocnemis unifilis</i>	0	20	4	24
<i>Podocnemis vogli</i>	52	84	28	164
Total general	79	105	32	216

En cuanto a la estructura poblacional, encontramos que para la especie *P. dumerilianus* la categoría de tamaño más capturada fueron los adultos; para *P. unifilis* adultos y sub-adultos fueron los tamaños más abundantes. En el caso de *P. vogli* se capturaron mayoritariamente adultos (Figura 1). En cuando al sexo de los individuos adultos para la especie *P. dumerilianus* el sexo más abundante fueron los machos, para *P. Unifilis*, todos los individuos capturados fueron hembras o juveniles de sexo indefinido, no se capturaron machos. En el caso de *P. vogli* las proporciones sexuales fueron similares entre machos, hembras y juveniles de sexo indefinido (Figura 2).

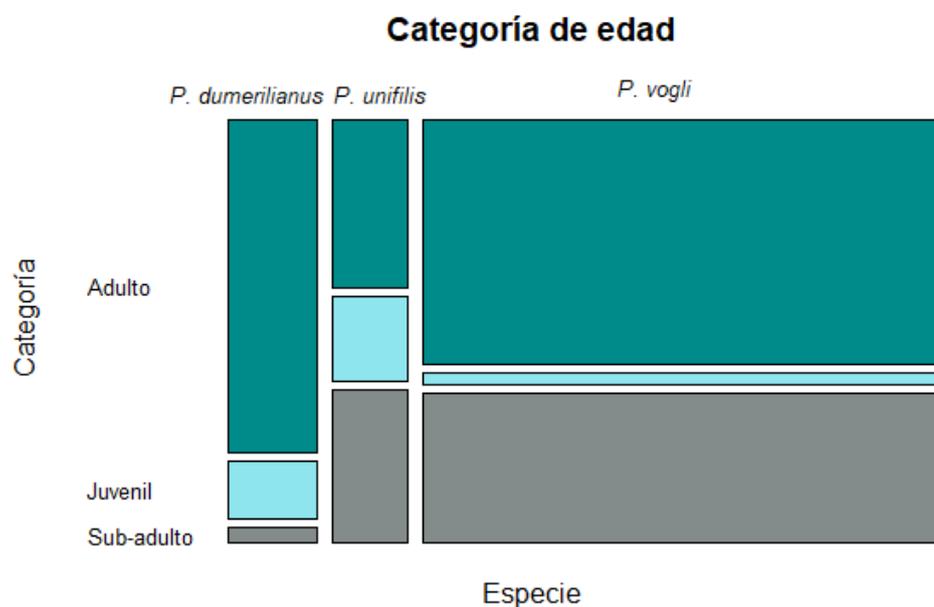


Figura 1. Gráfico de Mosaico para las categorías de edad en las tres especies de tortugas.

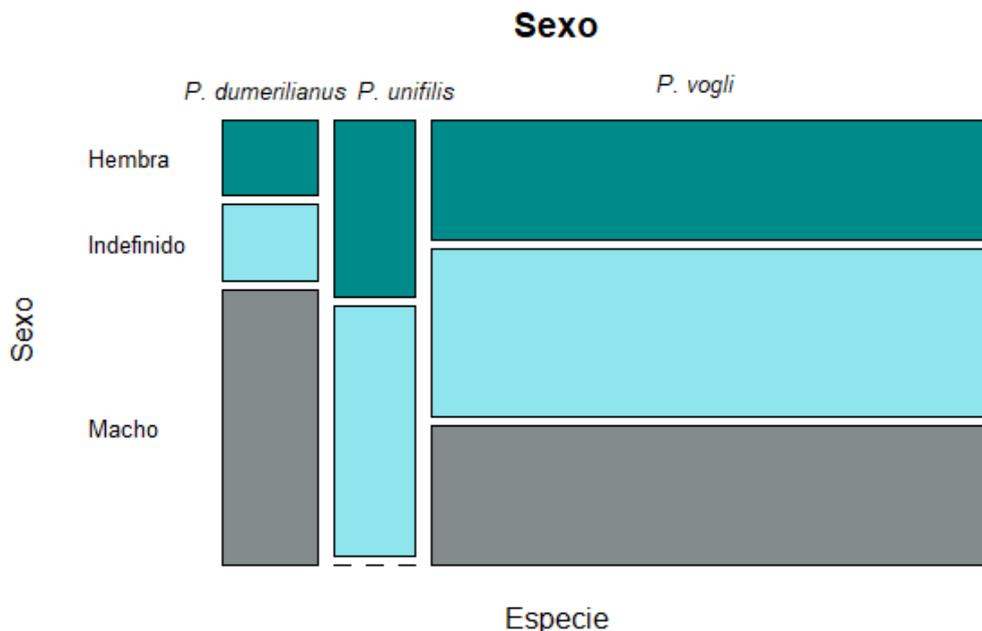


Figura 2. Gráfico de Mosaico para el sexo en las tres especies de tortugas.

De acuerdo con las variables ambientales medidas (Figura 3), se observa que el Caño es un macrohábitat que se caracterizó por tener el nivel más alto de transparencia (1.2 m en promedio) en comparación con la Laguna y el Río. El sitio con mayor porcentaje de cobertura vegetal fue el Caño, seguido por la Laguna y el Río. Siendo en promedio de 40% en el Río, 62% en la Laguna y 92% en el Caño. Al ser el Río y la Laguna cuerpos de agua con mayor cantidad de sedimentos, la transparencia en estos sitios fue mucho menor que la encontrada en el Caño.

Las mayores profundidades se encontraron en el Río (hasta 2.68 m). Por otro lado, la velocidad de la corriente fue menor en la Laguna (0.024 m/s en promedio), en comparación a los otros dos macrohábitats. La velocidad de la corriente presentó un gran número de valores atípicos en los tres sitios, dando indicios de que su medición pudo estar sesgada, posiblemente por el viento al ser una medición superficial.

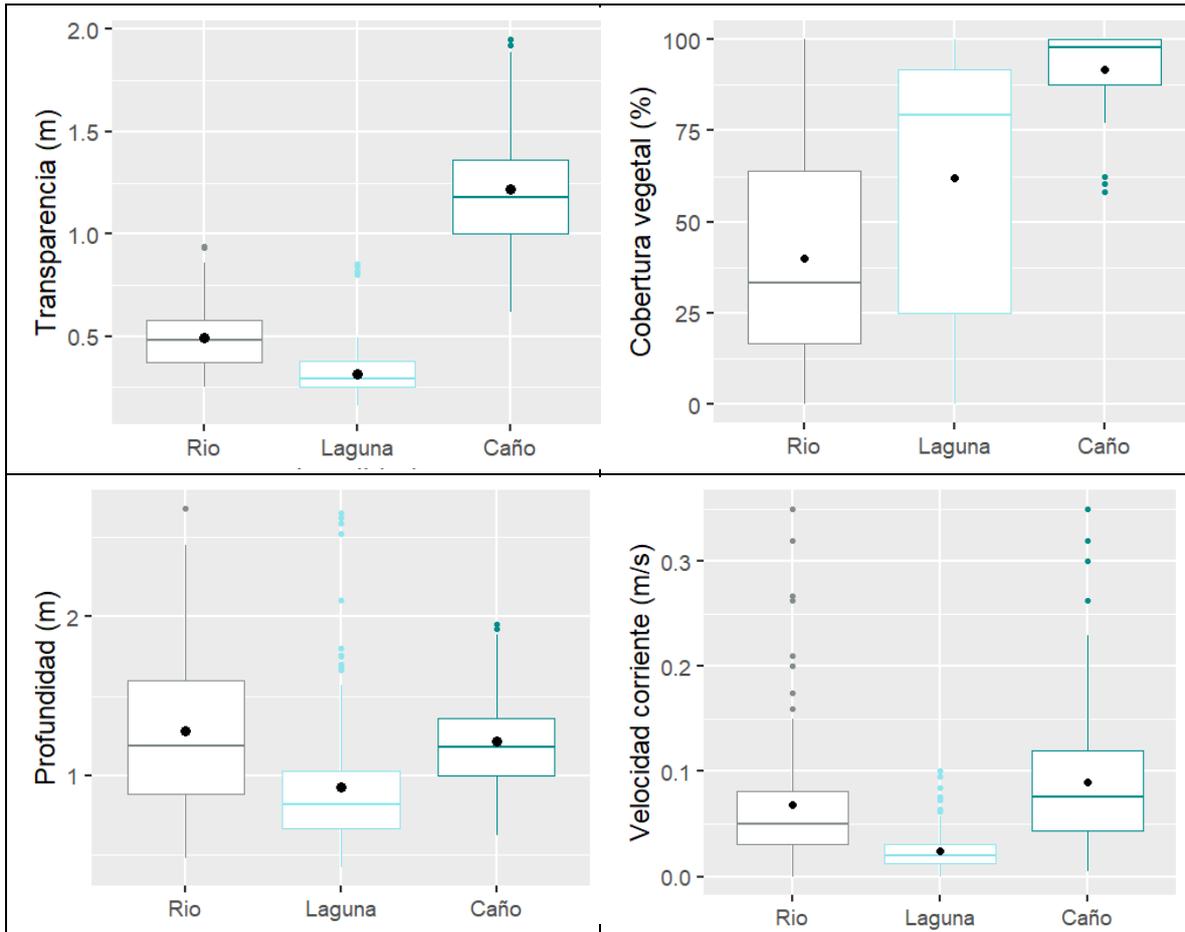


Figura 3. Gráfico de cajas y bigotes de las variables ambientales medidas en los tres macrohábitats.

En cuanto a las variables fisicoquímicas, el pH, la conductividad y los sólidos disueltos totales (TSD) siguen una misma tendencia siendo los valores más bajos en el Caño, seguidos de la Laguna y el Río con los valores más altos. Teniendo el Caño un pH principalmente ácido, mientras la Laguna y el Río un pH básico. En cuanto a la temperatura del agua, la Laguna y el Río tiene valores similares (en promedio 30.5°C y 30.6°C respectivamente) y más altos que en el Caño (en promedio 29.4°C). Finalmente, respecto al Oxígeno Disuelto, en la Laguna se obtuvieron los valores más bajos (en promedio 5.4 mg/L) (Figura 4).

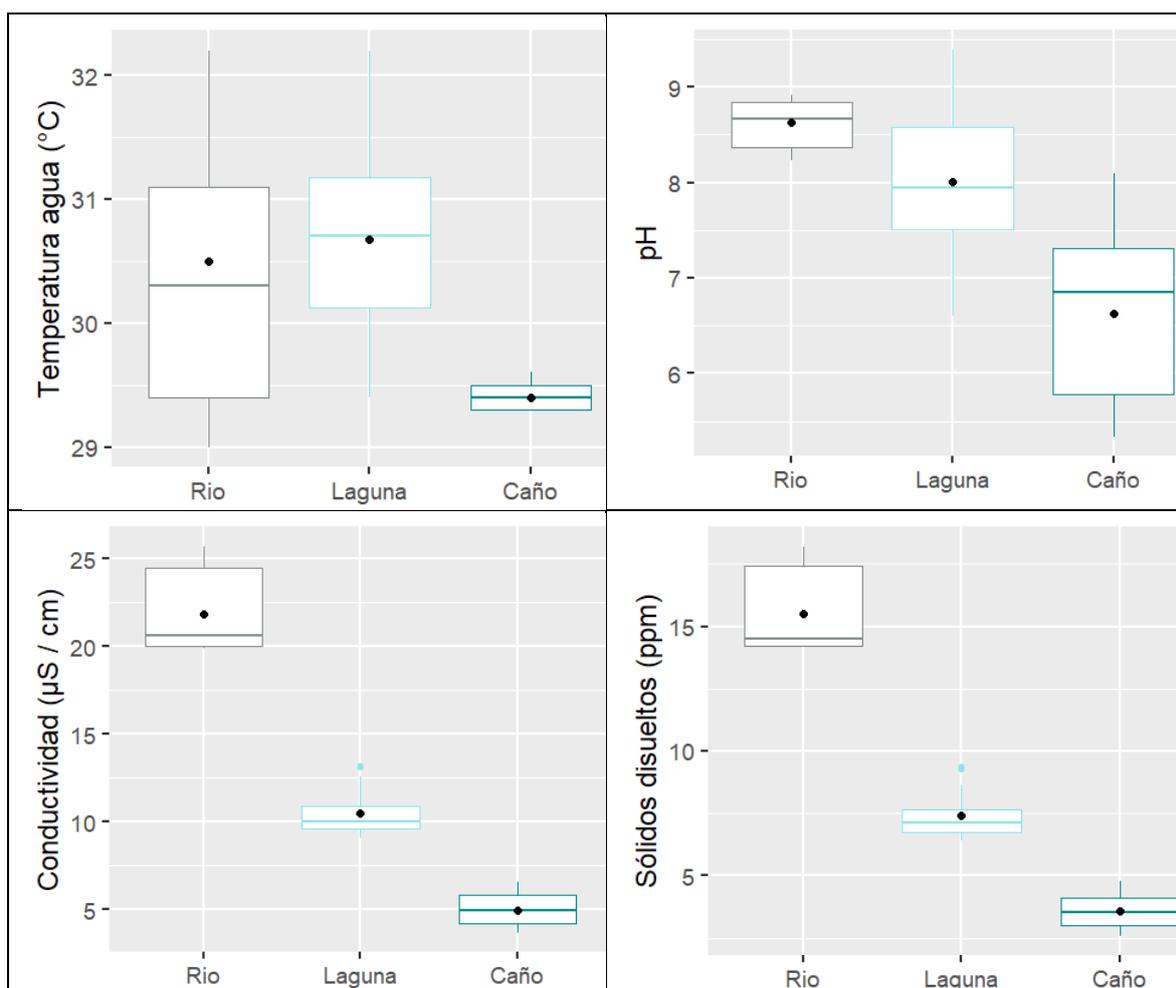


Figura 4. Gráfico de cajas y bigotes de las variables fisicoquímicas medidas en los tres macrohábitats.

Con la prueba de Kruskal-Wallis encontramos que todas las variables tanto ambientales y fisicoquímicas presentaron diferencias significativas ($p < 0.05$) entre los tres macrohábitats. Únicamente la profundidad no presentó diferencias significativas entre el Caño y el Río y la Temperatura no presentó diferencias significativas entre la Laguna y el Río y el Caño y el Río (Figura 5).

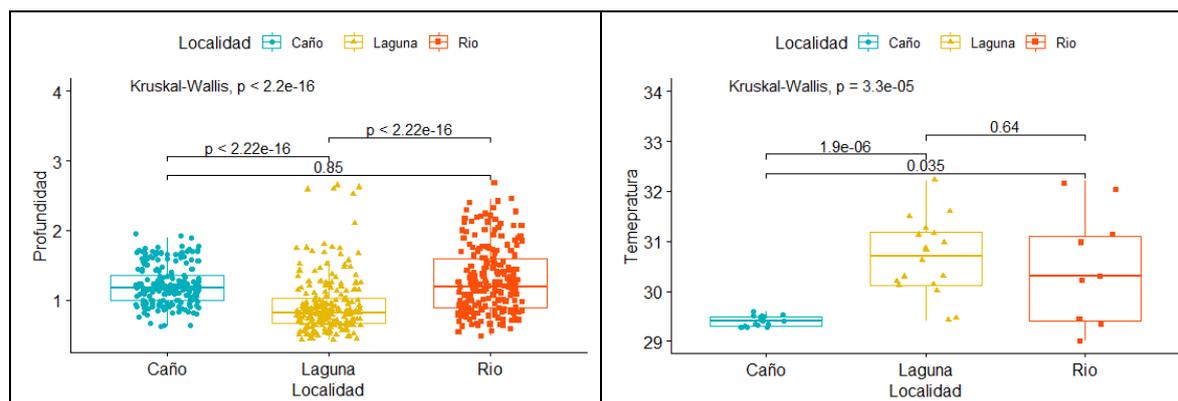


Figura 5. Resultados pruebas de hipótesis.

Con el análisis H_j-Biplot para cada macrohábitat se pueden evidenciar las correlaciones entre las variables ambientales y fisicoquímicas para cada sitio de estudio. Encontramos que para la Laguna se evidencia una correlación positiva fuerte para la Transparencia y el porcentaje de Cobertura Vegetal, mismas variables que se correlacionan de forma negativa con la Velocidad de la corriente. Igualmente, se puede evidenciar que las variables ambientales no se correlacionan con las variables fisicoquímicas en tres macro-hábitats (Figura 6).

En el caso del Caño, encontramos igualmente que las variables ambientales no se correlacionan con las variables fisicoquímicas. En este caso, la Transparencia y la Profundidad del cuerpo de agua se correlacionan de forma negativa con la Cobertura Vegetal y la Velocidad de la corriente. Todas las variables fisicoquímicas están altamente correlacionadas entre ellas. En este caso, en la dimensión 1 que es la que explica el 36% de la variación, la Conductividad y los Sólidos Disueltos Totales (TSD) son las variables con mayor peso (Figura 7).

En el caso del Río, encontramos que la Transparencia, la Velocidad y la Profundidad se correlacionan de forma positiva entre ellas, y de forma negativa con la Cobertura Vegetal. En este caso, el pH y la Temperatura están correlacionados de forma positiva entre ellos y de forma negativa con el TSD y la Conductividad. Igualmente no existen correlaciones entre las variables ambientales y fisicoquímicas. En este caso, en la dimensión 1 que es la que explica el 38% de la variación, también ocurre que la Conductividad y los Sólidos Disueltos Totales (TSD) son las variables con mayor peso (Figura 8).

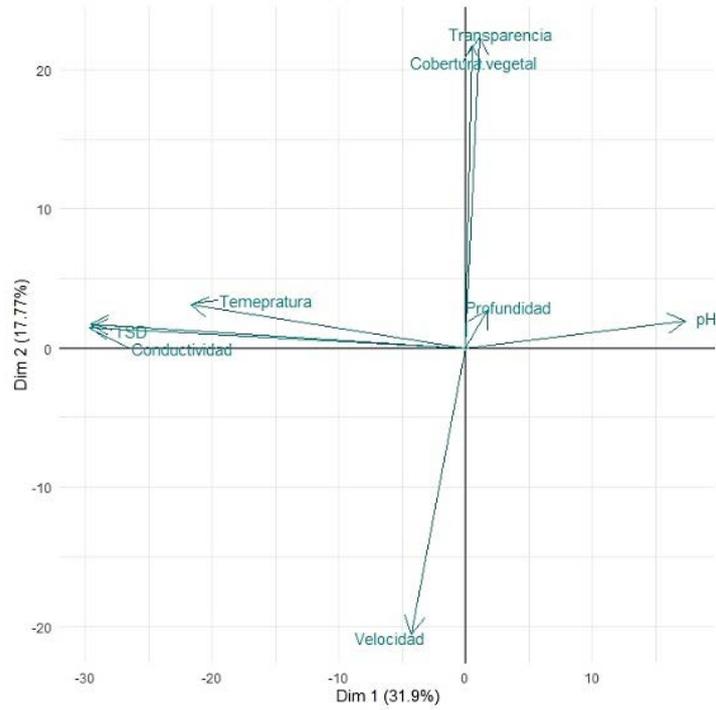


Figura 6. HJ-Biplot de las variables ambientales y fisicoquímicas para la Laguna.

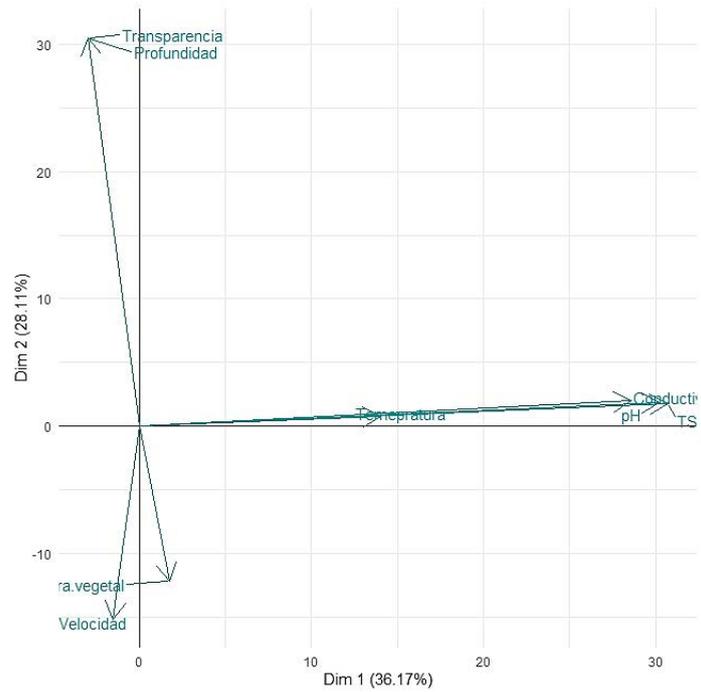


Figura 7. HJ-Biplot de las variables ambientales y fisicoquímicas para el Caño.

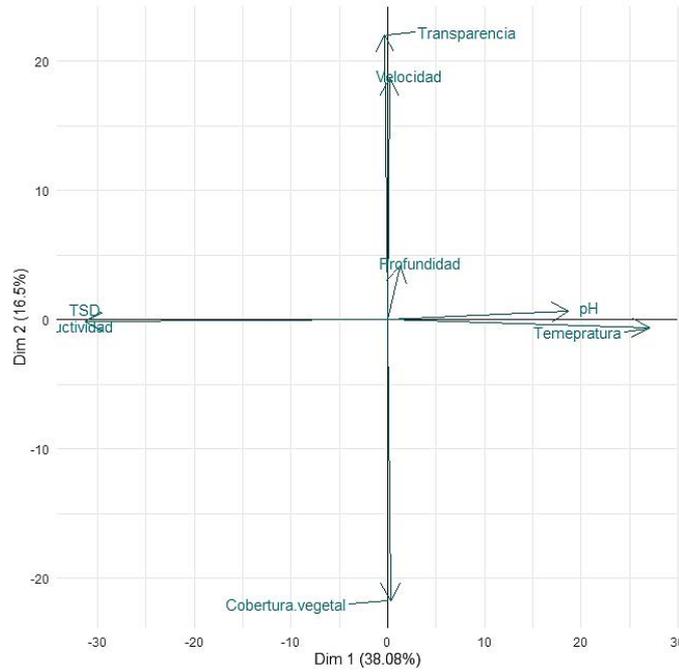


Figura 8. HJ-Biplot de las variables ambientales y fisicoquímicas para el Río.

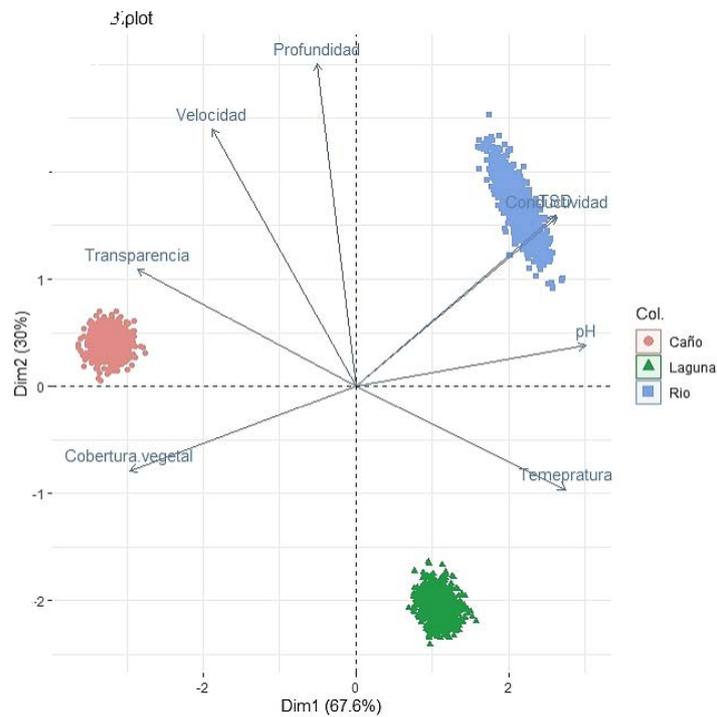


Figura 9. HJ-Biplot colapsado de las variables ambientales y fisicoquímicas para los tres macrohábitats.

Con el análisis Biplot colapsado (Figura 9), se observa que hay una segregación clara de los macro-hábitats en cuanto a las variables ambientales y fisicoquímicas. Se evidencia que para los caños las variables cobertura vegetal y transparencia del agua son las que más lo

diferencian de los otros macro-hábitats. A pesar de que todas las variables tienen un gran peso, la conductividad y la cobertura vegetal (Dim 1), son las variables responsables del 68% de la variación total entre los tres macro-hábitats.

Conclusiones y discusión

En este estudio los macro-hábitats fueron una laguna formada por inundación del río Orinoco, y dos caños que desembocaban, uno en la laguna y otro en el río, así que no fue sorpresa que entre ellos las variables ambientales fueran muy similares entre la Laguna y el Río, debido a que es un sistema interconectado durante la estación de lluvias o aguas altas. Es notable que los caños poseen características diferentes respecto a los otros macro-hábitats, al poseer aguas claras y, además, tener menores profundidades. Varios estudios han señalado que variables como velocidad de la corriente, cobertura vegetal, oxígeno disuelto y profundidad del agua influyen en la presencia y abundancia de los Podocnemídidos (González-Zárate *et al.* 2011, González-Zárate *et al.* 2014, Páez *et al.* 2015). En nuestro caso, las variables cobertura vegetal y transparencia del agua son las que más lo diferencian de los otros macro-hábitats.

P. vogli, de acuerdo con la literatura, posee una preferencia de hábitat por los cuerpos de agua someros, con poca corriente o sin corriente, con aguas turbias y fondos arcillosos (Rueda-Almonacid *et al.* 2007), tal como se presenta en los cuerpos de agua de la Reserva donde se encontró en mayores cantidades (Laguna y Caño). Sin embargo, fue la única especie que se encontró en todos los cuerpos de agua de la reserva. El uso de diferentes tipos de macrohábitats ya ha sido indicado (Pinzón *et al.* 2017), dado a que se adapta a los cambios bruscos ambientales (secos e inundados), lo cual le permite alcanzar densidades poblacionales altas en casi todos los ecosistemas que habita. En este trabajo se capturaron individuos de esta especie con frecuencia en los tres macrohábitats, indicando que dentro de la reserva esta especie se comporta de manera generalista. *P. vogli* es una especie excelente nadadora y tiene un caparazón plano e hidrodinámico y patas completamente palmeadas (Rueda-Almonacid *et al.* 2007, Morales-Betancourt *et al.*, 2012), lo que les permite tener mayores desplazamientos y habitar en diversos sistemas acuáticos.

P. dumerilianus se encontró en su gran mayoría en el macrohábitat Caño, esta especie se conoce por utilizar las solapas (pequeñas cuevas), zonas con poca corriente y los troncos sumergidos de los caños para refugiarse (Rueda-Almonacid *et al.* 2007). La mayor abundancia encontrada en los caños de aguas claras, puede asociarse al hecho de que estas tortugas parecen no ser buenas nadadoras debido al escaso desarrollo de membranas interdigitales y poseen un caparazón en domo lo que limita su capacidad de natación, por lo que, a pesar de ser acuáticas, se encuentran más asociadas a pequeños cuerpos de agua (Rueda-Almonacid *et al.* 2007; Rivas *et al.*, 2016) y pasan la mayor parte del tiempo caminando en el lecho de los ríos y asociadas al sustrato en zonas con poca o ninguna corriente (Medem, 1983). Este hábito bentónico es confirmado en este trabajo por las

observaciones subacuáticas. En este estudio no se capturó ningún individuo de esta especie en el río Orinoco. Con relación al uso de hábitat, Pritchard y Trebbau (1984), reportaron que el hábitat preferencial de estas tortugas son los ríos, igapós (bosques amazónicos inundados por aguas negras) y lagunas de aguas negras. Sin embargo, afirman que pueden ser encontradas, en menor número, en ríos de aguas blancas y claras. Según algunos autores *P. dumerilianus* habita principalmente en sistemas de aguas negras y ocasionalmente en blancas y claras (De La Ossa *et al.*, 2012). Sin embargo, al interior de la reserva, al menos durante la época seca, esta especie tendió a ser especialista, ya que la mayoría de los individuos fueron capturados en los caños de aguas claras.

P. unifilis ha sido reportada como una especie generalista, pues no tiene preferencias por algún tipo de agua y se asocia a diferentes ecosistemas acuáticos (Rueda-Almonacid *et al.*, 2007; Ferrara *et al.*, 2017), en este estudio no se detectó en los caños de aguas claras, pero si se observó que durante la época seca se asocia a grandes sistemas de aguas blancas como el Río y la Laguna.

Referencias

- De La Ossa, J. R. Bernhard y A. De La Ossa-Lacayo. 2012. *Platemys platycephala*. Pp. 261-265. En: Páez, V. P., M. A. Morales-Betancourt, C. A. Lasso, O. V. Castaño-Mora y B. C. Bock. (Eds.), V. Biología y conservación de las tortugas continentales de Colombia. Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia. Instituto de Investigación de los Recursos Biológicos Alexander von Humboldt (IAvH). Bogotá, D. C., Colombia.
- De La Ossa, J. V. y R. C. Vogt. 2011. Ecología poblacional de *Peltecephalus dumerilianus* (Testudines, Podocnemididae) em dois tributários do Rio Negro, Amazonas, Brasil. *Interciencia* 36: 53-58.
- Escalona, T., K. Conway-Gómez, M. A. Morales-Betancourt, F. Arbeláez y R. Antelo. 2012. *Podocnemis unifilis*. Pp. 387-397. En: Páez, V. P., M. A. Morales-Betancourt, C. A. Lasso, O. V. Castaño-Mora y B. C. Bock (Eds.), V. Biología y conservación de las tortugas conti-nentales de Colombia. Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH). Bogotá, D. C., Colombia.
- Ferrara, C. R., C. K. Fagundes, T. Q. Morcatty y R. C. Vogt. 2017. *Quelônios Amazônicos: Guia de identificação e distribuição*. Wildlife Conservation Society. Manaus, Brazil. 182 pp.
- Ferrara, C. R., Fagundes, C. K., Morcatty, T. Q., & Vogt, R. C. (2017). *Quelônios Amazônicos Guia de identificação e distribuição*. *Manaus, Brazil: Wildlife Conservation Society Brasil*.
- Forero-Medina, G., Páez, V. P., Garcés-Restrepo, M. F., Carr, J. L., Giraldo, A., & Vargas-Ramírez, M. (2016). Research and conservation priorities for tortoises and freshwater turtles of Colombia. *Tropical Conservation Science*, 9(4), 1940082916673708.

- Fundación Omacha. 2012. Reserva Natural Bojonawi. Available at: <http://www.omacha.org/2012-04-23-17-01-17/2012-05-23-17-21-05/reserva-natural-bojonawi>. Cited: 6 Jun 2017.
- González-Zárate, A., Montenegro, O. L., & Castaño-Mora, O. V. (2011). Habitat characterization of the river turtle *Podocnemis lewyana* in the Prado river, downstream of Hidroprado dam, Tolima, Colombia. *Caldasia*, 471-493.
- González-Zárate, A., Montenegro, O., Castaño-Mora, O. V., & Vargas-Ramírez, M. (2014). Abundancia, estructura poblacional y conservación de *Podocnemis lewyana* (Podocnemididae) en El Rio Prado, Colombia. *Acta Biológica Colombiana*, 19(3), 351-361.
- Lasso, C. A., Machado-Allison, A., & Taphorn, D. C. (2016). Fishes and aquatic habitats of the Orinoco River Basin: diversity and conservation. *Journal of fish biology*, 89(1), 174-191.
- Morales-Betancourt, M. A., C. A. Lasso, S. Bernal-Sierra, A. M. Sepúlveda-Seguro, K. D. Parra-Henao, M. T. Gómez-Rincón, V. P. Páez, B. Castañeda y B. Marín. 2020. Historia natural de las tortugas acuáticas de la Reserva Natural Bojonawi (Escudo Guayanés), Orinoquia, Vichada, Colombia. Pp. 431-453. En: Lasso, C. A., F. Trujillo y M. A. Morales-Betancourt (Eds.), VIII. Biodiversidad de la Reserva Natural Bojonawi, Vichada, Colombia: río Orinoco y planicie de inundación. Serie Editorial Fauna Silvestre Neotropical. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. Bogotá, D. C., Colombia.
- Morales-Betancourt, M. A., Lasso, C. A., Páez, V. P., & Bock, B. C. (2015). Libro rojo de reptiles de Colombia (2015) Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH), Universidad de Antioquia. Bogotá, DC, Colombia.
- Morales-Betancourt, M. A., Lasso, C. A., y Páez, V. P. 2012. Pp. 266-268. En: Páez, V. P., M. A. Morales-Betancourt, C. A. Lasso, O. V. Castaño-Mora y B. C. Bock. (Eds.), V. Biología y conservación de las tortugas continentales de Colombia. Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia. Instituto de Investigación de los Recursos Biológicos Alexander von Humboldt (IAvH). Bogotá, D. C., Colombia.
- Páez, V. P., Bock, B. C., Espinal-García, P. A., Rendón-Valencia, B. H., Alzate-Estrada, D., Cartagena-Otálvaro, V. M., & Heppell, S. S. (2015). Life history and demographic characteristics of the Magdalena River turtle (*Podocnemis lewyana*): implications for management. *Copeia*, 103(4), 1058-1074.
- Páez, V. P., Morales-Betancourt, M. A., Lasso, C. A., Castaño Mora, O. V., & Bock, B. C. (2012). V. Biología y conservación de las tortugas continentales de Colombia. Serie Recursos Hidrobiológicos y Pesqueros Continentales de Colombia. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Pinzón, M., C. Durán-Prieto, E. Izquierdo y P. Chaves-Hernández. 2017. Aves y tortugas estudiadas en el oleoducto de los Llanos Orientales. ODL S. A., Fundación Omacha. Bogotá, D. C., Colombia. 68 pp.
- Pritchard, P. C. H. y P. Trebbau. 1984. The Turtles of Venezuela. Oxford, Ohio. Society for the Study of Amphibians and Reptiles. 414 pp.
- Rivas, G.A., Barros, T.R., Hernández, O.E. (2016): *Peltocephalus dumerilianus* (Schweiger, 1812). In: Venezuela y sus Tortugas, p. 38–40. Trebbau, P., Pritchard P.C.H., Eds., Caracas, Venezuela, Oscar Todmann Editors.

- Rueda-Almonacid, J. V., Carr, J. L., Mittermeier, R. A., Rodríguez-Mahecha, J. V., Mast, R. B., Vogt, R. C., ... & Mittermeier, C. G. (2007). Las tortugas y los cocodrilianos de los países andinos del trópico. *Serie de guías tropicales de campo*, 6, 412-423.
- Spencer, R. J., & Janzen, F. J. (2010). Demographic consequences of adaptive growth and the ramifications for conservation of long-lived organisms. *Biological Conservation*, 143(9), 1951-1959.
- Turtle taxonomy Workin Group [Rhodin, A. G., Iverson, J. B., Bour, R., Fritz, U., Georges, A., Shaffer, H. B., & Van Dijk, P. P.]. 2017. Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (8th Ed.). In: Rhodin, A.G.J., Iverson, J.B., van Dijk, P.P., Saumure, R.A., Buhlmann, K.A., Pritchard, P.C.H., and Mittermeier, R.A. (Eds.). *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group*. Chelonian Research Monographs 7:1–292.

Relación entre algunos caracteres morfológicos y fisiológicos de odonatos y sus patrones de distribución geográfica

Juliana Sandoval Hernández¹, María Yesenia Zavaleta Sánchez², Cecilia Cruz López³

¹ Facultad de Estadística e Informática, Universidad Veracruzana, zS20000592@estudiantes.uv.mx

² Facultad de Estadística e Informática, Universidad Veracruzana, yzavaleta@uv.mx

³ Facultad de Estadística e Informática, Universidad Veracruzana, ceccruz@uv.mx

Resumen

Este proyecto, utilizando como modelo a las libélulas (Odonatos), se enfocó en una de las principales preguntas de la ecología: ¿Por qué las especies se encuentran en determinadas regiones? Se utilizó un análisis de componentes principales para explorar si los tamaños de los rangos de distribución (altitudinal, latitudinal, y el tamaño de distribución geográfica) de trece especies de libélulas tiene alguna relación con algunas características morfológicas y fisiológicas de los individuos. Este análisis resultó útil para entender que los caracteres más influyentes son la temperatura máxima, la longitud del ala y la masa de las libélulas, y nos lleva a afirmar que este método es bastante útil para detectar variables importantes en la distribución de estos organismos.

Palabras Clave: Componentes Principales, Odonata, Estadística Exploratoria, Libélulas, Biogeografía.

Introducción

Entender por qué las especies se encuentran en determinados lugares es una de las preguntas básicas de la biogeografía (Cox y Moore, 1993; Lomolino et al., 2010), y no por ser básica es la mejor comprendida. Para entender los patrones de distribución de las especies se han propuesto importantes reglas ecogeográficas relacionando variables como la diversidad, el tamaño del rango geográfico ocupado, la latitud y el tamaño corporal (Bergmann, 1847; Hawkins, 2001; Pintor et al., 2015; Stevens, 1992). Sabemos que la distribución de las especies está limitada por factores históricos, ecológicos y geográficos (Sandoval-H et al., 2017; Wiens y Donoghue, 2004). Entre los factores ecológicos más importantes se encuentran los ambientales, y es de aquí de donde surge la idea de Grinnell (1917), donde la presencia de una especie en cierto lugar dependería de la disponibilidad de condiciones ambientales adecuadas para el desarrollo y mantenimiento de las poblaciones.

Las condiciones ambientales cambian gradualmente con la latitud, la elevación y la estacionalidad, lo que hace que la variación de estas condiciones sea mucho menor en los trópicos a lo largo del año (Janzen, 1967; McCain y Grytnes, 2010). Lo anterior implica que los organismos que habitan en zonas más alejadas al trópico o que están distribuidas en amplios rangos de elevación presentarían adaptaciones para tolerar rangos de temperatura más amplios que los de las especies tropicales o de las que habitan rangos estrechos en los gradientes de elevación (Chavez et al., 2015; Ghalambor, 2006; Janzen, 1967; Shah et al., 2017; Suhling et al., 2015; Watson et al., 2019). Estos patrones ambientales han sido usados para explicar patrones de diversidad y distribución de los organismos con éxito, sin embargo,

la mayoría de pruebas a esta hipótesis se han hecho en grupos de animales vertebrados y todavía sigue sin entenderse si estas reglas operan en otros grupos como los insectos, en donde algunos estudios evidencian la importancia de estudiar la fisiología de las especies para entender sus patrones de distribución (García-Robledo et al., 2016; Hartley et al., 2010; Polato et al., 2018).

Otro factor que puede afectar la distribución de los organismos en el planeta es su capacidad de dispersión (Lomolino et al., 2010; Nathan, 2001), pues existen barreras geográficas que limitan el movimiento. La mayor parte de estudios en este campo se han hecho en plantas (Agnarsson et al., 2014; Hewitt y Kellman, 2002), en donde se han encontrado una gran variedad de estrategias morfológicas y fisiológicas que les permiten alcanzar amplios rangos de distribución geográfica. Dentro de los animales son las aves las que se han llevado la atención, y se atribuye su amplia capacidad de dispersión principalmente a algunas características morfológicas relacionadas con el vuelo (Böhning-Gaese et al., 2006; Claramunt et al., 2012; Fritz et al., 2012). Los insectos son otro grupo de gran interés, en donde la gran diversidad de especies ha sido relacionada con la capacidad que tienen estos organismos para dispersarse, aquellos que tienen mejores capacidades para volar han presentado rangos de distribución más amplios que los no voladores (Beck y Kitching, 2007; Benton, 1995; McCulloch et al., 2017; Rota et al., 2016).

Los odonatos (libélulas y caballitos) representan uno de los linajes más antiguos de insectos (Misof et al., 2014), una de sus características más llamativas es que fue uno de los primeros linajes en la historia de los animales en adquirir la capacidad de volar (Grimaldi y Engel, 2005). Su larga historia evolutiva implica que han superado varias de las grandes extinciones, muchas de las que se dieron por cambios significativos en la temperatura del planeta (Feulner, 2009). Actualmente existen 6340 especies conocidas de estos insectos en el planeta (Paulson y Schorr, 2020), distribuidas desde los trópicos hasta las regiones polares de todos los continentes (Kalkman et al., 2008). Algunas especies tienen altas capacidades de vuelo, que les permiten hacer largas migraciones que se reflejan en un amplio rango de distribución geográfica (Suárez-Tovar y Sarmiento, 2016); al contrario, otras especies se encuentran en áreas muy restringidas y cuentan con una alta especialización de hábitat (Kalkman et al., 2008), pero no es claro si esto se debe a las restricciones impuestas por su capacidad de dispersión, por sus hábitos o por su fisiología.

Teniendo en cuenta lo anterior, el objetivo de este trabajo es determinar si existe relación entre la distribución (tamaño de los rangos latitudinal y altitudinal, y tamaño del área de distribución geográfica) de 13 especies de odonatos y algunas características morfológicas y termofisiológicas propias de cada una de ellas.

Metodología

Objeto de estudio

Para responder las preguntas planteadas se seleccionaron 13 especies que se encuentran presentes en el área de estudio. Estas especies fueron elegidas porque además de tener rangos de distribución latitudinal y altitudinal contrastantes y preferencia por áreas abiertas o de bosque, se cuenta con datos de características morfológicas como la masa del adulto, y una colección de fotografías de sus alas (de la que fue posible extraer la longitud del ala posterior). También se cuenta con datos para algunos rasgos termofisiológicos (temperaturas mínimas, voluntarias y máximas) para cada especie. Las especies objeto de estudio son *Argia pulla*, *Argia medullaris*, *Argia indicatrix*, *Acanthagrion trilobatum*, *Erythrodiplax fusca*, *Hetaerina capitalis*, *Macrothemis pseudimitans*, *Epigomphus pechumani*, *Miocora aurea*, *Erythrodiplax abjecta*, *Erythrodiplax andagoya*, *Heteragrion aequatoriale* y *Heteropodagrion superbum*.

Obtención de los rasgos morfológicos y termofisiológicos

La longitud del ala posterior (mm) es una variable numérica continua que se obtuvo utilizando el software ImageJ v.1.52^a (National Institute of Health, USA). La masa para cada una de las especies fue obtenida a partir de lo reportado por Bota-Sierra (Cornelio Andrés Bota-Sierra, 2017).

Las variables termofisiológicas hacen referencia a un grupo de temperaturas (°C) que pueden ser utilizadas para determinar el funcionamiento óptimo de los seres vivos y son variables numéricas continuas. Éstas fueron obtenidas a partir de los datos recopilados por Bota-Sierra (2017), que son: la temperatura mínima de vuelo, la temperatura óptima para realizar actividades, y la temperatura máxima que soportan estos organismos antes de morir.

Rangos y área de distribución y hábitat

Todos los rangos de distribución fueron obtenidos a partir de los registros para cada especie provenientes de diversas fuentes, incluyendo publicaciones (Cornelio A. Bota-Sierra et al., 2019; Cornelio A Bota-Sierra et al., 2016; Pérez-Gutiérrez y Palacino-Rodríguez, 2011), plataformas de ciencia ciudadana en línea (GBIF.org, 2020; iNaturalist, 2020) y colecciones de especímenes de museo (Colección entomológica de la Universidad de Antioquia, Medellín, Colombia). El rango latitudinal corresponde a la distancia (en grados decimales) entre los datos registrados en la mayor y menor latitud. El rango de elevación corresponde a la diferencia entre el registro más alto y el más bajo (metros sobre el nivel del mar, m. s. n. m.) existente para cada especie. El área de distribución (Km²) se obtuvo generando polígonos envolventes convexos con el programa QGIS v.3.6.3 (QGIS Development Team, 2020), en donde también se calculó su área (Figura 1).

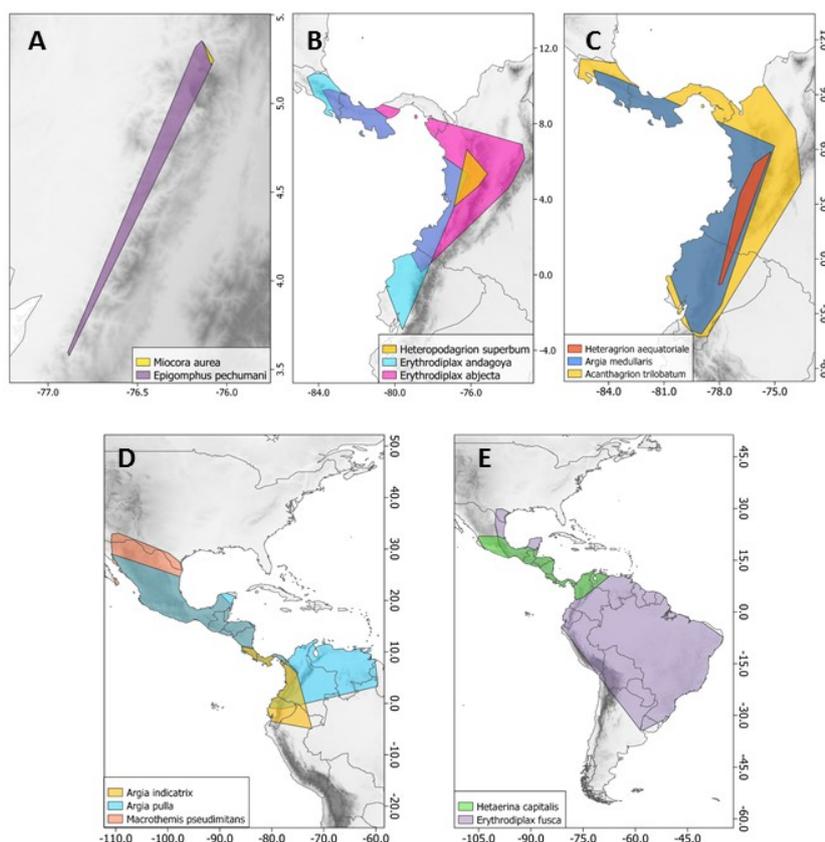


Figura 18. Mapas del área de la distribución geográfica para las 13 especies de interés: A) *M. aurea* y *E. pechumani*; B) *H. superbum*, *E. andagoya* y *E. abjecta*; C) *H. aequatoriale*, *A. medullaris* y *A. trilobatum*; D) *A. indicatrix*, *A. pulla* y *M. pseudimitans*; y E) *H. capitalis* y *E. fusca*.

Análisis estadístico

En primer lugar, se llevó a cabo un análisis exploratorio para buscar si existen correlaciones entre las variables de interés. Para saber si las características morfológicas y fisiológicas (independientes) de las especies influyen en sus rangos de distribución latitudinal, altitudinal y su área de distribución geográfica (dependientes), se utilizó un análisis de componentes principales (ACP) para cada una de las variables dependientes, así se determinaron cuál o cuáles de las variables independientes tienen mayor peso en los rangos de distribución latitudinal, de elevación y tamaño del área geográfica que ocupan. El ACP es una técnica multivariada que busca extraer las variables más importantes del set de datos compuesto por variables numéricas, aquí se generan nuevas variables denominadas componentes principales (Abdi y Williams, 2010; Kassambara, 2017) Las variables de distribución fueron categorizadas en los rangos “Estrecho”, “Medio”, y “Amplio” para determinar si hay una agrupación independiente de estas categorías (Tabla 1). Todos los análisis se hicieron en R Project v.4.0.5 (R Core Team, 2021).

Tabla 13. Se muestran los rangos de distribución latitudinal, de elevación y el área de distribución de las 13 especies y las categorías asignadas para cada rango.

Especie	Rangos			Categoría del rango		
	Latitudinal (°)	Elevación (m. s.n.m)	Área (Km ²)	Latitudinal	Elevación	Área
<i>E. fusca</i>	64,5	2238	1632875877	Amplio	Amplio	Amplio
<i>A. pulla</i>	30,15	1919	404625587	Amplio	Amplio	Amplio
<i>M. pseudimitans</i>	33,1597	1663	269970116	Amplio	Medio	Amplio
<i>E. andagoya</i>	13,61	493	18008117	Estrecho	Estrecho	Estrecho
<i>H. aequatoriale</i>	7,31	497	4909976	Estrecho	Estrecho	Estrecho
<i>H. superbum</i>	1,94	697	2779140	Estrecho	Estrecho	Estrecho
<i>E. pechumani</i>	1,77	1244	166105	Estrecho	Medio	Estrecho
<i>M. aurea</i>	0,13	1313	2001	Estrecho	Medio	Estrecho
<i>H. capitalis</i>	18,42	1964	170034343	Medio	Amplio	Medio
<i>A. indicatrix</i>	23,85	1525	79080276	Medio	Medio	Medio
<i>A. trilobatum</i>	15,26	1786	66832722	Medio	Medio	Medio
<i>A. medullaris</i>	25,02	1878	35752644	Medio	Medio	Medio
<i>E. abjecta</i>	9,686	2357	30690185	Estrecho	Amplio	Medio

Nota. Para el rango latitudinal (en grados decimales) se determinó un tamaño rango amplio cuando este va de 30 a 65°, medio de 15 a 30° y estrecho de 0,1 a 15°. En cuanto al rango de elevación (metros sobre el nivel del mar), el tamaño de rango amplio va de 1901 a 2500 m, el medio de 1001 a 1900 m y el estrecho de 100 a 1000 m. El tamaño del área de distribución geográfica (Km²) se categorizó como amplio las áreas entre 1,6e+9 a 2,7e+8 Km²; el medio de 1,7e+8 a 3,1e+7 y el estrecho de 1,8e+7 a 2e+3.

Resultados

Al hacer el análisis de correlación entre las variables, el mayor valor de correlación fue de 0.63 (Figura 2), por lo que se conservaron todas las variables. Las variables en donde se encuentra mayor correlación son la masa y la longitud del ala posterior (0.6) y las temperaturas voluntarias y máximas (0,63).

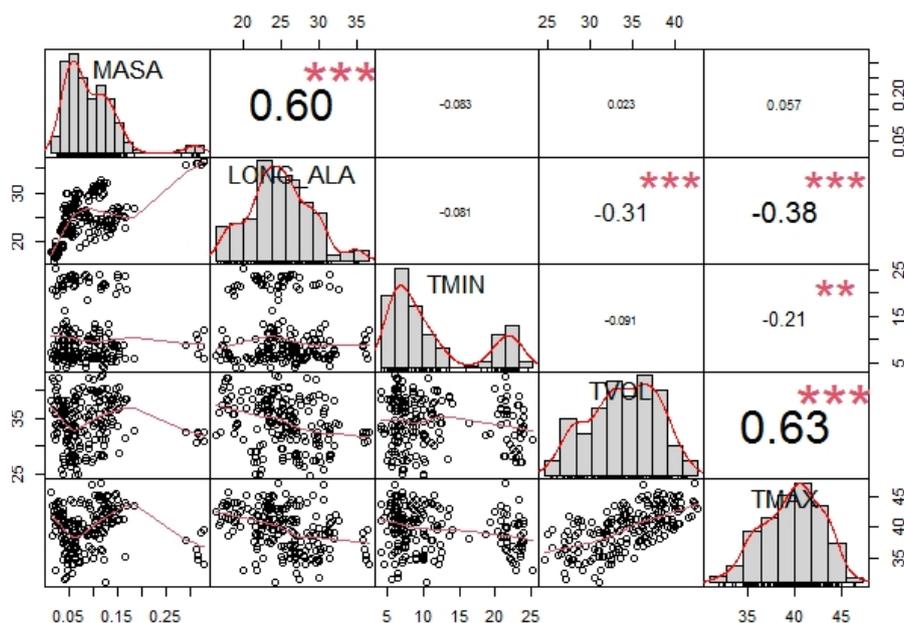


Figura 19. Gráfico que muestra el índice de correlación entre las variables y su distribución.

Análisis de componentes principales

Para este análisis exploratorio por componentes principales, la base de datos utilizada cuenta con datos para 238 individuos. Como las varianzas de las variables son muy diversas se reescalaron los datos para hacer el análisis, de manera que los datos quedaran con media 0 y varianza 1. El análisis arrojó que los primeros tres componentes principales explican el 87% de la variación. En el primer componente las variables con mayor peso son la longitud del ala y las temperaturas voluntaria y máxima, en el segundo componente la variable con más peso es la masa y en el tercero es la temperatura mínima (Tabla 2).

Tabla 2. Resultados del análisis de componentes principales

Componentes	PC1	PC2	PC3
Desviación estándar	1,41	1,21	0,95
Proporción de la varianza	0,40	0,29	0,18
Proporción acumulada	0,40	0,69	0,87
Variables			
Masa	-0,28	0,66	-0,32
Longitud del ala	-0,55	0,41	-0,06
Temperatura mínima	-0,10	-0,41	-0,90
Temperatura voluntaria	0,54	0,32	-0,28
Temperatura máxima	0,56	0,36	-0,11

Rango latitudinal

De acuerdo a la agrupación por tamaño de rango latitudinal, se puede observar que los tres grupos parecen ser independientes, lo que indica que nuestras variables morfológicas y fisiológicas son diferentes para los individuos que tienen rangos latitudinales de distribución amplios, estrechos, e intermedios (Figura 3). El grupo con un rango latitudinal más amplio tiene valores similares de temperatura máxima y voluntaria, que además están correlacionados, este grupo lo conforman las especies *E. fusca*, *A. pulla* y *M. pseudimitans*. El rango latitudinal estrecho (*E. andagoia*, *H. superbum*, *H. aequatoriale*, *E. pechumani*, *M. aurea* y *E. abjecta*) y medio (*H. capitalis*, *A. indicatrix*, *A. trilobatum* y *A. medullaris*) se aproximan más entre ellos.

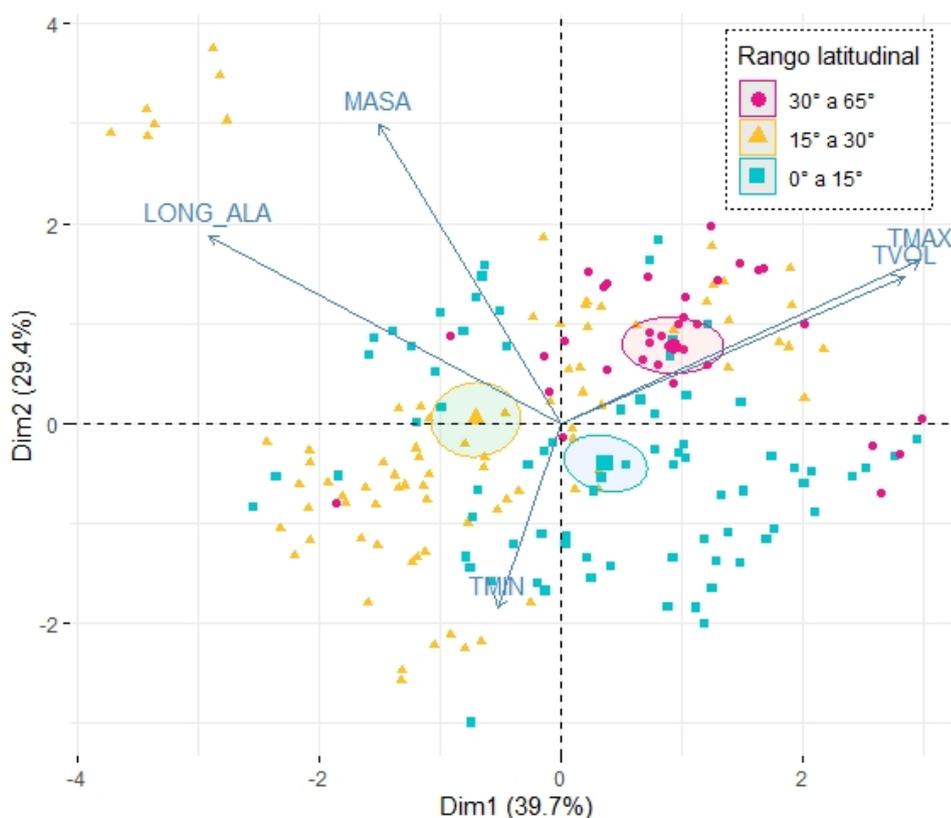


Figura 20. Resultado del PCA agrupado por el rango latitudinal de las especies.

Rango de elevación

Al igual que la agrupación por rango latitudinal, en la agrupación por tamaño de rango de elevación se observa que los tres grupos parecen ser independientes, lo que indica que las variables morfológicas y fisiológicas son diferentes para los individuos que tienen rangos latitudinales de distribución amplios, estrechos, e intermedios. Las especies con rango de elevación amplio (*E. fusca*, *A. pulla*, *H. capitalis* y *E. abjecta*) tienen mayores masas, seguido por las especies con rango medio (*M. pseudimitans*, *E. pechumani*, *M. aurea*, *A.*

indicatrix, *A. trilobatum* y *A. medullaris*), mientras que las especies de menor masa son aquellas con rangos estrechos (*E. andagoya*, *H. aequatoriale*, y *H. superbum*) (Figura 4).

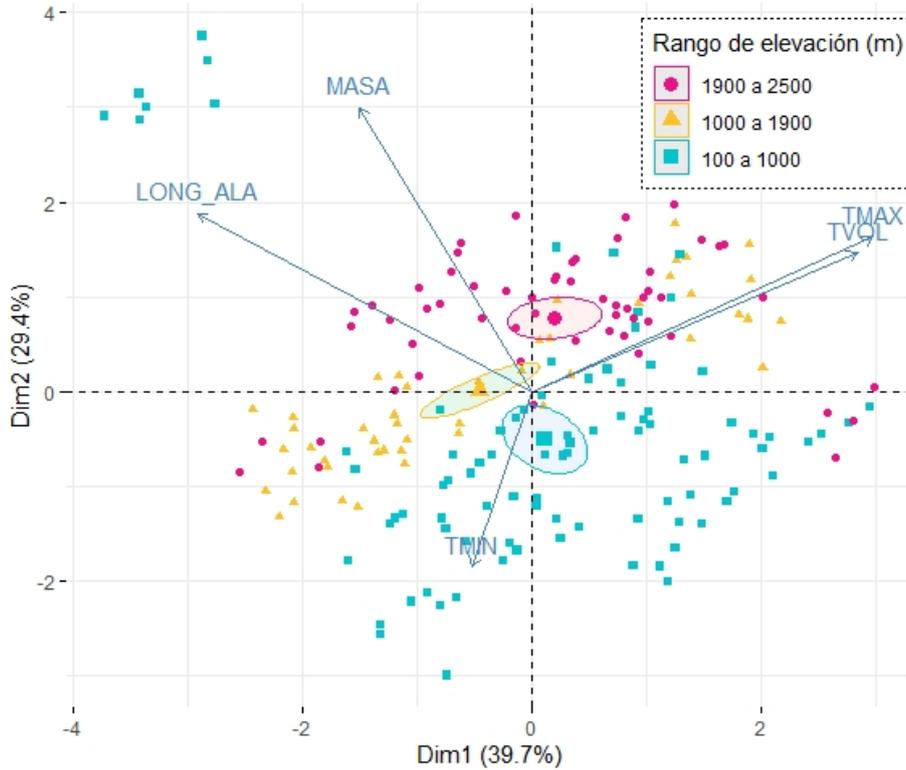


Figura 21. Resultado del PCA agrupado por el rango de elevación de las especies.

Área de distribución geográfica

Estos resultados son similares a los anteriores, parece que los datos agrupados en diferentes tamaños de área de distribución geográfica también presentan diferencias en sus variables morfológicas y fisiológicas. Al igual que para el rango de elevación, las especies con mayores áreas de distribución geográfica resultaron ser aquellas con mayores masas (*E. fusca*, *A. pulla*, y *M. pseudimitans*), seguidas por las de área de distribución intermedia (*H. capitalis*, *A. indicatrix*, *A. trilobatum*, *A. medullaris* y *E. abjecta*), mientras que aquellas con menores áreas de distribución son también las más pequeñas o de menor masa (*E. andagoya*, *H. aequatoriale*, *H. superbum*, *E. pechumani* y *M. aurea*) (Figura 5).

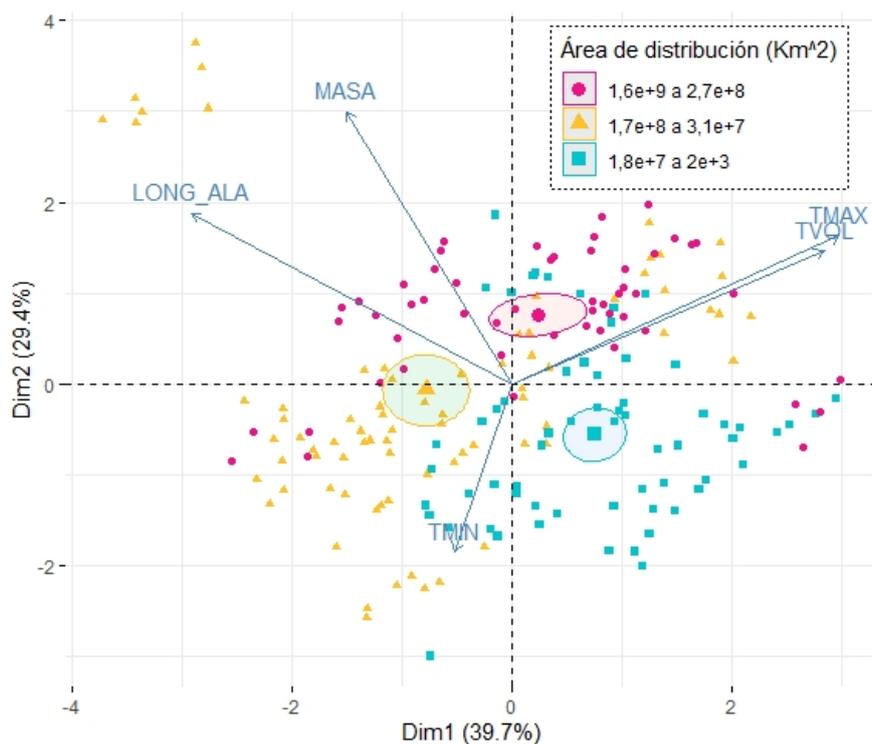


Figura 22. Resultado del PCA agrupado por el tamaño del área de distribución de las especies.

Conclusiones

De acuerdo con los resultados, se puede concluir que el análisis de componentes principales resultó útil para identificar cuáles de los caracteres utilizados pueden estar influyendo en la distribución de las 13 especies.

Se observa que los dos primeros componentes principales explican el 70% de la variación de nuestros datos. La mayor contribución en el primer componente principal la tienen las temperaturas máxima y voluntaria, que además están correlacionadas, y la longitud de las alas, siendo esta última opuesta a las dos primeras. Sin embargo, es el segundo componente principal el que parece dar más pistas que se pueden analizar desde una perspectiva biológica; los individuos agrupados en el rango de distribución más amplio (para las tres categorías; elevación, latitud y área de distribución) tienen mayores masas, y esto puede estar relacionado con que las especies más grandes utilizadas en este estudio pertenecen a la familia Libellulidae que habitan áreas abiertas, lo que sugiere que pueden desplazarse por distancias más amplias sin una mayor necesidad de refugiarse en bosques o vegetación densa.

Con la ayuda de este método estadístico se pueden identificar algunas variables que pueden estar influyendo en mayor medida en los eventos en los que la biología se interesa, en este caso, ¿qué está influyendo en los patrones de distribución de las especies? En este ejercicio puntual, se puede ver que existe una mezcla entre factores fisiológicos y morfológicos que se presentan en las especies con diferentes características de distribución. Posteriormente se podrían hacer las pruebas de hipótesis pertinentes, para comprobar si existe una influencia estadísticamente significativa entre las variables analizadas.

Referencias

- Abdi, H., y Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459. <https://doi.org/https://doi.org/10.1002/wics.101>
- Agnarsson, I., Cheng, R. C., y Kuntner, M. (2014). A multi-clade test supports the intermediate dispersal model of biogeography. *PLoS ONE*, 9(1), e86780. <https://doi.org/10.1371/journal.pone.0086780>
- Beck, J., y Kitching, I. J. (2007). Correlates of range size and dispersal ability: A comparative analysis of sphingid moths from the Indo-Australian tropics. *Global Ecology and Biogeography*, 16, 341–349. <https://doi.org/10.1111/j.1466-8238.2007.00289.x>
- Benton, T. G. (1995). Biodiversity and biogeography of Henderson Island's insects. *Biological Journal of the Linnean Society*, 56, 245–259. <https://doi.org/10.1111/j.1095-8312.1995.tb01089.x>
- Bergmann, C. (1847). Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. En *Göttinger Studien*.
- Böhning-Gaese, K., Caprano, T., Van Ewijk, K., y Veith, M. (2006). Range size: Disentangling current traits and phylogenetic and biogeographic factors. *American Naturalist*, 167(4), 555–567. <https://doi.org/10.1086/501078>
- Bota-Sierra, Cornelio A., Sandoval-H., J., Ayala-Sánchez, D., y Novelo-Gutiérrez, R. (2019). *Libélulas de la Cordillera Occidental Colombiana, una mirada desde el Tatamá/ Dragonflies of the Colombian Cordillera Occidental, a look from Tatamá* (1a ed.). Panamericana S.A.
- Bota-Sierra, Cornelio A, Maufray, B., Palacino-Rodríguez, F., Hofmann, J., Tennesen, K. J., Rache, L., y Tognelli, M. F. (2016). Estado de conservación de las libélulas de los Andes Tropicales. En M. F. Tognelli, C. A. Lasso, C. A. Bota-Sierra, L. F. Jiménez-Segura, y N. A. Cox (Eds.), *Estado de Conservación y Distribución de la Biodiversidad de Agua Dulce en los Andes Tropicales* (pp. 67–86). UICN. <https://doi.org/10.2305/IUCN.CH.2016.02.es>
- Bota-Sierra, Cornelio Andrés. (2017). *Odonatos de las Montañas Tropicales Bajo Cambio Climático: La Perspectiva Térmica de los Adultos*. Instituto de Ecología, A. C.
- Chavez, M. Y., Mabry, K. E., McCauley, S. J., y Hammond, J. I. (2015). Differential larval responses of two ecologically similar insects (Odonata) to temperature and resource variation. *International Journal of Odonatology*, 18(4), 297–304. <https://doi.org/10.1080/13887890.2015.1082946>
- Claramunt, S., Derryberry, E. P., Remsen, J. V., y Brumfield, R. T. (2012). High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733), 1567–1574. <https://doi.org/10.1098/rspb.2011.1922>
- Cox, C. B., y Moore, P. D. (1993). *Biogeography: an ecological and evolutionary approach*. (5a ed.). Blackwell Science Ltd. <https://doi.org/10.1080/10635150500541581>
- Feulner, G. (2009). Climate modelling of mass-extinction events: A review. *International Journal of Astrobiology*, 8(3), 207–212. <https://doi.org/10.1017/S1473550409990061>
- Fritz, S. A., Jönsson, K. A., Fjeldså, J., y Rahbek, C. (2012). Diversification and biogeographic patterns in four island radiations of passerine birds. *Evolution*, 66(1), 179–190. <https://doi.org/10.1111/j.1558-5646.2011.01430.x>
- García-Robledo, C., Kuprewicz, E. K., Staines, C. L., Erwin, T. L., y Kress, W. J. (2016).

- Limited tolerance by insects to high temperatures across tropical elevational gradients and the implications of global warming for extinction. *Proceedings of the National Academy of Sciences*, 113(3), 680–685. <https://doi.org/10.1073/pnas.1507681113>
- GBIF.org. (2020). *GBIF*. www.gbif.org
- Ghalambor, C. K. (2006). Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, 46(1), 5–17. <https://doi.org/10.1093/icb/icj003>
- Grimaldi, D., y Engel, M. S. (2005). *Evolution of Insects*. Cambridge University Press.
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*. <https://doi.org/10.2307/4072271>
- Hartley, S., Krushelnycky, P. D., y Lester, P. J. (2010). Integrating physiology, population dynamics and climate to make multi-scale predictions for the spread of an invasive insect: The Argentine ant at Haleakala National Park, Hawaii. *Ecography*, 33, 83–94. <https://doi.org/10.1111/j.1600-0587.2009.06037.x>
- Hawkins, B. A. (2001). Ecology's oldest pattern? *Trends in Ecology y Evolution*, 16(8), 470.
- Hewitt, N., y Kellman, M. (2002). Tree seed dispersal among forest fragments: II. Dispersal abilities and biogeographical controls. *Journal of Biogeography*, 29, 351–363. <https://doi.org/10.1046/j.1365-2699.2002.00679.x>
- iNaturalist. (2020). *iNaturalist*. www.inaturalist.org
- Janzen, D. H. (1967). Why Mountain Passes are Higher in the Tropics. *The American Naturalist*, 101(919), 233–249. <https://doi.org/10.1086/282487>
- Kalkman, V. J., Clausnitzer, V., Dijkstra, K. D. B., Orr, A. G., Paulson, D. R., y Van Tol, J. (2008). Global diversity of dragonflies (Odonata) in freshwater. *Hydrobiologia*, 595(1), 351–363. <https://doi.org/10.1007/s10750-007-9029-x>
- Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (Vol. 2). Sthda.
- Lomolino, M. V., Riddle, B. R., Whittaker, R. J., y Brown, J. H. (2010). *Biogeography* (4a ed.). Sinauer Associates, Inc.
- McCain, C. M., y Grytnes, J.-A. (2010). Elevational Gradients in Species Richness. En *Encyclopedia of Life Sciences*. <https://doi.org/10.1002/9780470015902.a0022548>
- McCulloch, G. A., Wallis, G. P., y Waters, J. M. (2017). Does wing size shape insect biogeography? Evidence from a diverse regional stonefly assemblage. *Global Ecology and Biogeography*, 26, 93–101. <https://doi.org/10.1111/geb.12529>
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Flouri, T., Beutel, R. G., Niehuis, O., y Petersen, M. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–768. <https://doi.org/10.1126/science.1257570>
- Nathan, R. (2001). Dispersal Biogeography. En *Encyclopedia of Biodiversity* (Vol. 2, pp. 127–152). Elsevier. <https://doi.org/10.1016/B0-12-226865-2/00073-0>
- Paulson, D. R., y Schorr, M. (2020). *World Odonata List*. University of Puget Sound. <https://www.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resources/dragonflies/world-odonata-list2/>
- Pérez-Gutiérrez, L. A., y Palacino-Rodríguez, F. (2011). Updated checklist of the Odonata known from Colombia. *Odonatologica*, 40(3), 203.
- Pintor, A. F. V., Schwarzkopf, L., y Krockenberger, A. K. (2015). Rapoport's rule: Do Climatic Variability gradients shape range extent? *Ecological Monographs*, 85(4), 643–659. <https://doi.org/10.4225/28/5536F28D28A5E>
- Polato, N. R., Gill, B. A., Shah, A. A., Gray, M. M., Casner, K. L., Barthelet, A., Messer, P.

- W., Simmons, M. P., Guayasamin, J. M., Encalada, A. C., Kondratieff, B. C., Flecker, A. S., Thomas, S. A., Ghalambor, C. K., LeRoy Poff, N., Chris Funk, W., y Zamudio, K. R. (2018). Narrow thermal tolerance and low dispersal drive higher speciation in tropical mountains. *Proceedings of the National Academy of Sciences of the United States of America*, 115(49), 12471–12476. <https://doi.org/10.1073/pnas.1809326115>
- QGIS Development Team. (2020). QGIS Geographic Information System. En *Open Source Geospatial Foundation Project*. <https://doi.org/http://www.qgis.org/>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (4.0.5). <https://www.r-project.org/>
- Rota, J., Peña, C., y Miller, S. E. (2016). The importance of long-distance dispersal and establishment events in small insects: Historical biogeography of metalmark moths (Lepidoptera, Choreutidae). *Journal of Biogeography*, 43, 1254–1265. <https://doi.org/10.1111/jbi.12721>
- Sandoval-H, J., Gómez, J. P., y Cadena, C. D. (2017). Is the largest river valley west of the Andes a driver of diversification in Neotropical lowland birds? *The Auk*, 134(1), 168–180. <https://doi.org/10.1642/AUK-16-91.1>
- Shah, A. A., Gill, B. A., Encalada, A. C., Flecker, A. S., Funk, W. C., Guayasamin, J. M., Kondratieff, B. C., Poff, N. L. R., Thomas, S. A., Zamudio, K. R., y Ghalambor, C. K. (2017). Climate variability predicts thermal limits of aquatic insects across elevation and latitude. *Functional Ecology*, 31(11), 2118–2127. <https://doi.org/10.1111/1365-2435.12906>
- Stevens, G. C. (1992). The elevational gradient in altitudinal range: an extension of Rapoport's latitudinal rule to altitude. *American Naturalist*. <https://doi.org/10.1086/285447>
- Suárez-Tovar, C. M., y Sarmiento, C. E. (2016). Beyond the wing planform: Morphological differentiation between migratory and nonmigratory dragonfly species. *Journal of Evolutionary Biology*, 29, 690–703. <https://doi.org/10.1111/jeb.12830>
- Suhling, F., Suhling, I., y Richter, O. (2015). Temperature response of growth of larval dragonflies – an overview. *International Journal of Odonatology*, 18(1), 15–30. <https://doi.org/10.1080/13887890.2015.1009392>
- Watson, J. E. M., Segan, D. B., y Tewksbury, J. (2019). Tropical Forests in a Changing Climate. En T. E. Lovejoy y L. Hannah (Eds.), *Biodiversity and Climate Change: Transforming the biosphere* (pp. 196–207). Yale University Press. <http://www.jstor.org/stable/j.ctv8jnzwl.26>
- Wiens, J. J., y Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. En *Trends in Ecology and Evolution*. <https://doi.org/10.1016/j.tree.2004.09.011>

Comparación de la mirmecofauna en agroecosistemas de café (*Coffea arabica L.*) con diferente estructura arbórea en el centro de Veracruz

Rafael Sánchez Gregorio¹, Cecilia Cruz López², Jorge Valenzuela González³, Gibrán Renoy Pérez Toledo⁴

¹ Facultad de Estadística e Informática, Universidad Veracruzana. uvemiliano@icloud.com

² Facultad de Estadística e Informática, Universidad Veracruzana. cecacruz@uv.mx

³ Instituto de ecología, A. C. Jorge.valenzuela@inecol.mx

⁴ Instituto de ecología, A. C. gibrán.perez@hotmail.com

Resumen

Los paisajes agrícolas de café con distintos grados de complejidad estructural arbórea dominan las áreas de cultivo en el centro de Veracruz, impactando de diversas formas a las hormigas tropicales que habitan estos agroecosistemas. Examinamos la riqueza de especies de hormigas utilizando diferentes métodos de muestreo (trampas, colecta directa y sacos mini-Winkler) para diversos estratos (hipógeo, epígeo, hojarasca y vegetación arbórea), además de medir distintas variables de la estructura vegetal (altura de árboles, área basal, cobertura del dosel, densidad de cafetos, densidad de árboles y riqueza de árboles) en 12 fincas de café distribuidas en los municipios de Coatepec y Huatusco en el centro de Veracruz. Los sitios de estudio se agruparon en dos clústeres con base a la estructura arbórea con el fin de obtener una mayor robustez en las pruebas estadísticas. En total se encontraron 196 especies de hormigas pertenecientes a ocho subfamilias, 15 tribus y 54 géneros. Los resultados obtenidos en una comparación de la riqueza de hormigas en función de la estructura vegetal arbórea por medio de una prueba de Wilcoxon mostraron que la estructura arbórea en su conjunto parece no influenciar la riqueza de hormigas. Esto puede deberse, a factores externos a las fincas (por ejemplo, el paisaje adyacente) el cual no se tomó en cuenta en el estudio y puede estar influyendo sobre la riqueza de especies. La utilización de modelos de regresión lineal múltiple para determinar los mejores predictores de la riqueza de hormigas indicaron que la altura de los árboles es un predictor positivo y la cobertura del dosel un predictor negativo.

Palabras Clave: Mirmecofauna, Biodiversidad, Manejo de agroecosistemas, Regresión múltiple.

Introducción

El café es uno de cultivos comerciales con mayor demanda en el mundo, así como una fuente de ingresos esencial para pequeños y grandes productores en el mundo, además de que su cadena de elaboración (tostado, empaquetado y venta) genera múltiples empleos (Donald, 2004). En México este grano se cultiva principalmente en áreas donde hacen contacto dos o más tipos de vegetación y su manejo abarca diferentes niveles de complejidad estructural arbórea. En México, estos paisajes agrícolas se distinguen en cinco sistemas de producción: dos tipos de agrobosques tradicionales, un sistema de sombra poliespecífico y dos sistemas de monocultivo (monocultivo sombreado y no sombreado) (Moguel & Toledo, 1999; Toledo & Moguel, 2012).

Se ha observado que los sistemas de cultivo de café con mayor complejidad estructural arbórea conservan mayormente a la biota nativa a diferencia de los sistemas simples o mono específicos (Escobar-Ramírez, Tscharrntke, Armbrecht, Torres, & Grass, 2020; Noponen, Hagggar, Edwards-Jones, & Healey, 2013). Los servicios ecosistémicos, como la polinización, el control de plagas, la captura de carbono y nutrientes también es mayor en las fincas con mayor riqueza arbórea (Jha et al., 2014), ya que en estos mosaicos agrícolas se alberga una mayor diversidad de enemigos naturales de insectos plaga (Aristizábal, Bustillo, & Arthurs, 2016). Sin embargo, este cultivo se ha visto afectado en los últimos años por un cambio hacia un manejo agrícola más intensificado, donde se ha transformado de un paisaje silvicultural ha sistemas simplificados monoespecíficos con mayor dependencia de insumos agrícolas (Toledo & Moguel, 2012; Tscharrntke et al., 2012).

Los insectos forman uno de los taxones más diversos en casi todos los ecosistemas del mundo. Entre ellos las hormigas generan la mayor parte de la biomasa de los insectos en las regiones tropicales (Hölldobler & Wilson, 1990). Las hormigas actúan como depredadores en los sistemas agrícolas, donde tienen efectos complejos en los nivelestróficos más bajos (Philpott, Perfecto, & Vandermeer, 2008). Estas forrajean sobre distintos estratos del cultivo abarcando desde el nivel subterráneo hasta el dosel de los árboles (Burd & Howard, 2005). Entre los diversos servicios ecológicos que brindan las hormigas en los sistemas de cultivo de café, se encuentran la acciones de actuar como ingenieras de ecosistema, dispersores de semillas, además de ser agentes biocontroladores de la broca del café (*Hypothenemus hampei*), una de las plagas de café con mayor importancia económica, lo que complementa las estrategias del manejo agrícola aplicado a las fincas de cultivo (Armbrecht & Gallego, 2007; Cammeraat & Risch, 2008; Philpott, 2006).

Las hormigas también pueden considerarse como organismos bioindicadores de la calidad de hábitat, debido a que son un grupo de insectos altamente sensibles a la perturbación de los sitios donde habitan (Andersen, Hoffmann, Muller, & Griffiths, 2002). En los sitios de cultivo de café donde se elimina la estructura arbórea para dar paso a sistemas más simplificados se ha observado que su diversidad disminuye notablemente (Perfecto & Snelling, 1995). Debido a esta sensibilidad a las interacciones antropogénicas en su hábitat las vuelve un taxón bastante efectivo para estudios en los cambios de hábitat, como las modificaciones en los paisajes agrícolas silviculturales hacia sistemas con una estructura arbórea más simple (Armbrecht, Rivera, & Perfecto, 2005).

Metodología

Área de estudio

El área de estudio se encuentra localizada en la región cafetalera del centro de Veracruz, México entre los rangos de 1010 a 1340 msnm. Los suelos dominantes son Acrisoles y Cambisoles ándicos, la precipitación media anual oscila entre 1350 y 2200 mm y el rango de las temperaturas es de 12 a 19 °C (Hietz & Hietz-Seifert, 1995). En la región se presentan tres estaciones durante el año. Una época fría de noviembre a marzo, una cálida seca de abril a mayo y por último una cálida húmeda de junio a octubre (Williams-Linera, Sosa, & Platas, 1995). El cultivo de café es una de las actividades de mayor importancia económica entre los pobladores de la región teniendo su mayor auge durante las décadas de los 60's y 70's (Griffith, Zamudio-Grave, Cortés-Viveros, & Cabrera- Cabrera, 2017).

Para los análisis estadísticos se seleccionaron en total 12 fincas de café (Tabla 1) con diferente manejo agrícola, ubicadas en las localidades de los municipios de Coatepec y Huatusco. El criterio de la selección de los sitios fueron los diferentes grados de complejidad de la estructura arbórea.

Tabla 1. Descripción de las 12 fincas de estudio.

Cafetales	Tamaño (Ha)	Municipio	Latitud (N)	Longitud (W)	Altura (msnm)
1	6.43	Coatepec	19°25'54"	96°57'2"	1,115
2	15.74	Coatepec	19°26'1"	96°52'32"	1,050
3	1.32	Huatusco	19°2'32"	96°54'45"	1,170
4	19.31	Coatepec	19°28'26"	96°57'18"	1,210
5	17.38	Coatepec	19°25'49"	96°52'55"	1,015
6	140.57	Huatusco	19°12'57"	96°53'7"	1,090
7	10.51	Huatusco	19°2'27"	96°55'23"	1,200
8	1.96	Coatepec	19°25'56"	96°57'50"	1,170
9	113.85	Huatusco	19°12'14"	96°52'44"	1,060
10	113.85	Huatusco	19°12'16"	96°53'15"	1,080
11	113.85	Huatusco	19°12'22"	96°53'4"	1,090
12	10.26	Coatepec	19°26'8"	96°53'55"	1,070

Caracterización de las variables estructurales de los sitios

La caracterización de la estructura vegetal arbórea de los sitios de estudio se tomó de los datos publicados por (López-Gómez, Williams-Linera, Manson, 2008), quienes realizaron las caracterizaciones en las mismas fincas donde se llevó a cabo el estudio. Para caracterizar los sitios se midieron cinco descriptores de los árboles con un dap (diámetro a la altura del pecho) mayor a cinco cm en las mismas fincas, además de medir la densidad de cafetos (plantas/ha). Los descriptores de los árboles seleccionados fueron: altura (metros), área basal (m^2/ha), cobertura del dosel (%), densidad de árboles (individuos/ha) y riqueza de especies arbóreas (especies/ha).

Muestreo de hormigas en sistemas de cultivo de café

Para el diseño de muestreo se seleccionaron 10 puntos por cada sitio estudiado, las colectas se realizaron dentro de un radio de 25 metros alrededor de cada punto utilizando diferentes técnicas, según lo necesitara cada estrato estudiado (subterráneo, epigeo, hojarasca y vegetación arbórea).

Subterráneo. Se utilizaron frascos de plástico con capacidad de 250 ml, a los cuales se le realizaron perforaciones sobre el contorno con el fin de permitir la entrada de las hormigas que forrajean bajo el nivel del suelo. Posteriormente se les colocó un cebo de atún en un frasco de menor tamaño en la parte central interior del frasco. Al frasco de mayor tamaño se le agregó propilenglicol diluido al 50% con el fin de preservar los individuos. Finalmente se taparon los frascos y se enterraron a una profundidad de 10 cm bajo el suelo.

Epigeo. Se utilizaron trampas de caída o trampas tipo “pit-fall”. Éstas consistieron en recipientes de plástico con capacidad de 500 ml, de un diámetro de 10 cm y una altura de 7.5 cm. Los recipientes contenían en su interior 100 ml de propilenglicol diluido al 50%. Las trampas fueron enterradas en un agujero para mantener la boca del recipiente al nivel de la altura de la superficie. Posteriormente, se cubrieron con un plato de unicel de 26 cm de diámetro a una separación de la trampa de cinco cm, para permitir el ingreso de las hormigas. Esto con el fin de evitar la contaminación de material externo, como las hojas de árboles y agua en caso de lluvia. Con el fin de obtener más información sobre los gremios tróficos, en cada punto se colocaron dos trampas, una con cebo y otra sin cebo.

Hojarasca. En cada punto de los sitios de estudio se colectó una muestra de hojarasca del suelo de $1 m^2$. Ésta se tamizó en campo con un cernidor de 1 cm de apertura y posteriormente se colocó en bolsas de plástico para trasladarlas al laboratorio, donde el contenido se introdujo en sacos mini-Winkler, durante un periodo de 48 horas con el fin de extraer el mayor número de hormigas de este estrato (Agosti, 2000).

Cafetales. Para la captura de las hormigas que forrajean sobre las plantas de café, se decidió emplear una manta cuadrada de caída tipo “Bignell” de 80 x 80 cm (Knudsen, 1972). En cada punto de muestreo de las fincas se seleccionaron tres plantas, donde se hicieron las colectas y con el fin de obtener información sobre las especies nectarívoras se agregaron al muestreo trampas de miel. Éstas consistieron en frascos con capacidad de 100 ml con perforaciones sobre el contorno superior, donde se introdujo 20 ml de miel de abeja.

Arbóreo. Se colectaron con la ayuda de pinceles y aspiradores bucales a todas las hormigas que forrajeaban en la base de los troncos de los árboles, a una altura de dos metros. Estas colectas se hicieron durante un periodo de 15 minutos sobre los árboles que se encontraban cercanos a cada punto. Todas las trampas fueron recuperadas después de 48 horas de exposición en campo.

Trabajo de laboratorio

Las hormigas fueron limpiadas, separadas y contabilizadas a través de un microscopio estereoscópico. Al finalizar su conteo fueron preservadas en frascos viales con alcohol al 70%, posteriormente se montaron en seco sobre triángulos de papel con alfileres entomológicos, para su futura identificación y preservación. Los especímenes se determinaron a nivel de género utilizando las claves de Mackay y Mackay (1989) y Bolton (1994). Para la identificación a nivel de especie se utilizaron varias claves dependiendo del grupo involucrado. Todos los individuos montados quedaron depositados en la colección entomológica IEXA del instituto de Ecología A.C. (Xalapa, Veracruz, México).

Comparación de la riqueza de especies de hormigas en función de la estructura vegetal arbórea de las fincas

Para la comparación de la riqueza de hormigas en función del manejo de la estructura vegetal arbórea en las fincas de estudio, se realizó una agrupación de los sitios, en dos clústeres haciendo uso de un análisis de k-medias. Esto con el fin de obtener el mayor número de repeticiones en cada grupo y aumentar la robustez estadística en los próximos análisis. Esta técnica de agrupación es un método eficaz que clasifica a los objetos dentro de un número de grupos que se asigna (Dubey, Dixit, Singh, & Gupta 2013), de modo que todos los datos de cada grupo comparten un rasgo en común partiendo de una matriz de distancias, en nuestro caso de tipo “euclidiano”. Posteriormente se evaluaron los supuestos de los conteos de las especies capturadas, haciendo uso de una prueba de normalidad Shapiro-Wilk (1965) y una prueba de Bartlett para la homocedasticidad. Como nuestros datos no siguieron una distribución normal se decidió utilizar una prueba no paramétrica Wilcoxon de Mann-Whitney (Hollander, M., Wolfe, D. A., & Chicken, E., 2013). Esto para la comparación de las medianas entre grupos independientes a un nivel de confiabilidad del 99 %. Estos análisis fueron realizados en el software estadístico R versión 3.6 (R Core Team, 2020).

Variables estructurales arbóreas como predictores de la riqueza de hormigas

Para determinar cuáles son los mejores predictores de la riqueza de especies de hormigas en función de la estructura vegetal arbórea, se decidió crear un modelo de regresión lineal múltiple. En este modelo los predictores fueron las mediciones que componen a la estructura vegetal arbórea de los sitios de estudio (altura de los árboles, cobertura del dosel, área basal, riqueza de especies de árboles y densidad de árboles). Para asegurar que los predictores fueran independientes y detectar multicolinealidad se realizó una matriz de correlación de Pearson con las variables estructurales iniciales. Para la selección del modelo final se eliminaron las variables que no aportaran al modelo a través de un método mixto, el cual combina las técnicas de selección y eliminación de variables “*Fordwar*” y “*Backward*” haciendo uso del paquete “*MASS*” del software estadístico R versión 3.6 (R Core Team, 2020). El criterio para la selección del mejor modelo que explicara de forma más eficiente la predicción de la riqueza de hormigas se basó en el valor matemático Akaike (AIC).

Resultados

Riqueza de especies de hormigas

Considerando las 12 fincas de estudio se colectaron en total 67,176 individuos pertenecientes a ocho subfamilias, 15 tribus, 54 géneros y 196 especies. La subfamilia con mayor representación fue Myrmicinae con cuatro tribus, 32 géneros y 107 especies, seguida por Formicinae con cuatro tribus, siete géneros y 31 especies, Ponerinae con 1 tribu, siete géneros y 18 especies, Dolichoderinae con dos tribus, cinco géneros y 12 especies, Pseudomyrmecinae con una tribu, un género y 12 especies, Dorylinae con dos tribus, siete géneros y 10 especies, Proceratinae con una tribu, dos géneros y tres especies y Ectatominae con una tribu, un género y tres especies. En cuanto al número de especies que se colectaron por género, el de mayor diversidad fue *Pheidole* con 23 especies, seguido por *Solenopsis* (14 spp), *Camponotus* (13 spp), *Pseudomyrmex* (12 spp) y *Nylanderia* (9 spp).

Efecto de la estructura arbórea sobre la riqueza de especies de hormigas

De los dos grupos que se generaron con el método de k-medias, (Figura 1), se observó que cinco fincas se agruparon en la categoría de estructura baja y siete la de estructura alta. La riqueza de especies fue mayor para el grupo con una estructura alta, con un promedio de 50.85 ± 11.89 especies a diferencia del grupo con una estructura baja, donde se obtuvo un promedio de 44.4 ± 9.96 especies. Los resultados de la prueba no paramétrica de Wilcoxon de Mann-Whitney muestran los valores ($w_1 = 23, p - value = 0.416$) para ambos grupos (Figura 2). Aunque hay una diferencia en la riqueza de especies de hormigas entre ambos grupos, esta no llega a ser significativa estadísticamente.

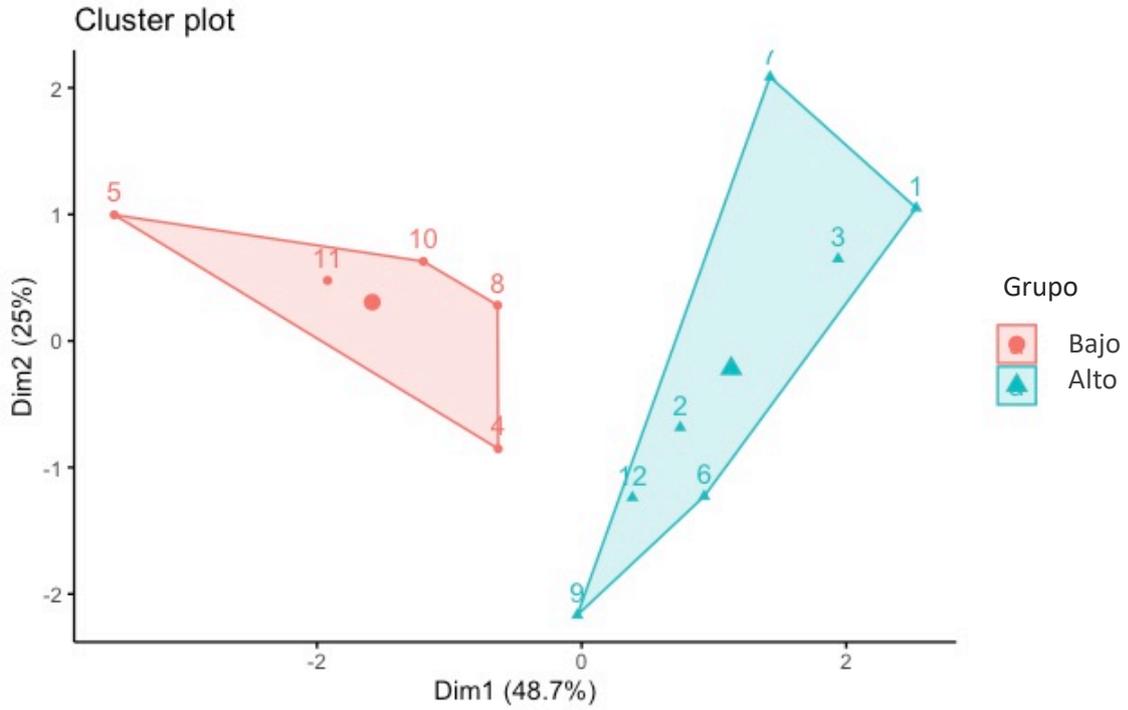


Figura 1. Agrupación de las 12 fincas de cultivo de café tomadas del cuadro 1, con un método de k-medias. La designación de los conglomerados se hizo basado e en la similitud de los sitios con los descriptores de la estructura vegetal arbórea.

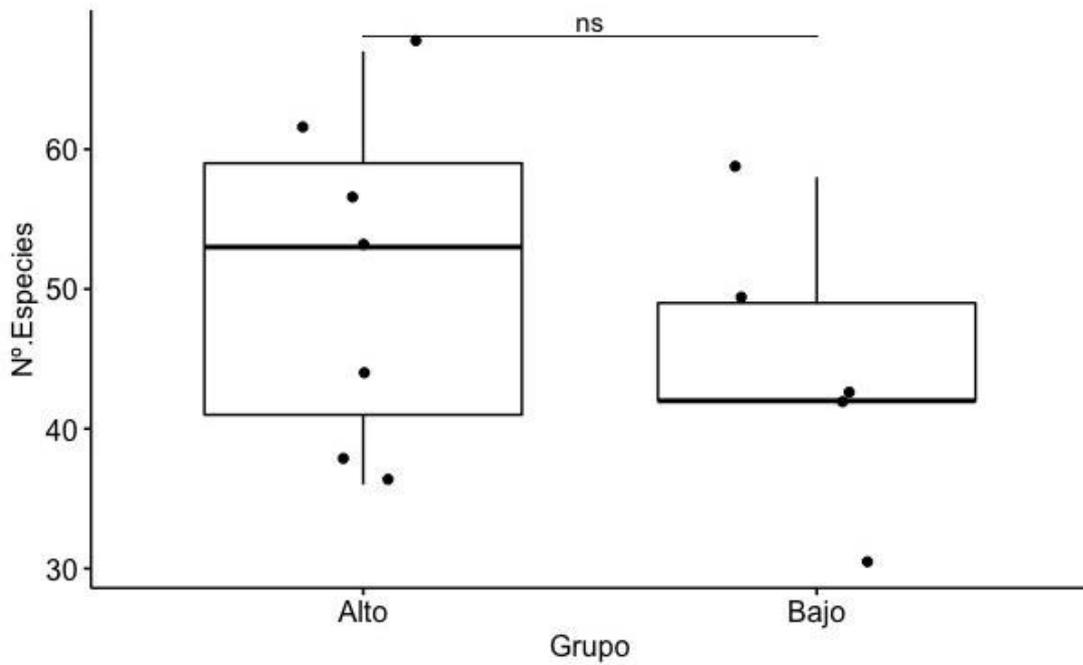


Figura 2. Comparación de las 12 fincas de cultivo de café tomadas del cuadro 1, con un método de k-medias. La designación de los conglomerados se hizo basado e en la similitud de los sitios con los descriptores de la estructura vegetal arbórea.

La especie con mayor abundancia para el grupo con una estructura baja (Figura 3) fue *Solenopsis geminata* seguida por *Solenopsis* sp1. y la especie con mayor abundancia para el grupo de estructura alta (Figura 3) también fue la especie *Solenopsis geminata*, seguida por *Forelius* sp1.

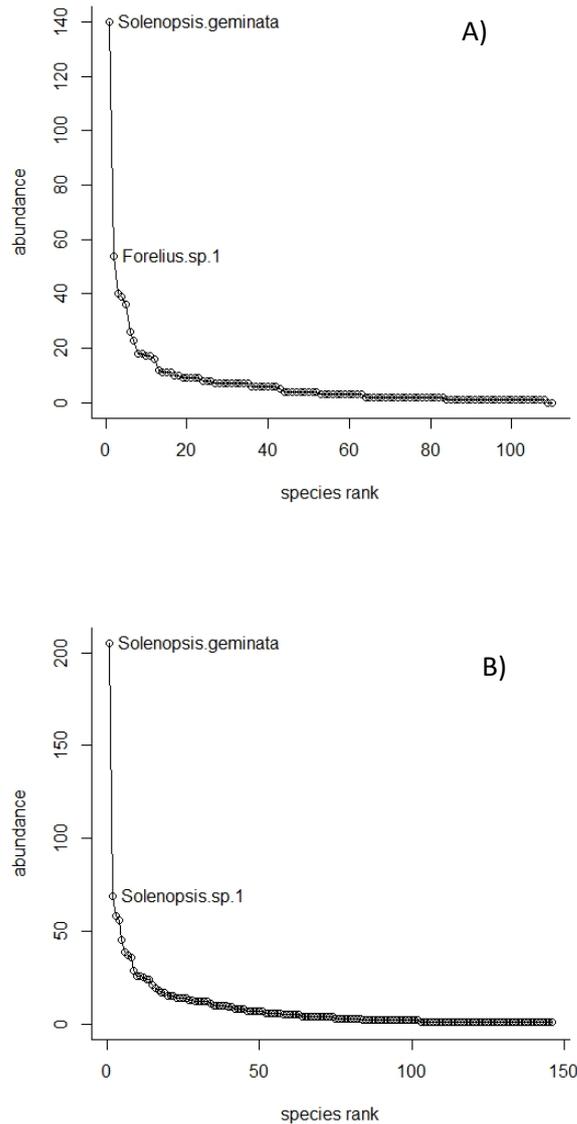


Figura 3. Curva de abundancias de especies de hormigas, A) grupo de fincas con estructura arbórea baja, B) grupo de fincas con estructura arbórea alta.

Predictores de la riqueza de especies de hormigas

De las seis variables de la estructura vegetal que se tomaron como posibles predictores sólo dos fueron significativas para el modelo. La altura de los árboles influyó como un predictor positivo con $p - \text{valor} = 0.009$ y la cobertura del dosel como un predictor negativo con un $p - \text{valor} = 0.041$. La ecuación final del modelo para la predicción de la riqueza de hormigas fue $\text{Especies de hormigas} = 68.80 + 2.92 \text{ Altura de los árboles} - 0.63 \text{ cobertura del dosel}$.

El modelo de la regresión lineal múltiple (Figura 4), fue estadísticamente significativo obteniendo los valores ($F_{2,9} = 6.132, p - \text{value} = 0.02$) y un ($R^2 = 0.576$), capaz de explicar el 57.6 de la variabilidad de la riqueza de hormigas.

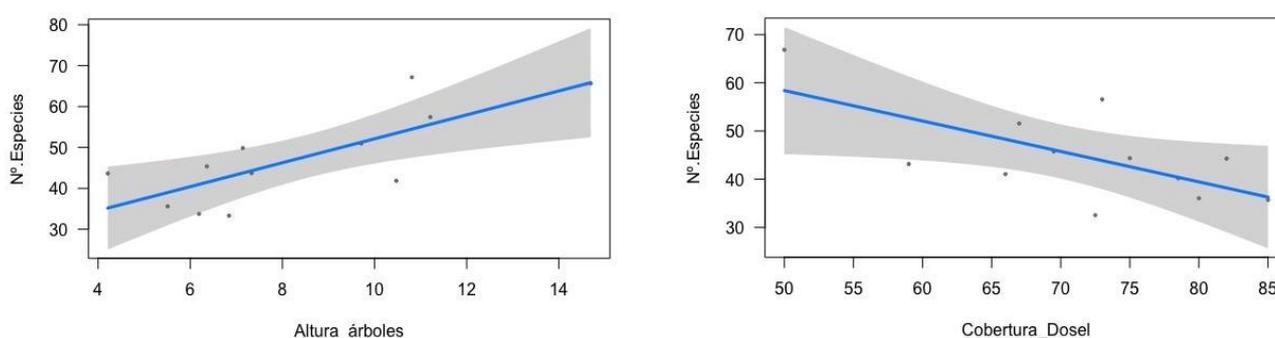


Figura 4. Regresiones parciales para la predicción del número de especies de hormigas en función de la altura de los árboles y la cobertura del dosel en 12 sitios de estudio.

Conclusiones

La pregunta inicial que se planteó fue si los agroecosistemas con mayor complejidad estructural arbórea son capaces de albergar mayor biodiversidad de hormigas, en contraste de los agroecosistemas con un manejo más intensivo donde la estructura arbórea disminuya por el manejo agrícola. La clasificación de las fincas de café se puede dar por factores como la estructura vegetal de los sitios, el manejo agrícola y el impacto hacia los productores (Moguel & Toledo, 1999). En nuestro estudio sólo se evaluó la calidad de la estructura arbórea. Los valores de las variables estructurales variaron entre ambos grupos, al igual que los promedios de las especies de hormigas. Sin embargo, nuestra hipótesis no fue respaldada estadísticamente. Estos resultados pueden deberse a factores externos a las fincas como la colindancia con otros cultivos, fragmentos de bosque o asentamientos urbanos que pueden estar influenciando en la composición poblacional de las hormigas.

Nuestros resultados no difieren con los resultados publicados por otros autores (De la Mora, Murnen, & Philpott, 2013), donde la diversidad de hormigas que anidan en hojarasca, ramas huecas y arbóreas tampoco cambia con respecto a las características de vegetación arbórea de las fincas. Los agroecosistemas con mayor diversidad de árboles aumentan la producción

de recursos como la aportación de material vegetal para diferentes invertebrados (Abebe Nigussie, 2012). Las fincas con mayor manejo agrícola como la poda de árboles desechan material vegetal al suelo como ramas y hojas aumentando los recursos para las hormigas que forrajean en el suelo, lo que simula un hábitat similar a las fincas con mayor número de árboles donde el material vegetal cae de forma natural por la saturación de plantas leñosas, lo que genera pocos cambios en la composición de especies de hormigas.

Los predictores de la estructura vegetal que influyen más sobre la riqueza de especies de hormigas fueron la altura de los árboles de forma positiva y la cobertura del dosel de forma negativa. Estos efectos pueden ser debido a que la altura de los árboles aumenta la disponibilidad de hábitat y recursos para las hormigas como néctar, flores y frutos, además de brindar sitios de anidación como estructuras huecas (Dejean, Corbara, Orivel, & Leponce, 2006). En cambio, el aumento de la cobertura del dosel puede evitar la entrada de la luz solar y disminuir la temperatura del microclima del hábitat, la cual influye en los aspectos de su biología, especialmente la tasa metabólica, el crecimiento y su desarrollo (Kadochová, & Frouz, 2013), lo que ocasiona que disminuya la actividad de las hormigas en estos sitios.

El género más abundante en nuestro estudio fue *Pheidole*, uno de los más diversos dentro de la familia Formicidae (Wilson, 2003), además de ser más abundante en la mayoría de los ecosistemas. La especie que representó la mayor abundancia fue *Solepnosis geminata*, la cual es perteneciente al grupo de las hormigas conocidas como hormigas de fuego. Este grupo tiene la característica de ser especies invasoras y oportunistas en sitios con cierta perturbación como los paisajes de cultivo (Trager, 1991).

Referencias

- Abebe Nigussie. (2012). The contribution of coffee agroecosystem to soil fertility in Southwestern Ethiopia. *African Journal of Agricultural Research*, 7(1). <https://doi.org/10.5897/AJAR11.1566>
- Agosti, D., Majer, J. D., Alonso, L. E., & Schultz, T. R. (2000). Standard methods for measuring and monitoring biodiversity. *Smithsonian Institution*, Washington DC, (9), 280.
- Andersen, A. N., Hoffmann, B. D., Muller, W. J., & Griffiths, A. D. (2002). Using ants as bioindicators in land management: Simplifying assessment of ant community responses. *Journal of Applied Ecology*, 39(1), 8-17. <https://doi.org/10.1046/j.1365-2664.2002.00704.x>
- Aristizábal, L., Bustillo, A., & Arthurs, S. (2016). Integrated Pest Management of Coffee Berry Borer: Strategies from Latin America that Could Be Useful for Coffee Farmers in Hawaii. *Insects*, 7(1), 6. <https://doi.org/10.3390/insects7010006>
- Armbrrecht, I., & Gallego, M. C. (2007). Testing ant predation on the coffee berry borer in shaded and sun coffee plantations in Colombia. *Entomologia Experimentalis et Applicata*, 124(3), 261-267. <https://doi.org/10.1111/j.1570-7458.2007.00574.x>
- Armbrrecht, I., Rivera, L., & Perfecto, I. (2005). Reduced Diversity and Complexity in the Leaf-Litter Ant Assemblage of Colombian Coffee Plantations. *Conservation Biology*, 19(3), 897-907. <https://doi.org/10.1111/j.1523-1739.2005.00062.x>
- Bolton, B. (1994). *Identification guide to the ant genera of the world*. Harvard University Press.

- Burd, M., & Howard, J. J. (2005). Central-place foraging continues beyond the nest entrance: The underground performance of leaf-cutting ants. *Animal Behaviour*, 70(4), 737-744. <https://doi.org/10.1016/j.anbehav.2004.12.012>
- Cammeraat, E. L. H., & Risch, A. C. (2008). The impact of ants on mineral soil properties and processes at different spatial scales. *Journal of Applied Entomology*, 132(4), 285-294. <https://doi.org/10.1111/j.1439-0418.2008.01281.x>
- De la Mora, A., Murnen, C. J., & Philpott, S. M. (2013). Local and landscape drivers of biodiversity of four groups of ants in coffee landscapes. *Biodiversity and Conservation*, 22(4), 871-888. <https://doi.org/10.1007/s10531-013-0454-z>
- Dejean, A., Corbara, B., Orivel, J., & Leponce, M. (2006). Rainforest Canopy Ants: The Implications of Territoriality and Predatory Behavior. *Functional Ecosystems and Communities*, 1(2), 105-120.
- Donald, P. F. (2004). Biodiversity Impacts of Some Agricultural Commodity Production Systems. *Conservation Biology*, 18(1), 17-38. <https://doi.org/10.1111/j.1523-1739.2004.01803.x>
- Dubey, S. R., Dixit, P., Singh, N., & Gupta, J. P. (2013). Infected Fruit Part Detection using K-Means Clustering Segmentation Technique. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2(2), 65. <https://doi.org/10.9781/ijimai.2013.229>
- Escobar-Ramírez, S., Tschardtke, T., Armbricht, I., Torres, W., & Grass, I. (2020). Decrease in B -diversity, but not in A -diversity, of ants in intensively managed coffee plantations. *Insect Conservation and Diversity*, 13(5), 445-455. <https://doi.org/10.1111/icad.12417>
- Griffith, D., Zamudio Grave, P., Cortés Viveros, R., & Cabrera Cabrera, J. (2017). Losing Labor: Coffee, Migration, and Economic Change in Veracruz, Mexico. *Culture, Agriculture, Food and Environment*, 39(1), 35-42. <https://doi.org/10.1111/cuag.12086>
- Hietz, P., & Hietz-Seifert, U. (1995). Structure and ecology of epiphyte communities of a cloud forest in central Veracruz, Mexico. *Journal of Vegetation Science*, 6(5), 719-728. <https://doi.org/10.2307/3236443>
- Hölldobler, B., & Wilson, E. O. (1990). *The ants*. Harvard University Press.
- Jha, S., Bacon, C. M., Philpott, S. M., Ernesto Méndez, V., Läderach, P., & Rice, R. A. (2014). Shade Coffee: Update on a Disappearing Refuge for Biodiversity. *BioScience*, 64(5), 416-428. <https://doi.org/10.1093/biosci/biu038>
- Kadochová, Š. & Frouz, J. (2013) Thermoregulation strategies in ants in comparison to other social insects, with a focus on red wood ants (Formica rufa group). *F1000Research*, 2, 280. <https://doi.org/10.12688/f1000research.2-280.v2>
- Knudsen JW. 1972. *Collecting and preserving plants and animals*. Harper and Row, New York. *Insects*, 128- 176.
- López-Gómez, A. M., Williams-Linera, G., & Manson, R. H. (2008). Tree species diversity and vegetation structure in shade coffee farms in Veracruz, Mexico. *Agriculture, ecosystems & environment*, 124(3-4), 160-172. <https://doi.org/10.1016/j.agee.2007.09.008>
- Mackay, W. P., & Mackay, E. (1989). Clave de los géneros de hormigas en México (Hymenoptera: Formicidae). In *Memorias II Simposio Nacional de Insectos Sociales*. Oaxtepec, Morelos (1-82).

- Moguel, P., & Toledo, V. M. (1999). Biodiversity Conservation in Traditional Coffee Systems of Mexico. *Conservation Biology*, 13(1), 11-21. <https://doi.org/10.1046/j.1523-1739.1999.97153.x>
- Noponen, M. R. A., Hagggar, J. P., Edwards-Jones, G., & Healey, J. R. (2013). Intensification of coffee systems can increase the effectiveness of REDD mechanisms. *Agricultural Systems*, 119, 1-9. <https://doi.org/10.1016/j.agsy.2013.03.006>
- Perfecto, I., & Snelling, R. (1995). Biodiversity and the Transformation of a Tropical Agroecosystem: Ants in Coffee Plantations. *Ecological Applications*, 5(4), 1084-1097. <https://doi.org/10.2307/2269356>
- Perfecto, I., Vandermeer, J., Hanson, P., & Cartiñ, V. (1997). Arthropod biodiversity loss and the transformation of a tropical agro-ecosystem. *Biodiversity and Conservation*, 6(7), 935-945. <https://doi.org/10.1023/A:1018359429106>
- Philpott, S. M. (2006). Ant patchiness: A spatially quantitative test in coffee agroecosystems. *Naturwissenschaften*, 93(8), 386-392. <https://doi.org/10.1007/s00114-006-0123-0>
- Philpott, S. M., Perfecto, I., & Vandermeer, J. (2008a). Behavioral Diversity of Predatory Arboreal Ants in Coffee Agroecosystems. *Environmental Entomology*, 37(1), 181-191. <https://doi.org/10.1093/ee/37.1.181>
- Philpott, S. M., Perfecto, I., & Vandermeer, J. (2008b). Effects of predatory ants on lower trophic levels across a gradient of coffee management complexity. *Journal of Animal Ecology*, 77(3), 505-511. <https://doi.org/10.1111/j.1365-2656.2008.01358.x>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591. <https://doi.org/10.2307/2333709>
- Toledo, V. M., & Moguel, P. (2012). Coffee and Sustainability: The Multiple Values of Traditional Shaded Coffee. *Journal of Sustainable Agriculture*, 36(3), 353-377. <https://doi.org/10.1080/10440046.2011.583719>
- Trager, J. C. (1991). A revision of the fire ants, *Solenopsis geminata* group (Hymenoptera: Formicidae: Myrmicinae). *Journal of the New York Entomological Society*, 141-198.
- Tscharntke, T., Tylianakis, J. M., Rand, T. A., Didham, R. K., Fahrig, L., Batáry, P., Bengtsson, J., Clough, Y., Crist, T. O., Dormann, C. F., Ewers, R. M., Fründ, J., Holt, R. D., Holzschuh, A., Klein, A. M., Kleijn, D., Kremen, C., Landis, D. A., Laurance, W., ... Westphal, C. (2012). Landscape moderation of biodiversity patterns and processes—Eight hypotheses. *Biological Reviews*, 87(3), 661-685. <https://doi.org/10.1111/j.1469-185X.2011.00216.x>
- Williams Linera G, Sosa V, Platas T. 1995. Structure and ecology of epiphyte communities of a cloud forest in central Veracruz, Mexico. *Journal of Vegetation Science* 6(5):719-728.
- Wilson, E. O. (2003). *Pheidole in the New World: a dominant, hyperdiverse hormiga genero* (Vol. 1). Harvard University Press.

EDITA: RED IBEROAMERICANA DE ACADEMIAS DE INVESTIGACIÓN A.C
DUBLÍN 34, FRACCIONAMIENTO MONTE MAGNO
C.P. 91190. XALAPA, VERACRUZ, MÉXICO.
CEL 2282386072
PONCIANO ARRIAGA 15, DESPACHO 101.
COLONIA TABACALERA
DELEGACIÓN CUAUHTÉMOC
C.P. 06030. MÉXICO, D.F. TEL. (55) 55660965
www.redibai.org
redibai@hotmail.com

Sello editorial: Red Iberoamericana de Academias de Investigación, A.C. (978-607-59328)
Primera Edición, Xalapa, Veracruz, México.
Presentación en medio electrónico digital
Formato PDF 20 MB
Fecha de aparición 27/10/2022
ISBN 978-607-59328-1-1

Derechos Reservados © Prohibida la reproducción total o parcial de este libro en cualquier forma o medio sin permiso escrito de la editorial.