

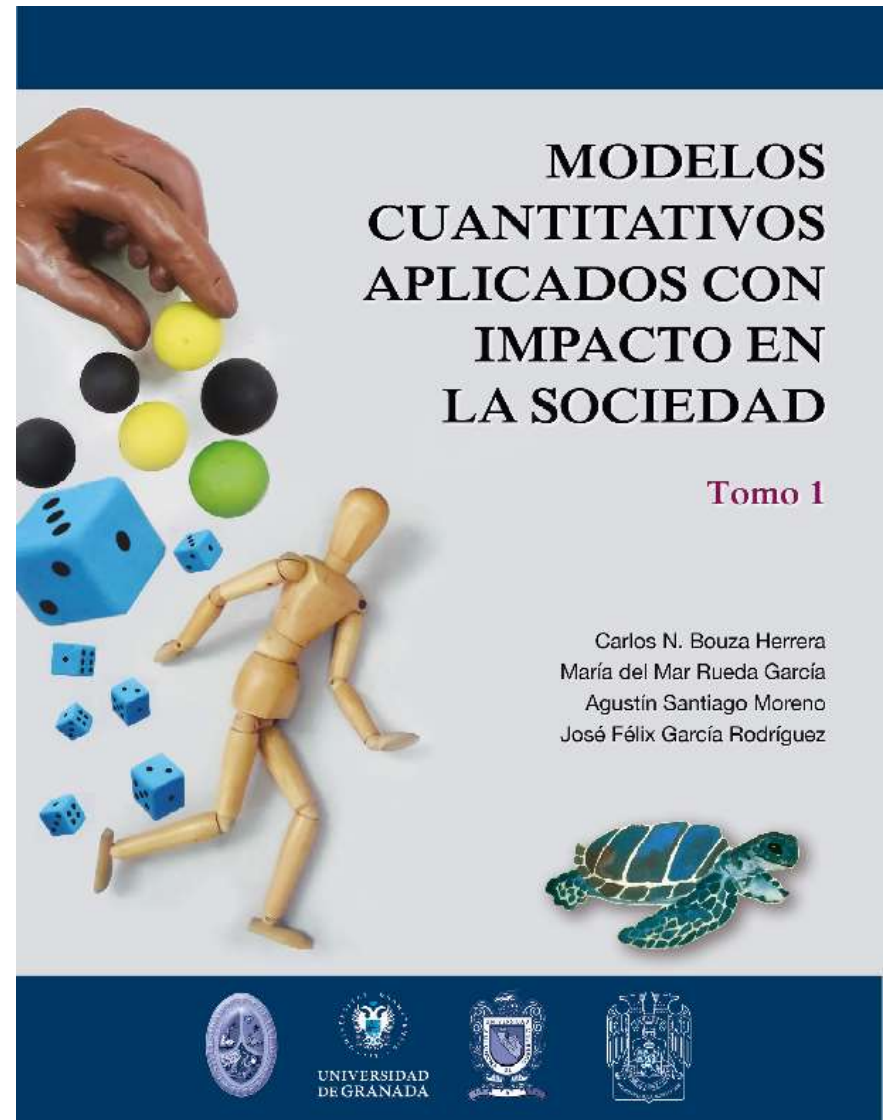
Modelos cuantitativos aplicados con impacto en la sociedad

Está compuesto de ocho trabajos de investigación elaborados por investigadores de los centros de investigación de México, Cuba y Colombia que colaboran en la Red Iberoamericana de Estudios Cuantitativos Aplicados (RIDECA). La presente obra da cuenta de la importancia de las investigaciones basadas en métodos cuantitativos para la explicación de situaciones problema e identificación de los métodos y modelos apropiados que contribuyen en su resolución y que finalmente permiten explicar la complejidad, las estrategias de mejora, y la predicción sobre los fenómenos de estudio.

De manera específica, en la obra se plantean investigaciones sobre patrones de distribución económica e innovación social en iniciativas de víctimas, costos y rentabilidad de la producción del limón persa, determinación de la probabilidad de que un cliente recomiende un hotel en seis ciudades latinoamericanas, la necesidad de implementar innovación en cursos introductorios de estadística: un estudio exploratorio en la educación superior, innovación y capital social: estudio de opinión universitaria, relación entre el desempleo docente y la satisfacción laboral de profesores de educación básica, modelos ocultos de Markov: aplicaciones en el ambiente y en las finanzas, una revisión de recientes enfoques en la identificación de clústeres en el análisis de grandes masas de datos. Estos trabajos dan cuenta de las intervenciones en lo local, nacional e internacional. En RIDECA se ha desarrollado un escenario propicio para la contribución científica, la divulgación y socialización de los saberes.

Carlos N. Bouza, Doctor en Ciencias por la Universidad de la Habana y profesor de la Facultad de Matemática y Computación; María del Mar Rueda García, Doctora en Matemáticas por la ugr y profesora en la Facultad de ciencias de la ugr; Agustín Santiago Moreno, Doctor en estadística e IO por la ugr y profesor en la Facultad de Matemáticas de la UAGro, México; José Félix García R. Doctor en Finanzas Públicas por la UV, México y profesor de la Universidad Juárez de Tabasco.

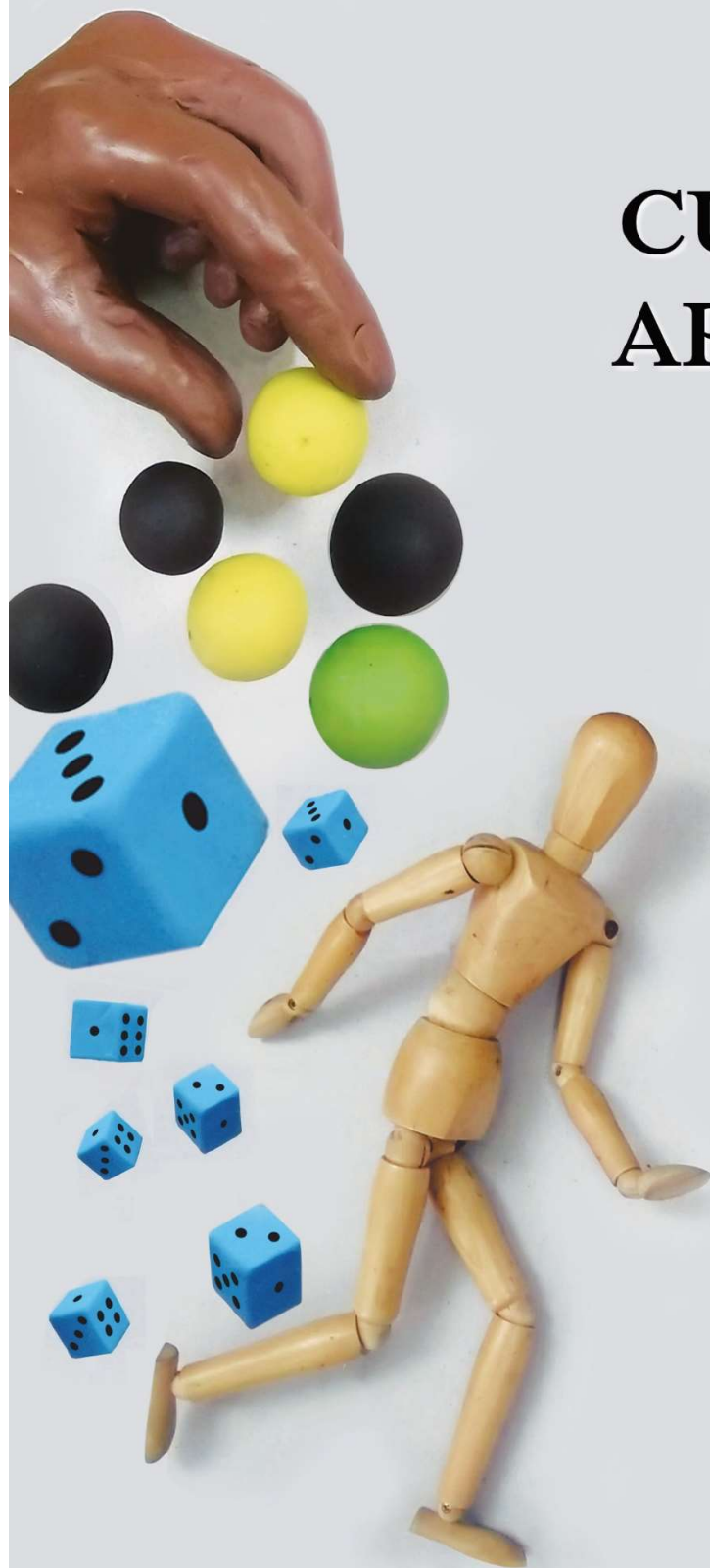
Editorial de la Universidad
Autónoma de Guerrero,
México, Año 2020.



MODELOS CUANTITATIVOS APLICADOS CON IMPACTO EN LA SOCIEDAD

Tomo 1

Carlos N. Bouza Herrera
María del Mar Rueda García
Agustín Santiago Moreno
José Félix García Rodríguez



Modelos Cuantitativos aplicados con impacto en la Sociedad

Tomo 1

Serie Métodos cuantitativos aplicados

Editores:

Carlos N. Bouza Herrera

María del Mar Rueda García

Agustín Santiago Moreno

José Félix García Rodríguez

Primera edición. Marzo del 2020.

© Universidad de la Habana, Cuba.
Universidad de Granada, España
Universidad Autónoma de Guerrero, México
Universidad Juárez Autónoma de Tabasco, México

ISBN: 978-607-9440-91-6

Editorial universitaria de la UAGro

El contenido de este libro es responsabilidad de los autores.
La presentación y disposición en conjunto de este libro son propiedad de los editores.
Ninguna parte de esta obra puede ser reproducida o transmitida, mediante ningún sistema o método, electrónico o mecánico, sin consentimiento de los autores.

Comité Científico:

Marco Negreiros, Universidad Estatal, Fortaleza, Brazil.

Pedro Monterrey, Universidad del Rosario, Colombia.

Amer Ibrahim Al-Omari / Ph.D. of Statistics Vice Dean of the Academic Research Department of Mathematics, Faculty of Science Al al-Bayt University, P.O. Box 130095, Mafrq 25113, Jordan. Mobile: 00962777906433

Marie Cottrell, Professeur émérite - Université Paris1. Tel et fax(prof): 33 1 44 07 89 22 SAMM, Université Paris 1, 90, rue de Tolbiac-75634, PARIS CEDEX 13-FRANCE. E-mail: cottrell@univ-paris1.fr

Jesús E. Sánchez García, departamento de física aplicada de ICIMAF, la Havana, Cuba.

Minerva Montero Díaz, departamento de física aplicada de ICIMAF, la Havana, Cuba.

Diseño de portada: Rosario Flores Sosa

Habana, Cuba, 5 marzo de 2020.

Índice de capítulos

CAPITULO	TÍTULO	AUTORES	PÁGINAS
1	PATRONES DE DISTRIBUCIÓN ECONÓMICA E INNOVACIÓN SOCIAL EN INICIATIVAS DE LAS VÍCTIMAS DE LA PALMA CUNDINAMARCA	Jenny Cubillos Díaz, Mercedes Gaitán Angulo, Jenny Paola Lis Gutiérrez	01-25
2	COSTOS Y RENTABILIDAD DE LA PRODUCCIÓN DEL LIMÓN PERSEA EN EL MUNICIPIO DE MISANTLA, VERACRUZ, MÉXICO	Ignacio Caamal Cauich, Verna Gricel Pat Fernández, José Félix García Rodríguez, Felipe Jerónimo Ascencio	27-48
3	DETERMINING PROBABILITY RECOMMENDATION OF HOTEL CUSTOMERS IN SIX LATIN AMERICAN CITIES	Carolina Heno, Andreina Moros, Yovanni Castro, Jenny Paola Lis-Gutiérrez, Manuel Ignacio Balaguera	49-60
4	SOBRE LA NECESIDAD DE IMPLEMENTAR INNOVACIÓN EN CURSOS INTRODUCTORIOS DE ESTADÍSTICA: UN ESTUDIO EXPLORATORIO EN LA EDUCACIÓN SUPERIOR MEXICANA	Cecilia Cruz López y Mario Miguel Ojeda Ramírez	61-91
5	INNOVACIÓN Y CAPITAL SOCIAL. ESTUDIO DE OPINIÓN UNIVERSITARIA EN EL SURESTE DE MÉXICO	José Félix García Rodríguez, Reyna Alejandra López Arias, Germán Martínez Prats, Ignacio Caamal Cauich, Aída Armenta Ramírez	93-102

6	RELACIÓN ENTRE EL DESEMPEÑO DOCENTE Y LA SATISFACCIÓN LABORAL DE PROFESORES DE EDUCACIÓN BÁSICA DE LA ZONA ESCOLAR 60, REGIÓN ACAPULCO-COYUCA	Tonantzin Martínez Bello, Elizabeth Ávila Cortes, Virginia García Cervantes y William de Jesús Merlán Avilés.	103-124
7	MODELOS OCULTOS DE MARKOV: APLICACIONES EN EL AMBIENTE Y EN LAS FINANZAS	María de Lourdes Sandoval Solís, Gladys Linares Fleites, Karla Mildred Cigarroa Alonso, Juan Badillo Iglecias	125-146
8	UNA REVISIÓN DE RECIENTES ENFOQUES EN LA IDENTIFICACIÓN DE CLÚSTERES EN EL ANÁLISIS DE GRANDES MASAS DE DATOS	Carlos N. Bouza Herrera, Agustín Santiago Moreno, Sira M. Allende Alonso, José M. Sautto-Vallejo	147-163

Nombre	Afiliación
Agustín Santiago Moreno	Facultad de Matemáticas, Universidad Autónoma de Guerrero, México.
Aída Armenta Ramírez	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
Andreina Moros	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
Carlos N. Bouza Herrera	Facultad de Matemática y computación, Universidad de la Habana
Carolina Henao	Corporación Universitaria Minuto de Dios (Bogotá, Colombia)
Cecilia Cruz López	Universidad Veracruzana
Elizabeth Ávila Cortes	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
Felipe Jerónimo Ascencio	Universidad Autónoma Chapingo (UACH).
Germán Martínez Prats	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
Gladys Linares Fleites	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México
Ignacio Caamal Cauch	Universidad Autónoma de Chapingo (UACH). Escuela de Economía
Jenny Paola Lis Gutiérrez	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
José Félix García Rodríguez	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
José M. Sautto Vallejo	Facultad de Matemáticas, Universidad Autónoma de Guerrero
Juan Badillo Iglecias	Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México
Karla Mildred Cigarroa Alonso	Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México
Manuel Ignacio Balaguera	Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)
María de Lourdes Sandoval Solís	Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México
Mario Miguel Ojeda Ramírez	Universidad Veracruzana

Reyna Alejandra López Arias	Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias Económico Administrativas (DACEA)
Sira M. Allende Alonso	Facultad de Matemática y computación, Universidad de la Habana
Tonantzin Martínez Bello	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
Verna Gricel Pat Fernández	Universidad Autónoma Chapingo (UACH).
Virginia García Cervantes	Coordinación General de Investigación Educativa, Secretaría de Educación en Guerrero, México.
William Merlán Avilés	Facultad de matemáticas de la Universidad Autónoma de Guerrero, México.
Yovanni Castro	Pontificia Universidad Javeriana (Bogotá, Colombia)

PRÓLOGO

El libro modelos cuantitativos aplicados con impacto en la sociedad, está compuesto de ocho trabajos de investigación elaborados por investigadores de los centros de investigación de México, Cuba y Colombia que colaboran en la Red Iberoamericana de Estudios Cuantitativos Aplicados (RIDECA). La presente obra da cuenta de la importancia de las investigaciones basadas en métodos cuantitativos para la explicación de situaciones problema e identificación de los métodos y modelos apropiados que contribuyen en su resolución y que finalmente permiten explicar la complejidad, las estrategias de mejora, y la predicción sobre los fenómenos de estudio.

De manera específica, en la obra se plantean investigaciones sobre patrones de distribución económica e innovación social en iniciativas de víctimas, costos y rentabilidad de la producción del limón persa, determinación de la probabilidad de que un cliente recomiende un hotel en seis ciudades latinoamericanas, la necesidad de implementar innovación en cursos introductorios de estadística: un estudio exploratorio en la educación superior, innovación y capital social: estudio de opinión universitaria, relación entre el desempleo docente y la satisfacción laboral de profesores de educación básica, modelos ocultos de Markov: aplicaciones en el ambiente y en las finanzas, una revisión de recientes enfoques en la identificación de clústeres en el análisis de grandes masas de datos. Estos trabajos dan cuenta de las intervenciones en lo local, nacional e internacional. En RIDECA se ha desarrollado un escenario propicio para la contribución científica, la divulgación y socialización de los saberes.

Dr. Armando Morales Carballo
Director de la Facultad de Matemáticas de la
Universidad Autónoma de Guerrero, México.

Marzo de 2020

CAPÍTULO 1

PATRONES DE DISTRIBUCIÓN ECONÓMICA E INNOVACIÓN SOCIAL EN INICIATIVAS DE LAS VÍCTIMAS DE LA PALMA CUNDINAMARCA

¹ Jenny Cubillos Díaz, Mercedes Gaitán-Angulo ¹, Jenny-Paola Lis-Gutiérrez¹

¹ Fundación Universitaria Konrad Lorenz, Bogotá D.C., Colombia

Jekacubillos@gmail.com
jenny.lis@konradlorenz.edu.co,
mercedes.gaitana@konradlorenz.edu.co

RESUMEN

Esta investigación tiene como propósito conocer cuál es la incidencia de las indemnizaciones económicas (producto de la restitución de derechos a las víctimas de la ley 1448 de 2011) en las iniciativas de innovación social impulsadas por las comunidades de víctimas del conflicto armado que apunten a la construcción de paz, en La Palma Cundinamarca. Para ello se realizó un instrumento de medición dividido en cuatro apartados; caracterización de la población, restitución de derechos económicos, patrones de distribución económica y por ultimo innovación social. Este instrumento fue aplicado a 100 participantes con el fin de analizar la incidencia y los usos que tuvo la reparación económica en este municipio y así determinar cuál es el desarrollo que las víctimas de la comunidad han hecho de los recursos obtenidos de la indemnización y si estos apuntaron a la construcción de paz establecida en la ya mencionada ley. Dentro de los principales hallazgos se identificó que las iniciativas de innovación social tienen poca relación con la reparación económica y por el contrario esta reparación al no contar con un acompañamiento por parte del estado termina por ser usado en gastos cotidianos o en mejoramiento de vivienda.

Palabras clave: Reparación económica, innovación social, víctimas del conflicto armado.

1. INTRODUCCIÓN

El conflicto armado colombiano ha sido uno de los conflictos latinoamericanos de más largo aliento y que ha dejado un gran número de víctimas en un periodo que data de aproximadamente 50 años. Hasta octubre del año 2018 se encuentran consignadas en el Registro Único de Víctimas (en adelante RUV) 8.746.541 personas por diferentes hechos de violencia, es decir uno de cada cinco colombianos fue víctima del conflicto armado, con la particularidad de que existen personas víctimas de más de un hecho victimizante (RUV, 2018). Este panorama refleja la complejidad del conflicto y evidencia la necesidad de establecer acciones que enmarcadas en el ámbito académico y social apunten a resolver las necesidades reales de esta población víctima¹.

Las consecuencias del conflicto armado son evidentes, las comunidades que se vieron afectadas de manera directa por los hechos de violencia, se encuentran en un periodo de retorno a la cotidianidad, enfrentándose a un contexto donde las dinámicas del conflicto, cambiaron la forma de relacionarse no solo entre ellos, sino con su territorio, cambiaron sus costumbres, sus actividades económicas entre otras acciones que definían su identidad. Esta transformación implica que las comunidades afrontan escenarios donde es vital la acción del estado y la acción de la comunidad para que lo ocurrido nunca vuelva a pasar y para que las gestiones que se realicen apunten a la construcción de un país en paz (Centro Nacional de Memoria Histórica, 2013).

Las poblaciones víctimas de hechos como el desplazamiento, masacres, torturas, desaparición forzada entre otras, por lo general estaban ubicadas en sectores estratégicos para los actores del conflicto y por ende padecían continuamente los enfrentamientos entre actores armados, viviendo en constante incertidumbre y miedo. Por lo general eran territorios rurales de difícil acceso que en su mayoría se encontraban en condición de pobreza y con poca presencia del estado. En este sentido es importante señalar que el conflicto armado interno acentuó las condiciones de vulnerabilidad a las cuales se veían expuestas las comunidades lo cual implicó que situaciones como: la desigualdad social, el poco acceso a salud y educación,

¹ Para esta investigación se utilizará la definición otorgada en la Ley 1448, 2011 la cual reconoce a las víctimas como: *“Aquellas personas que individual o colectivamente hayan sufrido un daño por hechos ocurridos a partir del 1° de enero de 1985, como consecuencia de infracciones al Derecho Internacional Humanitario o de violaciones graves y manifiestas a las normas internacionales de Derechos Humanos, ocurridas con ocasión del conflicto armado interno”*.

la falta de oportunidades, entre otras, se agudizaran (Centro Nacional de Memoria Histórica, 2017).

No obstante, no solo los sectores más alejados del centro del país vivieron de primera mano el conflicto. La violencia fue tan grande que llegó a las ciudades y a todos los estratos socioeconómicos del país. Fueron años en los que las acciones por terminar definitivamente el conflicto en ocasiones terminaron agudizándolo, generando nuevas condiciones de vulnerabilidad y acentuando las afectaciones sociales, económicas y ambientales que ya existían. Durante estos años las comunidades resistieron y sobrevivieron adaptándose al conflicto, movilizándose de su territorio, viviendo en zonas apartadas del conflicto y en ocasiones se vieron obligados a hacer parte del mismo, quebrando el tejido social que se había generado a lo largo de las generaciones.

Sin embargo, en la actualidad los colombianos se encuentran en un proceso de cambio, buscando alternativas para reparar los impactos de la guerra y construir un país en paz. Esta tarea involucra a todos los sectores inmersos en la sociedad, es una tarea de largo aliento que implica reconocer lo ocurrido y desarrollar estrategias que permitan a los sobrevivientes de la guerra tener una buena calidad de vida. Para ello estrategias como la atención integral, la construcción de la memoria histórica y el acceso a la verdad, son mecanismos que permiten una búsqueda real de justicia, reparación y garantías de no repetición (Benavides-Ponce, 2017)

Para dar respuesta a este panorama se crea la ley 1448 de 2011, la cual tiene como objetivo principal:

“Establecer un conjunto de medidas judiciales, administrativas, sociales y económicas, individuales y colectivas, en beneficio de las víctimas de las violaciones”... “dentro de un marco de justicia transicional, que posibiliten hacer efectivo el goce de sus derechos a la verdad, la justicia y la reparación con garantía de no repetición, de modo que se reconozca su condición de víctimas y se dignifique a través de la materialización de sus derechos constitucionales.” (Ley 1448, 2011. Art. 1).

De esta forma la ley nace como un mecanismo que reconoce un nuevo modo de abordar el conflicto donde se propende por un escenario de reivindicación social y dignificación de las víctimas el cual permite dar paso a acciones en pro de una reparación integral. La ley transforma una concepción que durante años denominaba vagamente a las víctimas como “población civil” o, peor aún, describía las acciones de violencia como “daños colaterales”. Por ende, la ley nace como la posibilidad de transformar no solo la concepción de las víctimas sino las acciones de dignificación y reparación reconociendo que lo ocurrido no debió pasar

y que el estado tiene una responsabilidad con restituir los derechos que fueron violados (Centro Nacional de Memoria Histórica, 2013).

Dentro de las medidas de reparación establecidas por la ley se encuentra las medidas de “restitución, indemnización, rehabilitación, satisfacción y garantías de no repetición” (Ley 1448 de 2011, artículo 25). La reparación económica establecida en acciones de indemnización ya sea individual o colectiva es una estrategia que busca satisfacer las necesidades económicas de las víctimas para suplir sus necesidades materiales y de esta forma dar condiciones para el restablecimiento de los derechos a las poblaciones para que puedan mejorar su calidad de vida. Lo que pretende esta indemnización es subsanar las necesidades materiales e intentar contribuir a la reparación integral.

El enfoque que tiene esta ley, tal como se estipula en los artículos 8, 11 y 12, apunta a la constitución de un escenario de reconciliación nacional y al establecimiento de una paz sostenible, en el cual la reparación integral es un elemento absolutamente indispensable. El carácter general de los principios que orientan esta ley, supone que lo establecido en ella sirve como base para la construcción de paz. Frente a este panorama general, una cuestión que parece presumirse con bastante facilidad es que los elementos incluidos en la reparación integral de las víctimas constituyen por sí mismos elementos para la adecuada construcción de paz.

Entendiendo que los principios que orientan la ley tienen esta pretensión, parece admitirse también que mecanismos como la indemnización económica de las víctimas, sirven inmediatamente a su reparación. Para fortalecer esta pretensión la ley estipula, dentro de los artículos 132, 133 y 134, las disposiciones para el uso de recursos otorgados por indemnización a las víctimas y establece algunas medidas de acompañamiento para la gestión adecuada de los recursos, con miras a la reconstrucción de los proyectos de vida de las víctimas. Sin embargo, lo que este estudio pretende es reconocer si ese acompañamiento es suficiente, y si la distribución que las víctimas realizan de los recursos obtenidos por indemnización toma un curso de inversión que permita al mediano y largo plazo la mejoría en su calidad de vida.

En este sentido, un estudio realizado por Villa e Insuasty en el 2015, sobre el significado en torno a la indemnización y la restitución de derechos a las víctimas, en el municipio de San Carlos, Antioquia, señala que existe un descontento colectivo en las poblaciones que han recibido beneficios restituidos, ya que muchas veces no se sienten conformes con los procesos de ejecución de los mismos y consideran que el acompañamiento que reciben no es suficiente. Esto hace que los recursos no sean destinados a la reparación ni a la reconstrucción de tejido social (Villa e Insuasty, 2015). Así pues, sin la pretensión de desconocer los esfuerzos nacionales por la reparación a víctimas del conflicto y los programas de

acompañamiento dentro de los mismos, resulta importante señalar cuáles métodos en la implementación de la ley han truncado el objetivo de la misma y cómo con el ánimo de construir un país en paz se pueden realizar cambios que permitan mejoras significativas en dichos procesos.

La Palma, Cundinamarca

La Palma es un municipio ubicado a 150 kilómetros, de la Capital del país, se encuentra situado en el departamento de Cundinamarca. Su fundación data de la época de la colonia, fundada por Antonio Toledo el 19 de noviembre de 1561 y su nombre hace alusión a la villa de nuestra señora de la Palma. Es una región que se caracteriza por sus tradiciones, por ser pionero en la producción de café, plátano y cacao. Es el centro de un extenso territorio que comprende los municipios de La Peña, El Peñón, Topaipí, Yacopí y Caparrapí (Meléndez, 2000). Ubicada en un lugar estratégico para el paso del centro del país a la zona andina ha sido escenario de enfrentamientos bélicos de toda índole que han hecho de este espacio un territorio en el que el conflicto armado ha dejado miles de víctimas y grandes secuelas por reparar.

Dentro de la historia documentada sobre este municipio durante los años más agudos del conflicto es posible hallar información sobre como el territorio ha estado en la mira de la guerra. Durante la década de los 50, la violencia bipartidista, provocó el desplazamiento de miles de Palmeros al centro del país, así como la muerte, desaparición y tortura de los que se quedaron. Además, para la época de los años 80, la población no estuvo alejada del conflicto con el narcotráfico, el terrorismo y la usurpación de tierras entre otros. Posteriormente en los años 60 la incursión del grupo armado FARC trajo consigo los rigores de la lucha armada.

Este grupo armado se ubicó en los municipios del norte de Cundinamarca, incluido La Palma, con el objetivo de financiar al Comando Central de las FARC ya que era ruta comercial que les permitía tener acceso a zonas centrales y económicamente importantes del país. Fueron años en los que el grupo armado tenía el control de la zona, se reclutaron hombres, mujeres y niños para hacer parte de las filas armadas, secuestros, extorsiones, boleteos, amenazas, desaparición forzada entre otros hechos que hicieron de la Palma un territorio en disputa.

No obstante, a mediados de 2001, fue cuando el conflicto se agudizó e inició la etapa más difícil y severa, ya que la disputa por la apropiación del territorio por parte de la guerrilla y las Autodefensas Unidas de Colombia (AUC) dejó a su paso un panorama de desolación, pobreza y abandono. Esto sin contar que en el territorio se puso en marcha la Fuerza de Despliegue Rápido (FUDRA), una unidad de combate de las fuerzas armadas cuya misión consistía en combatir los grupos al margen de la ley (Verdad Abierta, 2013). Este panorama dejó a su paso miles de víctimas de asesinatos selectivos, masacres, secuestros desapariciones

y sobre todo de desplazamiento forzado (Lizarazu, 2016). La mayoría de palmeros salieron en el año 2002. “Según cifras del Registro Único de Población Desplazada (RUPD) mientras en 2000 se presentaron 342 desplazamientos, en 2001 la cifra aumentó a 986 y en 2002 llegó a los 3.000 desplazados” (Verdad Abierta, 2013)

Los pobladores de La Palma fueron víctimas de la guerrilla, de los paramilitares y las fuerzas armadas. Fue una de las épocas más difíciles que tuvo que enfrentar el municipio ya que dentro de las consecuencias del conflicto además de la violación de los derechos a las víctimas estuvo el declive de la economía, y el abandono por parte del estado (Lizarazu, 2016). Cerca de 473 víctimas denunciaron ante la ley de Justicia y Paz que abandonaron sus fincas por la violencia que ocasionaron los paramilitares. La mayoría fueron desplazados masivamente del municipio por medio de acciones violentas. Según el Registro Único de Víctimas, 1.136 personas se vieron obligadas a salir en 2002 de este municipio. Es decir, que más de la tercera parte de los desplazados del departamento de Cundinamarca fueron desplazados de La Palma (RUV, 2018).

Frente al proceso de la restitución de los derechos de las víctimas y al proceso de construcción de paz en audiencia de imputación de cargos realizada en Bogotá, los ex paramilitares de las Autodefensas de Cundinamarca aceptaron haber desterrado a la gente de su tierra, por tal razón la Fiscalía los acusó de 148 hechos. Además de la imputación de cargos contra los ex paramilitares de las Autodefensas de Cundinamarca, las víctimas reiteraron que la Fuerza Pública no impidió que estos hechos ocurrieran, pues la base militar y la estación de Policía de La Palma nunca reportó ninguno de estos incidentes. (Verdad Abierta, 2013).

Sin embargo, debido a la magnitud de la violencia en este territorio, La Palma es un municipio priorizado en el proceso de reparación a víctimas del conflicto armado. Dentro de su cabecera municipal se encuentra una sede de la Unidad de Víctimas a la cual acuden miles de palmeros a hacer las respectivas denuncias de los hechos y a recibir las acciones de reparación específicas ordenadas por el estado dentro de la ley 1448, de 2011 anteriormente descrita.

El proceso de restitución ha sido largo y complejo debido a la cantidad de personas víctimas y sujetas de reparación. En los primeros años luego del retorno de la población al territorio, el estado incentivo ayudas humanitarias, las cuales consistían en un monto económico que tenían la pretensión de servir en una primera instancia para solventar las necesidades básicas de las personas. Estas ayudas fueron la base de la reparación integral, en la cual se continúa trabajando (Gobernación de Cundinamarca, 2016). Sin embargo, en la actualidad la Palma es un municipio que busca alternativas de progreso enfocadas en el trabajo conjunto de la comunidad y sobre todo en construir un territorio de paz que se preocupe por el desarrollo del campo y porque los años de violencia no vuelvan a ocurrir.

2. MARCO TEÓRICO

En la actualidad existe un creciente interés por buscar soluciones que, orientadas a reconstruir este tejido social –en donde se persigue que las poblaciones que han estado en vulnerabilidad, ya sea por contextos con condiciones precarias de desarrollo o con un historial de violencia, como es el caso del municipio la Palma– desarrollen iniciativas que propendan por mejorar la calidad de vida de las poblaciones. Las investigaciones actuales apuntan a encontrar soluciones alternativas que desde diferentes enfoques procuren generar un cambio en las dinámicas sociales y renueven el tejido social. La innovación social es uno de esos enfoques estratégicos, pues busca responder a las dificultades en cuanto a cómo crear y apoyar iniciativas de proyectos sostenibles que hagan frente a los problemas sociales que presentan las poblaciones, en este caso poblaciones que han sido víctimas del conflicto armado (Buckland y Murillo, 2014)

La *Innovación social* (en adelante IS) ha sido un tema central en las investigaciones académicas durante los últimos diez años. No solo en temas social, sino que se ha relacionado con el tema económico, ambiental, cultural, tecnológico entre otros. Este interés se debe a que la IS resulta ser una respuesta a muchas de las necesidades y problemáticas recientes en el mundo.

De esta manera el termino *innovación social* es entendido como innovación en bienes y servicios que abordan necesidades sociales y fallas del mercado, de esta forma apuntan a generar servicios y procesos de cambio en las relaciones sociales (van der Have y Rubalcaba, 2016; OCDE, 2011). Sin embargo estos cambios deben ser considerados como positivos ya que reconfiguran las prácticas sociales para responder a las necesidades y a los desafíos de la sociedad (Mulgan et al ., 2007), por ende estos cambios no solo involucran a la sociedad civil sino que buscan formas novedosas de establecer relaciones entre los miembros de la comunidad (empresas, el estado, las organizaciones sociales entre otros) en procesos que pueden ser potenciadores de cambios sociales y que beneficien a la gran mayoría (Grimm et al., 2013).

En la innovación social emergen varios factores que hacen que una iniciativa² sea favorable. No es lo mismo pensar en que una comunidad pueda realizar una iniciativa que favorece a

² La definición de una iniciativa de Innovación social adoptada por esta investigación corresponde a la planteada por Buckland y Murillo, (2014) la cual es entendida como: *Iniciativas, producto, proceso o programa que cambian profundamente las rutinas básicas, los recursos y las dinámicas de un sistema social a través de nuevas ideas ya sean productos, servicios y modelos que responden a las necesidades sociales y crean redes de colaboración y reciprocidad. (Pag. 6)*

las personas cercanas, que pensar en una asociación entre sectores que haga que los beneficiados de la misma sean a nivel local, regional y nacional. Las innovaciones sociales pueden ser desarrolladas por diversas organizaciones, (desde empresas orientadas a los beneficios corporativos hasta organizaciones netamente filantrópicas cuyo objetivo es mejorar el acceso a condiciones básicas de las comunidades en vulnerabilidad y luego como efecto secundario se enfocan en su rentabilidad (Andriesa, Daoub, y Verheydenc, 2018).

Por ende, diversas investigaciones a nivel mundial han enfatizado en la innovación social como una respuesta a las crisis económicas, sociales, ambientales y a los problemas de migración, pobreza, bienestar entre otros (Rogelja et al., 2018; Sabato et al., 2017). Sin embargo, para esta investigación se aborda la innovación social desde un enfoque sociológico (Cajaiba-Santana, 2014) y económico (Dawson y Daniel, 2010). En primer lugar, una conceptualización de la innovación social orientada sociológicamente reconoce que para alcanzar objetivos sociales que permitan mejorar la calidad de vida de los ciudadanos, es necesario que existan intenciones y acciones orientadas a impulsar un cambio social y eso como tal sería lo que genera en últimas la innovación social (Cajaiba-Santana, 2014).

En segundo lugar, un enfoque orientado hacia la economía reconoce que la innovación social tiene un potencial económico que permite dar solvencia a las necesidades monetarias de las poblaciones, por ende, la define como ideas, servicios, o transformaciones novedosas que involucran impactos sociales y beneficios generando asociaciones y colaboraciones entre sectores (empresas, organizaciones sociales, el estado entre otras) (Dawson y Daniel, 2010; Comisión Europea, 2013). Estas perspectivas de la innovación social se complementan. Autores como Ruiz y Parra (2013) combinan la perspectiva sociológica y económica argumentando que la innovación social es un proceso en el cual existe una fase de diseño y otra de implementación que junto con los procesos de difusión terminan promoviendo cambios en la organización social y generan fines económicos (Ruiz y Parra, 2013; Van der Have y Rubalcaba, 2016)

Sin embargo, la innovación social es un proceso que resulta difícil de medir y aún más difícil de promover. Varios autores coinciden en que la IS es algo que se emerge en las comunidades como un proceso automático de cooperación entre los distintos sectores de la comunidad y que su promoción resulta difícil en lugares donde no existe la asociación comunitaria. Por ende, autores como Dawson y Daniel (2010) y Buckland y Murillo (2014) coinciden en que al ser un proceso social, tiene inmerso dentro de sí variables difíciles de controlar que median

los resultados de las iniciativas y que son comunes en varios lugares en los que se ha intentado hacer un análisis cuantitativo del proceso de innovación social.

Estas características fácilmente identificables en las iniciativas de IS permiten identificar factores potencializadores de las mismas. Buckland y Murillo (2014) hacen un análisis exhaustivo de estas características las cuales denominan “variables de análisis de la Innovación Social”. La primera de ellas es *el impacto social* el cual mide hasta qué punto la iniciativa logra generar un impacto real que dé respuesta a las necesidades del contexto y logre transformar los desafíos a los cuales se enfrenta. En esta variable es importante saber cuál etapa de la iniciativa medir y cuál es el verdadero alcance al que se desea llegar con la iniciativa. La segunda característica es la *sostenibilidad económica* en la cual se mide el modelo de financiación de la iniciativa para ver si es sostenible a lo largo del tiempo y si existe un equilibrio entre el impacto económico y social de la iniciativa.

La tercera variable es *el tipo de innovación* en esta se mide las características propias de la iniciativa es decir si es de tipo abierto (personas externas pueden hacer parte) o cerrada (beneficia a un sector específico), si el grado de cambio es Incremental (pequeñas modificaciones sociales que no requieren gran inversión), Semirradical (Cambios significativos, pero no radicales) Radical (Cambios sustanciales que requieren grandes esfuerzos, inversión y alto riesgo). También hace referencia al alcance de la iniciativa, es decir si es de bienes o servicios, procesos, métodos de comercialización o de organización. La cuarta variable es *la colaboración intersectorial* esta pretende responder al tipo de agremiación y vínculo entre los sectores de las comunidades, es un análisis de los actores inmersos y cuál es su papel dentro de la iniciativa. Y por último la quinta característica es la escalabilidad y la replicabilidad, las cuales pretenden conocer en qué medida la iniciativa es replicable y si esta se puede ampliar y promover en contextos con características similares o diversos.

Sin embargo, la innovación social sigue siendo un tema al cual resulta difícil acceder en comunidades en condición de vulnerabilidad, el desarrollo de estrategias de medición y promoción de la IS requiere de la participación de la comunidad, el sector empresarial, el estado y las agremiaciones cercanas a la misma que puedan servir de apoyo para que las iniciativas surjan en contextos altamente complejos para su desarrollo y sobre todo para que estas iniciativas se mantengan a lo largo del tiempo, generando impactos positivos en la comunidad a lo largo de varias generaciones.

De este modo, tener un diagnóstico inicial sobre como se ve la IS en lugares como La Palma Cundinamarca es una estrategia útil para pensar en políticas públicas de desarrollo económico y social que permitan establecer en cuáles puntos se pueden enfocar las iniciativas y como solventar las dificultades que van atravesando. Por ende y de acuerdo con los

planteamientos señalados anteriormente, esta investigación pretende comprender ¿Cuál es la incidencia de las indemnizaciones económicas (producto de la restitución de derechos a las víctimas de la ley 1448 de 2011) en las iniciativas de innovación social impulsadas por las comunidades que apunten a la construcción de paz, en La Palma Cundinamarca?

Para dar respuesta a esta cuestión, se realizó un estudio de caso con las víctimas del conflicto armado de La Palma, Cundinamarca, con el objetivo de observar cuál es el uso que esta comunidad hace de los recursos obtenidos de la indemnización por parte del estado según la ley 1448 de 2011 y si este uso tiene o no relación con la innovación social. De igual forma realizó un breve rastreo de la dinámica estatal de acompañamiento y se examinó la posibilidad de promover alternativas de innovación social enfocada hacia la construcción de paz. Con base en este objetivo se plantearon las siguientes hipótesis:

- Hipótesis 1: La indemnización económica proveniente de la reparación integral a víctimas no cuenta con el suficiente acompañamiento para promover iniciativas de innovación social y por el contrario es usado en gastos diarios.
- Hipótesis 2: La efectividad de las medidas de restitución de derechos proviene de un trabajo en conjunto entre las iniciativas de la comunidad, la indemnización producto de la restitución de derechos a las víctimas y un acompañamiento integral
- Hipótesis 3. La innovación social inmersa en las iniciativas de las víctimas no depende de la restitución de derechos por indemnización

3. PROCEDIMIENTO

Para la realización de esta investigación, en una primera fase y con base las variables propuestas por los autores Buckland y Murillo (2014), se realizó un instrumento de medición con preguntas orientadoras para cada categoría con el fin de conocer si en un contexto como La Palma con las características históricas y sociales ya descritas es posible encontrar innovación social que promueva cambios significativos para responder a las necesidades actuales del contexto, en el cual se buscan soluciones a las secuelas producidas por el conflicto. Pero sobre todo si la innovación social puede ser una respuesta a las necesidades sociales de las víctimas en pro de su reparación integral.

Para ello el instrumento consistió en cuatro secciones. La primera de ella fue *Caracterización de la población*, en esta sección en esta categoría se pretendió ahondar en los principales componentes que caracterizan a la población, empezando por su ubicación demográfica, y

las características sociales de las que hacen parte. La segunda sección denominada *Restitución de derechos económicos* se quería conocer la existencia o no de restitución económica en la población y cómo ha sido el acompañamiento de estos procesos. La tercera parte *Patrones de distribución* quería conocer el uso que la población ha hecho de sus recursos ya sea por restitución de derechos o por sus ingresos mensuales. Y por último la sección *Innovación social* donde se pretendía dar respuesta a la presencia o no de iniciativas de innovación social y cuáles eran sus principales características dentro del territorio.

El cuestionario está compuesto por 30 preguntas de selección múltiple y 12 preguntas abiertas, para un total de 42 preguntas. Fue evaluado por expertos y por un comité de bioética ya que la recolección de datos de se haría de fuentes primarias a una comunidad en condición de vulnerabilidad. Por esta razón también se realizó un pilotaje del instrumento para corroborar que no tuviera inconvenientes en el lenguaje utilizado y que no causara re victimización. La recolección de datos se realizó en la cabecera municipal de La Palma, participaron 100 personas las cuales se encontraban registradas en el RUV. Es importante señalar que al realizar el pilotaje se pudo establecer que las iniciativas de innovación social eran escasas, por lo cual se desarrollaron seis entrevistas a personas de la comunidad para ahondar en las razones por las cuales se presentaba este fenómeno y rastrear las posibles alternativas para el desarrollo de las mismas.

MÉTODO

Participantes

En esta investigación se realizó un muestreo voluntario de 100 participantes tomado al azar.

Los participantes fueron hombres y mujeres mayores de edad, inscritos en el RUV, pertenecientes a los niveles socioeconómicos 1, 2 y 3 del municipio de La Palma Cundinamarca.

El cálculo de la muestra se realizó tomando como base un tamaño de población de $N=8.000$ personas habitantes del municipio, con un intervalo de confianza del 95%, y un error muestral para una variable de diseño cualitativa dicotómica bajo muestreo aleatorio simple asumiendo error absoluto del 9,75%, lo que lleva a un tamaño de muestra $n=100$ personas.

Diseño

Para la presente investigación se aplicó un instrumento de encuesta con validación de contenido efectuada por tres jueces expertos, y una prueba piloto aplicada a 12 personas para testear el nivel de comprensión del instrumento.

El instrumento fue aplicado por dos personas conocedoras del tema, nivel de formación de doctorado, con experiencia de manejo de personas en campo.

La encuesta se efectuó en forma presencial donde el entrevistador diligencio el instrumento, con una duración de 23.5 minutos. Los datos fueron digitadas y consolidados en un archivo de Excel por una joven investigadora profesional capacitada para este fin. Los datos se verificaron en cuanto a consistencia, depurando la información.

PROCEDIMIENTO

Cada uno de los participantes realizó la encuesta siguiendo la secuencia que se presenta a continuación:

Primero se hizo la lectura del consentimiento informado, segundo se Aplicación del filtro para validación de criterios de inclusión y exclusión, tercero lectura de las instrucciones generales y por último se aplicó el instrumento.

Análisis

Los datos obtenidos se procesaron mediante el analizaron mediante SPSS Statistics 25. A través de este se calcularon los estadísticos descriptivos.

Consideraciones éticas

De acuerdo a lo establecido en la ley 1090 del ejercicio profesional del psicólogo, desde el punto de vista ético y procedimental se cumplió con lo siguiente:

1. Se suministró información clara acerca del tipo de investigación realizada y los responsables de la misma y la institución para la cual se estaba llevando a cabo. Se informó también el carácter confidencial de la información suministrada y de uso exclusivo de los investigadores.

2. Se dio información suficiente en la que se explicó el propósito de la investigación.
3. Se informó la duración promedio de la prueba, así como se explicó la tarea a realizar, así como se informó que ésta no representa ningún efecto secundario o perjuicio de salud en los participantes.
4. Se dejó explícita la manifestación de participación voluntaria en la investigación y se informó acerca del incentivo recibido como retribución de su participación.
5. También se informó que se podía abandonar la prueba en cualquier momento de manera voluntaria.

RESULTADOS

Resultados Cuantitativos

A continuación, se presentan los resultados de la aplicación del instrumento:

Tabla 1. Lugar de nacimiento

Lugar	Frecuencia	Porcentaje
Bogotá	2	2,0
El Ato	1	1,0
La Palma	85	85,0
Pasca	1	1,0
Utica	1	1,0
Yacopi	10	10,0
Total	100	100,0

En esta tabla están descritos los lugares de procedencia de los encuestados, la mayoría de los participantes nacieron en el municipio de La Palma.

Tabla 2. Estrato socioeconómico

Estrato	Frecuencia	Porcentaje
1	45	45,0
2	50	50,0
3	5	5,0
Total	100	100,0

Los participantes pertenecían al en su mayoría al estrato socioeconómico 1 y 2

Tabla 3. Sexo

Sexo	Frecuencia	Porcentaje
Femenino	73	73,0
Masculino	27	27,0
Total	100	100,0

Dentro de los participantes un 73 % pertenecían a sexo femenino y 27% al masculino

Tabla 4. Lugar de residencia

Residencia	Frecuencia	Porcentaje
Cabecera Municipal	88	88,0
Zona Rural	12	12,0
Total	100	100,0

El 88% de las personas que participaron de la aplicación del instrumento pertenecían a la cabecera municipal de la Palma y el 12% a la zona rural.

Tabla 5. Ocupación

Ocupación	Frecuencia	Porcentaje
Ama de casa	39	39,0
Campesina	2	2,0
Comerciante	12	12,0
Desempleada	5	5,0
Empleado	6	6,0
Independiente	25	25,0
Obrero	9	9,0
Profesora	1	1,0
Promotora de Salud	1	1,0
Total	100	100,0

Dentro de las ocupaciones de los encuestados la mayoría (39%) tienen como ocupación ama de casa, seguido de Independiente en un 25%.

Tabla 6. Nivel de escolaridad

Escolaridad	Frecuencia	Porcentaje
Primaria	41	41,0
Secundaria	46	46,0
Técnico	13	13,0
Total	100	100,0

El nivel de escolaridad de los encuestados era bajo ya que la mayoría 46% habían culminado el bachillerato y 41% la escuela primaria.

Tabla 7. Hecho víctimizante

Hecho víctimizante	Frecuencia	Porcentaje
Amenaza	2	2,0
Desaparición forzada	1	1,0
Desplazamiento	97	97,0
Total	100	100,0

La tabla evidencia el tipo de hecho por el cual se encuentra registrada la persona ante el registro único de víctimas.

Tabla 8. Si ha sido reparado

Reparado	Frecuencia	Porcentaje
No	37	37,0
Si	63	63,0
Total	100	100,0

En la tabla se puede identificar cuantas personas han sido o no reparadas

Tabla 9. Tipo de medidas de reparación

Medidas de Reparación	Frecuencia	Porcentaje
No ha recibido medidas de reparación	37	37,0
Ayuda Humanitaria	52	52,0
Indemnización Económica	11	11,0
Total	100	100,0

La tabla evidencia las medidas de reparación que han recibido los participantes, el 52% fue beneficiario por ayuda humanitaria.

Tabla 10. Uso de la Reparación

Uso de la reparación	Frecuencia	Porcentaje
No ha recibido medidas de reparación	37	37,0
Arreglos de la casa	2	2,0
Ya ha hecho uso de estos recursos en gastos cotidianos	40	40,0
Ya ha hecho uso de estos recursos en otras inversiones	21	21,0
Total	100	100,0

Al preguntarle a los participantes si ya habían hecho uso de la reparación recibida la respuesta más frecuente con un 40% “Ya ha hecho uso de estos recursos en gastos cotidianos”

Tabla 11. Tipo de inversiones

Tipo de inversiones	Frecuencia	Porcentaje
Gastos cotidianos	79	79,0
Arreglos de la casa	10	10,0
Ayudar a mi familia	3	3,0
Compra animales	3	3,0
Negocio	4	4,0
Comprar un computador para mi hijo	1	1,0
Total	100	100,0

Dentro de los usos que la población ha hecho de la reparación económica esta “gastos cotidianos” como la respuesta más frecuente con 79%

Tabla 12. Acompañamiento

Acompañamiento	Frecuencia	Porcentaje
No ha recibido restitución	39	39,0
No	57	57,0
Si	4	4,0
Total	100	100,0

El acompañamiento recibido por parte del estado a las personas encuestadas ha sido poco de 100 participantes solo 4 han recibido algún tipo de acompañamiento.

Tabla 13. Tipo de acompañamiento

Tipo acompañamiento	Frecuencia	Porcentaje
No ha recibido ningún tipo de restitución	96	96,0
Acompañamiento	1	1,0
Asesoramiento	1	1,0
Atención Psicológica	1	1,0
Atención Psicológica por tres meses	1	1,0
Total	100	100,0

En la tabla se evidencia el tipo de acompañamiento que han recibido las personas encuestadas.

Tabla 14. Tipo de reparación

Tipo de reparación	Frecuencia	Porcentaje
No ha recibido restitución	37	37,0
Colectiva	4	4,0
Individual	59	59,0
Total	100	100,0

La tabla evidencia si la reparación recibida fue individual o colectiva

Tabla 15. Ingresos de la población

Ingresos	Frecuencia	Porcentaje
0 a 500,000	71	71,0
1,000,001 a 2,000,000	11	11,0
500,001 a 1,000,000	18	18,0
Total	100	100,0

Los ingresos de la población están en su mayoría entre 0 a 500.000 pesos (71%), seguidos por un 500,001 a 1,000,000 con un 18%.

Tabla 16. Egresos de la población

Egresos	Frecuencia	Porcentaje
0 a 500,000	61	61,0
1,000,001 a 2,000,000	12	12,0
500,001 a 1,000,000	27	27,0
Total	100	100,0

Los egresos de la población están en su mayoría entre 0 a 500.000 pesos (61%) seguidos por 27% entre 500,001 a 1,000,000.

Tabla 17. Deudas con entidades financieras

Tiene deuda	Frecuencia	Porcentaje
No	83	83,0
Si	17	17,0
Total	100	100,0

Dentro de la población encuestada el 83% no tiene deudas con entidades financieras

Tabla 18. Valor de la deuda

Valor Crédito	Frecuencia	Porcentaje
No tiene deudas con entidades financieras	82	82,0
de 1,000,000 a 5,000,000 pesos	12	12,0
De 50,000 a 500,000 pesos	3	3,0
Más de 5,000,000 de pesos	1	1,0
Más de 500,001 de pesos	2	2,0
Total	100	100,0

El 12% de la población tiene créditos de 1,000,000 a 5,000,000 pesos seguidos por un 3% que tiene deudas de 50,000 a 500,000 pesos.

Tabla 19. Uso del crédito en una iniciativa

Uso crédito iniciativa	Frecuencia	Porcentaje
No tiene deudas con entidades financieras	82	82,0
No	4	4,0
Si	14	14,0
Total	100	100,0

Solo el 14% de las personas encuestadas ha hecho uso del crédito en una iniciativa de Innovación social.

Tabla 20. Innovación social

Ha escuchado termino innovación social	Frecuencia	Porcentaje
No	91	91,0
Si	9	9,0
Total	100	100,0

Para la mayoría de las personas (91%) encuestadas el término Innovación social era completamente desconocido

Tabla 21. Definición coincide

Definición coincide	Frecuencia	Porcentaje
Coincide parcialmente	9	9,0
No aplica porque no tenía conocimientos previos	91	91,0
Total	100	100,0

Para el 9% de personas que conocían el término Innovación social este coincidía con la definición propuesta en el instrumento.

Tabla 22. Estadísticos

		Edad	personas hogar	personas dependientes
N	Valido	100	100	100
Mean		50,30	3,50	1,96
Std. Error of Mean		1,128	0,140	0,142
Median		52,00	4,00	2,00
Std. Deviation		11,277	1,403	1,421
Minimum		19	1	0
Maximum		75	6	5
Percentiles	10	35,10	2,00	0,00
	20	40,00	2,00	1,00
	25	44,00	2,00	1,00
	30	44,30	3,00	1,00
	40	48,40	3,00	1,00
	50	52,00	4,00	2,00
	60	55,00	4,00	2,00
	70	57,00	4,00	3,00
	75	60,00	4,75	3,00
	80	60,00	5,00	3,00
	90	62,90	5,00	4,00

Tabla 23. Número de personas del hogar

personas hogar	Frequency	Percent	Valid Percent	Cumulative Percent
1	8	8,0	8,0	8,0
2	20	20,0	20,0	28,0
3	19	19,0	19,0	47,0
4	28	28,0	28,0	75,0
5	17	17,0	17,0	92,0
6	8	8,0	8,0	100,0
Total	100	100,0	100,0	

Tabla 24. Número de personas dependientes

personas dependientes	Frequency	Percent	Valid Percent	Cumulative Percent
0	16	16,0	16,0	16,0
1	27	27,0	27,0	43,0
2	24	24,0	24,0	67,0
3	16	16,0	16,0	83,0
4	12	12,0	12,0	95,0
5	5	5,0	5,0	100,0
Total	100	100,0	100,0	

Resultados Cualitativos

Dentro de las entrevistas realizadas fue posible identificar algunos factores importantes en los procesos de reparación integral que vale la pena resaltar. Para algunas de las personas entrevistadas el proceso ha sido largo y no ha tenido los resultados que esperaban. “Los alcaldes de turno han reparado según la conveniencia. Los primeros favorecidos fueron las personas de la alcaldía o que estaban en cargos públicos” (Entrevistado 1). “Por amenazas el alcalde se fue en el 2002 y el municipio quedo abandonado la gente no sabía qué hacer, por eso ahora mucha gente no cree en las instituciones porque piensan que no les van a ayudar de verdad” (Entrevistado 2).

Los entrevistados señalan que el conflicto ha dejado secuelas que requieren más que ayudas económicas para ser reparadas “Se nota en algunos muchachos el impacto de lo que generaron los grupos armados. Porque son muchachos que quieren ser comandantes o irse al monte en lugar de trabajar” (Entrevistado 3) “Es que los muchachos veían que esa gente tenía mucha plata y poder entonces no quieren estudiar sino ser como ellos” (Entrevistado 4).

En este sentido para ellos el acompañamiento es vital y debe ser un proceso de largo aliento que les permita apropiarse de las ventajas de su territorio “Lo que yo veo en reparación es plata, acompañamiento no ha habido, en lo que se necesita de programas educativos no ha habido, ni en salud, es mas en recuperamiento de las parcelas tampoco, no hay un acompañamiento técnico” (Entrevistado 1).

Para los entrevistados el desarrollo del campo es vital, la mayoría concuerda en que hace falta un modelo educativo que se enfoque en recuperar el campo y aprovechar los recursos que brinda el territorio “aquí lo que usted quiera sembrar se da, el problema es que no hay

cooperación para sacar los productos y comercializarlos” (entrevistado 2) “Aquí hay mano trabajadora, hay personas que tienen un oficio y saben cómo hacerlo, pero no consiguen trabajo fácil, porque aquí hay muy pocas fuentes de trabajo y a la gente le toca irse a las ciudades a ver que pueden hacer” (Entrevistado 6).

Durante las entrevistas se preguntó a los participantes si ellos habían hecho uso de los recursos económicos producto de la indemnización y si recordaban la finalidad de este dinero, la mayoría concordó en que había sido para reparaciones del hogar o para gastos cotidianos, ninguno realizó inversiones o gastos diferentes a estos “esa platica la esperamos por varios años, pero cuando nos la dieron teníamos muchas deudas y por lo menos las terminamos de pagar y lo que quedo se fue yendo en la comida y los gastos diarios (entrevistado 2) “Nunca pensamos en invertirla porque no sabíamos cómo o en qué, entonces la usamos para arreglar la casa y comprar mercado” (Entrevistado 3)

Al preguntarle a los entrevistados por el desarrollo de iniciativas de innovación social y como perciben ellos la asociación y la cooperación, factores clave en la IS ellos señalan “Aquí no hay ninguna forma de asociatividad, aquí cada quien vive de su propio proyecto, no hay asociación comunitaria. La gente vive del pan coger o del comercio, pero hay muy poca colaboración entre vecinos” (Entrevistado 3) “Yo creo que tiene que ver con la desconfianza, con que hay muchos que no han recibido su indemnización y que no tiene un ingreso estable o una forma de colaborar” (Entrevistado 5).

Por otra parte, los entrevistados concuerdan en que las iniciativas han existido, aunque muchas de ellas vienen de políticas gubernamentales solo se mantienen hasta que la política cambia o hasta que se ven los primeros resultados “Si han tratado de hacer iniciativas que asociación, pero no se mantienen en el tiempo, son dadas por los intereses de las políticas del momento entonces uno va, se hace una reunión dan un refrigerio y ya no se vuelve a hablar del tema”

Al preguntarles por sus expectativas frente a los procesos de reparación los entrevistados son claros en que lo que necesitan es cooperación y apoyo, pero que no están de acuerdo con algunos procesos de asistencia económica que se han desarrollado en el municipio “Las acciones concretas consisten en hacer las cosas que, si son importantes, vías, un sistema de acueducto, un sistema bueno de basuras y proyectos que mejoren las condiciones de vida” (Entrevistado 5). “Me gustaría que los grandes estamentos del estado no estigmaticen más La Palma, desplazados no, no queremos regalos, queremos acciones concretas, porque esto nos ha llevado cada vez más a la pobreza, vemos un pueblo que perdió economía, perdió capacidad operativa, porque la gente se acostumbró a esperar los recursos económicos” (Entrevistado 1)

En contraste con lo anterior para algunos entrevistados la ayuda no ha sido suficiente o no la han recibido y esperan que esta reparación económica sea la que beneficie algunas ideas que quisieran desarrollar “ Yo solo espero que salga esa plata para poder hacer el taller de pintura, eso es lo único que yo quiero después de tantos años de sufrimientos” (Entrevistado 6) “A veces lo llaman a uno de la unidad o uno va a ver cómo va el proceso y le dicen que se demora o que mejor vaya después, entonces uno ya no espera esa plata sino que trata de trabajar y hacer lo más que pueda” (Entrevistado 2). “A mí nunca me dijeron que podía hacer con el dinero, aunque no era mucho, pero pues si era una inversión segura o algo en lo que pudiera trabajar yo la hubiera dado” (Entrevistado 4)

CONCLUSIONES

Es importante señalar que aunque la finalidad de este estudio era encontrar ¿Cuál es la incidencia de las indemnizaciones económicas (producto de la restitución de derechos a las víctimas de la ley 1448 de 2011) en las iniciativas de innovación social impulsadas por las comunidades que apunten a la construcción de paz, en La Palma Cundinamarca?, la aplicación inicial del instrumento termino por transformar la recolección de información enfocando los resultados hacia el ¿Por qué existen pocas iniciativas de innovación social y como estas pueden ser una solución para algunos de los problemas sociales que vive en la actualidad el municipio?

De este modo y en contraste con las hipótesis inicialmente planteadas se puede confirmar que “La indemnización económica proveniente de la reparación integral a víctimas no cuenta con el suficiente acompañamiento para promover iniciativas de innovación social y por el contrario es usado en gastos diarios. Esto fue evidente a lo largo de la investigación, tanto en los resultados cualitativos como cuantitativos.

Por otra parte y frente a la hipótesis de que “la efectividad de las medidas de restitución de derechos proviene de un trabajo en conjunto entre las iniciativas de la comunidad, la indemnización producto de la restitución de derechos a las víctimas y un acompañamiento integral” es difícil corroborar que una iniciativa producto del trabajo comunitario pueda ser la solución a las necesidades de la población ya que en cuanto a la innovación social es muy poca la información presente en el territorio.

Sin embargo, se pudieron evidenciar ciertos factores recurrentes en la falta de iniciativas de innovación social. La primera es la falta de ingresos y oportunidades laborales de la población, la mayoría de personas encuestadas tenían ingresos menores a 500.000 pesos y su ocupación era amas de casa u obreros. Por ende y en contraste con lo establecido en la ley de victimas las ayudas humanitarias y la indemnización económica no parecen ser suficientes para que exista una reparación integral. De las 100 personas encuestadas solo 4 habían recibido algún tipo de acompañamiento.

Y por último decir que “La innovación social inmersa en las iniciativas de las víctimas no depende de la restitución de derechos por indemnización” no aplica para este caso ya que las iniciativas encontradas dependen esencialmente de políticas públicas aplicadas al territorio. Sin embargo, la colaboración estatal es indispensable, el acompañamiento por parte de las entidades que pueden favorecer estrategias de cooperación comunitaria resulta vital en el desarrollo de iniciativas sostenibles y que generen impactos importantes.

Los resultados cualitativos permiten evidenciar un panorama en el cual la falta de oportunidades laborales y educativas hace que los ingresos de la población sean bajos y el nivel de escolaridad de los mismos también. A lo largo de las entrevistas se pudo ver que la

Es importante señalar que el instrumento utilizado queda a disposición para futuras investigaciones en lugares en los cuales exista una dinámica más asociada a la innovación social. No obstante, esto no quiere decir que en territorios como la palma se pueda impulsar iniciativas que promuevan la participación, la sostenibilidad económica y la interseccionalidad factores vitales en la innovación social.

REFERENCIAS

1. Andriesa, P., Daoub, L. y Verheydenc, P. (2018) Innovation as a vehicle for improving socially vulnerable groups' access to basic provisions: A research note on the development of a questionnaire module. *Research Policy*. pp. 1-8 DOI <https://doi.org/10.1016/j.respol.2018.08.017>
2. Benavides-Ponce, G. (2017). Programa de intervención: Atención integral a las víctimas del conflicto armado y excombatientes para la elaboración subjetiva de las catástrofes de guerra, reparación simbólica, construcción de memoria histórica y restauración del tejido social como aporte a la reconciliación nacional. *Universidad y Salud*, 19(2), 149-151. - Universidad de Nariño. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S012471072017000200149&lng=en&tlng=es.
3. Buckland, H, y Murillo, D, 2014. La innovación Social e América Latina. Marco conceptual y Agentes. Esade. Universidad Ramón Llull.
4. Cajaiba-Santana, G. (2014). Social innovation: Moving the field forward. A conceptual framework. *Technological Forecasting and Social Change*, 82(1), 42-51. doi:10.1016/j.techfore.2013.05.008
5. Centro Nacional de Memoria Histórica. (2013). ¡BASTA YA! Colombia: Memorias de guerra y dignidad. Bogotá Imprenta Nacional. Disponible en:

- <http://www.centrodememoriahistorica.gov.co/micrositios/informeGeneral/descargas.html>
6. Centro Nacional de Memoria Histórica (2017), Campesinos de tierra y agua: memorias sobre sujeto colectivo, trayectoria organizativa, daño y expectativas de reparación colectiva en la región Caribe 1960-2015. Introducción – metodología y conclusiones generales del trabajo de investigación, CNMH, Bogotá. Disponible en: http://www.centrodememoriahistorica.gov.co/descargas/informes-accesibles/campesinos-de-tierra-y-agua_accesible.pdf
 7. Comisión de la Unión Europea (2013). Guía de Innovación Social. Disponible en: https://www.asturias.es/Asturias/descargas/PDF_TEMAS/Asuntos%20Sociales/guia_innovacion_social.pdf
 8. Congreso de la República de Colombia (2011). Ley 1448 de 2011. Diario Oficial 48096. Bogotá: Imprenta Nacional. Disponible en <https://www.unidadvictimas.gov.co/sites/default/files/documentosbiblioteca/ley-1448-de-2011.pdf>
 9. Dawson, P., & Daniel, L. (2010). Understanding social innovation: A provisional framework. *International Journal of Technology Management*, 51(1), 9-21. doi:10.1504/IJTM.2010.033125
 10. Grimm, R., Fox, C., Baines, S. y Albertson, K. (2013) ¿Innovación social, una respuesta a los desafíos sociales contemporáneos? Localización del concepto en teoría y práctica. research note on the development of a questionnaire module. *Research Policy*. 26. pp. 436 - 455. DOI <https://doi.org/10.1016/j.respol.2018.08.017>
 11. Gobernación de Cundinamarca (2016). Plan de desarrollo Cundinamarca 2016 – 2020. Unidos podemos más.
 12. Lizarazu, B. (2016) La Palma ciudad de los pozos. Editorial Códice. Bogotá ISBN: 9789587520347
 13. Meléndez, J. (2000). Camino del Rionegro: historia regional y política de la ciudad La Palma (Cundinamarca). Editorial: Santafé de Bogotá: Universidad pedagógica Nacional. ISBN: 9589097456
 14. Mulgan, G., Tucker, S., Ali, R., & Sanders, B. (2007). What it is, why it matters and how it can be accelerated. The Young Foundation, Londres, Reino Unido. Disponible en: http://eureka.sbs.ox.ac.uk/761/1/Social_Innovation.pdf
 15. Organización para la Cooperación y el Desarrollo Económico (OCDE) (2011). Fomento de la innovación para abordar los desafíos sociales. París, Francia. Disponible en <https://www.oecd.org/sti/inno/colombia-innovation-review-assessment-and-recommendations-spanish.pdf>
 16. Rogelja, T., Ludvig, A., Weiss, G., & Secco, L. (2018). Implications of policy framework conditions for the development of forestry-based social innovation

- initiatives in slovenia. *Forest Policy and Economics*, 95, 147-155.
doi:10.1016/j.forpol.2018.07.011
17. Ruiz, C., y Parra, C. (2013). New forms of organization in knowledge-based societies: social innovation. *Non-profit Organizations and Social Entrepreneurship*, Routledge, New York, N.Y
 18. Sabato, S., Vanhercke, B., y Verschraegen, G. (2017). Connecting entrepreneurship with policy experimentation? the EU framework for social innovation. *Innovation*, 30(2), 147-167. doi:10.1080/13511610.2017.1282308
 19. Unidad de Víctimas (2018). Registro Único de Víctimas [Base de datos]. Bogotá: Unidad de Víctimas. Disponible en: <https://www.unidadvictimas.gov.co/es/registro-unico-de-victimas-ruv/37394>
 20. Van der Have, R. y Rubalcaba, L. (2016) Social innovation research: An emerging area of innovation studies? *Research Policy*. 45 pp. 1923 – 1935. doi: <https://doi.org/10.1016/j.respol.2016.06.010>
 21. Verdad Abierta (2013). El terror que los paramilitares sembraron en La Palma, Cundinamarca. Tomado de: <http://www.verdadabierta.com/justicia-y-paz/imputaciones/5070-el-terror-que-los-paramilitares-sembraron-en-la-palma-cundinamarca>
 22. Villa. J e Insuasty. A, 2015. Significado en torno a la indemnización y la restitución en víctimas del conflicto armado en el municipio de San Carlos, Antioquia. Medellín Colombia

CAPÍTULO 2

COSTOS Y RENTABILIDAD DE LA PRODUCCIÓN DEL LIMÓN PERSA EN EL MUNICIPIO DE MISANTLA, VERACRUZ, MÉXICO

Ignacio Caamal Cauich¹, Verna Grisel Pat Fernández¹, José Félix García Rodríguez²,
Felipe Jerónimo Ascencio¹

¹Universidad Autónoma Chapingo (UACH). Carretera México-Texcoco, Km. 38.5,
Chapingo, Estado de México, México. C.P. 56230.

²Universidad Juárez Autónoma de Tabasco (UJAT). Av. Universidad, Zona de la
Cultura, Col. Magisterial, Villahermosa, Centro, Tabasco, México. C.P. 86040.

RESUMEN

Los limones son los cítricos que generan mayor valor de la producción y exportación en México. Así mismo, México es el principal productor y exportador de limón persa en el mundo. Los principales estados productores en México son Veracruz, Oaxaca, Tabasco, Jalisco y Yucatán. El estado de Veracruz es el que aporta el mayor volumen de producción, con cerca del cincuenta y cuatro por ciento de la producción en México. El objetivo del trabajo fue analizar los costos y la rentabilidad de la producción del limón persa en el municipio de Misantla, Veracruz, calculándose los costos e indicadores de rentabilidad de la producción. Los resultados obtenidos indican que, en la estructura de los costos de producción por hectárea del limón persa, los mayores gastos se encuentran en la aplicación de fertilizantes granulados, labores culturales y mano de obra por cosecha. Los indicadores de rentabilidad precio-costo y tasa de ganancia fueron mayores que uno y cero, respectivamente, los cuales reflejan que el cultivo de limón persa en el municipio de Misantla, Veracruz, es rentable.

PALABRAS CLAVE: superficie, producción, rendimiento, costo, ingreso.

ABSTRACT

Lemons are the citrus crops that generate higher value of production and export in Mexico. Likewise, Mexico is the leading producer and exporter of Persian lime in the world. The major producing States in Mexico are Veracruz, Oaxaca, Tabasco, Jalisco y Yucatán. The State of Veracruz provides the largest volume of production, with about fifty-four percent of the production in Mexico. The objective of the work was to analyze the costs and the profitability of the production of the Persian lime in the municipality of Misantla, Veracruz, calculating the costs and profitability indicators of the production.

The results indicate that in the structure of production costs per hectare of Persian lime, the greatest costs are in the application of granulated fertilizers, cultural work and labor per harvest. The cost-price profitability and profit rate indicators were greater than one and zero, respectively, which reflects that the cultivation of Persian lime in the municipality of Misantla, Veracruz, is profitable.

KEYWORDS: surface, production, yield, cost, income.

1. INTRODUCCIÓN

El cultivo de cítricos se realiza en una franja geográfica que se extiende a 40° de latitud Norte y Sur del Ecuador, englobando regiones tropicales y subtropicales, donde concurren condiciones de suelo y clima favorables, en más de cien países del mundo (Saunt, 1992).

Importancia

México es el país que ocupa el primer lugar en la producción de limones y limas con 2,528,174 toneladas, seguido por la India con 2,364,000 toneladas, en tercer lugar se encuentra China continental con 2,316,876 toneladas y Argentina en cuarto lugar con 1,676,000 toneladas, en conjunto estos países aportan alrededor del 52% de la producción total de limones y limas (Figura 1).

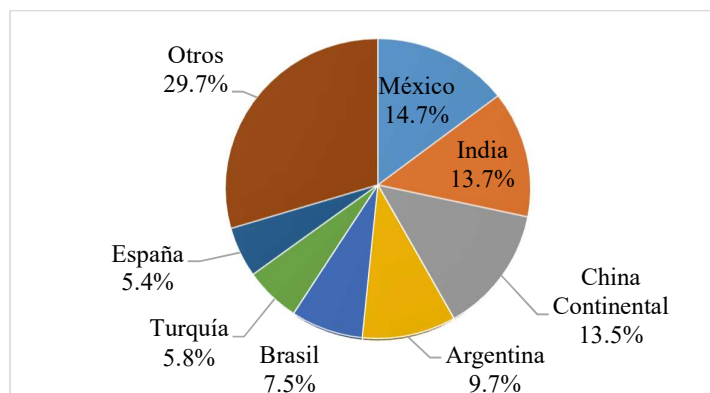


Figura 1: Distribución de la producción de limones y limas en el mundo, 2017.

Fuente: Elaborado con datos de FAOSTAT (2019).

La producción total de limones y limas en el mundo se ha venido incrementando, pasando de 8,082,508 a 17,218,173 toneladas durante el periodo de 1994 al 2017 (Figura 2), con una tasa de crecimiento de 113.0%.

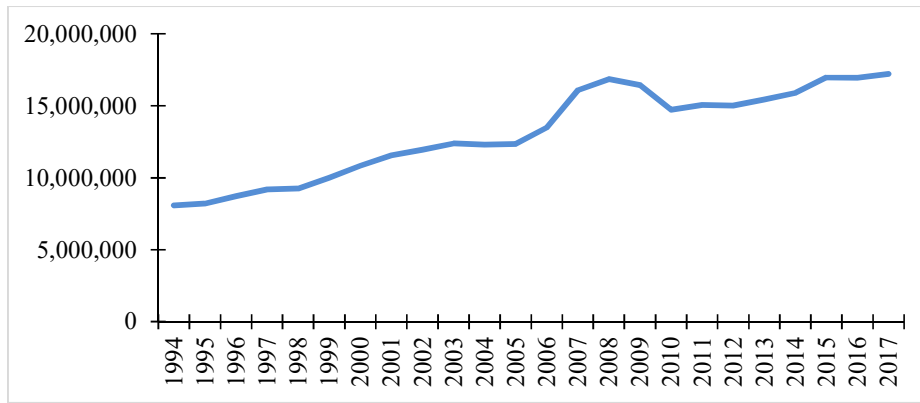


Figura 2: Comportamiento de la producción de limones y limas a nivel mundial, 1994-2017 (t).
Fuente: Elaborado con datos de FAOSTAT (2019).

Los principales países exportadores de limones y limas son México (667,572 t) España (547,155 t), Turquía (448,781 t) y Argentina (279,339 t), los cuales aportan alrededor del 62% de las exportaciones totales (Figura 3). México destaca como el primer productor y primer exportador mundial de limones y limas (FAOSTAT, 2019).

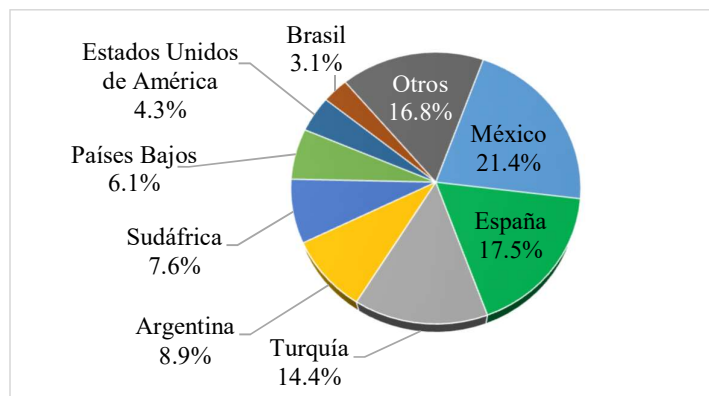


Figura 3: Distribución de las exportaciones de limones y limas en el mundo, 2016.
Fuente: Elaborado con datos de FAOSTAT (2019).

El volumen de las exportaciones totales de limones y limas a nivel mundial, tuvo un incremento de 147.3%, al pasar de 1,262,615 a 3,122,981 toneladas exportadas durante el periodo de 1994 al 2016 (Figura 4).

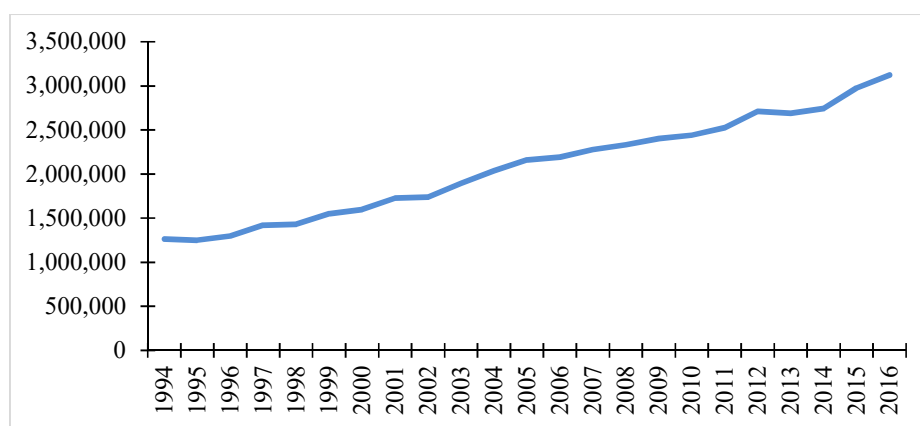


Figura 4: Comportamiento de las exportaciones de limones y limas a nivel mundial, 1994-2016 (t).

Fuente: Elaborado con datos de FAOSTAT (2019).

El limón persa es la variedad más importante de los limones y limas en México, con una superficie cosechada de 86,706 hectáreas y un rendimiento promedio a nivel nacional de 14.0, seguido por el limón agrio (mexicano) que tiene una superficie cosechada de 75,755 hectáreas, un rendimiento de 14.8 y que aportan una producción de 1,120,973 toneladas (Tabla 1).

Tabla 1: Variedades cultivadas de limón persa en México, 2017.

Variedad	Superficie sembrada (ha)	%	Superficie cosechada (ha)	%	Rendimiento (t/ha)	Producción (t)	%	Valor de la producción (millones de pesos)	%
Limón persa	93,313	47.7	86,706	50.3	14.0	1,217,250	48.1	5,729.5	45.2
Limón agrio	89,795	45.9	75,755	44.0	14.8	1,120,973	44.3	5,857.1	46.2
Limón italiano	9,612	4.9	7,344	4.3	21.0	153,980	6.1	878.3	6.9
Lima	1,691	0.9	1,537	0.9	9.62	14,783	0.6	63.6	0.5
Otros	1,067	0.5	913	0.5	23.2	21,188	0.8	160.6	1.3
Total	195,479	100.0	172,255	100.0	14.7	2,528,174	100.0	12,689.1	100.0

Fuente: Elaborado con datos de SIACON (2019).

En México se cultivan 93,313 hectáreas de limón persa en 23 entidades federativas, las cuales producen 1,217,250 toneladas por año, con un rendimiento promedio de 12.3 toneladas por hectárea y generando un valor de 5,729.5 millones de pesos (SIACON, 2019).

El limón persa es un producto generador de divisas, puesto que se exporta alrededor del 53% de la producción nacional y el resto se destina para el mercado nacional y la agroindustria, lo que implica una fuerte derrama económica en las zonas productoras de limón (SIACON, 2019).

El estado de Veracruz es el principal productor de limón persa en México, con una superficie cosechada de 44,853 hectáreas, una producción de 655,300 toneladas, un rendimiento promedio de 14.6 toneladas por hectárea, ligeramente superior a la media nacional de 14.0 toneladas por hectárea, y un valor de la producción de 3,225.6 millones de pesos (SIACON, 2019).

En el estado de Veracruz, el principal municipio productor es Martínez de la Torre, con una superficie cosechada de 14,659 hectáreas, una producción de 222,172 toneladas, un rendimiento de 15.2 toneladas por hectárea y un valor de la producción de 1,081.4 millones de pesos, mientras que el municipio de Misantla con una superficie cosechada de 1,057 hectáreas, un rendimiento de 18.0 toneladas por hectárea, una producción de 19,006 toneladas y un valor de la producción de 99.3 millones de pesos, ocupa el décimo lugar a nivel estatal (SIACON, 2019).

Marco teórico

La producción se refiere al proceso de creación de cualquier bien o servicio que la población requiere y puede adquirir para satisfacer sus necesidades. Cualquier proceso de producción requiere de una gran variedad de insumos, tales como trabajo, tierra, maquinaria y equipo, materias primas, entre otros (Ferguson y Gould, 1978). Los tipos y la cantidad de insumos utilizados permiten realizar la caracterización de las unidades de producción.

Los costos de producción representan las erogaciones (flujos de salida de recursos) y cargos (amortizaciones, depreciaciones, agotamientos) asociados clara y directamente con la producción de los bienes o la prestación de los servicios de los cuales el agente económico obtuvo sus ingresos (Heredia, 2008).

El costo total de producción es igual a la suma de los costos variables totales más los costos fijos totales. Si la producción es cero, el costo variable (CV) es cero y el costo total (CT) es igual al costo fijo total, sin embargo, cuando hay algún nivel de producción se emplean insumos variables, por lo tanto, los costos totales son iguales a la suma de los costos variables totales más los costos fijos totales (Ferguson y Gould, 1978).

El ingreso es el flujo positivo de riqueza generada como consecuencia de las actividades de la empresa en un periodo determinado y que, por lo tanto, producen un incremento del patrimonio neto. Los ingresos generan incrementos de los recursos económicos de la empresa como consecuencia de la venta de bienes o de la prestación de servicios que son objeto del tráfico de la empresa, así como de otros beneficios de operaciones económicas no corrientes (Muñoz, 2008).

La rentabilidad se refiere, en general, a la capacidad de producir o generar un beneficio adicional sobre la inversión o esfuerzo realizado, en otras palabras, es la capacidad que se tiene de obtener un valor superior a los costos de producción. La rentabilidad en la

actualidad es un indicador suficiente de la competitividad, es la mejor medida en un período prolongado (McFetridge, 1995).

Según Lerma (2000) la competitividad en términos comerciales significa la posibilidad de triunfar sobre otros oferentes cuando, al confrontarse con los productos sustitutos, se tiene alta posibilidad de salir victorioso, favorecidos por la compra del consumidor. La competitividad en el ámbito del comercio exterior depende de condiciones que están presentes en dos niveles: 1. Macro y estructural. Que comprende las condiciones básicas y generales del país o región necesarios para que las empresas puedan operar con eficiencia; y 2. Micro y estructural. En el que se encuentran las condiciones a nivel empresa y producto que puede hacer que la mercancía se desplace exitosamente en los mercados meta.

Para la Organización de Cooperación para el Desarrollo Económico (OCDE), la competitividad se define como el grado en que un país es capaz de producir, bajo condiciones de libre mercado, bienes y servicios acorde a las necesidades de los mercados internacionales, manteniendo o incrementando, paralelamente, los ingresos reales de la población en el largo plazo (OCDE, 1992).

El objetivo del trabajo fue analizar y determinar la rentabilidad y competitividad del limón persa en el municipio de Misantla, Veracruz. De este se desprenden: Identificar la tecnología empleada en la producción de limón persa, calcular la estructura de costos de limón persa, y determinar los indicadores de rentabilidad y competitividad en la producción de limón persa en el municipio de Misantla, Veracruz.

2. MATERIALES Y MÉTODOS

Definición del área de estudio

El estado de Veracruz se encuentra ubicado entre la Sierra Madre Oriental y el Golfo de México, en las coordenadas 17° 09' y los 22° 28' de latitud norte y los 98° 39' y los 93° 36' de longitud oeste. Limita al Norte con el estado de Tamaulipas y el Golfo de México; al Este con el Golfo de México, Tabasco y Chiapas; al Sur con los estados de Chiapas y Oaxaca, y al Oeste con Puebla, San Luis Potosí e Hidalgo (Figura 5). Cuenta con una superficie de 71,820.04 km², que representa el 3.7% del total de la República Mexicana, ocupando el 11° lugar por extensión territorial (INEGI, 2019).

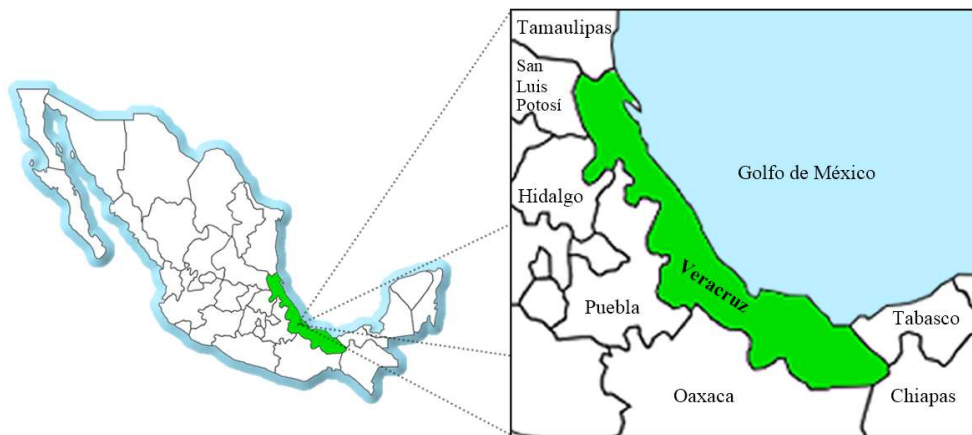


Figura 5: Mapa de localización del estado de Veracruz, México.

Fuente: Elaborado con información de INEGI (2019).

El municipio de Misantla se encuentra ubicado en la zona centro montañosa del Estado de Veracruz, en las coordenadas 19° 56' latitud Norte y 96° 51' longitud Oeste, a una altura de 300 metros sobre el nivel del mar. Limita al Norte con Nautla, al este con Yecuatla y Colipa, al sur con Chiconquiaco y Landero y Coss, al suroeste con Tenochtitlán, con Altotonga y Atzalán al Oeste, al noroeste con Martínez de la Torre (Figura 6). Su distancia aproximada al norte de la capital del Estado, por carretera es de 80 Km y tiene una superficie de 537.94 km²; que representa el 0.74% del total del territorio del estado (INEGI, 2015).



Figura 6: Mapa de localización del municipio de Misantla, estado de Veracruz.

Fuente: Elaborado con información de INEGI (2019).

Cálculo del tamaño de la muestra y obtención de información

El tipo de muestreo utilizado fue el muestreo bola de nieve, que consiste en elegir una pequeña muestra inicial y a cada uno de los individuos seleccionados se le solicita que sugieran a otros posibles individuos que podrían ser incluidos en la muestra, de acuerdo con requisitos previamente establecidos. También se le conoce como muestra en cadena o por redes, pues se identifica a unidades que llevarán a otras unidades que están

íntimamente relacionadas y por tanto permitirán analizar la cadena (Kleeberg y Ramos, 2009). Las comunidades seleccionadas fueron las que mostraban mayor superficie plantada de limón persa reportada en el padrón de citricultores del estado de Veracruz (CONCITVER, 2012).



Figura 7: Mapa de localidades seleccionadas en el municipio de Misantla, estado de Veracruz.

Fuente: Elaborado con datos de INEGI (2019).

La información se obtuvo a través de encuestas aplicadas con los productores directamente de campo. Para realizar la selección de ejidos en los que se levantarían las encuestas, solo se identificaron a las comunidades con la mayor cantidad de hectáreas sembradas de limón persa, las cuales se localizan principalmente en el Norte del municipio (Figura 7), quedando seleccionados La Defensa (187.69 ha), La Poza del Tigre (139.89 ha), Francisco Sarabia (135.62 ha), Ignacio Zaragoza (107.75 ha), Arroyo Hondo (99.6 ha), Troncones (92.75 ha), Palpoala Ixcan (80.85 ha) y Máximo García (La Guadalupe) (74.33 ha), descartándose las comunidades con superficie sembrada de limón persa menor a 70 hectáreas.

Se aplicaron 35 encuestas piloto y 36 adicionales, en total 71 encuestas, en el año 2012. Del análisis de la información se obtuvieron el promedio y la varianza del rendimiento por hectárea del limón persa. En la presente investigación se consideró el rendimiento de la producción de limón persa cuya varianza fue de 3.84 toneladas por hectárea.

La información que se obtuvo mediante la aplicación de las encuestas a productores fue sobre superficie en producción, edad de la plantación, edad de los productores, cantidad de fertilizantes, insecticidas, herbicidas aplicados durante un año de producción, además de los precios de cada uno de los insumos aplicados y los precios recibidos por la cosecha.

Los datos proporcionados por los productores entrevistados se sistematizaron en matrices para calcular los valores parciales, valores totales, estructuras porcentuales de variables y costos de producción e indicadores de rentabilidad de la producción de limón persa.

Estructura de costos y competitividad

Las estructuras que se calcularon son la de los factores tecnológicos y la de los costos de producción. El procedimiento de cálculo es:

$$P(\%) = \frac{VP}{VT} \times 100 \quad (1)$$

Donde:

P(%) = Proporción

VP=Valor parcial

VT=Valor total

Los indicadores de competitividad de la producción de limón persa a nivel de unidad productiva que se calcularon fueron retomados del modelo de competitividad de cadenas agroindustriales de Ibáñez y Caro (2001), los cuales son competitividad precio-costo y competitividad tasa de ganancia (Tabla 2).

Tabla 2: Modelo de estudio a nivel productor.

Indicador / Agentes	Ingreso total (IT)	Costo total (CT)	Costo unitario (Cu)	Precio-costos (P/Cu)	Tasa de ganancia (g)
Productores primarios	$P_x X$	$P^m M$	$a P^m$	$P_x / a P^m$	$(P_x - a P^m) / k_p$

Dónde: P_x =Precio del producto agrícola; X =Cantidad del producto agrícola; M =Cantidad de insumos agrícolas; P^m =Precio del insumo agrícola; $a=M/X$: Coeficiente de requerimientos de insumo por unidad de producto agrícola; $k_p=k_p/X$ =Coeficiente de requerimientos de capital por unidad de producto agrícola.

Fuente: Ibáñez y Caro (2001).

3. RESULTADOS Y DISCUSIÓN

Distribución de la producción de limón persa en México

Como se observa en el Figura 8, el estado con mayor superficie sembrada de limón persa en México es Veracruz con 46,122 hectáreas cultivadas, seguido por Oaxaca con 14,626 hectáreas, Tabasco con 7,223 hectáreas, Jalisco con 5,393 hectáreas, Yucatán con 2,942 hectáreas y Puebla con 2,642 hectáreas, los cuales aportaron alrededor del 85% de la superficie sembrada total (Figura 8).

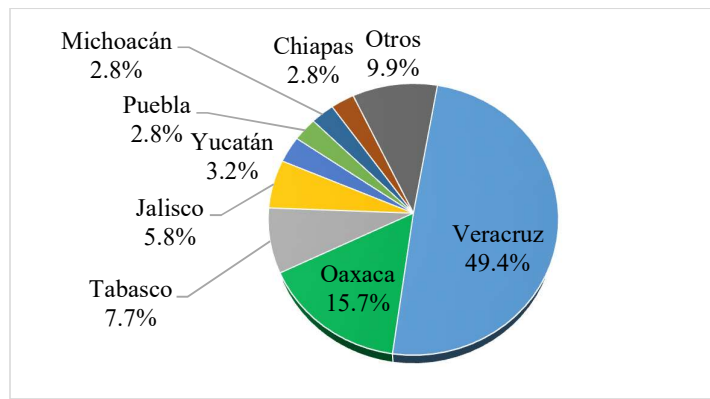


Figura 8: Distribución de la superficie sembrada de limón persa en el estado de Veracruz, 2017.
Fuente: Elaborado con datos de SIACON (2019).

El estado con la mayor superficie cosechada de limón persa en México es Veracruz, con 44,853 hectáreas, seguido por Oaxaca con 13,805 hectáreas, Tabasco con 7,192 hectáreas, Jalisco con 4,121 hectáreas, Yucatán con 2,773 hectáreas y Puebla con 2,473 hectáreas, los que en conjunto aportaron alrededor del 87% de la superficie cosechada total (Figura 9).

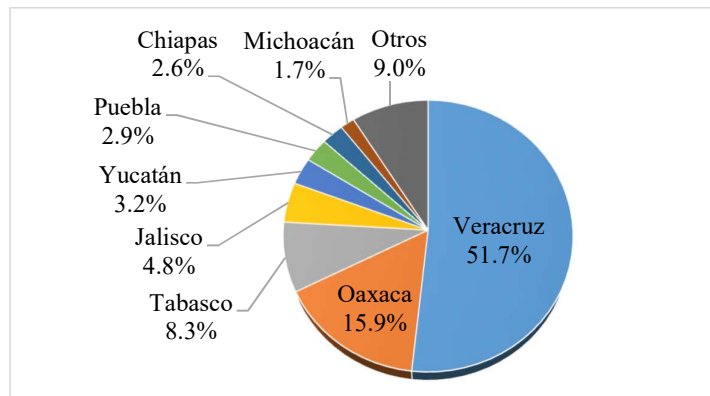


Figura 9: Distribución de la superficie cosechada de limón persa en el estado de Veracruz, 2017.
Fuente: Elaborado con datos de SIACON (2019).

Dentro de los principales estados productores de limón persa, los que tienen los rendimientos más altos, mayores a la media nacional, son Yucatán y Jalisco, debido a que cultivan la mayor parte del limón persa en condiciones de riego. En los estados de Puebla, Tabasco, Michoacán y Chiapas los rendimientos son menores a la media nacional, mientras que en los estados de Veracruz y Oaxaca los rendimientos son ligeramente superiores al rendimiento promedio (Figura 10).

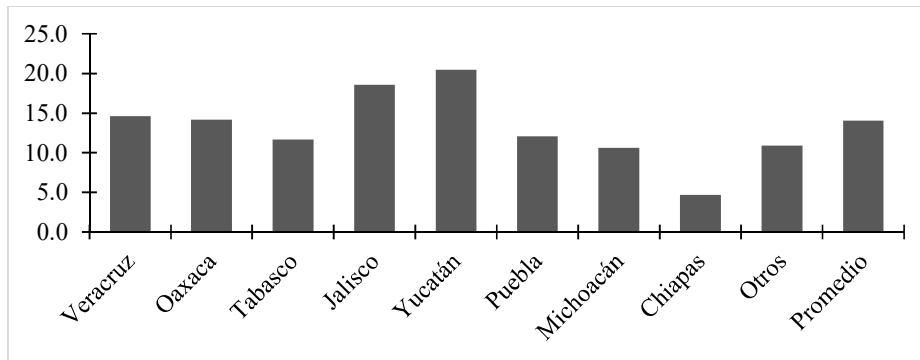


Figura 10: Rendimiento del limón persa de los principales estados productores en México, 2017 (t/ha).

Fuente: Elaborado con datos de SIACON (2019).

Los principales estados productores de limón persa en México son Veracruz (655,300 t), Oaxaca (195,904 t), Tabasco (83,971 t), Jalisco (76,478 t) y Yucatán (56,764 t), los cuales aportaron el 88% de la producción total (Figura 11).

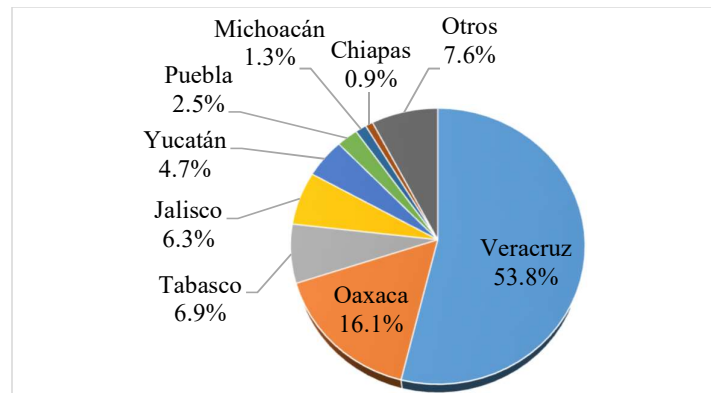


Figura 11: Distribución de la producción de limón persa en México, 2017.

Fuente: Elaborado con datos de SIACON (2019).

El estado de Veracruz, con un rendimiento cercano a la media nacional y por tener la mayor superficie sembrada, aporta una producción significativamente mayor, con una participación de más del 50%, mientras que los estados de Jalisco y Yucatán, que tienen los rendimientos más altos, aportaron en conjunto alrededor del 11% de la producción total, respectivamente (Figura 11).

Comportamiento de la producción de limón persa en México

La superficie sembrada de limón persa en México se ha venido incrementado en el periodo de 1996 al 2017, al pasar de 13,357 a 93,313 hectáreas sembradas con una tasa de crecimiento en el periodo de 598.6% y con una tendencia creciente (Figura 12), lo que significa que es una actividad económica rentable y competitiva.

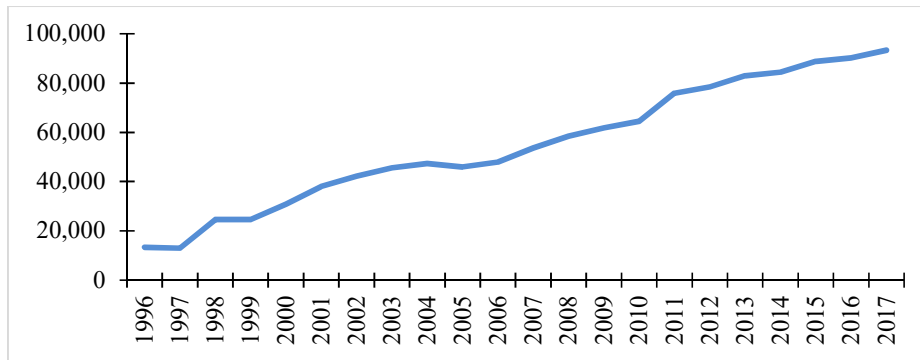


Figura 12: Comportamiento de la superficie sembrada de limón persa en Méx., 1996-2017 (ha).
Fuente: Elaborado con datos de SIACON (2019).

La superficie cosechada de limón persa en México también se ha venido incrementando a nivel nacional, presentando una tasa de crecimiento de 564.6%, al pasar de 13,047 a 86,706 hectáreas cosechadas, durante el periodo de 1996 a 2017 (Figura 13).

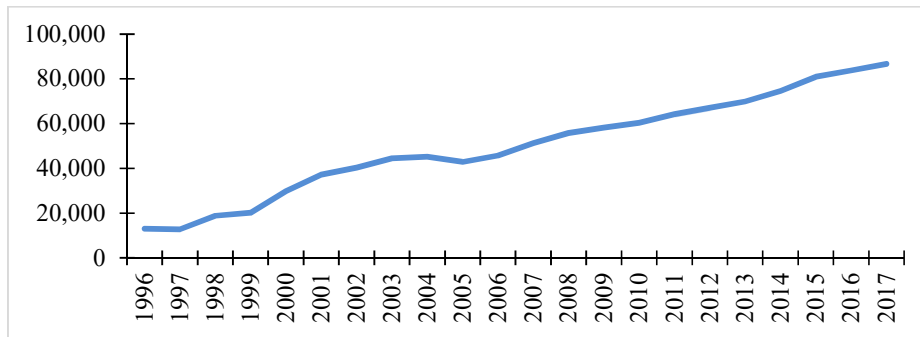


Figura 13: Comportamiento de la superficie cosechada de limón persa en Méx., 1996-2017 (ha).
Fuente: Elaborado con datos de SIACON (2019).

El rendimiento de limón persa en México ha crecido en alrededor del 6.2%, al pasar de 13.2 a 14.0 toneladas por hectárea en el periodo de 1996 al 2017 (Figura 14).

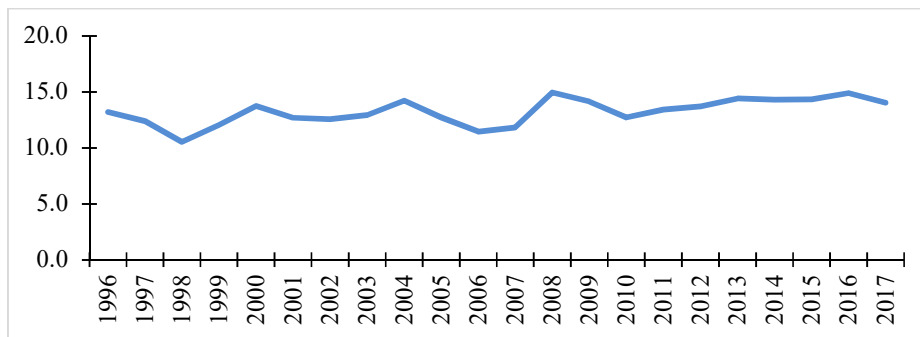


Figura 14: Comportamiento del rendimiento promedio de limón persa en México, 1996-2017 (t/ha).

Fuente: Elaborado con datos de SIACON (2019).

La producción total de limón persa en México ha mostrado una tendencia creciente durante el periodo de 1996 a 2017, pasando de 172,409 a 1,217,250 toneladas producidas, lo que representa un incremento de 606.0% (Figura 15).

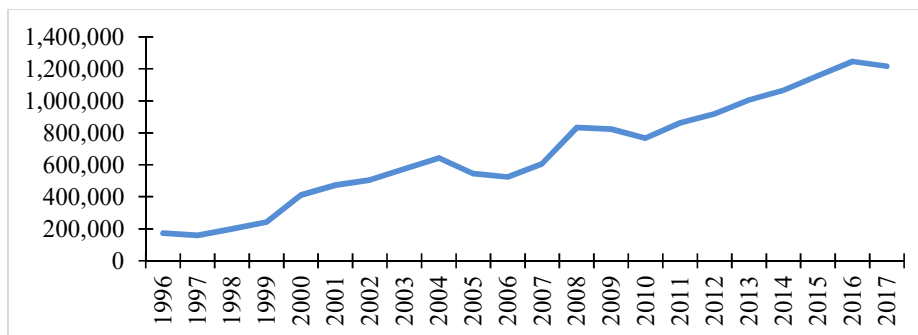


Figura 15: Comportamiento de la producción de limón persa en México, 1996-2017 (t).
Fuente: Elaborado con datos de SIACON (2019).

La baja tasa de crecimiento del rendimiento del cultivo del limón persa, refleja que la producción se ha incrementado en mayor medida por la expansión de la superficie sembrada y cosechada, puesto que el incremento de estas dos variables fue mayor que el incremento del rendimiento.

Distribución de la producción de limón persa en el estado de Veracruz

Los principales municipios con la mayor superficie sembrada de limón persa en el estado de Veracruz son Martínez de la Torre con 15,233 ha, seguido de Atzalan con 5,748 ha, San Rafael con 4,545 ha, Tlapacoyan con 3,743 ha y Papantla con 1,875 ha, los cuales en conjunto aportaron el 67.5% de la superficie sembrada total (Figura 16). El municipio de Misantla ocupa el décimo lugar por la superficie cultivada.

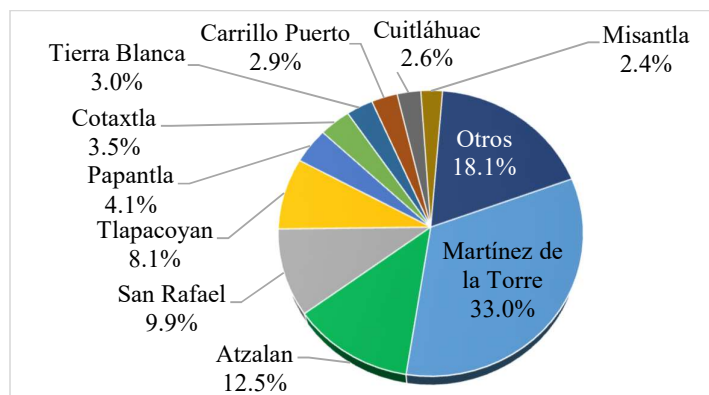


Figura 16: Distribución de la superficie sembrada de limón persa en el estado de Veracruz, 2017.

Fuente: Elaborado con datos de SIACON (2019).

La superficie cosechada de limón persa en los principales municipios productores, mantiene una proporción similar que la superficie sembrada. Los municipios con la mayor superficie cosechada son Martínez de la Torre con 14,659 ha, Atzalan con 5,658 ha, San Rafael con 4,420 ha, Tlapacoyan con 3,668 ha y Papantla con 1,840 ha, los que en conjunto aportaron el 67.4% de la superficie cosechada total del estado (Figura 17).

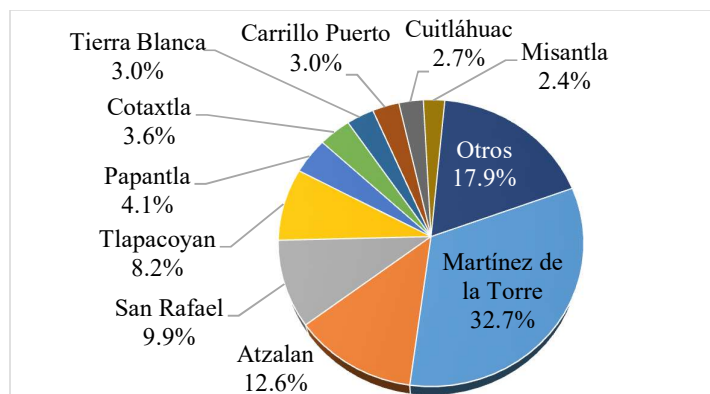


Figura 17: Distribución de la superficie cosechada de limón persa en el estado de Veracruz, 2017.

Fuente: Elaborado con datos de SIACON (2019).

El rendimiento de limón persa alcanzando en Martínez de la Torre, principal productor, en el año 2017, fue de 15.2 t/ha, en Tlapacoyan de 15.8 t/ha y en Papantla de 15.1 t/ha, los cuales son superiores al rendimiento promedio del estado, que es de 14.6 t/ha, le siguen San Rafael con 13.6 t/ha y Atzalan con 13.5 t/ha (Figura 18).

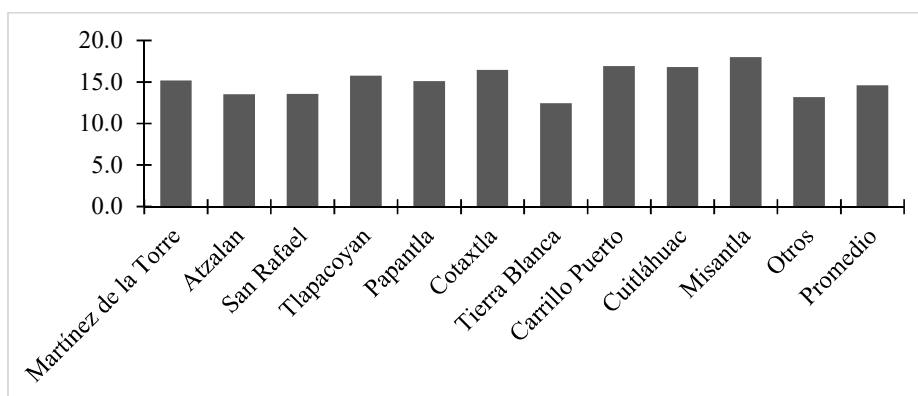


Figura 18: Rendimiento del limón persa de los principales municipios productores en el estado de Veracruz, 2017 (t/ha).

Fuente: Elaborado con datos de SIACON (2019).

Los municipios con mayor volumen de producción de limón persa en el estado de Veracruz son Martínez de la Torre (222,172 t), Atzalan (76,383 t), San Rafael (59,899 t), Tlapacoyan (57,788 t) y Papantla (27,784 t), los cuales aportaron el 68% de la producción

estatal (Figura 19). El municipio de Misantla ocupa el noveno lugar con una producción de 19,006 toneladas (SIAP, 2019).

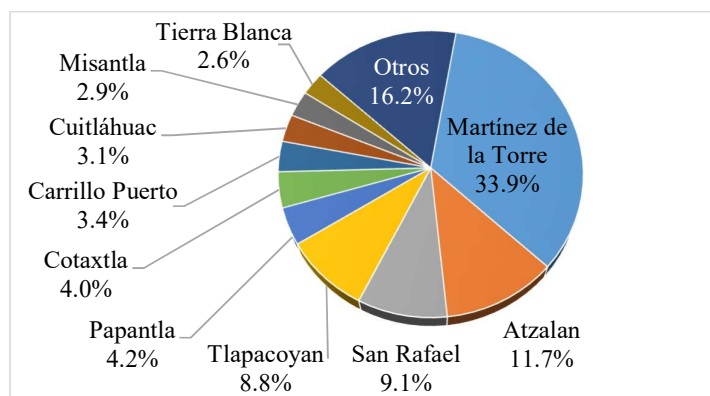


Figura 19: Distribución de la producción de limón persa en el estado de Veracruz, por municipio, 2017.

Fuente: Elaborado con datos de SIAP (2019).

Comportamiento de la producción de limón persa en el estado de Veracruz

La superficie sembrada de limón persa en el estado de Veracruz se ha venido incrementando durante el periodo de 2003 al 2017, con una tasa de crecimiento de 279.0%, al pasar de 12,168 a 46,122 hectáreas sembradas, observándose una tendencia creciente (Figura 20), lo que significa que es una actividad económica rentable y competitiva.

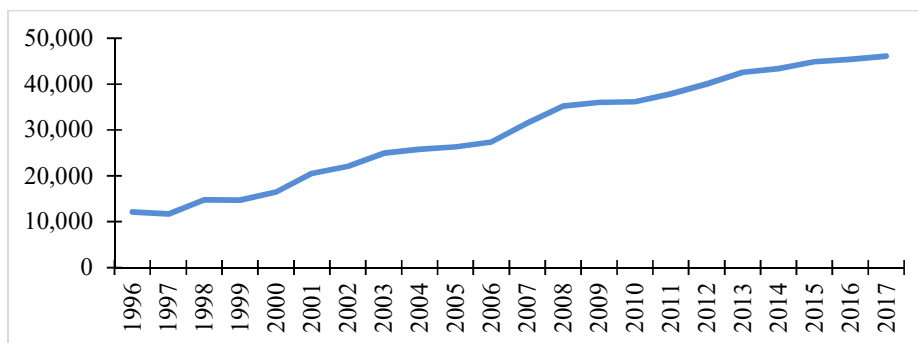


Figura 20: Comportamiento de la superficie sembrada de limón persa en el estado de Veracruz, 1996-2017 (ha).

Fuente: Elaborado con datos del SIACON, 2018.

La superficie cosechada de limón persa en el estado de Veracruz, tuvo un incremento de 269.8%, al pasar de 12,130 a 44,853 ha en el mismo periodo analizado, observándose una tendencia creciente (Figura 21).

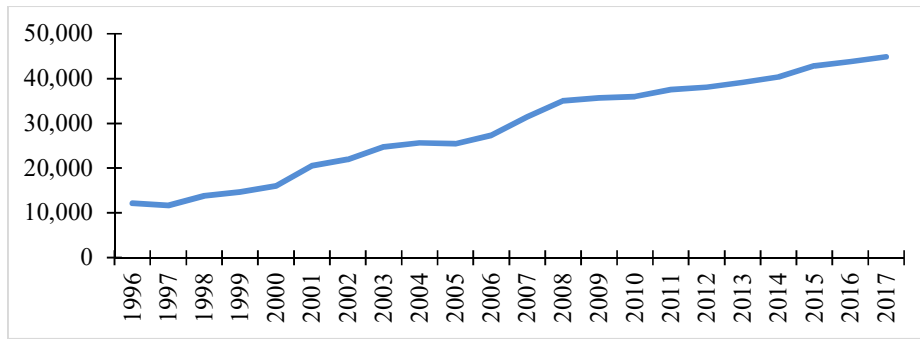


Figura 21: Comportamiento de la superficie cosechada de limón persa en el estado de Veracruz, 1996-2017 (ha).

Fuente: Elaborado con datos del SIACON, 2018.

El rendimiento promedio de limón persa en el estado de Veracruz, registró un incremento de 9.1% en el periodo de 1996 al 2017, al pasar de 13.4 a 14.6 t/ha, con una tendencia ligeramente creciente y con algunos altibajos (Figura 22).

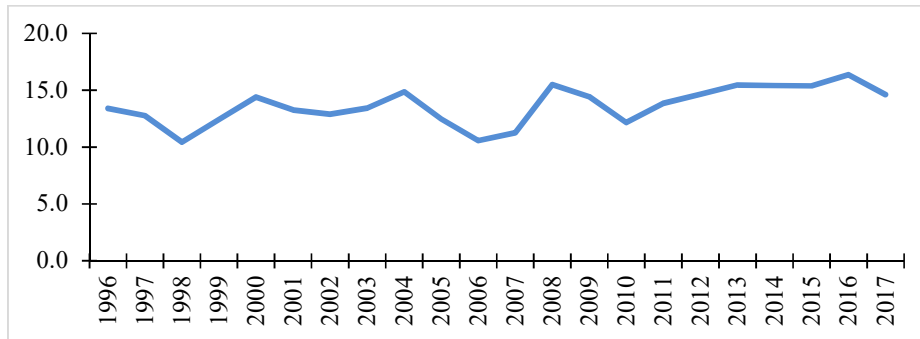


Figura 22: Comportamiento del rendimiento promedio de limón persa en el estado de Veracruz, 1996-2017 (t/ha).

Fuente: Elaborado con datos de SIACON (2019).

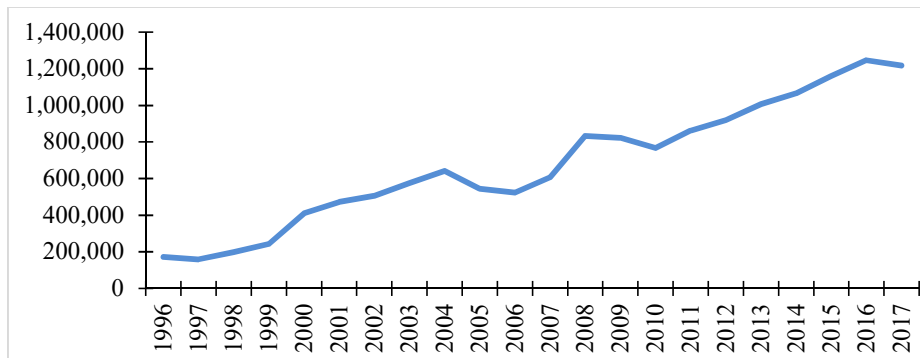


Figura 23: Comportamiento de la producción de limón persa en el estado de Veracruz, 1996-2017 (t).

Fuente: Elaborado con datos de SIACON (2019).

La producción de limón persa en el estado de Veracruz se ha venido incrementado, al pasar de 162,493 a 655,300 toneladas en el periodo de 1996 a 2017, con una tasa de crecimiento de 303% y una tendencia creciente (Figura 23).

La producción de limón persa que México exportó en 2017, 650,556 t (53.4%), se destinó principalmente a Estados Unidos de América con un volumen de 587,483 t lo que representó alrededor del 90% de las exportaciones totales de limón persa, seguido de Europa con 48,042 t, Canadá con 2,615 t y Japón con 2,019 t (SIAVI, 2019), el volumen de producción restante, 566,694 t (46.6%), se distribuye en el mercado nacional.

Producción de limón persa en el Municipio de Misantla, Veracruz

Caracterización de las variables de producción

El municipio de Misantla cuenta con 44 comunidades productoras de limón persa que reporta el Consejo Estatal Citrícola A. C. de Veracruz (CONCITVER, 2012), con una superficie sembrada de 1,103 ha y cosechada de 1,057 ha y una producción de 19,006 t, en el año 2017 (SIACON, 2019).

En la zona productora de limón persa en el municipio de Misantla, la superficie promedio por productor es pequeña. La cantidad de productores dedicados al cultivo de limón es de 626 productores, distribuidos en 44 localidades con 1448.92 hectáreas sembradas y un promedio de 2.31 hectáreas por productor, para el 2012 (Tabla 3).

Tabla 3: Cantidad de productores y superficie sembrada por localidad en el municipio de Misantla, Veracruz, 2008.

Localidad	No. de productores	Total de hectáreas	Promedio (ha/productor)
La Defensa	59	187.69	3.18
La Poza del Tigre	53	139.89	2.64
Francisco Sarabia	26	135.62	5.22
Ignacio Zaragoza	51	107.75	2.11
Arroyo Hondo	28	99.6	3.56
Troncones	55	92.75	1.69
Palpoala Ixcan	23	80.85	3.52
La Guadalupe	47	74.33	1.58
Santa Cecilia	26	68.75	2.64
Coapeche	28	48.32	1.73
Otros	230	413.37	1.80
Total	626	1,448.92	2.31

Fuente: Elaborado con datos de CONCITVER (2012).

Las variables analizadas son número de productores, superficie cosechada, aplicación de fertilizantes, herbicidas y pesticidas y cantidad de jornales que emplean para las distintas actividades en un año de cosecha.

La superficie sembrada media en el municipio de Misantla es de 3.16 hectáreas por cada productor, mientras que la superficie cosechada es de sólo 2.9. El rendimiento medio obtenido por los productores de limón persa fue de 12.63 toneladas por hectárea y el volumen promedio por productor fue de 38.93 toneladas anuales por hectárea (Tabla 4).

Tabla 4: Datos generales de la producción de limón persa en el municipio de Misantla, Veracruz.

Concepto	Cantidad
Superficie sembrada media (ha)	3.16
Superficie sembrada máxima (ha)	22.00
Superficie sembrada mínimo (ha)	0.50
Superficie en producción media (ha)	2.90
Superficie en producción máxima (ha)	22.00
Superficie en producción mínima (ha)	0.50
Rendimiento promedio (t/ha)	12.63
Rendimiento máximo (t/ha)	21.40
Rendimiento mínimo (t/ha)	4.70
Volumen de producción medio (t)	38.93
Volumen de producción máximo (t)	213.64
Volumen de producción mínimo (t)	6.07

Fuente: Elaborado con datos de la encuesta aplicada a los productores en el año 2012.

Manejo técnico de la producción

El uso de fertilizantes es considerable, el 94.44% de los productores aplica fertilizantes granulados, mientras que un 61.11% realiza aplicación de fertilizantes foliares. En el caso de los herbicidas un 94.44% del total de los productores realizan ésta actividad mientras que una parte apenas superior al 50% realizan aplicación de algún tipo de pesticida (Tabla 5). Los fertilizantes granulados están más relacionados con el rendimiento y los fertilizantes foliares con el color del fruto del limón persa.

Tabla 5. Uso de agroquímicos en la producción de limón persa en el municipio de Misantla, Veracruz.

Insumo/Productores	Usan	%	No usan	%
Fertilizantes granulados	68.00	94.44	4.00	5.55
Fertilizantes foliares	44.00	61.11	28.00	38.88
Herbicidas	68.00	94.44	4.00	5.55
Pesticidas	57.00	79.16	15.00	20.83

Fuente: Elaborado con datos de la encuesta aplicada a los productores en el año 2012.

Estructura de costos

Las unidades de producción de limón persa encuestadas tuvieron costos promedio cercanos a los 16,468.35 pesos, pero la mayoría de los productores no consideran el costo de oportunidad de la tierra, sin el cual no es posible determinar el costo real de producción. Considerando una renta de 2,000.00 pesos, el costo real de producción es de 18,468.35 pesos.

Si se dividen los costos de cada uno de los insumos empleados entre el costo total de producción, la mano de obra es la que tiene el mayor porcentaje de los gastos realizados,

en promedio con el 56.62% de los costos totales de producción. Los fertilizantes granulados son los insumos que emplean el 14.33% en promedio de los costos. Los fertilizantes son los insumos más importantes puesto que de ellos depende el crecimiento del árbol y que los frutos puedan tener las características que el mercado requiere, tales como el color y el tamaño adecuado.

Los tres insumos que tuvieron menor participación fueron los fertilizantes foliares, herbicidas y pesticidas, que registraron 2.17, 2.98 y 1.95%, respectivamente (Tabla 6).

Tabla 6: Estructura de costos de producción del limón persa en el municipio de Misantla, Veracruz.

Concepto	Cantidad (\$/ha)	%	Costo (\$/t)
Fertilizantes granulados	2,646.03	14.33	209.84
Fertilizantes foliares	401.17	2.17	31.81
Herbicidas	550.22	2.98	43.63
Pesticidas	359.86	1.95	28.54
Gasolina	2,054.74	11.13	162.95
Mano de obra	10,456.33	56.62	829.21
Subtotal	16,468.35	89.17	1,305.98
Renta	2,000.00	10.83	158.60
Total	18,468.35	100.00	1,464.58

Fuente: Elaborado con datos de la encuesta aplicada a los productores en el año 2012.

En síntesis, el costo de producción de una hectárea plantada de limón persa en el municipio de Misantla es de \$18,468.35 considerando la renta de la tierra y de \$16,468.35 sin considerar la renta de la tierra.

Ingreso

El ingreso promedio obtenido por los productores, con base en el rendimiento promedio y al precio pagado en el municipio de Misantla, estado de Veracruz, que fueron de 12.61 toneladas y 1,845.16 pesos, respectivamente, asciende a 23,267.47 pesos anuales (Tabla 7).

Tabla 7: Ingreso promedio del productor de limón persa en el municipio de Misantla, Veracruz.

Rendimiento	Precio (\$/t)	Ingreso (\$)
12.61	1,845.16	23,267.47

Fuente: Elaborado con datos de la encuesta aplicada a los productores en el año 2012.

Rentabilidad y competitividad

Los indicadores de rentabilidad y competitividad calculados son la competitividad precio-costo y la tasa de ganancia. La relación precio costo permite medir los efectos que tiene los costos de los factores de la producción en la rentabilidad y competitividad del producto. Por otra parte, la tasa de ganancia, determina la eficiencia en función de los beneficios relativos, se define como la proporción de la diferencia porcentual entre las ventas totales y costos totales.

El indicador precio-costo obtenido indica que los productores de limón persa del municipio de Misantla, estado de Veracruz, son competitivos, ya que la relación precio costo es superior a la unidad, reflejando en general que se están utilizando eficientemente los diferentes recursos para la producción de limón (Tabla 8).

Tabla 8: Indicadores de competitividad con costos por hectárea y por tonelada de limón persa con y sin renta en el municipio de Misantla, Veracruz.

Concepto	Por ha		Por t	
	Sin renta	Con renta	Sin renta	Con renta
Ingreso (\$)	23,230.11	23,230.11	1,842.20	1,842.20
Costo (\$)	15,412.41	17,412.41	1,240.95	1,408.68
Beneficio (\$)	7,817.70	5,817.70	604.21	433.52
Precio-costos (u)	1.52	1.34	1.52	1.34
Tasa de ganancia (u)	0.52	0.34	0.52	0.34
CPMe (t)	8.37	9.48	8.37	9.48

Fuente: Elaborao con datos de la encuesta aplicada a los productores en el año 2012.

La tasa de ganancia de los productores de limón persa antes y después de pagar la renta de la tierra es de 0.52 y 0.34, respectivamente, los cuales indican que la inversión se incrementa en 52% en el primer caso y en 34% en el segundo caso. La cantidad de producción mínima económica que se debe producir para recuperar los costos de producción es de 8.37 sin renta de la tierra y de 9.48 con renta de la tierra.

El 2.78% de los productores no son rentables, ya que presentan valores negativos, debido a que tienen plantaciones jóvenes y no se encuentran en plena producción, el 34.72% tienen una rentabilidad inferior a la media y el 47.91 una rentabilidad igual o superior a la media, por lo que son más competitivos.

4. CONCLUSIONES

La superficie sembrada y cosechada de limón persa en el estado de Veracruz tuvieron un incremento superior al doscientos por ciento, el volumen de producción tuvo un crecimiento mayor al trescientos por ciento y los rendimientos tuvieron un ligero crecimiento, del nueve por ciento. En el estado de Veracruz y en el municipio de Misantla, el limón persa es uno de los cítricos más importantes, ocupando el segundo lugar en

superficie sembrada, superficie cosechada y valor de la producción, solamente superado por la naranja valencia.

Los rubros que absorben la mayor proporción de los costos de producción de limón persa son la mano de obra y los fertilizantes. Las actividades con mayores costos de la mano de obra para los productores fueron la poda, la aplicación de fertilizantes granulados y la labor de cosecha. Los indicadores de competitividad precio-costo unitario y tasa de ganancia reflejan que los productores son competitivos y están usando adecuadamente los factores que intervienen en las diferentes etapas de la producción, son productores eficientes y rentables.

REFERENCIAS

- CONSEJO ESTATAL CITRÍCOLA DE VERACRUZ A. C. (CONCITVER). (2012): Portal de internet. [en línea] Disponible en Web: <<http://www.concitver.com/PADRON/p-mtzdelatorre.html>>
- FAOSTAT. (2019): Datos. [en línea]. FAO. Disponible en Web: <<http://www.fao.org/faostat/es/#data>> Consultado 5, 3, 2019.
- FERGUSON, C.; GOULD, J. (1978): Teoría Microeconómica. E. L. Suárez, Trad. México, D.F.
- HEREDIA G., D.C. (2008): Metodología de costeo basado en las actividades para confecciones. En Económicas CUC. Col. 29, Núm. 29. EDUCOSTA-Corporación Universidad de la Costa, Barranquilla, Colombia.
- IBÁÑEZ, C.; CARO T. (2001): *Algunas teorías e instrumentos para el análisis de la competitividad*. IICA: San José Costa Rica.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI). (2015). Tabulados de la Encuesta Intercensal 2015. INEGI: México. [en línea] INEGI: México. Disponible en Web: <<https://www.inegi.org.mx/programas/intercensal/2015/>>, Consultado 3, 7, 2019.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI). (2019): Marco Geoestadístico. [en línea] INEGI: México. Disponible en Web: <<https://www.inegi.org.mx/datos/?t=0160000000000000>>, Consultado 9, 7, 2019.
- KLEEBERG H., F.; RAMOS R., J. C. (2009): Aplicación de las técnicas de muestreo en los negocios y la industria. *Ingeniería Industrial*, núm. 27.
- LERMA K., A. (2000): Comercio y Mercadotecnia Internacional: Metodología para la formulación de estudios de competitividad empresarial. México: Ediciones International Thomson.
- MCFETRIDGE D., G. (1995): Competitiveness: Concepts and Measures, Occasional Paper No. 5, Carleton University, Canada.
- MUÑOZ J., J. (2008): Contabilidad financiera. Pearson Educación S.A. España.
- ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS (OCDE) (1992): Technology and the Economy: The key relationships. París: OCDE.

- SAUNT, J. (1992): *Variedades de cítricos del mundo*. Sinclair Internacional. Valencia, España.
- SERVICIO DE INFORMACIÓN ARANCELARIA VÍA INTERNET (SIAVI). (2019): *Estadísticas anuales*. [en línea] Secretaria de Economía: México. Disponible en Web: <<http://www.economia-snci.gob.mx/>>, Consultado 2, 5, 2019.
- SISTEMA DE INFORMACIÓN AGROALIMENTARIA DE CONSULTA (SIACON). (2019): *Agrícola estatal*. [en línea]. SIAP-SAGARPA: México. Disponible en Web: <<https://www.gob.mx/siap/documentos/siacon-ng-161430>>. Consultado 02, 05, 2019.

CAPÍTULO 3

DETERMINING PROBABILITY RECOMMENDATION OF HOTEL CUSTOMERS IN SIX LATIN AMERICAN CITIES

Carolina Henao^{1}, Andreina Moros², Yovanni Castro³, Jenny Paola Lis-Gutiérrez²,
Manuel Ignacio Balaguera²*

¹ Corporación Universitaria Minuto de Dios (Bogotá, Colombia)

² Fundación Universitaria Konrad Lorenz (Bogotá, Colombia)

³ Pontificia Universidad Javeriana (Bogotá, Colombia)

SUMMARY

This chapter aims to identify the determinants of probability that a customer recommended by the hotel we have stayed. The study was conducted using data from 400 surveys of guests from different hotels in six Latin American cities: Cartagena, Montevideo, Tijuana, Temuco, San Sebastian and Bilbao. For this purpose, was estimated a model Logit and chose the model that minimized the loss of information with the criterion of Akaike, then took the test of kindness and adjustment of Pearson and concluded that the model presented a good fit. To detect whether the model was correctly specified, linear predicted value and the square linear predicted value as predictors was calculated to reconstruct the model and found that there was no additional predictor was statistically significant except by chance. Empirical testing showed that the existence of elements technology in the rooms increases the likelihood that the individual recommended by the hotel, while the failure promised, lack of visually appealing physical elements and little quickly to meet the needs of customers, decreases the likelihood of recommending the hotel. Finally, variables like the number of stars the hotel, the existence of updated technological means and reasons for travel were not significant when recommending the hotel.

Keywords: Service Quality; Servquality; Recommendation, Logic Model

* Corresponding Author address
Email: linda.henao@uniminuto.edu.co

INTRODUCTION

This paper aims to identify the determinants of probability that a customer hotel recommended by the hotel where have stayed. To do this, the data obtained from the measurement of perceived quality of service in a total sample of 400 customers in six Latin American cities were used by an adaptation of the world famous Servqual, which measures one hand expectations that the client an ideal hotel and on the other, the same variables regarding the service received at the hotel which is evaluating these variables are grouped into five dimensions: Tangibles; reliability; Answer's capacity; Security and empathy. The Servqual has been a tool used in various service sectors [1] they used the five dimensions to measure perception quality and its impact on satisfaction customers; [2] used four of them as a base and added the communication Green and reduction of Energy in his model GLSERV; [3] and [4] used the five dimensions the Servqual measure which one was plus important in hotels 2/3 and 4/5 stars respectively. Thus, the review of the literature shows that the Servqual has been used as a basis in research seeking to measure the perception of service quality in the tourism sector. For this reason, [5] performed an adaptation of Servqual which called Cultur, which maintains the original five dimensions of Servqual. Then, it is considered important to highlight some characteristics of cities in which the instrument was applied, so were chosen for analysis of the tourism sector:

Cartagena (Colombia)

Cartagena is a coastal city in Colombia, it has more than 14,000 lodging establishments and lodging assets, of which approximately 700 are in the Department of Bolivar it belongs Cartagena, which in turn has approximately 331 establishments equivalent to 11,434 rooms and 54% of these are hotels ranked four and five star [6]. According to the [7] city tourism is the main generator of employment with 14,574 employees and 2,836 companies formally constituted, besides their hotel occupancy is one of the highest in Colombia with 66.20% in 2017.

Temuco (Chile)

Temuco is the capital of the region of Araucania, which it has a rich natural and cultural heritage highlighting its landscape, the multi-entnicidad, marked by the presence of the "Mapuche" people, the arrival of Swiss, German, French and Italian as colonists and, of course, a Creole identity that determine cuisines, handicrafts. The area has developed special interest tourism, which is done in Mapuche's communities to show the ancestral culture highlighting cultural, historical, economic, social aspects in traditional houses called "rukas". Coordinator of the Economic and Social Observatory belonging to the Laboratory for Research in Applied Social Sciences (LICSA) at the University of La Frontera (OES-UFRO), analyzed the figures who experienced the 2016 special interest tourism which reached 64%.

Tijuana (Mexico)

According to Ministry of Tourism [8], the city Tijuana is popularly known as the "world's most visited border", being a charming destination that surprises visitors with its wide variety of attractions and activities to do. It has a first class hotel infrastructure and great restaurants where you can enjoy the delicious regional dishes. It has been recommended by the New York Times as the eighth place to visit in 2017. You can walk the streets to soak in its multifaceted and vibrant culture; It has its own charm and is a melting pot of regional diversities of Mexico and other cultures beyond its borders as it has been favored by the proximity to the United States.

Montevideo (Uruguay)

Montevideo is the capital of Uruguay and turn concentrates almost 50% of the total population, is the destination with the highest number of visitors to the country. Thus, tourism is mainly managed by the Department of Tourism of the Municipality of Montevideo, systematically articulating with private sector actors through the various associations: Uruguayan Chamber of Tourism (CAMTUR), Association of Hotels and Restaurants of Uruguay and the Bureau of Montevideo, public - private partnership whose mission is to promote the development of the city and facilitation in attracting conferences and events.

According to the annual report of [9], in recent years there has been a substantial improvement in infrastructure and tourism services mainly provided by private organizations. private.

San Sebastian and Bilbao (Basque Country-Spain)

Basque Country (or Euskadi), one of the 17 regions that make up Spain, and consists of three provinces: Araba or Alava (Vitoria-Gasteiz Capital, welcomes around 15% of the population); Bizkaia or Vizcaya (capital Bilbao-Bilbo, hosts around 52% of the population) and Gipuzkoa or Guipúzcoa (San Sebastián-Donostia capital city, hosts around 33% of the population) [10]. The Basque Government has a Strategic Plan for Tourism, Trade and Consumer Affairs [11], aligned with the Strategy of the European Union for growth and employment - Europe 2020, and has as one of its three global objectives "A strong tourism growth with more and better jobs".

Importantly, the accommodation in the Basque Country has grown significantly in recent years and has done so in a more important way in Bilbao where there has been further growth in the number of visitors. This growth will be enhanced with increasing due 1000 rooms over the "imminent arrival of six new establishments" [12].

Thus, it summarizes the most important in each of the cities in the sample aspects: Cartagena (Colombia); Temuco (Chile); Tijuana (Mexico); Montevideo, Uruguay); San Sebastian and Bilbao (Basque Country-Spain) and decided that the contribution of this work is based on identifying: What are the determinants of probability recommendation of hotel customers in six Latin American cities? To answer this research question, the data obtained in the study by [5] and decided to use a logistic regression with binary response variables were used. Thus, the chapter is composed of the introduction, methodology, results and conclusions and recommendations.

METHODOLOGY

As already mentioned in the previous section, the data of this study were taken from the results of applying Caltur which is an adaptation of Servqual, consisting of five dimensions and 16 items in total. Each dimension includes, for a part, expected service levels in general (the expectation that was created) and, moreover, the same items but referred to the perceived level of service delivered by the hotel is evaluating (the actual customer's perception of the stated feature). These items are evaluated using a scale of seven-point Likert ranging from 1 (totally disagree) to 7 (completely agree). Thus, the Figure 1 you can see the dimensions that compose the Caltur and Table 1 you can observe the distribution of the variables in each dimension.

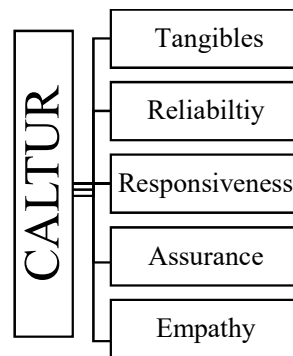


Figure 1. Dimensions of perceived quality of hotel service Caltur

Source: Self made

Table 1. Caltur dimensions according to the corresponding number on the instrument

Dimensions	No. OF ITEMS	No. QUESTION
Tangibles	5	From 8.1 to 12.2
Reliability	3	From 13.1 to 15.2
Responsiveness	5	From 16.1 to 20.2
Security	1	From 21.1 to 21.2
Empathy	2	From 33.1 to 23.2

Source: Self made

The instrument was applied to a voluntary sample of 400 participants, men and women over age, all customers who received hotel service (more than three stars) in the participating cities: Cartagena (Colombia); Temuco (Chile); Tijuana (Mexico); Montevideo (Uruguay) and San Sebastian and Bilbao (Spain) in 2017. At this time, to validate the sample size was used the software G * d Power proposed by [13]. For this calculation a probability of committing Type I (alpha) error of 0.05 and a probability of committing a type II error (Beta) 0.2 and a size of the high effect of 0.8 as reference values cited by [14]. Thus, the calculation resulted in a size of 21 subjects per group (men and women)

It was applied in each city by the members of each research group in different cities, experts on the subject, most led by people with level of training doctoral and experienced management people in the field and the data were entered in each city by the same

researchers, which were checked for consistency information stripped and only valid data for analysis were used.

By having assessment of hotel customers in these six Latin American cities in each of the items that make up the model of perception of quality of service (Caltur), this paper raises a statistical analysis to meet the objective of identify the determinants of possible recommendation to perform these customers. To do this, a logit model where the dependent variable is a dichotomous variable that takes the value 1 if the individual replied that recommend the hotel where he stayed and 0 otherwise would recommend estimated. And is the random variable corresponding to a response variable observed X, then the link function performs the transformation and the model logit equation is:

$$g \{E(\text{and / recommendation})\} = E \{ \log(\text{and / recommendation}) \} - \log \{1 - E(\text{and / recommendation})\}$$

$$P_i \{E(x_i = 1 / L2)\} = 1 / (1 + e^{-z_i}) = \pi_i$$

Pi being the probability of poor considered: $1 - \pi_i = 1 / (1 + e^{z_i})$

$$\pi_i / (1 - \pi_i) = (1 + e^{z_i}) / (1 + e^{-z_i}) = e^{z_i}$$

So, they are taken into account assessments of perception and estimation of the model shown in Table 2.

Table 2. Logit Model
 Logistic regression Number of obs = 386
 LR chi2 (33) = 93.40
 Prob> chi2 = 0.0000

Log Likelihood = -88.178332 Pseudo R2 = 0.3462

recommendation	Odds Ratio	Std. Err.	z	P>	z	[95% Conf. Interval]
----- + -----						
p82medios						
2	.0625711	.1357049	-1.28	0,201	.0008918	4.390051
3	6.596331	14.99924	0.83	0,407	.07652	568.6304
4	1.579615	2.482266	0.29	0,771	.0726008	34.36855
5	.6308121	.9190324	-0.32	0,752	.0362895	10.96525
6	1.101199	1.650945	0.06	0,949	.0583073	20.79738
7	.6438648	.9315996	-0.30	0,761	.0377748	10.97456
p102elementos						
2 1 (empty)						
3	1.249614	1.88231	0.15	0,882	.0652528	23.93053
4	9.686589	13.41283	1.64	0,101	.6419913	146.1546
5	8.72119	10.34826	1.83	0,068	.8522547	89.24463
6	31.85902	38.47711	2.87	0,004	2.986831	339.8243
7	36.05218	41.31608	3.13	0,002	3.814589	340.7338
p112fisico						

```

1 | 1 (empty)
2 | .0387707 .0591519 -2.13 0,033 .0019491 .7712041
3 | .2346875 .2732532 -1.24 0,213 .0239559 2.299152
4 | .2143897 .1790132 -1.84 0,065 .0417311 1.101408
5 | .5469631 .3900335 -0.85 0,397 .1351996 2.212793
6 | .6877639 .4558923 -0.56 0,572 .1875904 2.521554
7 | 1 (omitted)
|
stars |
3 | 2.834766 3.781663 0.78 0,435 .2074859 38.72985
4 | 1.243414 1.675178 0.16 0,872 .0886835 17.43367
5 | .6959016 .9893131 -0.26 0,799 .0428999 11.28859
|
p7motivo |
2 | 2.25583 2.726618 0.67 0,501 .2110866 24.1075
3 | 1 (empty)
4 | 1.161605 1.409142 0.12 0,902 .1077622 12.52133
5 | 2.137752 1.422226 1.14 0,253 .5803117 7.875051
6 | 1.091055 1.049069 0.09 0,928 .1657299 7.182772
7 | .4335806 .2690914 -1.35 0,178 .1284671 1.463348
|
p142cumplimiento |
1 | 1 (empty)
2 | 1 (empty)
3 | .0274087 .0388175 -2.54 0,011 .0017075 .4399493
4 | .0720245 .0639717 -2.96 0,003 .0126313 .410688
5 | .6481364 .5149179 -0.55 0,585 .1365909 3.075468
6 | 1.138183 4.052362 3196802 .7374324 0.842 0.20
7 | 1 (omitted)
|
p182rapidez |
1 | 1 (empty)
2 | .3695941 .8164164 -0.45 0,652 .0048692 28.0537
3 | 3.445592 7.774291 0.55 0,583 .0413712 286.9658
4 | .1974109 .1832893 -1.75 0,081 .0319933 1.218099
5 | .4061699 .3318441 -1.10 0,270 .0818972 2.014403
6 | .2675223 .1704633 -2.07 0,039 .0767327 .9326945
7 | 1 (omitted)
|
_cons | 1.4707 2.954872 0.19 0,848 .0286632 75.46131

```

Source: Self made

The dependent variables are perceived responses guests calling from 1 to 7, the following:

- P 8.2 Updated technological means: reservation system, electronic record, global banking for ATM payments.
- P10.2 Technological elements in the rooms: key, illuminated, TV, internet, digital safe in the rooms.

- Q11.2 Physical elements visually appealing: rooms, recreational areas, restaurant, reception, linen, crockery, brochures.
- P14.2 Meet the promises at the right time.
- Q18.2 Employees to respond quickly to customer needs.

In addition, the following additional variables were taken:

- Stars: the number of stars the hotel.
- P7 Reason the reasons for travel (tourism wellness (1), tourism wellness (2), religious tourism (3) Nature (4) Business tourism (5), academic events (6) and family events and / or personal.

The previous model aims to identify the determinants of probability that a customer recommended by the hotel we have stayed. To do this, the model that minimized the loss of information with the Akaike criterion, which is a measure of the relative quality of the model was chosen (Table 3).

Table 3. Akaike's Information Criterion and Bayesian Information Criterion

Model	Obs	ll (null)	ll (model)	df	BIC	AIC
.	386	-134.8801	-88.17833	34	244.3567	378.8551

Note: N = Obs used in calculating BIC; see [R] BIC note.

Source: Self made

To assess the predictive power of the model the ROC curve was made, since the predictive ability of a logistic regression model is based on the concept of sensitivity ($P(= 1 | y = 1)$) the concept of specificity: $P(= 0 | y = 0)$. A curve of this type is a graph showing the sensitivity function Specificity is shown at a 45 ° line represents a model no predictive power, as shown in Figure 2 the area under the curve is 0.8840, so it can be concluded that the model has a high predictive power.

Number of observations = 386
Area under ROC curve = 0.8840

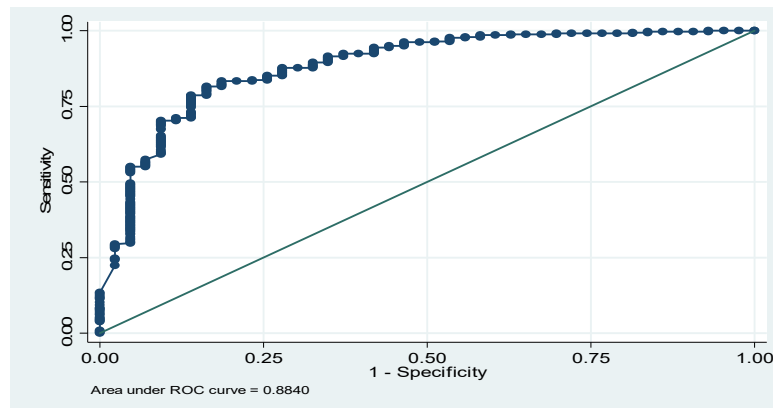


Figure 2. ROC curve

Source: Self made

In the table 4 it is evident that the overall correct classification rate is estimated at 91.97%, with 98.54% of the group recommends properly classified hotel (specificity) and only 39.53% of the group does not recommend the hotel, correctly classified (sensitivity).

Table 4. Logistic model for recommendation

----- True
Classified D ~ D Total
+ ----- + -----
+ 338 26 364
- May 17 22
+ ----- + -----
Total 343 43 386
Predicted Pr classified + if (D) >= .5
True D defined as recommendation! = 0

Sensitivity Pr (+ D) 98.54%
Specificity Pr (- ~ D) 39.53%
Positive predictive value Pr (D +) 92.86%
Negative predictive value Pr (~ D -) 77.27%

False + rate for true ~ D Pr (+ ~ D) 60.47%
False - rate for true D Pr (- D) 1.46%
False ++ Pr rate for classified (~ D +) 7.14%
False - rate for classified - Pr (D -) 22.73%

91.97% correctly classified

Source: Self made

Then test kindness and Pearson adjustment was made and it was concluded that the model had a good fit and it was shown that the number of covariate patterns is not close to the number of observations, making the applicability This test appropriate (Table 5).

Table 5. Logistic model for recommendation goodness-of-fit test

Number of observations = 386
Number of patterns covariate = 266
Pearson chi2 (232) = 231.46
Prob> chi2 = 0.4977

Source: Self made

When grouping the data ordering predicted probabilities and then forming 10 groups of equal size, again it accepted that the model has good fit (Table 6).

Table 6. Logistic model for recommendation goodness-of-fit test in ten groups
(Table Collapsed on quantiles of Estimated Probabilities)

Group	prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.6539	16	16.4	23	22.6	39
2	0.8489	31	30.0	8	9.0	39
3	0.9182	32	33.7	6	4.3	38
4	0.9485	40	39.3	2	2.7	42
5	0.9640	34	33.5	1	1.5	35
6	0.9767	38	37.8	1	1.2	39
7	0.9847	39	38.3	0	0.7	39
8	0.9898	67	68.2	2	0.8	69
9	0.9912	11	10.9	0	0.1	11
10	0.9968	35	34.8	0	0.2	35

Number of observations = 386
Number of groups = 10
Hosmer-Lemeshow chi2 (8) = 4.25
Prob> chi2 = 0.8339

Source: Self made

Finally, to detect whether the model is correctly specified, you should not find any additional predictor that is statistically significant except by chance. It used the linear predicted value (hat) and the square linear predicted value (hatsq) as predictors to rebuild the model. Thus, the variable hat should be a statistically significant predictor, since it is the predicted value of the model. In addition, if the model is correctly specified, the variable hatsq should not have much predictive power except by chance as hatsq not significant, there is no evidence have been omitted variable(s) relevant(s) or link function is not specified correctly (Table 7).

Table 7. Logistic regression

Number of obs = 386
 LR chi2 (2) = 93.58
 Prob> chi2 = 0.0000
 Pseudo R2 = 0.3469
 Log Likelihood = -88.090658

recomendation	Coef.	Std. Err.	z	P>	z	[95% Conf. Interval]
_ Hat	.9234694	.2181043	4.23	0,000	.4959928	1.350946
_ hatsq	.0268877	.0636651	1,516,691	0.42	0.673	-.0978937
_ cons	-.0042897	2,728,687	5,305,231	-0.02	0.987	-.5391024

Source: Self made

RESULTS

Empirical testing showed that the existence of elements technology in the rooms as key lighting system, TV, internet, digital safe, increases the likelihood that the individual recommended by the hotel, since customers make the assessment of perception on a scale from 1 to 7 in this item, a significant variable since scored 5 hereinafter also estimated beta is higher in increasing the rating level.

While; the lack of visually appealing physical elements such as rooms, recreational areas, restaurant, reception, linen, crockery, brochures; Failure to comply with promised at the right time and little quickly to meet the needs of customers; decrease likelihood to recommend the hotel.

Finally, variables like the number of stars the hotel, the existence of updated technological means and reasons for travel were not significant when recommending the hotel.

CONCLUSIONS AND RECOMMENDATIONS

This work allows us to conclude that the tourism sector is in a process of evolution worldwide. The six Latin American cities where data were taken, have specific characteristics that can be exploited for development, with tourism as an important source of income for countries: Colombia, Chile, Mexico, Uruguay and Spain, so it is recommended to continue making research with results that can be disseminated not only in the scientific community but also in the real sector for designing strategies.

On the other hand, it was determined that the variables used to measure the perception of service quality in these cities are suitable to analyze which are decisive for possible recommendation of the hotel in which you received the service, using the statistical model Logit.

The results showed that the item 10.1 "has technological elements in the rooms: key, illuminated, TV, internet, digital safe" found within the dimension "Tangibles" increases the likelihood that the customer recommend the hotel. On the other hand, a low rating in item 14.1 "deliver the goods at the right time" corresponding to the dimension of "Reliability" and in item 18.1 "Employees respond quickly to customer needs" corresponding to dimension "responsiveness" decrease the likelihood of recommending the hotel evaluated.

REFERENCES

- [1] Ahmad, S. Z., Ahmad, N., & Papastathopoulos, A. (2018). Measuring service quality and customer satisfaction of the small-and medium-sized hotels (SMSHs) industry: lessons from United Arab Emirates (UAE). *Tourism Review*, 74(3), 349-370. <https://doi.org/10.1108/TR-10-2017-0160>
- [2] Lee, W. H., & Cheng, C. C. (2018). Less is more: A new insight for measuring service quality of green hotels. *International Journal of Hospitality Management*, 68, 32-40.
- [3] Lestari, Y. D., & Saputra, D. (2018). Market study on hospitality sector: evidence from 4/5-star hotel in Bandung city Indonesia. *International Journal of Business & Society*, 19(1).
- [4] Lestari, Y. D., & Laode, M. I. (2018). Service Innovation of 3/2 Star Hotel in Bandung. *The Journal of Asian Finance, Economics and Business (JAFEB)*, 5(3), 73-80.
- [5] Moros Ochoa, M.A., Lis-Gutiérrez, J. P., Castro Nieto, G.Y., Vargas, C. y Rincón-Vásquez, J.C. (2019). La percepción de calidad de servicio como determinante de la recomendación: una predicción mediante inteligencia artificial para los hoteles en Cartagena. En: G.A. Campos Avendaño, M.A. Castaño Hernández, M. Gaitán Ángulo y V. Sánchez Mendoza (Compiladores). *Diálogos sobre investigación: avances científicos Konrad Lorenz* (pp. 143-162). Bogotá: Konrad Lorenz Editores. Disponible en https://blogs.konradlorenz.edu.co/files/book_interactivo_dialogos_en_investigacion.pdf
- [6] Bello, C. A. (2015). Tourist indicators of Cartagena de Indias 2014-2015. Retrieved from: www.colombiatraderade.com.co (2015/09/02).
- [7] Chamber of Commerce of Cartagena (2017). Cartagena in figures. Retrieved from: https://www.cccartagena.org.co/sites/default/files/publicaciones/cartagena_en_cifras-abril_2017_0.pdf (2018/02/08).
- [8] SECTUR - Ministry of Tourism of Mexico (2016). Indicadores turísticos de Baja California. Retrieved from: <https://www.bajanorte.com/files/estudios/resultados-secture-2016.pdf> (2018/10/07).
- [9] Uruguay XXI. (2017). Informe Sectorial de Turismo. Montevideo.
- [10] Basque Government (2015). Euskadi in figures. Retrieved from: <https://turismo.euskadi.eus/guia-sobre-euskadi/cifras/> (2018/07/20)
- [11] Basque Government (2017). Strategic Plan tourism, trade and consumption 2017-2020. Retrieved from: <http://www.euskadi.eus/plan-gubernamental/13-plan-de-turismo-comercio-y-consumo-2017-2020/web01-s2tumerk/es/> (2018/07/20)

[12] Vázquez, S. (2017). The boom of the hotels in Bilbao: 1,000 rooms more. The Mail, August 14th. Recovered from: <https://www.elcorreo.com/bizkaia/frenesi-hotelero-bilbao-20170814212806-nt.html> (2018/07/20)

[13] Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39 (2), 175-19. doi: 10.3758/BF03193146

[14] Castro, J. M. C., & Martini, H. A. (2014). Potencia estadística y cálculo del tamaño del efecto en G* Power: complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & sociedad*, 5(2), 210-244.

CAPÍTULO 4

SOBRE LA NECESIDAD DE IMPLEMENTAR INNOVACIÓN EN CURSOS INTRODUCTORIOS DE ESTADÍSTICA: UN ESTUDIO EXPLORATORIO EN LA EDUCACIÓN SUPERIOR MEXICANA

Cecilia Cruz López¹ y Mario Miguel Ojeda Ramírez¹

¹Universidad Veracruzana

Resumen

Se presenta una propuesta de innovación en cursos introductorios de estadística, argumentando sobre la necesidad de realizar un cambio orientado a la formación del pensamiento estadístico en las diferentes carreras universitarias; se introduce, para este fin, la pertinencia sobre la referencia de las llamadas Metas de Aprendizaje de la Estadística (MAE), propuestas por Gal y Garfield (1997). Para identificar qué tanto se utilizan las metas de aprendizaje se colectaron 410 programas de asignatura de estadística en diversas instituciones de educación superior (IES) de México que se encontraban disponibles en internet. Se realizó un análisis exploratorio, usando porcentajes y representaciones gráficas, para identificar las MAE que identificaron como las más y las menos atendidas en los programas de asignatura; se construyeron índices por meta y con un análisis de correspondencia múltiple se identifican los cursos que menos atienden las MAE en las IES del país. Se encontró que los cursos de áreas de ciencias de la salud fomentan menos el uso de las MAE.

Palabras clave: Educación estadística, Cursos de servicio, Educación superior, Análisis de correspondencia múltiple.

Abstract

A proposal for innovation in introductory courses in statistics is presented, arguing about the need to make a change oriented towards the formation of statistical thinking in the different university careers; For this purpose, the relevance of the reference of the so-called Statistics Learning Goals (MAE), proposed by Gal and Garfield (1997), is introduced. To identify how much the learning goals are used, 410 statistical course programs were collected in various higher education institutions (HEIs) in Mexico that were available online. An exploratory analysis was carried out, using percentages and graphic representations, to identify the MAEs that they identified as the most and least attended in the subject programs;

Indexes were constructed by goal and with a multiple correspondence analysis, the courses that the MAE least attend in the HEIs of the country are identified. It was found that courses in health sciences areas encourage less the use of MAEs.

Keywords: Statistical Education, Service courses, Higher education, Correspondence analysis.

Introducción

La estadística es una disciplina cuya consolidación data de los primeros cincuenta años del Siglo XX, que es cuando se fundan la mayoría de los departamentos y facultades de esta disciplina en las principales universidades del mundo; es así como se hace visible una academia mundial de la estadística; también es por esta época que se acepta que la estadística es una herramienta metodológica necesaria para las profesiones y las diferentes áreas de la ciencia, la ingeniería e incluso para las humanidades. Este reconocimiento para la estadística es el resultado de un movimiento a nivel mundial, en el que se coronan los esfuerzos de una comunidad científica y académica, pero que tiene su principal justificación en el devenir del desarrollo social, científico y de la educación superior; es decir, son las necesidades y los problemas que atiende la estadística los que hacen ver su importancia y propician su reconocimiento. Ante esto, los planes de estudio de la inmensa mayoría de las carreras se aprestan a incluir al menos un curso de estadística. Es de esta manera como la enseñanza de la estadística logra ocupar a una gran cantidad de profesores y estudiantes de la educación superior. Ante esta masificación se empiezan a analizar las formas en que esta disciplina se debe enseñar y se identifican los problemas que se presentan en el ámbito de los cursos introductorios de esta materia, sobre todo para estudiantes de otras disciplinas diferentes de la estadística, que por otra parte son la inmensa mayoría. Es así que, ya en la década de los años ochenta del siglo pasado, se acepta que la problemática del aprendizaje de la estadística es un fenómeno generalizado. Se habla de una materia clave, de elevada complejidad, en la que se conjugan muchos elementos asociados a las dificultades del aprendizaje en general (algunas dosis de matemáticas, razonamiento inductivo, lógica de la investigación, algoritmos, uso de recursos tecnológicos -calculadoras y después programas de cálculo estadístico, los llamados paquetes-, etc.). Esto pronto atrae la atención de los psicólogos educativos, de los que se dedican a la pedagogía, de los mismos profesores de estadística, y finalmente de los investigadores educativos. Muchas de las asociaciones profesionales y científicas de estadísticos empiezan a crear secciones de “educación estadística”, y es cuando aparece la investigación en esta temática. Pronto las revistas empiezan a dedicar espacio especial al tema, se hacen números especiales, e incluso empiezan a aparecer las revistas especializadas en la temática de educación estadística.

Innovación en cursos de estadística

Existe un reconocimiento de que la estadística es una metodología que se utiliza para los estudios técnicos y en la investigación en todas las áreas; asimismo, ayuda a la formación del pensamiento científico porque fomenta el análisis y la toma de decisiones a partir de información extraída de datos pertinentes. Se dice que esta manera de pensar y de actuar está sustentada en un pensamiento estadístico, el cual es deseable que se eduque. Los grandes movimientos a nivel mundial en este orden se agrupan en el tema de educación estadística. Batanero y Godino (2005) hacen hincapié en que la enseñanza de la estadística se ha incorporado crecientemente en todos los niveles educativos, principalmente por el valor que el desarrollo del razonamiento estadístico tiene en las grandes cantidades de información en el que actualmente se vive, así como por la frecuente necesidad de la toma de decisiones en ambientes de incertidumbre.

La enseñanza de la estadística es una cuestión que en las últimas décadas ha tenido una gran relevancia dentro del campo de la estadística, esto se debe a que internacionalmente ha habido cambios con las reformas educativas, por lo que la disciplina está intentando evolucionar en su proceso de enseñanza; al respecto Hassad (2003) menciona que la enseñanza y el aprendizaje a la estadística continúa generando un gran debate sobre el contenido y la pedagogía en medio de una reforma que se está usando. En las últimas décadas ha aumentado el número de países que se han preocupado por esta situación, debido a que se han dado cuenta de lo importante que es para un país que se tenga una cultura estadística básica a temprana edad; asimismo, Batanero (2001) menciona que la relación entre el desarrollo de un país y el grado en que su sistema estadístico produce estadísticas fiables es fundamental para la toma de decisiones acertadas que traen beneficios económicos, sociales y políticos, por tanto se puede decir que la estadística es un motor de desarrollo. De igual forma, Cravero, Redondo y Santellán (2011) enfatizan que en todos los ámbitos de la vida ciudadana se deben tomar decisiones, por lo que consideran necesaria una alfabetización científica del ciudadano y específica que ésta debe estar fundamentada en que se logre estar estadísticamente alfabetizado. Holmes no sólo enfoca la importancia de la estadística en niveles básicos de educación, sino que enfatiza que la estadística es útil en muchas profesiones que precisan conocimientos básicos del tema y el estudio de ésta ayuda al desarrollo personal ya que fomenta el razonamiento crítico basado en la valoración de la evidencia objetiva (citado por Batanero, 2001). Algunas de las primeras medidas que se están tomando es terminar con los cursos matematizados y que los conceptos estadísticos se enseñen como recetas. Moore (1993) habla también de que algunos movimientos fuertes han estado sucediendo, tanto en la investigación educativa como en la reforma a la educación estadística, y hace hincapié en que la enseñanza debe fomentar la actividad del estudiante en lugar de simplemente apuntar al conocimiento en una dirección general.

Diversas asociaciones internacionales han sido creadas con el propósito de fomentar el desarrollo de la educación estadística en el mundo, una de las más importantes

es la International Association of Statistical Education (IASE) creada en 1991 por el International Statistics Institute (ISI), esta asociación año con año organiza eventos mundiales que concientizan a los países en desarrollo a que tomen importancia del papel que juega la estadística dentro de sus procesos de enseñanza-aprendizaje. Países como Estados Unidos tienen ya varios años fomentando el pensamiento estadístico no sólo en sus niveles básicos de educación sino desde su sistema preescolar. En México, la inclusión de conceptos estadísticos en niveles básicos es reciente, los nuevos programas educativos de la Secretaría de Educación Pública (SEP) ya muestran desde los primeros grados de educación primaria contenidos temáticos que desarrollan la alfabetización estadística. Sin embargo, la educación estadística en México no es de reciente creación; ésta se ha venido incluyendo años atrás a partir del sistema educativo medio superior dentro de los cursos de matemáticas, y en casi todos los programas de educación superior se tiene al menos un curso básico de estadística.

Chance y Garfield (2002) enfatizan en que en la última década ha habido un movimiento pedagógico en educación estadística destinada a cambiar el enfoque de la enseñanza fuera de la teoría y recetas y encaminadas hacia el desarrollo del pensamiento estadístico, usando datos reales, la comprensión conceptual, y el aprendizaje activo. Además, aseguran que gran parte de este movimiento ha sido motivado por la investigación en psicología de la educación, la psicología, la educación matemática y la educación científica (Garfield, 1995). Asimismo, se está intentando que el estudiante deje de ser un ente pasivo que sólo escucha al profesor y se involucre de forma eficiente en el aprendizaje, involucrándose desde el inicio del curso en un proyecto que le permita aplicar todo el proceso de la metodología estadística, desde generar sus propios datos hasta presentar resultados.

No obstante, a pesar de los numerosos esfuerzos para que el aprendizaje de la estadística se desarrolle significativamente a nivel internacional, esto no ha sido posible, existen numerosas investigaciones que hablan de la problemática en los procesos de aprendizaje de la estadística (Garfield, 1995; Behar, Grima, Ojeda y Cruz (2013); Snee, 1993). Behar (2001) menciona que la problemática actual de la educación está enfocada en la metodología para enseñar más que para entender y actuar en el cuánto y de qué forma se aprende. En la mayoría de los cursos tradicionales de estadística no se están logrando los resultados esperados y hay insatisfacción en los estudiantes por la dificultad que tienen de aprender a usar y aplicar la estadística al terminar los cursos.

En este sentido, es necesario impulsar la educación estadística en México con el fin de incrementar la calidad de los cursos y con esto ir disminuyendo los niveles de insatisfacción de los jóvenes al término de ellos, para ello es necesario fomentar la innovación estadística en los cursos. Batanero (2001) indica que el uso de recursos informáticos facilita el aprendizaje de la estadística, ya que gracias a esto se desarrollan habilidades de solución de problema e interacción con los datos. De tal manera que el uso

de recursos tecnológicos puede ser visto como una herramienta de innovación educativa. Al respecto Palomo, Ruiz y Sánchez (2006) enfatizan que el carácter innovador de los profesores, se ve influenciado por el uso de las tecnologías de la información y la comunicación, además de que es un recurso para motivar al alumnado y con base a ello se replantean su papel como docentes y buscan nuevas formas de enseñar. Por esta razón, es necesario incluir el uso de tecnologías en cursos de estadística no sólo para introducir innovación en ellos sino para que el estudiante se vea motivado y adquiera un aprendizaje significativo.

Ben-Zvi y Gardfield (2004), recomiendan innovar en cursos de estadística incorporando las siguientes recomendaciones: Incorporar más datos y conceptos; que lo datos sean reales; centrarse en el desarrollo de la alfabetización, el razonamiento y el pensamiento estadístico; siempre que sea posible automatizar los cálculos apoyándose en herramientas tecnológicas; fomentar el aprendizaje activo a través de diversas alternativas a la docencia; fomentar una mayor variedad de actitudes, incluyendo una apreciación de la importancia que tienen los procesos estadísticos, oportunidad, aleatoriedad y rigor investigativo y un interés de convertirse en un evaluador crítico de resultados estadísticos; utilizar métodos de evaluación alternativa para comprender mejor el aprendizaje de los estudiantes;

Otra estrategia de innovación para cursos introductorios es presentada por Gal y Garfield (1997): son llamadas Metas de Aprendizaje de la Estadística (MAE), las cuales se detallan a continuación.

Las MAE como referente de evaluación de cursos

En el nivel universitario la mayoría de los estudiantes se encuentran con cursos introductorios de estadística, también llamados cursos de servicio; algunos de estos cursos introducen la estadística dentro de las necesidades y aplicaciones de un campo específico, como la psicología, negocios, economía o ingeniería. La estadística también sigue siendo enseñada como parte del plan de estudios básico en la mayoría de los departamentos de matemáticas. Estas versiones de los cursos introductorios de estadística en la educación superior difieren un poco en sus metas para los estudiantes, por el nivel de formación matemática necesaria, la cobertura de la teoría de la probabilidad, y el uso de la tecnología. Sin embargo, existe un descontento general por parte de los estudiantes; esto se pudo determinar a través de los resultados arrojados por un grupo de trabajo, encargado por la Asociación Americana de Matemática (Cobb, 1992). Este informe ofrece recomendaciones para la "reforma" de la mayoría de los cursos introductorios de la estadística. El mensaje básico era "más datos y conceptos, menos teoría y un menor número de recetas". También se instó a que los profesores de estadística incluyan oportunidades de un aprendizaje más activo y disminuir la cantidad de conferencias.

Gal y Garfield (1997) especifican que existen objetivos comunes a la enseñanza de los diversos niveles y contextos educativos. Y piensan en términos generales acerca de qué es lo que se quiere que los estudiantes aprendan y qué deben ser capaces de hacer con sus conocimientos; esto es lo que sería un objetivo único para la educación estadística. Este objetivo es que cuando los estudiantes terminen un curso de estadística, se convierten en ciudadanos informados capaces de: 1) Comprender y hacer frente a la incertidumbre, la variabilidad y la información estadística en el mundo que les rodea, y participar efectivamente en una sociedad cargada de información. 2) Contribuir o participar en la producción, interpretación y comunicación de datos pertenecientes a los problemas que se enfrentan en su vida profesional. Estas dos aseveraciones son una visión de conocimiento en un sentido amplio, y cuyo logro puede extenderse por varios años o niveles de escolaridad de los cursos de estadística.

Como parte de la consecución de esta visión amplia, se presentan ocho metas básicas para el aprendizaje de la estadística; estas metas se describen como "básicas", ya que se relacionan con la primera fase en que los estudiantes se enfrentan a la estadística en la mayoría de los niveles de enseñanza. Las MAE están orientadas al diseño y desarrollo de los cursos introductorios; esta propuesta ha sido tomada como una técnica de innovación internacional, debido a que en ellas se establecen los objetivos comunes que deben cumplirse en cursos de estadística tanto de nivel medio superior como a nivel universitario. Estas metas están orientadas a que el estudiante desarrolle el razonamiento y pensamiento estadístico. Las metas se aplican para cualquier disciplina e indican que un estudiante que haya llevado al menos un curso de estadística debe dominar ocho puntos que se desarrollan a continuación.

La primera meta es que el estudiante entienda el propósito y la lógica de las investigaciones estadísticas; esta aseveración hace énfasis en la comprensión del por qué se lleva a cabo una investigación estadística y que a su vez sea capaz de entender las "grandes ideas" que subyacen el enfoque de las investigaciones basadas en datos. Estas ideas son: la existencia de la variación en los datos; la necesidad de describir poblaciones a partir de éstos; la necesidad de resumir datos para identificar tendencias o patrones, la necesidad de estudiar la muestra en lugar de la población y hacer inferencia a partir de ésta; la lógica detrás de los procesos de muestreo (necesidad de representatividad, aleatoriedad, entre otras...); la noción del error de medición e inferencia; finalmente la necesidad de encontrar formas de estimar y controlar dichos errores, la necesidad de encontrar los procesos o factores causales que expliquen la variación de los datos; y la lógica detrás de los tipos de estudios (tales como experimentos o estudios observacionales) para determinar los procesos causales.

La segunda meta fue definida para que el estudiante entienda el proceso de la investigación estadística y las consideraciones que se deben tomar en cuenta en los diseños para recolectar los datos. En que pueda reconocer cómo, cuándo y porqué las

herramientas estadísticas deben ser usadas para ayudar en un proceso de investigación. A que puedan estar familiarizados con las etapas de una investigación estadística: La formulación de las preguntas de la investigación o el estudio; la planeación general del estudio (enfoque y diseño en general, diseño de muestreo, herramientas para la recolección de datos, etc.); la fase de recolección y organización de los datos; la importancia de saber resumir, explorar y analizar correctamente los datos; interpretar los resultados a la luz de las preguntas de investigación; discutir las conclusiones e implicaciones en base a los resultados, identificando los temas para futuras investigaciones.

La tercera meta se refiere a que el estudiante adquiera destrezas en los procedimientos estadísticos básicos; habilidades que se puedan usar en el proceso de investigación estadística. Estas destrezas incluyen; organizar los datos (construir la matriz de datos, tabularlos, manejarlos en software estadístico); construcción de tablas y gráficas, figuras útiles (ya sea a mano o asistidas por la tecnología); calcular estadísticas descriptivas (media, mediana, porcentajes, varianzas, desviación típica, etc.).

La cuarta meta indica que el alumno debe comprender las relaciones matemáticas, desarrollando un entendimiento intuitivo o formal de las principales ideas matemáticas que fundamentan los resúmenes, procedimientos gráficos o conceptos estadísticos. Particularmente: sobre la interpretación geométrica de los estadísticos descriptivos; las ideas matemáticas asociadas a los gráficos (concepto de funciones y relaciones); para explicar cómo la media se ve afectada a valores extremos; deducción de fórmulas (sumatoria y sumatoria de diferencias al cuadrado, etc.).

El quinto aspecto o meta se orienta a que razone el azar y la probabilidad, para tener una comprensión informal de la probabilidad y un razonamiento de la inferencia estadística, donde se destaca que la probabilidad es una medida de la incertidumbre. Particularmente: usando ejemplos de juegos de azar; usando simulación computacional; desarrollando modelos y usándolos para simular eventos (como una forma para generar datos y calcular probabilidades); discutiendo e ilustrando los conceptos y las palabras que estén relacionadas con los conceptos de azar, incertidumbre y probabilidad que aparecen en nuestra vida cotidiana (particularmente en los medios de comunicación); ejercitando el entendimiento de los eventos que sucedan en el mundo, así como la información en los medios de comunicación; demostrando que algunas veces nuestra intuición es incorrecta y nos puede llevar a conclusiones erradas con respecto a la probabilidad y eventos al azar.

La sexta meta hace énfasis en que el estudiante desarrolle destrezas interpretativas y cultura estadística, a fin de que sea capaz de interpretar los resultados de una investigación estadística y estar al tanto de identificar posibles sesgos o limitaciones en la generalización que se pueden manifestar en los datos. Se piensa que la mayoría de los estudiantes son consumidores más que generadores de datos y raramente recolectarán los

datos. Aunque así fuera, necesitarán ser aptos para entender lo ya publicado en artículos y que usen este conocimiento para lo que requieran. Por lo tanto, el estudiante necesita aprender a realizar la interpretación de resultados de una investigación estadística para que pueda tener una posición crítica y pueda hacer preguntas reflexivas sobre argumentos que se refieran a un resumen estadístico o a datos reportados en los medios de comunicación o en algún reporte de proyecto o incluso en trabajos de clase de sus compañeros. Ejemplos de preguntas que deberían aprender a hacer en cada contexto: ¿qué tan confiables son sus instrumentos de medición?, ¿qué tan representativa es la muestra?, ¿hasta dónde son válidas las conclusiones del estudio?, etc.

La séptima meta se refiere a que el estudiante obtenga habilidades para la comunicación estadística: se necesita buena redacción y habilidades de discurso, para que se tenga una comunicación efectiva acerca de investigaciones estadísticas y fenómenos de procesos probabilísticos. La comprensión de lectura y las habilidades de comunicación son requeridas para que pueda discutir de manera efectiva; que tengan un criterio bien sustentado para hacer afirmaciones basadas en los datos. Asimismo, debe estar capacitado para usar la terminología estadística o probabilística adecuada de manera correcta y así transmitir resultados de manera convincente y también construir sus propios argumentos basados en los resultados. Igualmente deberán saber discutir razonadamente sobre la validación de datos presentados por otras personas y plantear preguntas acerca de las generalizaciones hechas con base a una muestra.

La última meta indica desarrollar aprecio por la estadística: el estudiante debe desarrollar una valoración por el papel del azar y la aleatoriedad en el mundo, por los métodos estadísticos y la planeación de experimentos como herramientas científicas indispensables, como medios poderosos para tomar de manera sustentada decisiones personales, sociales y de negocios; debe dársele una postura de seguridad para pensar y actuar ante la incertidumbre. Deben darse cuenta de que el proceso de investigación estadística puede llevar a mejorar conclusiones; si se mantiene una postura en datos que no tienen fundamento o de basándose en sus propias experiencias subjetivas o intuiciones, todo esto abona en contra del pensamiento estadístico. Además, deberá estar preparado para adoptar posturas de cuestionamiento cuando se encuentre con argumentos que propongan basarse en datos o en un reporte de resultados o conclusiones de una investigación estadística, un estudio o una búsqueda empírica.

Las MAE pueden dividirse en dos grupos: aquellas que enfatizan en la generación de estadísticas; es decir, en el saber hacer, en el uso de las técnicas, la lógica que siguen los procedimientos, la comprensión de los propósitos, etc., este primer grupo lo definen las 5 primeras metas; el segundo enfoca las actividades hacia la importancia de las decisiones, a las destrezas de comunicación, del reporte correcto de los resultados, así como a la reflexión y al cuestionamiento de las conclusiones propias y de otras

investigaciones, etc.; este grupo lo definen las metas de 6, 7 y 8. Podemos denominar a los grupos en metas asociadas a “saber hacer” y metas para “saber pensar”.

Estos ocho propósitos destacan un cambio de punto de vista de la enseñanza tradicional tratando de que los cursos terminen con el enfoque matemático que se les da aún a muchos de ellos y que sólo usen a la matemática como una ciencia de apoyo. Moore (1992) especifica que la estadística no es una rama de las matemáticas y que ha surgido claramente como una disciplina con derecho propio. En la evolución de la educación estadística se ha ido observando cómo el desarrollo de la enseñanza de la estadística en países de América Latina está basado específicamente en el uso de cálculo diferencial y álgebra, cursos que tienen influencia europea. Este hecho es una constante en el desarrollo de la enseñanza de la estadística, porque permite darse cuenta que esta disciplina en la educación superior ha sido orientada matemáticamente y a pesar de que existen reformas y movimientos pedagógicos que generan acciones de cambio, éstos se han hecho aisladamente y no se tiene un esfuerzo conjunto que genere un cambio a nivel mundial.

Contenidos temáticos de un curso de servicio

Los contenidos temáticos de un curso de servicio en estadística tienen el propósito que el alumno adquiera los conocimientos básicos de estadística, bajo la aplicación de las técnicas de uso común para la aplicación adecuada de la metodología estadística siguiendo los principios de esta disciplina. Con el objetivo de demostrar que mediante la práctica y la aplicación de dichas técnicas y herramientas se puedan adquirir las competencias fundamentales del método científico y se incremente el pensamiento estadístico. Asimismo, se busca que mediante un proyecto el alumno tenga la capacidad de desarrollar una investigación que le permita entender el proceso de la metodología estadística en la solución de problemas.

El objetivo fundamental de un curso de servicio de estadística es que el estudiante adquiera las competencias metodológicas de la estadística que le permitan desarrollar una investigación, desde el planteamiento del problema hasta la fase de comunicación de los resultados. El contenido temático que se sugiere está encaminado a dar una solución a los problemas que se están presentando con respecto a la educación estadística.

Ojeda y Sahai (1994) recomiendan comenzar en la primera unidad del curso con la importancia de la recolección de datos, la comprensión de la variación, la exploración y visualización gráfica de los datos, los principios de inferencia estadística y modelación de datos. Con la comprensión de estos temas se logrará desarrollar el pensamiento estadístico que permitirá la solución de problemas reales. Esto da pauta a que al mismo tiempo que se resuelven problemas se desarrolle una apreciación hacia la estadística. Dándole este enfoque al curso, el estudiante obtiene una vista panorámica sobre los

objetivos, metas y propósitos en el uso de técnicas y métodos estadísticos que estarían encaminados a resolver problemáticas en los que estén involucrados desde el inicio.

La segunda unidad del curso debe estar enfocada al planteamiento del problema que sea del interés del estudiante, plantearse preguntas de investigación que a su vez estén relacionadas con los objetivos del estudio, asimismo, se debe dar a conocer los diferentes tipos de estudios que existen en la estadística, para él decida cuál es el que necesita para desarrollar su proyecto. En esta unidad se deben ver los métodos de muestreo porque es necesario llevar a cabo la recolección de datos.

La unidad tres se debe enfocar a temas clásicos para el análisis exploratorio de datos, es aquí donde se enseña la estadística descriptiva, la unidad cuatro introduce al estudiante en temas de inferencia estadística pero de una manera informal, a través de problemas que lo motiven y no que lo desmotiven; es decir, que los conceptos involucrados no se enseñen mediante tediosas fórmulas matemáticas que lo único que propician es su descontento, lo importante es que ellos comprendan la utilidad del método para resolver sus proyectos. Asimismo, los modelos estadísticos se deben presentar en el contexto del mundo real, haciendo referencia a problemas en los que se verán involucrados en futuras actividades profesionales. Se ilustrarían los conceptos donde se recolectan los datos y el profesor debe mantener un equilibrio entre el procesamiento de éstos y la discusión de los conceptos que se incluyen en los contenidos para este tema. Dependiendo de la disciplina o de los proyectos que se desarrollen en el curso se determinarán si se incluyen temas adicionales o de eliminan algunos de los contenidos en esta unidad (Ojeda & Sahai, 1994).

Las dos últimas unidades deben estar enfocadas a la presentación y comunicación de los resultados, desde la elaboración del reporte, normas de citación hasta la presentación del proyecto y elaboración del discurso de presentación. Ojeda (2011) indica que se deben preparar conferencias para dar de manera rápida los contenidos. Estas conferencias deben estar soportadas por materiales necesarios para realizar el proyecto y el análisis detallado de los temas. Se sugiere dar asesorías individuales a los estudiantes por lo que es importante planear desde el inicio cada sesión del curso y las actividades que se desarrollarán. Asimismo, para todos los temas que involucren técnicas estadísticas se debe usar software estadístico.

Teoría del pensamiento estadístico

El pensamiento estadístico es un área fundamental que debe desarrollarse en el estudiante si se quiere lograr un aprendizaje significativo de la estadística; diversos autores hacen hincapié en que si éste se desarrolla generaría un gran impacto en el mundo (Moore, 1990; Wild & Pfannkuch, 1999; Batanero, 2000; Ojeda & Sahai, 2003; Méndez & Ojeda, 2006).

Sin embargo, al hablar de la importancia de esta concepción es necesario que nos detengamos a reflexionar qué es el pensamiento estadístico.

Snee (1991, p.11) lo define como “procesos del pensamiento, bajo los cuales se reconoce que la variación está en nuestro alrededor y presente en cualquier cosa que hacemos; en todo trabajo hay una serie de procesos interconectados tales como identificar, caracterizar, cuantificar, controlar y reducir la variación que proveen oportunidades para el mejoramiento”. Observemos que esta definición de versa en torno al reconocimiento de la variación.

Más tarde la Sociedad Americana de Calidad (ASQ, siglas en inglés) en 1996 define el pensamiento estadístico como una filosofía de aprendizaje basada en tres principios fundamentales: 1) Todo trabajo ocurre en un sistema de procesos interconectados, 2) La variación existe en cualquier proceso y 3) Lograr entender y reducir la variación de los procesos es la clave del éxito.

Mallows (1998, p.3) propone que “El pensamiento estadístico tiene que ver con la relación de datos cuantitativos a un problema del mundo real, a menudo en presencia de variabilidad e incertidumbre. Intenta hacer preciso y explícito lo que los datos dicen sobre el problema de interés.” Mallows, a diferencia de Snee y de la ASQ, no basa su definición en términos de calidad, sino que la plantea de una forma más general, aplicable en cualquier ámbito.

Wild y Pfannkuch (1999, p.223) especifican que “el pensamiento estadístico es la encarnación del sentido común, se reconoce cuando se ve, su ausencia es evidente y para la mayoría es producto de la experiencia”

Las MAE de Gal y Garfield (1997) -presentadas en el capítulo dos y base fundamental de esta investigación- tienen como principal objetivo desarrollar el pensamiento estadístico en estudiantes de cursos introductorios, cada uno de los aspectos establecidos en las metas está pensado en fomentar en el alumno este razonamiento requerido por los usuarios de la estadística. Los pioneros del pensamiento estadístico son Moore (1990) y Wild y Pfannkuch (1999).

Moore (1990) indica que el pensamiento estadístico debe ser parte de la cultura de cualquier persona y señala que los elementos de este pensamiento son:

- La omnipresencia de la variación, este principio establece que todo es variable, incluso las medidas repetidas en un mismo individuo son variables.

- La necesidad recabar datos en los procesos. La prioridad es generar los datos y no consumirlos.
- Al generar los datos, se debe tener en cuenta la variación. Saber que existen fuentes de variación no controlada. La selección de los datos se debe hacer aleatoriamente.
- La cuantificación de la variación. La variación aleatoria se describe matemáticamente a través de la probabilidad.
- La explicación de la variación. Con el análisis exploratorio de los datos se buscan efectos sistemáticos detrás de la variación aleatoria.

Moore indica que el pensamiento estadístico está asociado a la vida cotidiana, pero desafortunadamente no va a ser aprendido por los estudiantes si no se incluye en los planes de estudio. El objetivo de desarrollar este pensamiento consiste en crear la capacidad para enfrentar de manera inteligente a la variación y a la incertidumbre, si éste se instruye enfocándose en aplicar la estadística en la vida cotidiana, entonces la adquisición del pensamiento se hace más evidente (Nisbett et al., citado por Moore, 1990).

Para Wild y Pfannkuch (1999) los elementos planteados por Moore no son suficientes y sugieren 4 dimensiones, basados en un estudio empírico realizado a profesores y estudiantes, éstos son:

Dimensión 1: El ciclo investigativo es un modelo de pensamiento: se plantea como una forma para pensar una cuestión en diferentes circunstancias o situaciones. Consiste en definir y delimitar el problema, porque cuando se aborda un problema, éste no aparece como definido ni delimitado y el investigador tiene que ocupar el pensamiento para la definición del problema, porque a partir de aquí es donde comienza todo el proceso de investigación y a partir de allí se plantea un modelo llamado PPDAC (Problema, Planeación, Datos, Análisis y Conclusiones) planteado por MacKay y Oldford (1994). El término planear no solamente consiste en el asunto del diseño, sino también en pensar en cuestiones de cómo se va a medir, qué se va a medir y cuando ya se tienen los datos, ahora se debe pensar qué se va a hacer con ellos, incluso antes de obtener los datos, se debe pensar: qué se va a hacer con ellos, porque de otra manera no habría un camino que seguir. Ya cuando se tienen los datos, hay ciertos procesos que cuidar, como depurarlos, como manejarlos y ya después viene el análisis, la exploración, el desarrollo de sistemas y procesos más elaborados, esto serie de elementos que se realizan durante análisis llevan a la elaboración de las conclusiones. Esta dimensión es fundamental en el proceso de investigación. Por eso la estadística se tiene que impartir siempre ligada al este proceso de investigación. En pocas palabras es específicamente solucionar un problema a través de la metodología estadística

Un ciclo de investigación PPDAC, deberá ser realizado para lograr cada una de las metas de aprendizaje. El conocimiento ganado y las necesidades identificadas, se convierten en puntos de partida para nuevos ciclos PPDAC. Esta dimensión la podemos asociar con la segunda meta del aprendizaje de la estadística que precisamente consiste en que el alumno comprenda la planeación de una investigación, desde plantear las preguntas de investigación hasta la elaboración de conclusiones de una manera crítica.

Dimensión 2: Tipos de pensamiento. En esta dimensión se plantean el conjunto de estrategias de pensamiento, por ejemplo, el pensamiento estratégico, es un estilo de pensamiento en el que antes de hacer cualquier acción se debe pensar a donde va a conducir, así es posible anticiparse a una solución. Siempre existen distintas formas de pensar en la que el investigador se siente más cómodo hacia la búsqueda de explicaciones, es como la modelización que consiste en buscar dentro de las acciones que se tienen para pensar, cuál es la que mejor funciona para esta problemática, de esta manera los estadísticos usan modelos de pensamiento y los adaptan a diferentes problemas.

La lista de pensamientos a la que finalmente llegaron fue la siguiente:

- a) Reconocimiento de la necesidad de datos: El reconocimiento de lo inapropiado de las experiencias personales y de la evidencia anecdotal, que conducen al deseo de basar las decisiones sobre datos colectados. La necesidad de pasar de las anécdotas a las afirmaciones soportadas por datos válidos. Este tipo de pensamiento se asemeja a uno de los aspectos de la quinta meta del aprendizaje de la estadística que establece que el estudiante comprenda que algunas veces nuestra intuición es incorrecta y nos puede llevar a conclusiones erradas con respecto a la probabilidad y eventos al azar. Otro aspecto asociado a este tipo de pensamiento es el aspecto tres de la meta 8 que indica que el estudiante debe tener una apreciación por el proceso de investigación estadística que puede llevar a mejorar conclusiones (Wild & Pfannkuch, 1999).

- b) Transnumeración. Es la capacidad para obtener información de los datos, es la capacidad de adquirir la competencia que se debe desarrollar en un estadístico para poder leer cuestiones relevantes en los datos, a partir de observar los resultados, ver las herramientas gráficas, pero bajo la condición que tengan sentido en la problemática que se está estudiando, es lo que se dice convertir las evidencias o lo que los datos muestran en un discurso del contexto real. Wild y Pfannkuch especifican que la transnumeración ocurre cuando se encuentra la forma de obtener los datos (a través de medición o clasificación) y se capturan los elementos significativos del problema real. Esta dimensión se relaciona con el tercer aspecto de la meta 1, que especifica que el estudiante debe entender la necesidad de resumir los datos para identificar tendencias o patrones. Transnumeración es ir más allá de los números.

- c) Variación. La variación es lo más estudiado en esta investigación, la omnipresencia de la variación indica que se puede trabajar con ella, existen dos tipos de variación, la variación explicada y la no explicada, los autores indican que existen distintas discusiones de cómo los estadísticos observan la variación y que a veces se centran tanto en ella que el análisis de ésta lo llevan hasta las últimas consecuencias, aun cuando no era necesario abundar tanto en ella, porque la variación a veces no es tan mala como algunos investigadores piensan. Wild y Pfannkuch mencionan que se aborda el estudio de la variabilidad, con el propósito de explicar, predecir o controlar un proceso. Esto se asocia con dos aspectos de la primera meta de aprendizaje, el primer aspecto especifica que el estudiante debe entender la existencia de variación y el otro es la necesidad de identificar procesos causales o factores que expliquen la variación en los datos.

- d) Un conjunto específico de modelos. En este aspecto los autores especifican que toda forma de pensamiento hace uso de modelos. El aspecto dos de la meta cinco es el asociado a este pensamiento que trata de que el estudiante use la simulación computacional, para desarrollar modelos y aplicarlos para simular eventos (como una forma para generar datos y calcular probabilidades).

- e) Conocimiento contextual, que plantean como la integración de lo estadístico con el contexto; es decir, debe de haber una integración alta de la estadística con el problema, el problema estadístico no puede funcionar aislado del contexto. El valor de la consultoría, como los estudiantes se deben apropiar de ciertas formas de pensamiento al resolver un problema, cómo interactúan con los clientes que en este caso son investigadores que requieren de la estadística, a veces se crea conocimiento con aspectos que los clientes comparten de su pensamiento, esto plantea que existen procesos de aprendizaje cuando se da interacción con los propietarios de los problemas con los que se trabaja y el estadístico.

Dimensión 3: Ciclo interrogativo. El pensamiento estadístico consiste también en hacer preguntas, a veces en una clase de estadística no se capacita a los estudiantes para hacer preguntas, sino que los profesores son los que hacen las preguntas. Por lo cual se deben crear estrategias para desarrollar preguntas, así se dará el ciclo interrogativo, esto permite desarrollar las ideas e ir interactuando con la parte del contexto del problema. Si al estudiante se le capacita y se le da énfasis a la importancia de que estudie, de que pregunte, de que pida explicaciones, de que veas aspectos relevantes, de que tenga siempre una visión de preguntar, ¿Qué sucedería si sucede esto? o ¿Qué sucedería si no sucede? ¿Es importante o no es importante?, ¿lo puedo considerar o no?, a partir de ejercicios de este tipo que implican estar pensando todo el tiempo sobre el problema se logra que un estudiante piense. Hay muchos elementos para diseñar estrategias para detonar el pensamiento, lo importante es integrarlas para el aprendizaje del estudiante.

Wild y Pfannkuch también aportan algunas componentes del ciclo de interrogación con más detalle:

- a) Generar. Esto significa la capacidad de imaginar y proponer posibles soluciones, este trabajo se puede realizar individual o en equipo. Puede aplicarse a la búsqueda de causas, explicaciones y mecanismos, a las partes de un sistema que pueden interrelacionarse, y a otros bloques de construcciones mentales y modelos estadísticos. Puede también aplicarse a diversos tipos de información en el cual se busca completar datos que se requieren para verificar una idea o para planear un enfoque que permita resolver un problema. La generación de posibilidades puede ser del contexto, datos o conocimiento estadístico, aplicados al presente problema o puede referirse a futuras investigaciones. Este apartado se puede asociar con algunos aspectos de la segunda meta de aprendizaje, donde el alumno debe generar tanto sus preguntas de investigación, como los recursos necesarios para recolectar sus propios datos.
- b) Buscar. Este enfoque precede a la generación, que consiste en la búsqueda de información, ya sea de manera interna o externa. Para la búsqueda interna, podemos observar a la gente pensando "...yo se algo de esto..." y hurgando en sus memorias conocimiento relevante o la estructura cognitiva con la que cuentan.
- c) Interpretar. Este proceso se aplica a todas las formas de información que se conocen como son: gráficos, resúmenes y otros productos del análisis estadístico. "Conectar" se puede llamar al punto final de "interpretar", esto se refiere a la interconexión de las nuevas ideas que se han creado con los modelos mentales existentes y la prolongación de los modelos mentales para acoplar o encajar estas interrelaciones.

Dimensión 4: Disposición o actitud. La superación de diferentes Disposiciones del investigador en cuanto a la actividad desarrollada tales como la curiosidad, la imaginación o el escepticismo, entre otras, que dan paso al investigador a la entrada en las otras dimensiones.

La existencia de variación es un concepto que muchas veces ni el mismo profesor comprende, por tal razón no se especifica en la mayoría de los programas y se sabe que la existencia de la variación es la razón de ser de la estadística, es un concepto básico que se debe dominar; sin variación no habría estadística.

Diversos autores especifican la importancia de este concepto en los cursos de estadística, entre ellos tenemos a Moore (1990) quien señala que los elementos fundamentales del pensamiento estadístico son:

- La Omnipresencia de la variación, en contraposición a la visión determinista.
- La necesidad de los datos en los procesos. La primera prioridad es buscar en los datos.
- El diseño de la producción de datos, teniendo presente la variación.
- La cuantificación de la variación. La variación aleatoria se describe, matemáticamente, por la probabilidad.
- Explicación de la variación. El análisis estadístico busca efectos sistemáticos detrás de la variación aleatoria.

Del mismo modo, Jiménez e Inzunza (2011) indican que en cursos de servicio de estadística a nivel universitario se debe desarrollar el pensamiento estadístico y esto se lleva a cabo a través de estudios de investigación cuantitativa rigurosos donde es importante el uso de la estadística, específicamente para fundamentar formalmente si los resultados se deben a un efecto verdadero o son producto de la variabilidad aleatoria, la que no se puede explicar por fuentes identificables, controlables.

Las 4 dimensiones del pensamiento estadístico presentadas por Wild y Pfannkuch están contempladas en las metas de aprendizaje de la estadística, para resumir esta asociación se presenta la tabla 3.1.

Tabla 1. Relación entre las dimensiones del pensamiento estadístico y las metas de aprendizaje de la estadística.

Dimensión	Meta
1. Ciclo de investigación	2
2. Tipos de pensamiento	1, 3, 4, 5
3. Ciclo de cuestionamiento	2, 6
4. Disposición o actitud	7, 8

En este contexto, el problema a investigar en este estudio se centra de contextualizar la enseñanza de la estadística en instituciones de educación superior (IES) del país a través de la evaluación de los programas de asignatura para ver qué tanto se consideran en ellos las metas de aprendizaje de la estadística. La interrogante que se plantean es ¿qué debe hacerse para mejorar el aprendizaje de la estadística en las IES del país? para responder a esta pregunta se debe determinar ¿qué tanto se incorporan las MAE en los cursos de las IES del país a través de los programas de asignatura?

Materiales y métodos

La metodología que se siguió para conseguir los programas de asignatura fue recolectar aquellos que estaban disponibles en internet; para ello se ubicó una página web que

contiene la lista más completa de universidades en México por región¹, de esa lista se consiguieron las direcciones electrónicas de las páginas oficiales de todas las instituciones tanto públicas como privadas del país, en sus modalidades de universidades y tecnológicos; mediante una búsqueda exhaustiva se recolectaron 410 programas. La lista de programas colectados por institución se presenta en la Tabla 2.

Tabla 2. Programas recolectados de diversas instituciones del país.

Institución	Número de programas
Instituto Politécnico Nacional	19
Instituto Tecnológico Autónomo de México	10
Instituto Tecnológico de Aguascalientes	9
Instituto Tecnológico de Cancún	6
Instituto Tecnológico de Chapala	10
Instituto Tecnológico de Chetumal	9
Instituto Tecnológico de Ciudad Guzmán	12
Instituto Tecnológico de Coacalco	18
Instituto Tecnológico de Comitán	14
Instituto Tecnológico de Durango	5
Instituto Tecnológico de El Grullo	10
Instituto Tecnológico de Ensenada	7
Instituto Tecnológico de Mexicali	2
Instituto Tecnológico de Nuevo León	8
Instituto Tecnológico de Pabellón de Arteaga	2
Instituto Tecnológico de Tijuana	11
Instituto Tecnológico de Tuxtla Gutiérrez	9
Instituto Tecnológico el Llano	3
Universidad Autónoma de Baja California	29
Universidad Autónoma de Chihuahua	16
Universidad Autónoma de Colima	14
Universidad Autónoma de Guadalajara	20
Universidad Autónoma de San Luis Potosí	21
Universidad Autónoma Metropolitana	10
Universidad de Michoacán	13
Universidad de Sonora	29
Universidad Emiliano Zapata	10
Universidad Nacional Autónoma de México	37
Universidad Veracruzana	47
Total programas	410

¹ http://www.altillo.com/universidades/universidades_mex.asp

Para el análisis de los programas se diseñó una lista de cotejo que mide qué tanto son consideradas las MAE; para ello se utilizaron como dimensiones las ocho metas y cómo variables los diversos aspectos que en ellas se presentan; asimismo, se propusieron tres categorías de clasificación: No se considera en ninguna parte del programa, Se considera levemente en el programa y Se considera suficientemente en el programa. De tal forma que se construyeron indicadores que permitieron medir el grado de consideración de las metas tal como se describe en la Tabla 3.

Tabla 3. Variables definidas y construidas para la estrategia de análisis de datos.

Variable	Descripción	Escala
Estado	Muestra el estado de la República Mexicana al que pertenece la Institución. Los estados se enumeraron conforme el investigador lo determinó.	Nominal 1: Aguascalientes 2: Baja California 3: Chiapas ... 15: Veracruz
Curso	Se hizo una clasificación de los cursos, porque había una gran variedad; se crearon 4 categorías.	Nominal 1: Bioestadística 2: Estadística 3: Probabilidad y estadística 4: Métodos estadísticos
Área	Indica el área a la que pertenece el programa.	Nominal 1: Humanidades 2: Técnica 3: Económico-administrativa 4: Biológico-agropecuaria 5: Ciencias de la salud
Sum_A*	Es la suma de los valores de los ítems de la meta 1.	Numérica
IM1**	Es la calificación creada con la variable Sum_A. En una escala de 0 a 100.	Numérica
Ind_IM1**	Es el índice de consideración de la meta 1 creado con base a la variable IM1, codificado en 3 categorías: 1=Bajo, 2=Medio, 3=Alto.	Ordinal 1: 70 o menos 2: 71 a 89 3: 90 o más
Sum_Total	Es la suma de los valores de todas las variables Sum.	Numérica
IMT	Es la calificación creada con la variable Sum_Total. En una escala de 0 a 100.	Numérica
Ind_Total	Es el índice de consideración de las metas creado con base a la variable IMT, codificado en 3 categorías: 1=Bajo, 2=Medio, 3=Alto.	Ordinal 1: 70 o menos 2: 71 a 89 3: 90 o más

* A corresponde a la Meta 1, lo mismo se hizo para las otras 7 metas.

** IM1 es la calificación creada para la Meta 1, lo mismo se hizo con las otras 7 metas.

*** Ind_IM1 es el índice de la Meta 1, lo mismo se hizo para las otras 7 metas.

Para el análisis de los datos se realizó primero un análisis exploratorio para identificar patrones en los datos, posibles errores y determinar la estrategia de análisis. La Tabla 3 muestra las variables que se crearon con dicha estrategia.

Se realizó un Análisis de Correspondencia Múltiple (ACM), por lo que primero se calcularon las sumas con las respuestas de valores de los ítems, obteniéndose una suma global por meta; al final se tuvieron 8 sumas globales, cada una representando a una meta. Cada meta se transformó en una calificación en la escala de 0 a 100; finalmente se construyeron 8 índices con 3 categorías: Bajo, para aquellos programas que tuvieran una calificación menor o igual a 70; Medio, para los que tuvieran una calificación entre 71 y 89; y Alto a los programas que tuvieran una calificación de 90 o más. Posteriormente se construyó un índice global, con la suma global de los valores de todas las metas y siguiendo el mismo procedimiento de categorización descrito anteriormente. Con la construcción de estos índices se pudieron identificar las carreras en las que los programas no contemplan el uso de las MAE y específicamente cuáles son los aspectos que no se contemplan, asimismo qué universidades contemplan menos las metas.

Resultados

La distribución de los programas recolectados por área académica: Exactas 51%, Económicas 25%, Agropecuarias y Sociales 9% ambas y 7% restante corresponde al área Ciencias de la Salud. La distribución de la clasificación de los cursos se presenta en la Figura 1. En la Figura 2 se presenta la distribución de frecuencias de programas por institución.

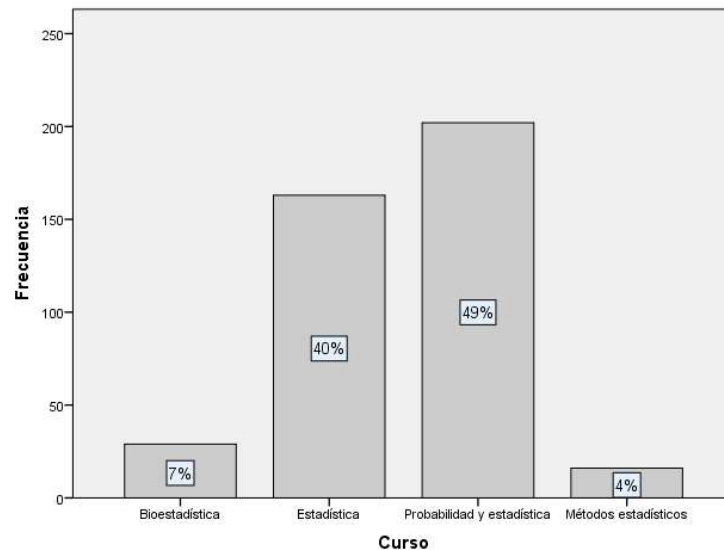


Figura 1 Distribución de los tipos de curso en que se clasificaron los programas.

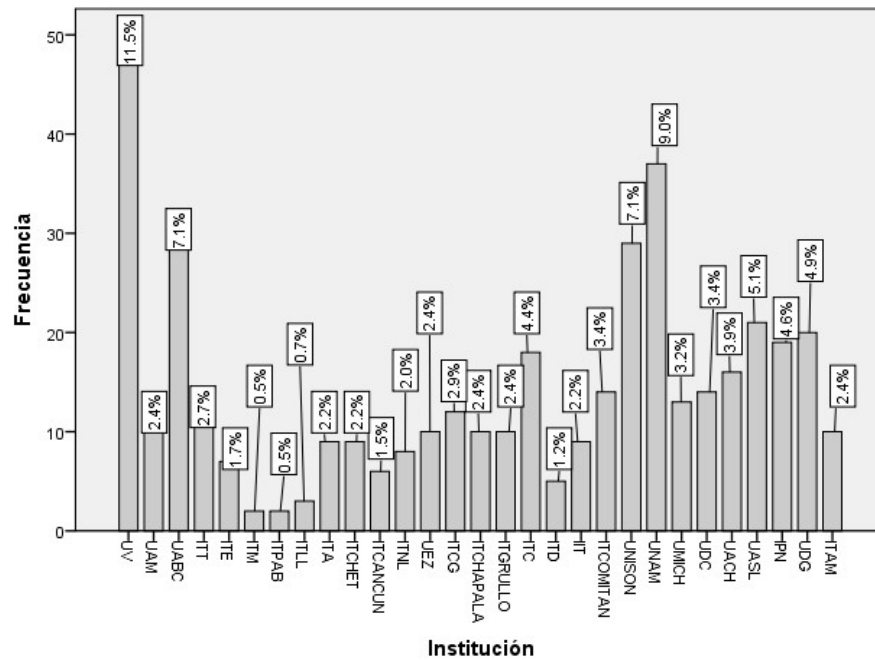


Figura 2 Gráfico de barras de las instituciones y el número de programas que se analizó.

Para evaluar las MAE en los programas de las asignaturas primero se obtuvieron las frecuencias de los valores de cada uno de los ítems para ver qué aspecto es el que no se atiende en cada meta, tomándose como no atendidos aquellos que tenían más del 30% en la categoría de No se considera en ninguna parte del programa. Los resultados se muestran en la Figura 3.

En la Figura 3 no se encuentra la meta 3, porque su categoría de No se considera en ninguna parte del programa, no rebaso más del 30%, por tal razón no se incluyó. Se recuerda al lector que la meta 3 es la que trata sobre las habilidades de calcular las medidas descriptivas y la realización de gráficos y esto la mayoría de los programas resultó que si lo consideran; es decir, esta es una meta totalmente atendida.

En la Figura 3 aparecen sólo los aspectos no considerados dentro de los programas; se puede observar el aspecto no tomado en cuenta y su frecuencia. Se destaca que los aspectos con frecuencias altas de No se consideran en el programa son: El aspecto 6, 7 y 8 de la meta 1; aspecto 1 y 6 de la meta 2; todos los aspectos de la meta 4; aspectos 2, 4 y 5 para la meta 5; 2, 3 y 4 para la meta 6; todos los aspectos de la meta 7; y finalmente el aspecto 4 de la meta 8; lo que indica que se está omitiendo de los programas fomentar el pensamiento estadístico en los estudiantes. Se enfatiza que el pensamiento estadístico planteado por Wild y Pfannkuch (1999) está dividido en dimensiones: la dimensión 1 consiste en el actuar y pensar del investigador, se está omitiendo de los programas dejando fuera aspectos importantes y básicos de la investigación como lo son: la formulación de preguntas y la discusión de los resultados, lo que indica que sólo se está enseñando al estudiante a operacionalizar los datos y por consiguiente la operacionalización carece de

sentido para la investigación, porque el estudiante no sabe para qué está haciendo el análisis de los datos. En la dimensión 2 se hace mención a que el pensamiento estadístico se relaciona con la toma de decisiones bajo incertidumbre y la causa de esta incertidumbre es la variación, si se omite el entendimiento de estos conceptos desde el programa de estudio del curso, el estudiante no logrará tomar decisiones bajo procesos de incertidumbre. Con esto se determina que sólo se están enfocando los cursos al cálculo de probabilidades, sin que los estudiantes comprendan la importancia de los procesos estocásticos y la utilidad que tienen en la vida cotidiana. Asimismo, en esta misma dimensión se hace énfasis de que la materia prima con la cual trabaja el pensamiento estadístico es la información contenida en los datos, por lo que es necesario producir implicaciones, indicios y conjeturas sobre la situación de interés, por esta razón es necesaria la comunicación efectiva de los resultados, para poder establecer conexiones entre el conocimiento existente y los resultados del análisis, de esta manera se obtendrán conclusiones significantes, pero estos aspectos contenidos en la meta 7 se están dejando totalmente fuera de los programas en los cursos de estadística. Respecto a la dimensión 3 del pensamiento estadístico, se plantea un ciclo de cuestionamiento que indica que el investigador permanece siempre en un estado de interrogación y a medida que encuentra respuesta a unas interrogantes del proceso, le surgen nuevas preguntas; por lo que es necesario la interpretación de resultados con posición crítica, aspectos que se están descartando en los programas de asignatura en la meta 6. Finalmente, la dimensión 4 nos habla de disposición y actitud, que se asocia a las metas como el aprecio por la estadística (meta 8) y que tampoco se está tomando en cuenta dentro de los programas; la consideración de esta dimensión da paso a las otras formas de pensamiento estadístico, por esa razón si se excluye de los programas esto dará pauta a que desde un inicio no se logró el objetivo en los cursos que es desarrollar en el estudiante el pensamiento estadístico, que finalmente consiste en formar ciudadanos estadísticamente alfabetizados capaces de realizar investigaciones con ayuda de la estadística.



Figura 3. Aspectos de las MAE que no son considerados en los programas.

Se realizó también un índice de consideración global para poder categorizar el total de los programas en tres categorías de consideración; los resultados arrojan que 16.3% de los programas prácticamente no consideran las MAE, 65.9% las consideran levemente y, únicamente, el 17.8% las consideran suficientemente. Esto da un panorama bastante desolado de la innovación en los programas de las materias de estadística.

Para identificar qué instituciones son las que indican niveles bajos de consideración de las metas, se realizó un ACM en las 7 metas mencionadas anteriormente que son las que presentan porcentajes más altos de desatención.

Se destaca que las variables para realizar el ACM son Área académica, Curso, Institución y Meta. La Figura 4, correspondiente a la meta 1, muestra que existe asociación (círculo azul) entre las siguientes categorías: meta No la contiene, área Ciencias de la salud, curso Bioestadística e institución: UEZ y UNISON. Otra relación

que se puede observar en el análisis (círculo amarillo) es en las categorías: meta Levemente, áreas Ciencias sociales, Agropecuarias, instituciones: ITAM, UNAM, UACH, UDC. Una tercera asociación (círculo azul) se da entre las categorías: meta Suficientemente, áreas Económicas y Exactas, curso Probabilidad y Estadística, instituciones: tecnológicos, de Aguascalientes, Baja California, Chiapas, Jalisco, Nuevo León y Quinta Roo.

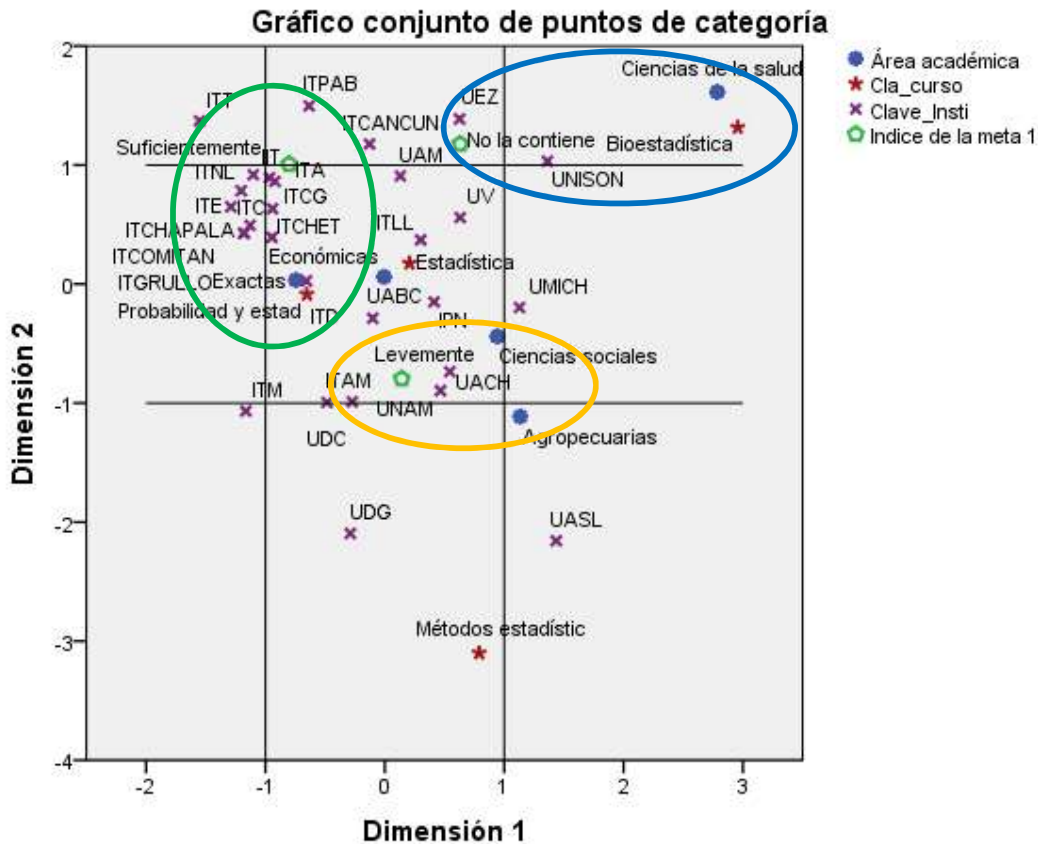


Figura 4. Análisis de correspondencia múltiple entre el Área, Curso, Institución y Meta 1.

El significado de las asociaciones marcadas en los círculos de colores de la Figura 4 significa que se identifican los siguientes grupos:

Un grupo de programas del área Ciencias de la Salud que en su mayoría tienen cursos de Bioestadística, que no toman en cuenta la meta que especifica la lógica detrás de la estadística, como son los conceptos de variación, muestreo, inferencia, experimentación, dentro de estas universidades encontramos a la UNISON y la UEZ.

Otro grupo que se vislumbra es el de las áreas de Ciencias Sociales y el área de Agropecuarias que consideran levemente la meta que especifica la lógica detrás de la estadística, estas universidades son el ITAM, la UNAM y la UACH.

Otro grupo identificado y el más numeroso, es el que se forma de las áreas de Ciencias Económicas y área de Ciencias Exactas con cursos como probabilidad y estadística que consideren suficientemente la meta 1 en sus programas; es decir, los programas toman en cuenta suficientemente la enseñanza de conceptos importantes como: el entendimiento de la variación, la importancia de la inferencia, los estudios de muestreo y experimentación, estas instituciones son los Tecnológicos de Nuevo León, de Chapala, de Chetumal, de Aguascalientes, de Tijuana, entre otros.

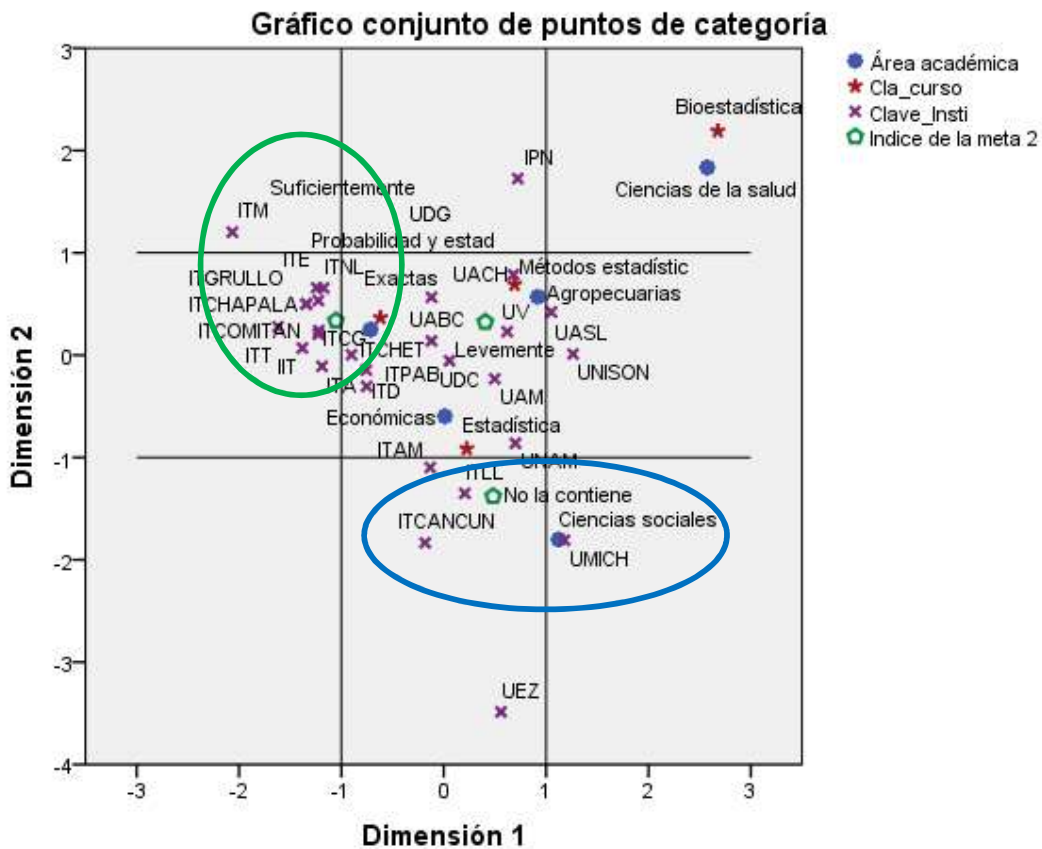


Figura 5. Análisis de correspondencia múltiple entre el Área, Curso, Institución y Meta 2.

Para la segunda meta de aprendizaje donde el análisis de correspondencia múltiple también resultó significativo, se observan dos grupos, uno son del área de ciencias sociales con el curso de estadística, donde no se toma en cuenta dentro del programa la meta que especifica que el alumno debe poder plantear un problema de investigación, así como sus preguntas, objetivos, recolectar sus datos y elaborar sus interpretaciones a la luz de las preguntas de investigación, dentro de este grupo se encuentran universidades como la UMICH, el Instituto Tecnológico de Cancún, el Instituto de Llano y la UNAM (Figura 5). El otro grupo que se identifica es en el área de Ciencias Exactas en los cursos

de probabilidad y estadística, se considera suficientemente esta meta, las principales instituciones que se aglomeran en este grupo son los tecnológicos.

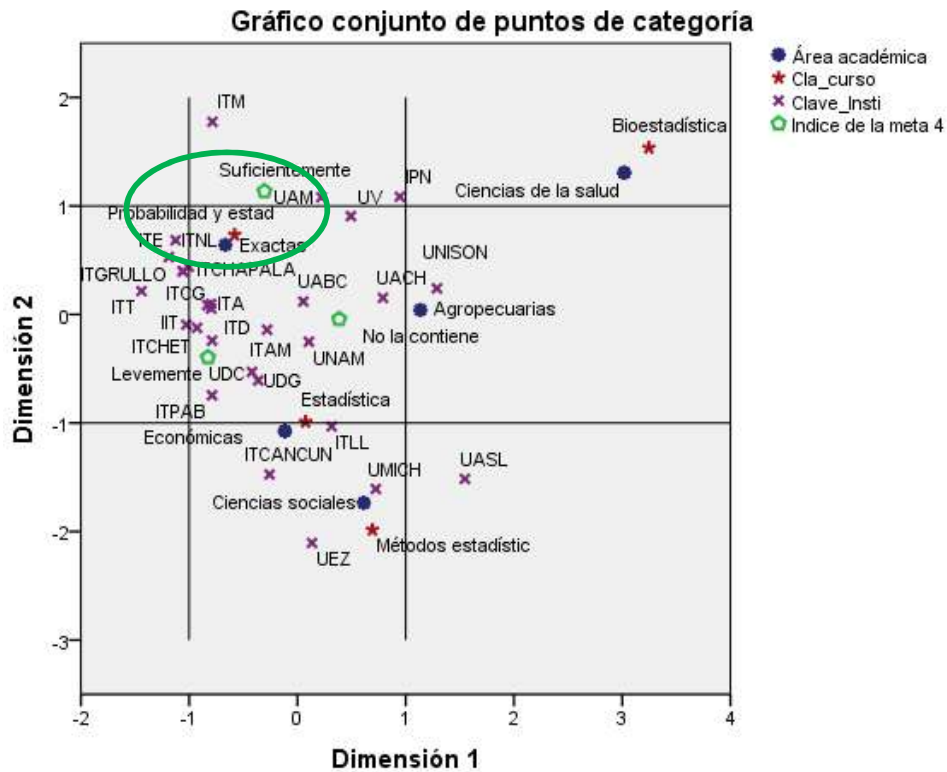


Figura 6. Análisis de correspondencia múltiple entre el Área, Curso, Institución y Meta 4.

Para la meta 4, sólo hubo una asociación (ver Figura 6), entre las categorías: meta Suficientemente, área Exactas, curso probabilidad y estadística, instituciones UAM y UV (círculo verde). Lo que significa que para la cuarta meta se observa solamente un grupo bien definido que es para el área de Ciencias exactas en su curso de probabilidad y estadística, en los cuales se toma lo suficientemente en cuenta la meta 4 que indica que los estudiantes entiendan las relaciones matemáticas, desarrollando un entendimiento intuitivo o formal de las principales ideas que fundamentan los resúmenes, procedimientos gráficos o conceptos estadísticos. En este grupo se ubican universidades como la UAM, la UV y el tecnológico de Llano.

En cuanto a la meta 5, presenta el mismo comportamiento que la 4 sólo se encontró una asociación en la categoría de meta Suficientemente, áreas Económicas, curso Estadística, instituciones ITCANCUN, ITLL, ITPAB (círculo verde, Figura 7). Esto significa que en la quinta meta se identifica un grupo que son de las áreas económicas en los cursos de estadística donde la meta 5 se considera lo suficientemente dentro del programa, esto significa que se pone énfasis en que entiendan el azar y la probabilidad, para tener una comprensión informal de la probabilidad y un razonamiento de la

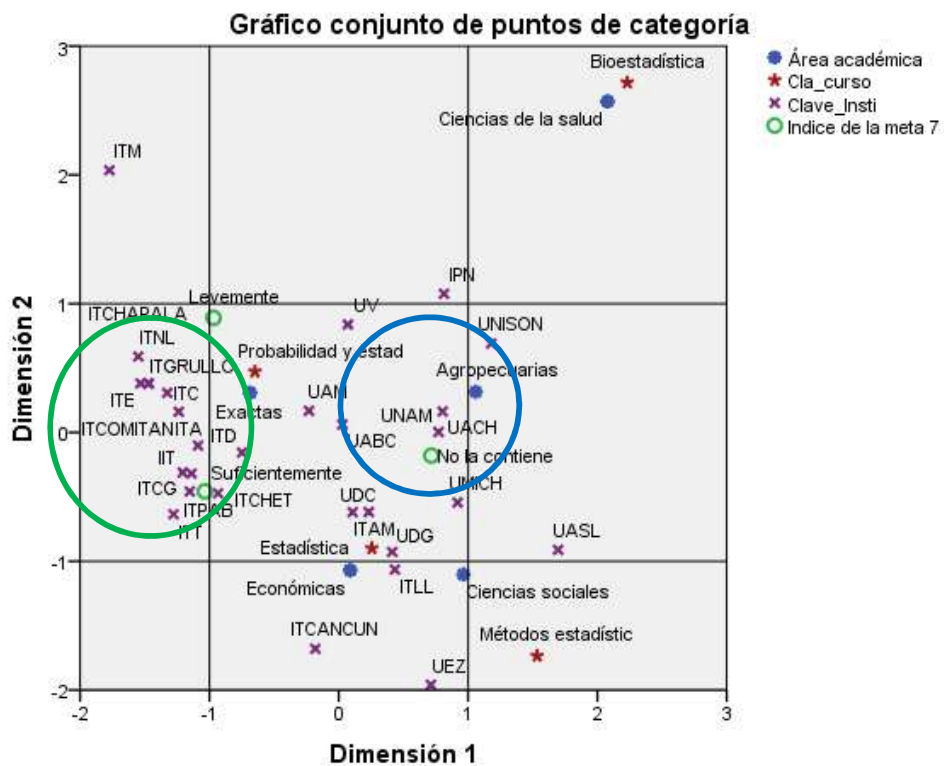


Figura 8. Análisis de correspondencia múltiple entre el Área, Curso, Institución y Meta 7.

Las metas 3, 6 y 8 no tuvieron ninguna asociación significativa por lo que no se incluyen en los análisis.

Conclusiones

La filosofía educativa y la sociología de la educación nos dotan de elementos para que se pueda mejorar el aprendizaje en cualquier sistema educativo, sin embargo en la educación superior la enseñanza tradicional ha caído en un crisis profunda sin importar la disciplina, se habla de reformas y cambios que se establecen mediante teorías que muy lejanamente se llevan a cabo dentro de las aulas, afectando no sólo al estudiante sino a la sociedad misma que recibe dentro de su campo laboral a profesionistas poco preparados y con grandes carencias de conocimiento. Este esquema no es ajeno a la educación estadística, que atraviesa por el mismo panorama que la educación superior en general, pero con más desventajas que otras disciplinas por ser una ciencia rechazada por la mayoría de los estudiantes y el motor fundamental para desarrollar la investigación en cualquier país.

El problema es que la mayoría de los actores involucrados en el proceso de enseñanza aprendizaje de la estadística no se percatan de eso, no reflexionan sobre la gran necesidad de tener un buen aprendizaje o brindar una buena enseñanza, hasta que se ven envueltos en situaciones que los harán reflexionar y que quizá será tarde para remediarlo.

Gran parte de este trabajo se desarrolló con el fin de crear conciencia en que para mejorar la educación estadística en el país, de que es necesario trabajar en equipo -

directivos, maestros, alumnos, profesionales de la estadística y servidores públicos de la educación- enfocándose en una misma dirección para así lograr un cambio y que exista un reconocimiento de la disciplina, esperemos que esta reflexión cumpla su cometido.

El análisis de los programas de asignatura evaluados en este trabajo pone de manifiesto que existe una problemática que debe tratar de erradicarse y encaminar poco a poco a la educación estadística a dar ese impulso que logre motivar tanto a profesores como estudiantes a salir del rezago en el que se encuentra debido a la actitud que toman hacia ella. Coincidimos con Batanero (2000) en que debe llevarse a cabo el análisis del currículo para poder tomar acciones de mejora, si desde las academias de profesores se reestructuran los programas de asignatura de estadística y se definen involucrando y tomando en cuenta las Metas de Aprendizaje de la Estadística para que se logre desarrollar el pensamiento estadístico y a su vez se logren las competencias necesarias que permitan el uso de la metodología estadística, se generaría un gran cambio, porque hay profesores que basan sus cátedras en lo que marca el 100% del programa y con una institución que comience el cambio se hará una bola de nieve de desencadenará en un avance significativo en la educación estadística.

Los resultados reflejan que la mayoría de los programas presentan el siguiente comportamiento. En programas de Ciencias de la salud no se contemplan las Metas de Aprendizaje de la Estadística. En áreas de Ciencias sociales y Agropecuarias se consideran levemente y en áreas de Ciencias Exactas y Administrativas se consideran suficientemente; esto se determinó principalmente con el análisis de correspondencia múltiple. Con el análisis de los datos se pudo observar que el comportamiento descrito era lo que se esperaba: en áreas donde la estadística no es muy usada las metas de aprendizaje se contemplan en menor proporción, incluso aunque este reflejado en el programa de la asignatura hay ocasiones que se omiten. Se coincide también con lo expuesto por Gal y Garfield (1997) que especifican que las metas se dividen en dos grupos las de “hacer” y las de “pensar” y la mayoría de los cursos concentran sus contenidos en la categoría de sólo “hacer”, dejando de lado el valioso pensamiento estadístico que es necesario desarrollar en el estudiante, tal como lo presenta Wild y Pfannkuch (1999).

Se resalta que en el sistema de Institutos Tecnológicos del país el programa de la asignatura de estadística ha sido diseñado e innovado con el enfoque de competencias, por lo que durante el estudio se observó que estos programas son los más completos y que en la mayoría de sus contenidos contemplaban las ocho MAE. Sería muy pertinente replicar la experiencia de estas IES que han estandarizado sus programas. También se requiere organizar foros para concientizar a los profesores y para que la problemática identificada pueda revertirse.

Una vez evaluados los cursos en cuanto a las metas, se deben hacer recomendaciones a los profesores, para que utilicen los enfoques innovadores planteados por Garfield y Gal (1999), y así impulsar el cambio en la educación estadística del país, a fin de que el profesorado sea capaz de generar un conocimiento significativo basado en el pensamiento estadístico y la alfabetización estadísticas, lo que finalmente redundará en

la mejor preparación de los futuros profesionales. También se recomienda a las instituciones que ofrezcan a los profesores cursos de capacitación en metodología estadística, lo que puede ser un factor que impacte en el aprendizaje de los estudiantes. Los retos que deberán enfrentar los docentes de los cursos de estadística de las instituciones será modificar los enfoques y contenidos para garantizar una preparación que les permita lograr el aprendizaje efectivo en los estudiantes.

Referencias

- Batanero, C. (2000). ¿Hacia dónde va la educación estadística? *Blaix*, 15, 2-13.
- Batanero, C. (2001). *Didáctica de la estadística*. Granada: Universidad de Granada.
- Batanero, C., & Godino, J. (2005). Perspectivas de la educación estadística como área de investigación. *Líneas de investigación en Didáctica de las Matemáticas*, 203-226.
- Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza aprendizaje de la estadística (Tesis Doctoral)*. Universidad Politécnica de Cataluña. Barcelona, España.
- Behar, R., Grima, P., Ojeda, M.M. & Cruz, C. (2013). Educación Estadística en cursos introductorios a nivel universitario: Algunas reflexiones. En A. Salcedo (Ed.), *Educación Estadística en América Latina: Tendencias y Perspectivas* (343-360). Venezuela: Programa de Cooperación Interfacultades. Universidad Central de Venezuela.
- Ben-Zvi, D. & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. B. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 3-16). Netherlands: Springer.
- Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, 1(2), 38-41.
- Cobb, G. (1992). Teaching Statistics. In L. Steen (Ed.). *Heeding the Call for Change* (3-34). Washington: Mathematical Association of American.
- Cravero, M., Redondo, Y. & Santellán, S. (2011). Competencias en Educación Estadística: de una Alfabetización Estadística hacia una Alfabetización Científica. XIII

Conferencia Interamericana de Educación Estadística. 26 al 30 de junio. Recife, Brasil.

Gal, I. & Garfield, J. (1997). Curricular Goals and Assessment Challenges. In I. Gal and J. Garfield (Eds.). *The Assessment Challenge in Statistics Education* (1-14). The Netherlands: IOS Perss, ISI, Voorburg.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63, 25-34.

Garfield, J. & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67(1), 1-12.

Hassad, R. A. (2003). Teaching Introductory Statistics in the Social & Behavioral Sciences: Approach & Rationale. ERIC: Online Submission, 1783-1788.

Jiménez, J. V., & Inzuna, S. (2011). Razonamiento y pensamiento estadístico en estudiantes universitarios. In XIII Conferencia interamericana de Educación Matemática. Recife, Brasil.

MacKay, R. J., & Oldford, W. (1994). Stat 231 course notes fall 1994. Waterloo: University of Waterloo.

Mallows, C. (1998). The zeroth problem (1997 Fisher Memorial Lecture). *The American Statistician*, 52, 1-9.

Méndez, M. A. & Ojeda, M. M. (2006). Evaluación de profesores de nivel superior usando metas de aprendizaje en estadística. *Revista Heurística*, 13, 53-60.

Moore, D. S. (1990). Uncertainty. In L. Steen, (Ed.) *On the Shoulders of giants: New Approaches to numeracy* (95-137). Washington, DC: National Academy Press.

Moore, D. S. (1992). Teaching Statistics as a Respectable Subject. In F. Gordon and S. Gordon (Eds.). *Statistics for the Twenty-First Century* (14-25). Washington: Mathematical Association of America.

Moore, D. S. (1993). A Generation of Statistics Education: An Interview with Frederick Mosteller. *Journal of Statistics Education*, Electronic journal. Recuperado de www.stat.ncsu.edu

Ojeda, M. M. (2011). Aprender Estadística con Proyectos: Memoria de una experiencia replicable. Xalapa: Universidad Veracruzana.

- Ojeda, M. M. & Sahai, H. (1994). A general proposal for teaching undergraduate statistics service courses. In Proceedings of the Statistical Education Section of the American Statistical Association, 311-316.
- Ojeda, M. M. & Sahai, H. (2003). A multidisciplinary graduate level Project-based program in applied statistics. *International Journal of Mathematical Education in Science and Technology* (34)1, 57-63.
- Palomo, R., Ruiz, J., & Sánchez, J. (2006). Las TIC como agentes de innovación educativa. Sevilla: Junta de Andalucía, Consejería de Educación, Dirección General de Innovación Educativa y Formación del Profesorado.
- Snee R. (1993). What's missing in Statistical education? *The American Statistician*, 47, 149-154.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review/Revue Internationale de Statistique*, 223-248.

CAPÍTULO 5

INNOVACIÓN Y CAPITAL SOCIAL. ESTUDIO DE OPINIÓN UNIVERSITARIA EN EL SURESTE DE MÉXICO

José Félix García Rodríguez¹, Reyna Alejandra López Arias¹, Germán Martínez Prats¹,
Ignacio Caamal Cauich², Aída Armenta Ramírez¹

¹ Universidad Juárez Autónoma de Tabasco (UJAT). División Académica de Ciencias
Económico Administrativas (DACEA).

² Universidad Autónoma de Chapingo (UACH). Escuela de Economía

RESUMEN

Los estados del sureste de México se caracterizan por su rezago económico y social. Entre ellos, Tabasco sobresale por tener la más alta tasa de desempleo nacional. La investigación tuvo como objetivo conocer el grado de apropiación de los procesos de innovación tecnológica y capital social por los estudiantes de una universidad tecnológica del sureste de México. Se desarrolló una investigación cuantitativa, descriptiva y transversal, habiéndose aplicado un cuestionario estructurado a una muestra aleatoria de 28 alumnos. Se identificó un escaso nivel de conocimiento sobre innovación tecnológica, así como un bajo nivel de cohesión social, cooperación y ayuda mutua.

PALABRAS CLAVE: Innovación, Capital social, Empleo, Redes sociales

ABSTRACT

The southeastern states of Mexico are characterized by their economic and social lag. Among them, Tabasco stands out for having the highest national unemployment rate. The objective of the research was to know the degree of appropriation of the processes of technological innovation and social capital by the students of a technological university in southeastern Mexico. A quantitative, descriptive and cross-sectional research was developed, with a structured questionnaire applied to a random sample of 28 students. A low level of knowledge about technological innovation was identified, as well as a low level of social cohesion, cooperation and mutual help.

KEYWORDS: Innovation, Social capital, Employment, Social networks

1. INTRODUCCIÓN

La región sur de México, integrada por los Estados de Veracruz, Tabasco, Chiapas Campeche, Yucatán y Quintana Roo se caracteriza por su elevado rezago socioeconómico en comparación con la región norte y centro del país, prevaleciendo entre otros problemas el estancamiento económico, desempleo, desigualdad, pobreza e inseguridad. Ello no obstante su riqueza en materia turística y petrolera, así como su potencial en agricultura, ganadería, silvicultura y pesca. Las razones del rezago son diversas, pero entre todas sobresalen la crisis petrolera reciente y la falta de crecimiento económico. De la misma manera, la ausencia de procesos de innovación tecnológica determina que las materias primas y los productos regionales cuenten con un escaso valor agregado, y por lo tanto la industria existente sea incipiente e incapaz de generar fuentes de empleo bien remunerado. Todo lo anterior se agrava con la falta de aprovechamiento del capital social disponible.

Particularmente, el estado de Tabasco se caracteriza en el plano nacional por contar de manera sostenida durante los últimos años con la más alta tasa de desempleo, así como por sus elevados índices de pobreza, desigualdad e inseguridad. Esta investigación tiene como propósito indagar la percepción de los futuros egresados de una universidad tecnológica del sureste de México, respecto a dos aspectos centrales del éxito económico de los países desarrollados y emergentes: la innovación y el aprovechamiento del capital social y las sinergias productivas que éste es capaz de generar.

INNOVACIÓN

El constructo innovación está asociado a la introducción de un nuevo o significativamente mejorado producto, servicio, proceso, método de comercialización o método organizativo en las prácticas internas de una empresa, institución, u organización (OCDE, 2005). También puede considerarse como cualquier cambio o creación que genera y aporta valor. En este sentido, la innovación lleva implícita la incorporación de nuevas tecnologías y conocimientos capaces de transformar en el largo plazo la productividad de una empresa u organización, y la economía y el bienestar de una región o país (Venture Institute, 2013).

La innovación es un proceso que ha estado presente en la historia de la civilización. A raíz del avance en el conocimiento de las leyes de la naturaleza y las propiedades de muchos de sus elementos, se fueron gestando innovaciones en los modos de producción de las primeras sociedades humanas. La constante evolución en las técnicas de producción, herramientas y conocimientos, le fueron permitiendo al hombre alcanzar avances significativos en los sistemas de producción y en los niveles de bienestar. Actualmente la innovación se estudia de manera sistemática y organizada por la mayoría de los países del mundo, sus instituciones públicas y empresas privadas, en virtud de las evidencias relacionadas con los efectos positivos de las actividades innovadoras. Entre tales efectos se encuentran el crecimiento económico, desarrollo social, modernización y generación de riqueza, empleo y bienestar social.

En concordancia con la etimología del término innovación, ésta tiene que incluir un elemento de novedad. El concepto de novedad se configura bajo tres modalidades: nuevo para la empresa, nuevo para el mercado y nuevo para el mundo entero. El criterio mínimo para que se considere la existencia de la innovación en una empresa u organización es que el cambio sea nuevo para la misma, no importando que la innovación haya sido implementada en alguna otra parte del mundo e incluso en otra empresa perteneciente al mismo mercado. De acuerdo a este criterio, la innovación expande sus fronteras a aquellas empresas que no gastan cantidades importantes en Investigación y Desarrollo (I+D), pero que aprovechan los efectos de desbordamiento tecnológico que se producen en los mercados. También la innovación puede ser nueva para el mercado, cuando la empresa es la primera en lanzarla al mismo. El mercado se define simplemente como la empresa y sus competidores, y puede referirse a una región geográfica o a una gama de productos. La innovación también puede presentarse como nueva para el mundo entero cuando la empresa es la primera en lanzarlo en todos los mercados y los sectores de actividad, nacionales e internacionales (OCDE y EUROSTAT, 2005).

Según Porter (2015), las innovaciones tecnológicas constituyen una causa determinante del cambio en la estructura industrial de los países. La innovación del producto o servicio es fundamental para el incremento de su mercado, tanto interno como externo, así como para su diferenciación e impulso del crecimiento de la industria nacional. El estado de la innovación en México aún es débil, ya que de acuerdo al Índice Mundial de Innovación 2016, se ubicó en el lugar número 61 de 128 países evaluados, incluso por debajo de economías latinoamericanas como la de Chile y Costa Rica, que ocupan el lugar 44 y 45 respectivamente (Cornell University, INSEAD y WIPO, 2016).

CAPITAL SOCIAL

No existe una definición precisa de capital social, sin embargo, la literatura especializada establece diversas aproximaciones y enfoques que permiten entender el capital social como una función de la estructura social con que cuenta un individuo para producir un beneficio. Así, Bourdieu (1998) lo conceptualiza como la suma de los recursos actuales o virtuales que acumula un individuo o un grupo en virtud de su posicionamiento en una red social. Para Putman (2002), el capital social se refiere a las características de la organización social como la confianza, normas, y redes que pueden mejorar eficientemente la sociedad por acciones facultades y coordinación entre los individuos. El estudio del capital social es reciente y aún se debate sobre su definición debido a la amplitud de su significado, que varía según el enfoque con que se observe e incluso desde qué disciplina se estudia: sociología, ciencias políticas, economía, teoría organizacional, etc. Dicho esto, una de las definiciones que tratan de conceptualizar el capital social de manera muy general, es que se refiere a la buena voluntad que se genera como producto de las relaciones sociales, y que ayuda a concretar actividades o a generar acciones.

La caracterización y construcción teórica de la noción de capital social tiene una perspectiva teórica predominantemente sociológica (Portes, 1999). En líneas generales, se pueden distinguir dos aproximaciones de estudio. La primera se refiere a la discusión sobre la insuficiencia del modelo de desarrollo imperante para generar bienestar social. La segunda perspectiva de análisis plantea un cuestionamiento más profundo, estableciendo que lo que se encuentra en crisis es la propia naturaleza y, por tanto, los paradigmas subyacentes al concepto de desarrollo y la lógica de mercado estrechamente asociada a él. Desde este punto de vista, la tesis es que ningún nuevo concepto o enfoque que tienda a introducirse como otra dimensión del desarrollo, sin un análisis crítico al modelo de desarrollo vigente, puede colaborar efectivamente al logro de mayor equidad económica y social (Dávila, 2013).

A fines de la década de los setenta, las principales teorizaciones sobre la noción de capital social corresponden a Bourdieu (1980) y Loury (1977, 1981). Desde la sociología de la cultura, el primero sistematiza el concepto desde un abordaje fundamentalmente instrumental, centrándose en los beneficios que reciben los individuos, en función de su participación en grupos y en la construcción deliberada de la sociabilidad con el objetivo de crear ese recurso. A finales de los años ochenta y principios de los noventa, la construcción teórica del concepto adquiere real visibilidad en el debate sobre el desarrollo. Es así que a partir del trabajo desarrollado por Loury, se profundiza en el análisis del rol del capital social en la generación de capital humano. En la década de los noventa, Putnam (2002) es uno de los teóricos que alcanza más notoriedad en este campo de estudio, y es quien desarrolla formalmente el concepto de capital social, asumiendo que éste se halla compuesto fundamentalmente por el grado de confianza existente entre los actores sociales de una sociedad, las normas de comportamiento cívico practicadas y el nivel de asociatividad que caracteriza a esa sociedad, y que estos elementos evidencian la riqueza y la fortaleza del tejido social interno de una sociedad.

Por su parte, Fukuyama (citado en Zumbado, 1998) va más allá y habla del asociacionismo horizontal y vertical e integra varias teorías en sus planteamientos sobre el capital social. Su enfoque destaca lo que él denomina las virtudes sociales de una sociedad o grupo. Su propósito es explicar cómo a partir de ellas se crea prosperidad económica, pues considera que la vida económica de una nación, simultáneamente refleja y modela el entorno. Indica que las instituciones políticas y económicas dependen de una sociedad civil dinámica y activa. Es difícil establecer una definición puntual del constructo capital social, dada su naturaleza multidisciplinaria y compleja, hecho en el que la mayor parte de la comunidad científica que se dedica a su estudio está de acuerdo. De esta manera, tanto Bourdieu (1980) como Putman (2002), llevan a cabo amplias referencias al capital social como un atributo estrechamente ligado a grupos sociales, colectividades y comunidades, en el cual las instituciones juegan un rol fundamental.

Así, para Bourdieu (1980), capital social constituye el conjunto de recursos reales o potenciales ligados a la posesión de una red durable de relaciones más o menos institucionalizadas de reconocimiento común. Como puede apreciarse, Bourdieu pone énfasis en las redes sociales y en la necesidad de impulsar institucionalmente las

relaciones grupales dentro de una comunidad. En este sentido, el capital social se integra por el conjunto de recursos socio–estructurales, mismos que representan un activo de capital y facilitan las acciones de los individuos que conforman dicha estructura. En este contexto, Putnam (2002) menciona que el capital social se constituye con los elementos de las organizaciones sociales, como son redes, normas y confianza, mismas que facilitan la acción y la cooperación para beneficio social.

En términos instrumentales, el capital social constituye un recurso valioso para impulsar el desarrollo y el bienestar individual y colectivo, ya que facilita a los individuos la consecución de sus propios intereses, es inherente a la estructura de las relaciones sociales, a partir de la realización de apoyos recíprocos el capital social adopta la forma de obligación, y es a través de ella que el individuo alcanza sus propios objetivos. La presencia de capital social queda sujeta a diversos factores, como son el grado de confiabilidad del entorno social, las necesidades actuales de los individuos, y el grado de cercanía de las redes sociales (Forni, 2004).

2. MATERIAL Y MÉTODOS

Se desarrolló una investigación con enfoque cuantitativo de tipo descriptivo y transversal, con el propósito de conocer el grado de conocimiento prevaleciente entre los estudiantes de una institución educativa tecnológica de nivel superior del estado de Tabasco, acerca de los procesos relacionados con la innovación tecnológica y el capital social, asumiendo que en estas instituciones prevalece un modelo educativo enfocado a la innovación tecnológica, el emprendimiento y el aprovechamiento de las potencialidades locales. El universo poblacional de estudio correspondió a los 2,938 alumnos que cursan las diversas carreras que ofrece la institución; el marco muestral lo constituyeron los 265 alumnos inscritos en la carrera de Técnico Superior Universitario en Química, área de Tecnología Ambiental. Mediante muestreo aleatorio se obtuvo una muestra de 28 estudiantes, a quienes se aplicó un instrumento de captación de información de campo, cuestionario diseñado a partir de la encuesta de capital social (ENCAS, 2011) del Programa de las Naciones Unidas Para el Desarrollo (PNUD), y en función de los objetivos de la investigación.

3. RESULTADOS

Con relación al desarrollo de procesos asociados a la innovación tecnológica en la institución, se detectó una baja participación de los estudiantes en proyectos de innovación, pues solo el 29% de ellos estaba involucrado en alguna investigación de este tipo, en tanto que el 71% respondió que no está relacionado con ningún proyecto de innovación (figura 1).

El hecho de que solo 3 de cada 10 alumnos de esta universidad tecnológica estuvieran participando en algún proyecto de innovación revela un área de oportunidad para la

institución en materia de revisión y actualización de su programa académico en materia de innovación.

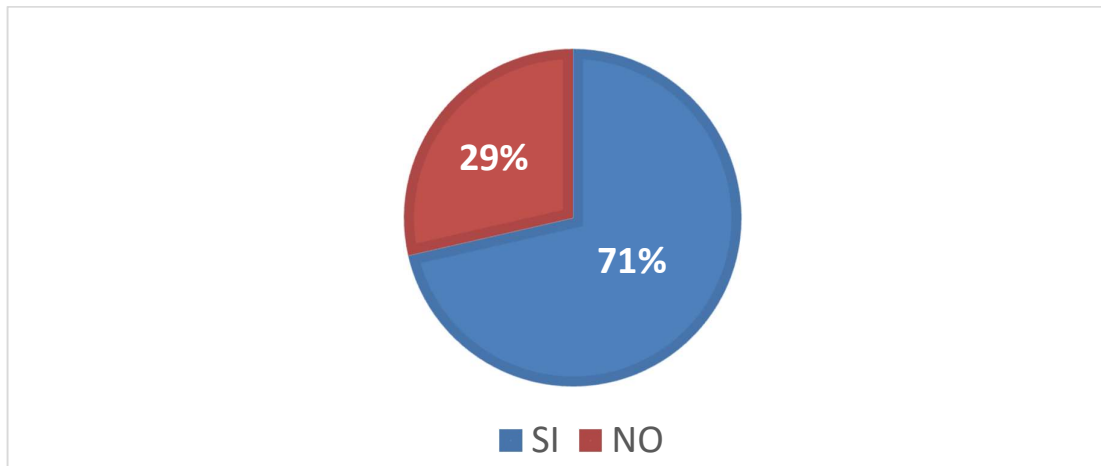


Figura 1. Estás involucrado o lo has estado en algún proyecto de innovación tecnológica
Fuente: Resultados del análisis estadístico

Con relación al fomento de una actitud innovadora, llama la atención que solo el 29% de la población estudiada consideró haber sido motivada para emprender algún proyecto que implicara procesos de innovación, aspecto que el 39% del total valoró como importante para un emprendimiento innovador (figura 2). La baja motivación que reciben los alumnos podría explicar la escasa participación de estudiantes en proyectos de innovación, tal como se señaló anteriormente.

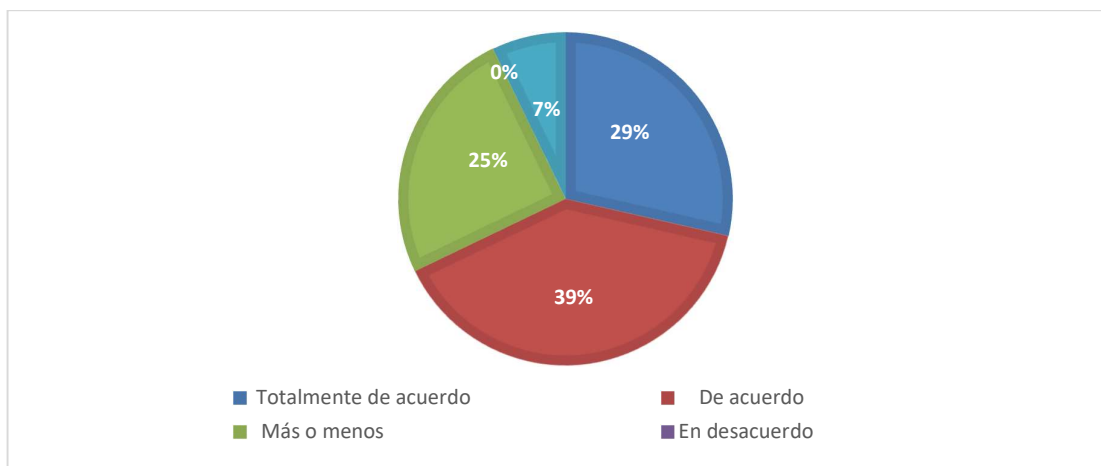


Figura 2. En tu universidad se te motiva a innovar o a emprender algo novedoso
Fuente: Resultados del análisis estadístico

Por otro lado, y partiendo del hecho que el fundamento del capital social es el aprovechamiento de los lazos de colaboración y apoyo mutuo presentes en un grupo social u organización, se encontró que existe una amplia área de oportunidad para la institución

en este sentido, pues únicamente el 21% de los entrevistados reportó tener más de seis compañeros cercanos involucrados en procesos de innovación. Llama la atención que el 39% del total solo contaba con uno o dos compañeros relacionados con esta actividad, y el 18% manifestara no tener ningún conocido vinculado a estos procesos (figura 3).

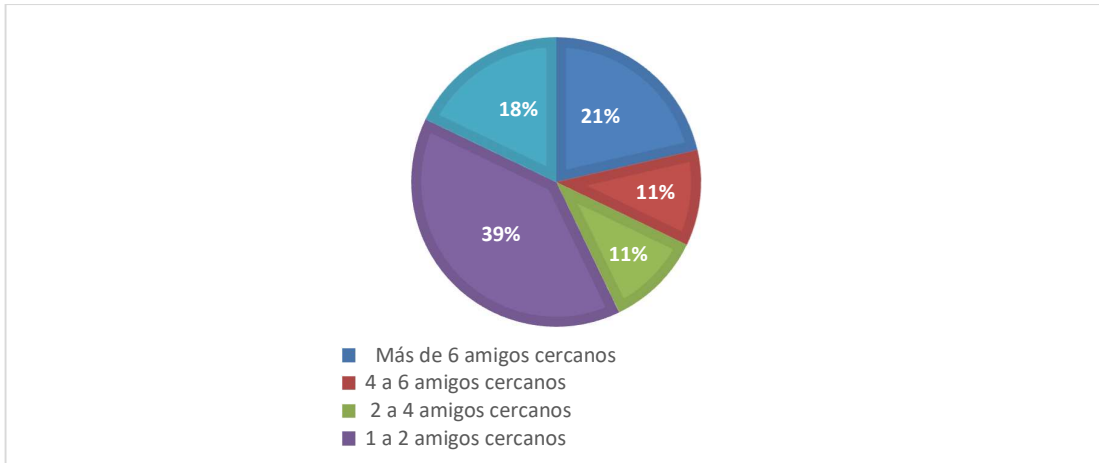


Figura 3. Cuantos amigos tienes que estén relacionados con la innovación tecnológica
Fuente: Resultados del análisis estadístico

La presencia de redes de colaboración y/o grupos de apoyo constituyen parte importante del capital social existente en la organización y facilitan su aprovechamiento con propósitos explícitos. En lo que toca a innovación, el 74% de los entrevistados manifestó conocer la existencia en la institución de un club o grupo social que fomente el emprendimiento y la innovación tecnológica, en tanto que el 19% desconoce su presencia (figura 4).

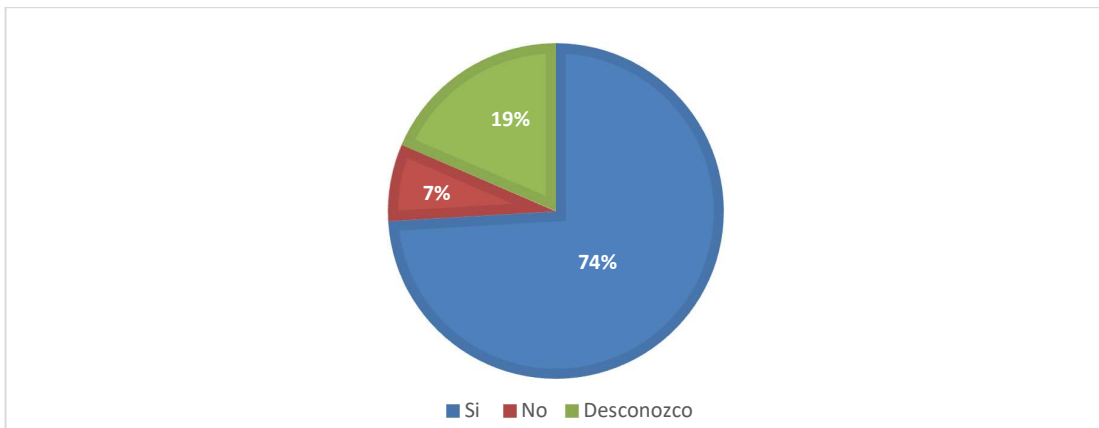


Figura 4. Existe alguna organización que fomente el emprendimiento en tu universidad
Fuente: Resultados del análisis estadístico

La colaboración mutua, la ayuda y apoyo son también sustento importante del capital social. Respecto a los procesos de innovación, el 53% de los entrevistados manifestó haber recibido apoyo de los maestros, el 18% de sus amigos o contactos claves y el 4%

de su familia. Llama la atención que el 25% del total considerara no haber recibido ayuda al respecto, y que nadie manifestara recibir apoyo por parte de una incubadora de empresas o algún programa de emprendimiento, lo que pone en evidencia la necesidad de impulsar la difusión de dichos instrumentos de apoyo a la innovación en la institución (figura 5).

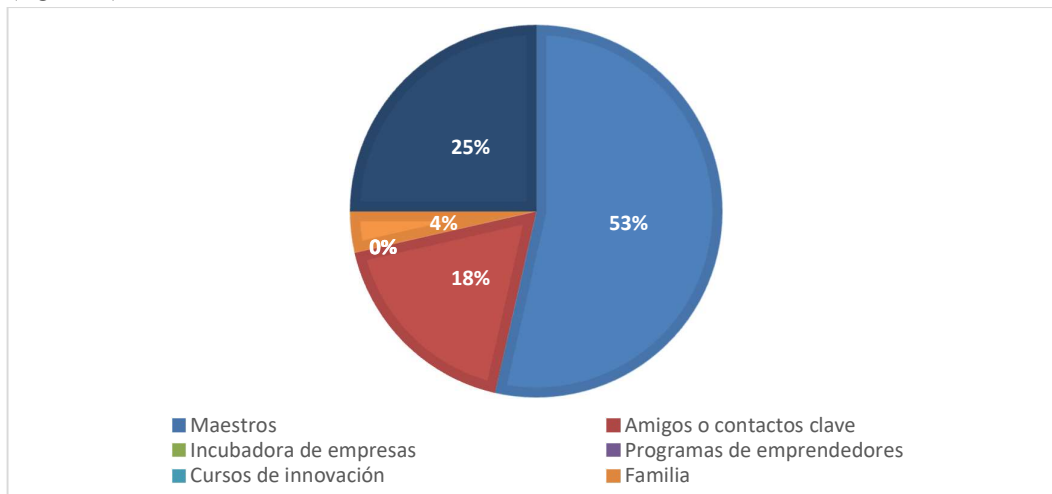


Figura 5. Si ha solicitado ayuda de quien la ha recibido

Fuente: Resultados del análisis estadístico

La difusión de las actividades relacionadas con la innovación, ciencia y tecnología resultan fundamentales para la creación de una actitud innovadora en los jóvenes, no solo en los estudiantes de esta institución sino de todo el Estado. Sin embargo, dicha actividad es muy limitada y necesita ser impulsada por las instancias públicas responsables de la misma. Esto resulta relevante, toda vez que el 96% de los estudiantes de esta universidad tecnológica manifestaron no haber participado en algún evento de divulgación científica organizado por el Consejo de Ciencia y Tecnología de Tabasco (CCYTET), organismo público responsable de la divulgación de la ciencia en el Estado (figura 6). Lo mismo sucede a nivel estatal, pues el 86% del total manifestó no conocer alguna institución u organización que fomente la innovación tecnológica en Tabasco (figura 7).

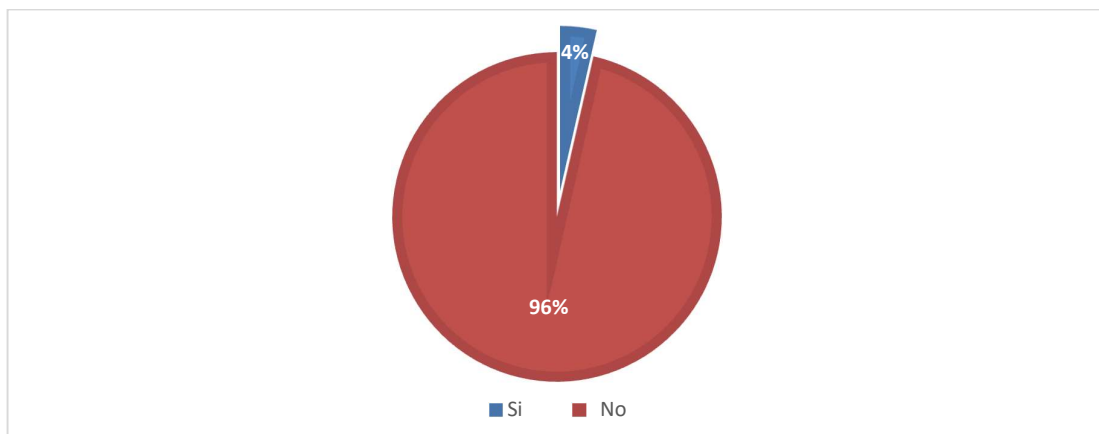


Figura 6. Has participado en algún evento organizado por el CCYTE

Fuente: Resultados del análisis estadístico

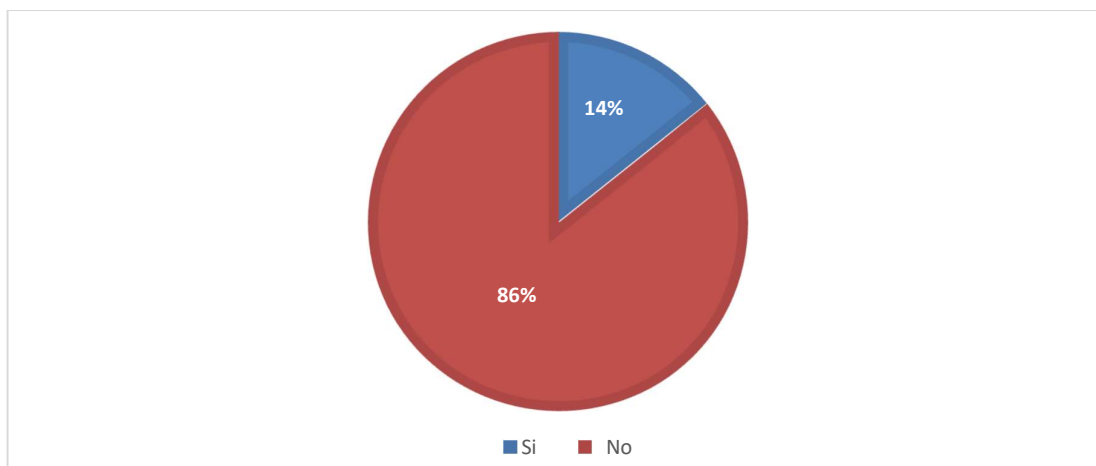


Figura 7. Conoce la existencia en Tabasco de algún organismo que impulse la innovación tecnológica.

Fuente: Resultados del análisis estadístico

4. DISCUSIÓN Y CONCLUSIONES

La innovación y el aprovechamiento del capital social disponible son procesos determinantes de las condiciones de productividad, empleo y bienestar de un grupo social, región o país. A pesar de ello, en nuestros países poco se presta atención a las experiencias exitosas de otras naciones, por lo que se siguen produciendo materias primas con escaso valor agregado y baja productividad, problemas asociados al bajo nivel de innovación tecnológica, lo que determina también bajos niveles de empleos y baja calidad de los mismos. Asimismo, es escaso el aprovechamiento de los vínculos de colaboración y ayuda mutua entre los individuos y grupos sociales, así como la presencia de redes de apoyo con propósitos de colaboración.

En lo que corresponde a innovación y capital social, en la presente investigación enfocada a una muestra de alumnos de una universidad tecnológica del sureste del país, se encontró que existe una baja participación estudiantil en procesos de innovación, ya que solo tres de cada diez alumnos estaban involucrados en proyectos de innovación, y que la institución no les motivaba ni proporcionaba los conocimientos básicos para una actitud innovadora y emprendedora; de la misma manera, la difusión y divulgación de la ciencia, tecnología e innovación es muy limitada y de bajo impacto, pues los alumnos manifestaron una mínima participación en eventos de este tipo, así como amplio desconocimiento de la existencia de instituciones y organizaciones vinculadas a estas actividades.

De esta manera, la Institución educativa requiere impulsar los procesos tendientes a la formación de una cultura innovadora y emprendedora, así como aprovechar el capital

social presente. En síntesis, para la transición a una sociedad de conocimiento, en México es fundamental impulsar en las universidades públicas y privadas la innovación tecnológica, así como hacer de la innovación, la educación y el capital social los pilares del desarrollo y el progreso económico.

REFERENCIAS

1. Bourdieu, P. 1980, “Los tres estados del capital cultural”, *Sociológica*, Vol. 2, pp.5.
2. Bourdieu, P. 1998, *Capital cultural, escuela y espacio social*, Siglo XXI, México.
3. Cornell University, INSEAD, and WIPO, 2016, *The Global Innovation Index 2016: Winning with Global Innovation*, Ithaca, Fontainebleau, and Geneva, Editor,
4. Dávila, O. y Honores, C. 2013, “Capital Social Juvenil y Evaluación Programática hacia Jóvenes”, *Última década*, versión On-line, vol.11 n.18, pp. 175-198. <http://dx.doi.org/10.4067/S0718-22362003000100009> consultado el 17 de agosto de 2017.
5. Forni, P. 2007, “¿Cómo generar capital social en contextos de exclusión? Experiencias de organizaciones comunitarias y redes sociales”, *Temas Sociológicos*, N° 12, pp. 169 – 19.
6. Loury, G. 1977, “A Dynamic theory of racial income differences”, en P. Wallace y A. La Mond (comps.), *Women, Minorities, and Employments Discrimination*, Heath, Lexington, Massachusetts.
7. OCDE-EUROSTAT, 2005, *Manual de Oslo* (3^{ra} Ed.). Traducción española (2006), Editor, España.
8. Portes, A. 1999, “Capital social: sus orígenes y aplicaciones en la sociología aplicada”, en J. Carpio e I. Novacovsky (compiladores), *De igual a igual. El desafío del Estado ante los nuevos problemas sociales*, FCE, Buenos Aires.
9. Putman, R. 2002, *Democracies in Flux. The Evolution of Social Capital in Contemporary Society*, Oxford University Press, New York.
Recuperado el 10 de enero de 2017, de <http://index.institute.vc/reporteINI.pdf>
10. Venture Institute, 2013, *Índice nacional de innovación*, Editor.
11. Zumbado, C. 1998, *Desarrollo y capital social: redescubriendo las riquezas de las naciones*, Instituto Internacional de Gobernabilidad de Cataluña, Barcelona.

CAPÍTULO 6

RELACIÓN ENTRE EL DESEMPEÑO DOCENTE Y LA SATISFACCIÓN LABORAL DE PROFESORES DE EDUCACIÓN BÁSICA DE LA ZONA ESCOLAR 60, REGIÓN ACAPULCO-COYUCA

Tonantzin Martínez Bello¹, Elizabeth Ávila Cortes¹, Virginia García Cervantes¹
William de Jesús Merlán Avilés².

¹Coordinación General de Investigación Educativa

²Facultad de matemáticas de la Universidad Autónoma de Guerrero

RESUMEN

En el área educativa, la satisfacción laboral de los docentes ha sido objeto de múltiples estudios, con frecuencia se busca conocer la relación tan estrecha que existe entre la satisfacción laboral y la eficacia en las escuelas, así como analizar el efecto de la satisfacción en el desempeño docente y en menor proporción algunos autores buscan indagar y comprender como se desarrolla la satisfacción laboral en los docentes y de que depende. El estudio se realizó en la zona escolar número 060 del nivel de educación básica específicamente del sector primaria, cuenta con 15 escuelas 11 del turno matutino y 4 del turno vespertino, 11 se encuentran ubicada en la zona suburbana y 4 en la zona urbana y cuenta con 141 profesores, por lo que la población objetivo estará integrada por 141 profesores. Dado que la población es pequeña, se realizó la investigación sobre todos los profesores presentes al momento de realizar la entrevista, excluyendo a aquellos que por razones personales no deseen participar en el estudio. Se utilizó un cuestionario estructurado para evaluar la satisfacción laboral y el desempeño docente, dentro de cada cuestionario se consideraron dimensiones compuestas por varios ítems, para hacer la medición se consideró una escala tipo Likert, pero, como se ha demostrado en la literatura, existe sesgo en las respuestas, por lo que se decidió utilizar un procedimiento alternativo, basado en la media de respuesta y su desviación típica, realizando una clasificación a partir de las nuevas categorías e interpretando los resultados de acuerdo con los datos obtenidos.

Palabras clave: desempeño docente, satisfacción laboral, escala tipo Likert, validación de cuestionarios.

ABSTRACT

In the educational area, teacher job satisfaction has been the subject of multiple studies, it is often sought to know the close relationship between job satisfaction and effectiveness in schools, as well as analyze the effect of satisfaction in the teaching performance and to a lesser extent some authors seek to investigate and understand how job satisfaction

develops in teachers and what it depends on. The study was carried out in the 060 school zone of the basic education level specifically of the primary sector, it has 15 schools 11 in the morning shift and 4 in the afternoon shift, 11 are located in the suburban area and 4 in the urban area and has with 141 teachers, so the target population will consist of 141 teachers. Since the population is small, research was conducted on all teachers present at the time of the interview, excluding those who for personal reasons do not wish to participate in the study. A structured questionnaire was used to evaluate job satisfaction and teacher performance, within each questionnaire dimensions were considered composed of several items, for the measurement a Likert scale was considered, but, as has been demonstrated in the literature, there is bias in the answers, so it was decided to use an alternative procedure, based on the average response and its standard deviation, making a classification based on the new categories and interpreting the results according to the data obtained.

Keywords: teacher performance, job satisfaction, Likert scale, questionnaire validation.

1. INTRODUCCIÓN

1.1 Antecedentes.

Actualmente la política educativa en México, demanda una educación equitativa, inclusiva y de excelencia; por lo cual son trascendentales las decisiones y las prácticas educativas que ocurren al interior de la escuela. En la Ley General de Educación, Artículo 8, Cap. III refiere que el Estado está obligado a prestar servicios educativos con equidad y excelencia; así también lo marca esta ley en su Título Segundo, Cap. I, Art. 11 el cual reitera el compromiso del Estado a través de la Nueva Escuela Mexicana, la cual buscara la equidad, la excelencia y la mejora continua en la educación, para lo cual colocará al centro de la acción pública el máximo logro de aprendizaje de las niñas, niños, adolescentes y jóvenes¹.

En el área educativa, la satisfacción laboral de los docentes, ha sido objeto de múltiples estudios, con frecuencia se busca conocer la relación tan estrecha que existe entre la satisfacción laboral y la eficacia en las escuelas, así como analizar el efecto de la satisfacción en el desempeño docente y en menor proporción algunos autores buscan indagar y comprender como se desarrolla la satisfacción laboral en los docentes y de que depende.

En este sentido, (Freitez, 1999) afirma que un clima organizacional adecuado en una institución educativa motiva a los docentes para lograr los objetivos deseados. Además, una moral elevada facilita la satisfacción en el trabajo y contribuye con la obtención de metas propuestas y a mejorar la eficiencia en la labor encomendada, generando

¹ Las maestras y los maestros son agentes fundamentales del proceso educativo y, por tanto, se reconoce su contribución a la transformación social. Tendrán derecho de acceder a un sistema integral de formación, de capacitación y de actualización retroalimentado por evaluaciones diagnósticas, para cumplir los objetivos y propósitos del Sistema Educativo Nacional. Art. 3 de la Constitución Política Mexicana. Adicionada mediante Decreto publicado en el Diario Oficial de la Federación el 15 de mayo de 2019

disposición y colaboración responsables en las diversas funciones. (Sáenz & Lorenzo, 2016) entienden la satisfacción docente como:

Una experiencia gozosa de crecimiento psicológico, producida por el logro de niveles cada vez más altos en: a) la calidad de su trabajo, b) de reconocimiento por lo que hace, c) de responsabilidad, d) de creación del saber, e) de libertad científica y f) de disfrute en el trabajo mismo. (p. 22) (Martinez-Garrido, 2017)

Con el fin de obtener una visión de las variadas acepciones sobre el tema; recuperamos la información de algunos autores en la siguiente tabla. (Tabla 1)

Tabla 1. Definiciones de satisfacción laboral frente a satisfacción profesional

Autor	Definición	Fuente
Robbins y Coulter (1996: 181).	Definen la satisfacción laboral “como la actitud general de un individuo hacia su trabajo. Una persona con un alto nivel de satisfacción en el puesto tiene actitudes positivas hacia el mismo; una persona que está insatisfecha con su puesto tiene actitudes negativas hacia él”.	Administración. Prentice-Hall Hispanoamericana SA, 1987.
Kreitner, y Kinicki, (1997).	La satisfacción laboral “es una respuesta afectiva o emocional hacia varias facetas del trabajo del individuo”.	<i>Comportamiento de las organizaciones</i> . Madrid: McGraw-Hill.
Sáenz y otros, (1993).	Definen la satisfacción laboral “como una experiencia gozosa de crecimiento psicológico, producida por el logro de niveles cada vez más altos en la calidad de su trabajo, de reconocimiento por lo que hace, de responsabilidad, de creación del saber, de libertad científica, de disfrute en el trabajo mismo”	La satisfacción del profesorado universitario. Granada: Servicio de Publicaciones de la Universidad de Granada.
Loitegui (1990: 83)	“La satisfacción laboral es un constructo pluridimensional que depende tanto de las características individuales del sujeto cuanto de las características y especificidades del trabajo que realiza”.	Tesis doctoral inédita, Facultad de Psicología, Universidad Complutense de Madrid.
Peiró, 1996	“La satisfacción es una actitud o conjunto de actitudes desarrolladas por la persona hacia su situación de trabajo”.	Tratado de psicología del trabajo. Síntesis S. A. Madrid.
Muñoz Adánez, (1990: 76)	La satisfacción laboral como “el sentimiento de agrado o positivo que experimenta un sujeto por el hecho de realizar un trabajo que le interesa, en un ambiente que le permite estar a gusto, dentro del ámbito de una empresa u organización que le resulta atractiva y por el que percibe una serie de compensaciones psico-socio-económicas acordes con sus expectativas”.	Tesis doctoral inédita, Facultad de Psicología, Universidad Complutense de Madrid

Carrión y cols. (2000)	La satisfacción es “la valoración cognitiva personal que realiza la persona de la vida y de los dominios de la misma, atendiendo a la calidad de vida, a las expectativas y aspiraciones, a los objetivos conseguidos, basada en los propios criterios de la persona y que se realiza de forma favorable”.	Estudio de la satisfacción personal según la edad de las personas. <i>Anales de psicología</i> 2000, vol. 16, n.º 2, 189-198
Díaz (2005). Citado por Iznola y Gabriel (2008:30)	La satisfacción profesional de los docentes puede entenderse como “el resultado de comparar lo que la realidad le invita a ser y lo que según sus expectativas debería ser”.	La satisfacción del profesorado de educación física. <i>Revista Educación física y deporte</i> , n.º 27-2, 27-35
Caballero, (2003:202)	“La satisfacción tiene que ver con las situaciones específicas de su labor docente y con las características de su propia personalidad, en cuanto que todo ello repercute en su estabilidad emocional, creando tensión, estrés, y produciendo malestar, tanto desde una perspectiva personal como profesional”.	Satisfacción de los secretarios de los centros educativos. <i>RELIEVE</i> , v. 9, n. 2, 198-235.
Torres González (2010:28)	“La satisfacción profesional de los docentes puede entenderse como el resultado de comparar lo que la realidad le invita a ser y lo que según sus expectativas debería ser”.	Análisis del grado de satisfacción del profesorado de educación secundaria en el desarrollo de su labor docente. <i>CONTEXTOS EDUCATIVOS</i> , 13, 27-41
Brüggemann (citado por Padrón, 1994)	“La satisfacción profesional es el resultado de una comparación entre lo que hace y lo que debería de hacer: a medida que la distancia disminuye, la satisfacción aumenta”.	La satisfacción profesional del profesorado. Tesis doctoral, Universidad de la Laguna, Canarias.
Anaya y Suárez (2007)	“La satisfacción laboral se entiende como un estado emocional positivo que refleja una respuesta afectiva al trabajo”.	Satisfacción laboral de los profesores de educación infantil, primaria y secundaria. <i>Revista de Educación</i> , 344. Septiembre-diciembre 2007, 217-243

Fuente: (Mayo & Martínez, 2016). La satisfacción laboral y profesional de los profesores. *Revista Lasallista de Investigación*, 13(1).

Así encontramos que las múltiples acepciones referentes a la temática de este trabajo se relacionan con factores cognitivo, actitudinal y afectivo; sin embargo, en la revisión de las publicaciones españolas sobre satisfacción profesional de los docentes muestran un

fuerte componente institucional: CIS, INEE, INCIE, CDIE, COFAPA y por diferentes universidades (Cantón Mayo & Tardif, Identidad profesional docente, 2018)

Con base en las diversas conceptualizaciones, coincidimos que la satisfacción laboral se genera a partir de las características personales, del grado de eficacia profesional y de las condiciones y estímulos que el contexto provea o limiten; es decir la satisfacción/insatisfacción del docente dependen del auto concepto, del desempeño y eficacia en sus funciones y en las condiciones institucionales, sociales y normativas que las autoridades y el sistema educativo establezcan.

En la vertiente negativa de la satisfacción encontramos sentimientos de frustración de los profesores al ver como constantemente están cambiando los currículos educativos, sin solucionar el problema de base para conseguir una educación de calidad, que es centrarse en el bienestar del docente y proporcionarle, por ejemplo una buena formación académica (Cantón Mayo & Telléz Martínez, La satisfacción en el desempeño profesional de los docentes de educación infantil y primaria . Un estudio de caso, 2007).

Algunos estudios ya realizados nos proporcionan los siguientes datos: el grado de satisfacción de los docentes es de medio-alto siendo los profesores de primaria y nivel superior los satisfechos; y los menos satisfechos los de secundaria, algunos otros estudios los resultados fueron de nivel medio otros autores refieren un nivel bajo de satisfacción. Cabe mencionar que encontramos un estudio que compara el nivel de satisfacción laboral de los docentes con otro de años anteriores que detecta que la satisfacción laboral del profesorado ha descendido (Mayo & Martínez, 2016)

Con respecto al desempeño docente encontramos el siguiente significado: el desempeño docente es un concepto que lo determinan en forma particular las acciones o practicas inherentes a la profesión docente los componentes que impactaran en la formación de los alumnos dentro de la gestión educativa, así como fortalecer e intervenir en el diseño, implementación y evaluación de políticas educativas locales y nacionales para promover aprendizajes, desarrollo de competencias y habilidades para la vida.

En la actualidad diferentes gobiernos consideran que la calidad educativa es un problema latente como consecuencia de un desempeño docente deficiente que forma parte de la cotidianidad, así mismo, el desempeño docente nos remite a nociones como función docente, capacidad docente, perfil docente, competencias docentes y desarrollo profesional.

Profesionalismo docente

La profesionalización docente es sin duda el paradigma que actualmente genera más consenso al abordar la labor docente, esto se explica porque ha encontrado la manera de flexibilizar sus límites y permite la entrada de caracterizaciones docentes diversa, sin embargo, esta asimilación no ha logrado ser sustantiva, en el mejor de los casos es nominal, lo que trae consigo la simplificación y/o malinterpretación de conceptualizaciones sobre docentes en sus orígenes, la labor del maestro se ubica en relación con la vocación; esta forma de mirar a los docentes tiene tres componentes: innatismo (el docente nace no se hace) desintereses o gratitud (no cobraras por ello, si naciste para ser maestro entonces debes de ejercer tu misión lejos de cualquier racionalidad instrumental) dignidad reconocida (es consecuencia de los otros componentes)

Perfil docente

La labor docente como una profesión que deviene en el otorgamiento de un servicio clave para el estado, surge la necesidad de distinguir lo que define a todo profesional de la enseñanza; es aquí donde el perfil docente toma lugar como el referente bajo el cual se expresa el conjunto de características que determinan lo que el docente, como profesional debe de llevar a cabo en su labor de enseñar; de acuerdo a los referentes teóricos consultados, se logra identificar que el perfil constituye la función docente de acuerdo con ciertas orientaciones, que en muchas ocasiones se articulan entre sí, pero en lo general retoman los siguientes aspectos:

- Conocimiento para la enseñanza, incluye los contenidos disciplinares, pedagógicos y didácticos que se ponen en práctica para propiciar el aprendizaje
- Saber práctico, lo que ejecuta el docente al resolver y tomar ciertas decisiones, corresponde a lo que presenta y adapta para atender las necesidades educativas de los alumnos.
- Competencias docentes, representan la capacidad que tiene de movilizar varios recursos cognitivos a fin de hacer frente a una determinada situación educativas
- Compromiso por la docencia es el conjunto de principios y facultades que orientan su actuación y la disposición que tiene ante su responsabilidad profesional

Formación docente

La formación inicial y la formación continua, son parte de un proceso de profesionalización que se concreta en el desempeño docente, son variables complementarias de un mismo proceso de formación que orienta el logro de un buen desempeño, se reconoce que la formación docente responde a la necesidad que tiene el docente para afrontar los cambios y demandas sociales que le obligan a mantener una

constante preparación sobre sus capacidades y conocimientos que propicien obtener el desempeño esperado durante el inicio de la profesión docente hasta el final de la formación inicial. El “éxito” de la formación inicial se puede ubicar en un momento crucial; se trata del ingreso al magisterio. Es hasta entonces que se demuestra que sus egresados cuentan con un perfil idóneo, listos para ejercer la profesión docente a partir de los conocimientos y capacidades para las que se fueron preparados durante esta etapa de su formación (Revista Educativa, 2017)

1.2 Planteamiento del problema

El desempeño y las condiciones laborales de la población docente son fenómenos relevantes en el ámbito educativo, que influyen e impactan en el logro de los aprendizajes, en los resultados académicos y en la participación social y comunitaria de las instituciones educativas. Desde nuestra perspectiva, consideramos que el docente en su formación debe desarrollar capacidades y habilidades para indagar, investigar, conocer a sus alumnos y desarrollar las estrategias que incidan en el logro educativo de sus educandos. Así también, si queremos una educación equitativa, inclusiva y de excelencia, debemos revisar el clima organizacional que permea en las escuelas, las condiciones laborales y el nivel de satisfacción/insatisfacción laboral que prevalece en las y los docentes de nuestra región.

El hecho educativo, al igual que variadas actividades humanas, presenta una dinámica compleja que dificulta aprehenderlo como un hecho descriptivo y de sencilla explicación. Para ello es menester recurrir a la sistematización, como acto que se define paulatinamente, que progresa conforme se pone a prueba de manera práctica y se mejora. Pero hay un componente más que permea la función directiva y docente: la satisfacción laboral como una de sus características personales en relación con la escuela, sus compañeros, su remuneración, sus logros con sus alumnos y con sus familias, así también consideramos relevantes algunos factores prácticos que suelen interrelacionarse para tener un status de satisfacción laboral en el enfoque por competencias; Para la valoración del director agrega una más: competencias para dirigir procesos administrativos y de gestión.

La organización, las condiciones de trabajo, la realización personal, las capacidades pedagógicas, la responsabilidad, la emocionalidad y las relaciones interpersonales son factores relevantes que están íntimamente relacionados con la satisfacción laboral y que impactan en el éxito y desempeño de los docentes y directores en todo plantel educativo.

Nos dice que una sociedad avanzada como la nuestra debe luchar por conseguir profesionales de la enseñanza laboralmente satisfechos, dado que los beneficios al nivel individual y organizacional proceden de la calidad de la educación proporcionada a los educandos y revierten en reputación para nuestras instituciones educativas. (Anaya, 2014)

Desde esta perspectiva, es conveniente cuestionar de qué forma y por medio de cuales acciones las autoridades educativas propician las condiciones favorables que garanticen el cumplimiento del mandato constitucional que se plantea en el artículo tercero de nuestra Constitución Política Mexicana. Por lo anterior, es necesario hacernos las siguientes preguntas: ¿Los docentes están satisfechos con su desempeño y condiciones laborales actuales?, ¿Los docentes de nuestra región educativa, alcanzan un status de satisfacción laboral? y lo más importante ¿Cuál es la relación entre desempeño docente y satisfacción laboral?

Realizar esta investigación conlleva el interés de identificar y reconocer las expectativas y motivación de las y los profesores, ya que el grado de satisfacción laboral de éstos, revierte directamente en sus alumnos y en los resultados y logros educativos. Por lo cual en este trabajo a través de dos cuestionarios uno de satisfacción laboral y otro de desempeño docente, se revisaron factores importantes como son: la organización del docente, las condiciones de trabajo, la realización personal, las capacidades pedagógicas, la responsabilidad, la emocionalidad y las relaciones interpersonales. Así mismo pretendemos identificar la relación entre desempeño docente y satisfacción laboral; ya que consideramos que existe una relación muy estrecha entre ambas variables.

1.3 Objetivos

Objetivo General.

Identificar la relación entre las variables de satisfacción laboral y desempeño profesional en la población docente del nivel primaria, sector 06, zona escolar 60 en la región Acapulco-Coyuca de Benítez.

Objetivos Específicos:

- Conocer el nivel de satisfacción laboral de la población objeto de estudio
- Revisar y analizar los factores y/o condiciones que generan o limitan el status de satisfacción laboral
- Contrastar la relación entre satisfacción laboral y desempeño docente.

1.4 METODOLOGÍA.

2.1 Población objetivo.

El estudio se realizó en la zona escolar número 060 del nivel de educación básica específicamente del sector primaria, cuenta con 15 escuelas 11 del turno matutino y 4 del turno vespertino, 11 se encuentran ubicada en la zona suburbana y 4 en la zona urbana y cuenta con 141 profesores, por lo que la población objetivo estará integrada por 141 profesores en la zona y en el nivel considerado. Dado que la población es pequeña, se realizará un censo, es decir, se realizará la investigación sobre todos los profesores

presentes al momento de realizar la entrevista, excluyendo a aquellos que por razones personales no deseen participar en el estudio. Se decidió por el censo, puesto que un muestreo no tiene sentido cuando se tiene acceso a toda la población.

2.2 Cuestionario.

Para investigar sobre la relación entre el desempeño docente y la satisfacción laboral, se adaptó un cuestionario estructurado, ver (Salluca Salluca, 2010), en el que se exploraron dimensiones y dentro de ellas, variables en forma de afirmaciones que, según la experiencia personal de cada entrevistado, se califican con una escala de acuerdo o aprobación tipo Likert, con las opciones **Nunca, Casi nunca, De vez en cuando, Casi siempre y siempre**, que claramente es una escala ordinal que fue codificada con los valores 1, 2, 3, 4 y 5, respectivamente. La satisfacción laboral se estudia a partir de cuatro dimensiones a saber, Organización, Condiciones de trabajo, Realización personal y Beneficios. La primera dimensión se explora con 6 ítems, la segunda con 5, la tercera con 5 y la cuarta con 4.

Para el desempeño docente se consideran las dimensiones Capacidades pedagógicas, Emocionalidad, Responsabilidad en el desempeño de sus funciones y relaciones interpersonales. Para cada una de estas dimensiones se consideraron 6 ítems que ayudan a explorarlas.

2.2.1 Fiabilidad

Todo instrumento de medición debe cumplir con tres requisitos fundamentales, los cuales son la validez, confiabilidad y objetividad. La confiabilidad de un instrumento se traduce en “el grado en que su aplicación repetida al mismo sujeto u objeto produce resultados iguales” (Corbetta, 2007), (Hernández-Sampieri & Fernández-Collado, 2010), o en otras palabras, la confiabilidad sería el grado en que una prueba puede generar resultados consistentes, estables y fiables (Kerlinger, F. & Lee, H.), (Namakforoosh, 2007).

Se comprende que todos los datos obtenidos a través de un instrumento de medición contienen errores, por lo cual se hace necesario calcular el grado de error en la medición para comprender de forma clara si los datos que nos entrega son fiables o no (Kerlinger, F. & Lee, H.).

La confiabilidad de nuestro instrumento fue probada a través del estadístico *alfa* de Cronbach, siendo esta una forma de medida de la coherencia interna del instrumento, la cual se considera significativo si su puntuación es superior a 0.75 en escala de 0 a 1 (Namakforoosh, 2007), aunque en trabajos de las ciencias sociales suele aceptarse un coeficiente superior a 0.65.

2.2.2 Validez

La validez es definida como el grado en que un instrumento mide realmente la variable que pretende medir (Hernández; Fernández, y Baptista, 2010, Namakforoosh, 2005). Se comprende que un instrumento puede ser confiable, pero no necesariamente válido, debido a esto se hace necesario someter al instrumento a los diferentes tipos de validez existentes para asegurar que las características de la variable que se presente medir queden plasmadas en los resultados del instrumento (Hernández; Fernández, y Baptista, 2010, Namakforoosh, 2005; Kerlinger y Lee, 2008).

Se trabajó en el análisis de la evidencia relacionada con el constructor, en nuestro caso, dimensión, o en otras palabras en la validez de constructo del instrumento a través del análisis factorial de los datos (Hernández; Fernández, y Baptista, 2010, Namakforoosh, 2005). La validez de constructo busca probar el grado en que un instrumento representa el modelo teórico subyacente (Kenlinger y Lee, 2008).

Por último, se trabajó en la validez de expertos, la cual consiste en someter el instrumento y sus reactivos al criterio de profesionales asociados al tema de investigación (Hernández; Fernández, y Baptista, 2010). Esta validez fue específicamente abordada en el proceso de adaptación del instrumento.

2.2.3 Objetividad

La objetividad es entendida como el grado de permeabilidad del instrumento ante la subjetividad de los investigadores, tanto en su aplicación, calificación e interpretación de resultados (Hernández; Fernández, y Baptista, 2010). En este caso la objetividad fue salvaguardada a través del uso de indicaciones estandarizadas, lo cual implica usar las mismas instrucciones para todos los participantes (Hernández; Fernández, y Baptista, 2010). Estas indicaciones fueron dispuestas de forma escrita en cada uno de los cuestionarios auto-administrados.

2.3 Análisis estadístico.

Debido a las características del instrumento de recogida de información, primeramente, se calculó en coeficiente *alfa de Cronbach*, con la finalidad de verificar la consistencia interna del cuestionario y de cada una de sus dimensiones o constructos. Una vez verificada la fiabilidad del instrumento se hace un análisis descriptivo de los resultados del cuestionario, se hallan las puntuaciones por ítem y por dimensión, además de la puntuación total para el cuestionario que explora satisfacción laboral y para el que explora desempeño docente. Estas puntuaciones se utilizan para establecer la relación que existe entre el desempeño docente a partir de las puntuaciones obtenidas en el cuestionario sobre satisfacción laboral. Dado que se trata de puntuaciones, es posible hacer un análisis de tendencia usando un modelo lineal o modelo lineal generalizado si consideramos otras variables tales como el género, nivel de estudios, situación laboral y antigüedad en el trabajo.

3. RESULTADOS Y DISCUSIÓN.

3.1 Fiabilidad del cuestionario

El análisis de fiabilidad se realizó a tres niveles, en el primer nivel se consideró todo el instrumento, es decir, el que incluye el bloque de dimensiones para satisfacción laboral y el de desempeño docente. El segundo análisis se realizó a los cuestionarios por separado y en la última etapa se realizó el análisis para cada una de las dimensiones dentro de cada cuestionario.

La fiabilidad del instrumento total fue realizada a partir del procesamiento de los 117 casos, de los cuales 117 fueron válidos y se excluyeron 0.0, como se puede observar en la tabla 1. De los datos analizados se obtuvo un alfa de cronbach de 0.921 lo cual indica que los resultados del instrumento son confiables, o, en otras palabras, la consistencia global del cuestionario es muy buena, como consecuencia la aplicación repetida del cuestionario en un mismo sujeto dará resultados similares y consistentes. Ocurre lo mismo para los cuestionarios analizados por separado, es decir, satisfacción laboral, cuyo alfa de cronbach es de 0.865 y el desempeño docente con 0.902, que resultan ser satisfactorios (véase tabla 2).

Tabla 1: Resumen del procesamiento de los casos

	N	%
Casos Válidos	117	100.0
Excluidos ^a	0	.0
Total	117	100.0

a. Eliminación por lista basada en todas las variables del procedimiento.

Tabla 2: Calidad global del instrumento

No.	Cuestionario	Ítems	Alfa de Cronbach
C1	Satisfacción laboral	1-20	0.865
C2	Desempeño docente	21-44	0.902
C1+C2	Completo	1-44	0.921

Ahora, al realizar un análisis para cada dimensión dentro del instrumento de satisfacción laboral, se encuentra que la consistencia interna por dimensión tiene valores superiores a 0.73, con excepción de la dimensión 4, cuyo alfa de cronbach es de 0.60, por abajo del nivel que aceptamos como válido, por lo que habrá que revisar la validez o realizar un análisis factorial y conservar los ítems de mayor carga factorial. En caso de existir una mejora en la consistencia interna de ésta dimensión, se desechan los ítems de menor carga factorial. En este caso no existe mejora en el alfa de cronbach, por lo que se toma la decisión de conservar todos los ítems (ver tabla 3).

Tabla 3: Calidad del instrumento para satisfacción laboral

No.	Satisfacción laboral	Ítems	Alfa de Cronbach
D1	Organización	1-5	0.751
D2	Condiciones de trabajo	6-11	0.730
D3	Realización personal	12-16	0.753
D4	Beneficios	17-20	0.6

La consistencia interna de las dimensiones dentro de desempeño docente es satisfactoria, puesto que el alfa de cronbach es superior a 0.72, superior al valor de referencia.

Tabla 4: Calidad del instrumento para Desempeño docente

No.	Satisfacción laboral	Ítems	Alfa de Cronbach
D5	Capacidades pedagógicas	1-6	0.721
D6	Emocionalidad	7-12	0.808
D7	Responsabilidad en el desempeño de sus funciones	13-18	0.751
D8	Relaciones interpersonales	19-24	0.809

Por los resultados obtenidos, se considera que el cuestionario cumple con los criterios de validez y fiabilidad que se demandan en este tipo de trabajos de investigación.

3.2 Resultados descriptivos.

Tabla 5: Distribución porcentual de las respuestas para satisfacción laboral.

N=Nunca, CN= Casi nunca, DVE= De vez en cuando, CS=Casi siempre y S= Siempre.

Satisfacción laboral	N	CN	DVE	CS	S
Participa en el diseño de organización en su institución educativa	2.56	.85	2.56	34.19	59.83
Realiza su planeación en tiempo y forma	.9	1.7	3.4	43.6	50.4
Tiene posibilidad de asistir a talleres que contribuyan a su preparación académica	1.7	5.1	19.7	38.5	35.0
Participa de la actualización permanente que ofrece la S.E.G.	.9	5.1	17.1	39.3	37.6
Tiene el asesoramiento y apoyo de expertos en su región ATP	18.8	11.1	23.1	18.8	28.2
Dispone de suficiente tiempo después de su horario para organizar sus clases	2.6	1.7	28.2	41.9	25.6

Se traslada con facilidad entre su hogar y el centro de trabajo	1.7		7.7	39.3	51.3
Dispone de suficiente tiempo para la familia		7.7	22.2	37.6	32.5
El horario de su trabajo le permite realizar otras actividades	10.3	10.3	27.4	32.5	19.7
Está satisfecho con la seguridad social que le proporciona su trabajo (ISSSTE)	16.2	23.1	25.6	17.9	17.1
Su centro de trabajo le brinda seguridad	4.3	12.8	17.9	35.0	29.9
Siente que está realizando algo valioso en su centro de trabajo		.9	3.4	30.8	65.0
El trabajo le ayuda a lograr sus metas profesionales	0	1.7	5.1	42.7	50.4
Tiene reconocimiento de la calidad de sus actuaciones profesionales por parte de los alumnos y/o sus padres de familia	1.7	3.4	20.5	40.2	34.2
Su trabajo es adecuado a sus habilidades y talentos	0.9	.9	3.4	44.4	50.4
Tiene buena relación con los padres de sus alumnos		.9	1.7	38.5	59.0
Reconoce económicamente su rendimiento laboral	17.9	12.8	20.5	26.5	22.2
Tiene libertad de elegir el método de trabajo	0.9	1.7	3.4	29.1	65.0
Está satisfecho con el salario que recibe	25.6	11.1	23.9	23.9	15.4
Tiene la disponibilidad para realizar el examen para una posible promoción	4.3	1.7	11.1	31.6	51.3
Total	2.44	4.45	14.39	34.31	40

Fuente: Elaboración propia

Según la distribución de las respuestas, se observa que la satisfacción laboral del docente se concentra en las dos categorías superiores, es decir, clasifican en las categorías consideradas como satisfacción laboral Media-Alta, que representan el 74.31%. En la literatura se ha establecido que esto obedece al sesgo introducido por la escala tipo Liker que se ha utilizado, ya que las respuestas positivas siempre superan a las negativas, además de que hay estudios que indican que los encuestados tienden a contestar “de acuerdo”, puesto que implica un menor esfuerzo mental a la hora de contestar la encuesta, ver tabla 6.

Se observa, además, que el 50% de los docentes tiene el asesoramiento y apoyo de expertos en su región ATP nulo o casi nulo; aproximadamente el 65% se encuentra insatisfecho con el servicio que le brinda en ISSSTE; el 35% percibe que su centro de trabajo no le brinda seguridad; aproximadamente el 51% no se le reconoce económicamente por su rendimiento laboral y aproximadamente el 61% no se siente satisfecho por el salario que recibe.

En la tabla 6 se observa que el comportamiento de las respuestas, por categoría, refleja la misma tendencia, es decir, se privilegian las respuestas positivas sobre las negativas, por lo que existe sesgo en las respuestas. Se observa que, aproximadamente el 50% de los docentes participan poco o muy poco en proyectos de investigación, además de que el 24% tiene muy poca participación en actividades de formación y actualización docente y de su disciplina. En general, tienen medio-alto desempeño docente, aproximadamente, el 91% de los docentes, es decir, 9 de cada 10, lo cual es realmente alto

Tabla 6: Distribución porcentual de las respuestas para desempeño docente.

N=Nunca, CN= Casi nunca, DVE= De vez en cuando, CS=Casi siempre y S= Siempre.

Desempeño docente	N	CN	DVE	CS	S
Usted domina la temática en sus clases			4.3	39.3	56.4
Usted hace que sus clases sean interesantes			6.0	47.0	47.0
Comunica en forma clara sus ideas y reflexiones	1.7		1.7	47.0	49.6
Utiliza diferentes formas de trabajo en clase que favorezcan el aprendizaje		.9	2.6	47.0	49.6
Utiliza diferentes formas de evaluación (reportes, participación en clases, trabajo en equipo, proyectos, entre otros)	1.7		2.6	42.7	53.0
Con qué frecuencia les llama la atención a sus alumnos cuando observa que se están aburriendo y se distraen		5.1	16.2	38.5	40.2
Lo aprendido en sus talleres y cursos lo aplica en sus alumnos			6.8	36.8	56.4
Genera en los estudiantes interés por realizar su propio aprendizaje	.9		3.4	39.3	56.4
El tema que va a tratar le motiva a buscar información adicional			7.7	35.9	56.4
Promueve los valores de honestidad, respeto, responsabilidad y colaboración entre sus alumnos			6.0	17.9	76.1
Está disponible para resolver dudas o consultas de sus estudiante			1.7	29.9	68.4
Promueve con responsabilidad y respeto el cuidado del medio ambiente	.9		5.1	20.5	73.5
Realiza proyectos de investigación educativa	6.8	11.1	32.5	24.8	24.8
cumple con el horario de clase establecido			1.7	13.7	84.6
Contribuye con aportaciones al logro de los objetivos de su institución	.9		5.1	35.9	58.1
Participa en los comités, consejos, grupos de trabajo y demás comisiones cuyo fin es el mejoramiento de la vida institucional		.9	6.8	33.3	59.0

Participa en actividades (cursos, congresos, seminarios, foros, entre otros) de formación y actualización docente y de su disciplina	.9	5.1	19.7	39.3	35.0
Planifica sus sesiones de clases de forma secuencial	1.7	.9	4.3	40.2	53.0
Demuestra respeto por las ideas de sus alumnos			1.7	12.8	85.5
Utiliza el espacio de clases para que el alumno realice investigación de acuerdo con sus propios intereses			12.0	33.3	54.7
Propicia un ambiente adecuado			1.7	28.2	70.1
Mantiene buenas relaciones humanas con el grupo de estudiantes	3.4		.9	19.7	76.1
Propone actividades que le permite desarrollar algunas de estas habilidades: analizar, comparar, clasificar, pensar de manera crítica o ser creativo	.9		3.4	31.6	64.1
Sus clases se desarrollan en un ambiente de cordialidad	.9		2.6	28.2	68.4
	0.22	0.29	6.52	32.62	59.02

Fuente: elaboración propia.

Por lo anterior, si hacemos un análisis distinto de los datos concentrándonos en las puntuaciones obtenidas, se puede observar que se encuentran ligeramente “cargadas” hacia la derecha, es decir, existe sesgo a la izquierda

3.3 Puntuaciones medias obtenidas.

En la escala tipo Likert se suman las puntuaciones asignadas por los sujetos a cada uno de los ítems y, bajo el supuesto de que todos los individuos responden, la mínima puntuación que pueden asignar a cada ítem es 1 y la máxima es 5. Dado que el desempeño laboral se mide a partir de 20 ítems, la puntuación mínima que se obtiene es de 20, bajo el supuesto de que todos califican con 1 cada ítem y la máxima será de 100, bajo el supuesto de que todos califican con 5, por lo que las puntuaciones se encontrarán en el intervalo [20, 100] y para el cuestionario sobre desempeño docente las puntuaciones variarán entre [24, 120].

Las puntuaciones medias para satisfacción laboral y desempeño docente se muestran en la siguiente tabla 7. Observe que los profesores entrevistados califican por arriba de la puntuación mínima y al menos uno de ellos ha calificado con la puntuación máxima.

Tabla 7: Puntuaciones medias para las variables de interés.

	N	Mínimo	Máximo	Media	Desv. típ.
Satisfacción Laboral	117	46	100	79.50	10.546
Desempeño docente	117	79	120	107.50	9.127
N válido (según lista)	117				

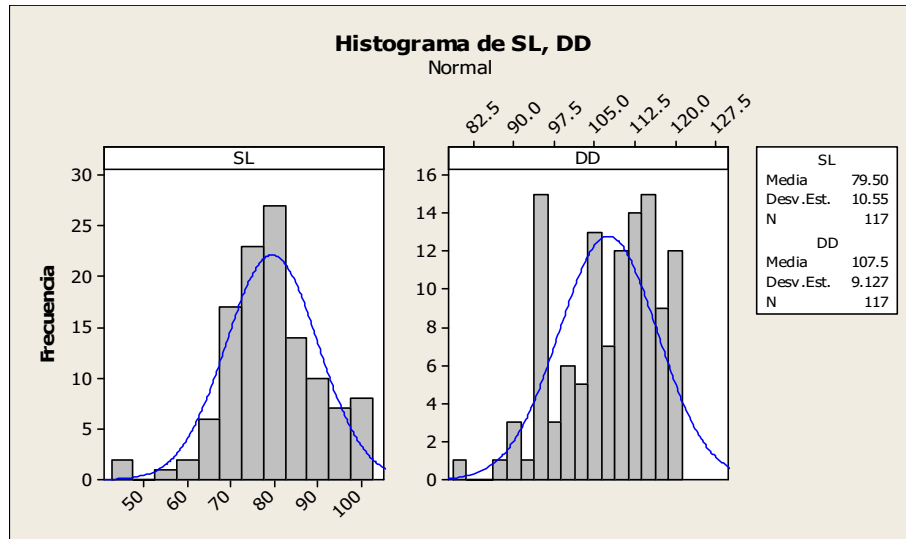


Ilustración 1: Gráfico de las puntuaciones obtenidas para SL y DD

Un criterio para clasificar a los profesores en las categorías Alto (A), Medio (Med), malo (M) y muy malo (MM), consiste en formar cuatro intervalos tomando como referencia la puntuación media y la desviación típica de la siguiente manera:

$$\begin{array}{c} | \quad \text{MM} \quad | \quad \text{M} \quad | \quad \text{Med} \quad | \quad \text{A} \quad | \\ \hline \bar{x} - \sigma \quad \quad \bar{x} \quad \quad \bar{x} + \sigma \end{array}$$

- MM contiene los profesores con puntuaciones por debajo de $\bar{x} - \sigma$
- M contiene los profesores con puntuaciones mayor o igual que $\bar{x} - \sigma$ y menor que \bar{x}
- Med contiene los profesores con puntuaciones mayores o iguales que \bar{x} y menores que $\bar{x} + \sigma$
- A contiene los profesores con puntuaciones mayores o iguales que $\bar{x} + \sigma$

Con este procedimiento es posible clasificar a cada uno de los profesores en cada una de estas categorías, además de la frecuencia relativa, si es que realizamos un conteo de profesores en cada categoría. Los intervalos serán, $MM < 68.954$, $68.954 \leq M < 79.5$, $79.5 \leq Med < 90.046$ y $A \geq 90.046$ para satisfacción laboral y para desempeño docente serán $MM < 98.373$, $98.373 \leq M < 107.5$, $107.5 \leq Med < 116.627$ y $A \geq 116.627$.

De acuerdo con la anterior clasificación, los resultados que se obtienen son los siguientes:

Tabla 8: distribución de la satisfacción laboral.

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	Muy baja	13	11.1	11.1
	Baja	48	41.0	52.1
	Media	40	34.2	86.3
	Alta	16	13.7	100.0
	Total	117	100.0	

Observe que, aproximadamente el 52% de los docentes clasifican en satisfacción laboral baja y muy baja y solo el 48% clasifica en media-Alta, lo cual indica que es posible que la mayoría de docentes se encuentren insatisfechos laboralmente. Una forma de confirmarlo sería mediante una prueba de hipótesis sobre igualdad de proporciones, bajo el supuesto de muestra grande.

Ahora, con relación al desempeño docente, según los resultados obtenido, aproximadamente 45% de los profesores se encuentran en la categoría malo y muy malo desempeño docente, por lo que el 55% de ellos tendrían un desempeño docente medio-Alto. Con esto podemos afirmar que la mayoría de los docentes tienen un desempeño considerado como bueno.

Tabla 9: Distribución de desempeño docente

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	Muy malo	24	20.5	20.5
	Malo	29	24.8	45.3
	Medio	43	36.8	82.1
	Alto	21	17.9	100.0
	Total	117	100.0	

Ahora, ¿el desempeño docente se ve influido por la satisfacción laboral? Para contestar a esta pregunta debemos buscar algún tipo de asociación entre éstas variables, lo cual puede realizarse con el coeficiente de correlación de Pearson, utilizando las puntuaciones totales asignadas por los profesores a cada ítem o con las variables categorizadas, usando un coeficiente de correlación de Spearman o Tau_b de Kendal.

Tabla 10: Matriz de correlaciones de Pearson.

		Satisfacción Laboral	Desempeño docente
Satisfacción Laboral	Correlación de Pearson	1	.614**
	Sig. (bilateral)		.000
	N	117	117
Desempeño docente	Correlación de Pearson	.614**	1
	Sig. (bilateral)	.000	
	N	117	117

** . La correlación es significativa al nivel 0,01 (bilateral).

El coeficiente de correlación de Pearson es de 0.614, por lo que el coeficiente de determinación es de 0.38, es decir, el desempeño docente es explicado por la satisfacción laboral en 38% y el resto es explicado por otras fuentes de variación u otras variables en un modelo lineal. Entonces, si se utiliza este criterio para explicar la relación entre la satisfacción laboral de los profesores y su desempeño laboral, estamos en condiciones de afirmar que la contribución de la satisfacción laboral al desempeño docentes es del 38% y el resto es explicado por otras fuentes de variación, si se supone un comportamiento lineal.

Ahora, si consideramos coeficientes para variables cualitativas obtenemos lo que se indica en la tabla 11.

Tabla 11: Matriz de correlaciones para las variables categóricas SL y DD

			Satisfacción laboral	Desempeño docente
Tau_b de Kendall	Satisfacción laboral	Coefficiente de correlación	1.000	.548**
		Sig. (bilateral)	.	.000
		N	117	117
	Desempeño docente	Coefficiente de correlación	.548**	1.000
Sig. (bilateral)		.000	.	
	N	117	117	
Rho de Spearman	Satisfacción laboral	Coefficiente de correlación	1.000	.628**
		Sig. (bilateral)	.	.000
		N	117	117
	Desempeño docente	Coefficiente de correlación	.628**	1.000
Sig. (bilateral)		.000	.	
	N	117	117	

** . La correlación es significativa al nivel 0,01 (bilateral).

Si consideramos el coeficiente Tau b de Kendal, cuyo valor es 0.548, dado que el pseudo coeficiente de determinación es 0.30, es decir, aproximadamente el 30% del desempeño docente es explicado por la satisfacción laboral, el resto es explicado por otras fuentes de variación. Ahora, si consideramos el coeficiente Rho de Spearman (0.628), el pseudo coeficiente de determinación es de 0.39, es decir, el 39% del desempeño docente es explicado mediante la satisfacción laboral, el resto, es explicado por otras fuentes de variación.

El resultado anterior nos indica que, el desempeño docente no solo es explicado por la satisfacción laboral, sino que se deben considerar otros factores que seguramente inciden en el desempeño de los profesores, por ejemplo, no se ha considerado la preparación profesional, la actualización de profesores, el salario percibido, los estímulos laborales o variables subjetivas como la vocación, el compromiso institucional, etc.

Con la información que tenemos, podemos suponer un modelo lineal, en el entendido que el coeficiente de correlación es apenas de 0.614 y que el modelo nos permitirá tener una idea aproximada sobre el cambio en el desempeño docentes, cuando la satisfacción laboral cambia una unidad.

Supongamos que

$$DD = a + b * SL + \varepsilon$$

Donde a se llama ordenada al origen e indica el valor medio del desempeño laboral, en ausencia de las otras variables; b es la pendiente de la recta e indica el cambio de valor en DD cuando SL cambia una unidad y, finalmente ε es un error aleatorio, en el cual se concentran todos los factores que no controla el investigador. Se estiman los parámetros a y b utilizando el método de mínimos cuadrados ordinarios y los resultados se indican en las siguientes tablas:

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
Dimensión 1 0	.61 0 ^a	.372	.367	.806

a. Variables predictoras: (Constante), Satisfacción laboral

ANOVA^b

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
1 Regresión	44.399	1	44.399	68.263	.000 ^a
Residual	74.797	115	.650		
Total	119.197	116			

a. Variables predictoras: (Constante), Satisfacción laboral

b. Variable dependiente: Desempeño docente

En ésta tabla se puede observar que la satisfacción laboral es explicada por el modelo de regresión lineal solo en un 44.4%, la variabilidad restante es explicada por otras fuentes de variación.

Coefficientes^a

Modelo	Coefficients no estandarizados		Coefficients tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	.735	.229		3.213	.002
Satisfacción laboral	.713	.086	.610	8.262	.000

a. Variable dependiente: Desempeño docente

De acuerdo con estos resultados, el modelo de regresión puede escribirse como

$$\widehat{DD} = 0.735 + 0.713 * SL$$

dado que se rechaza la hipótesis de que los coeficientes son cero, como se puede apreciar en las últimas dos columnas de la tabla anterior. Entonces, en ausencia de satisfacción laboral, el desempeño docente estimado es de 0.735 y la tasa de cambio del desempeño docente, cuando la satisfacción laboral cambia una unidad es de 0.713, es decir crece menos rápido el desempeño docente que la satisfacción laboral.

4. CONCLUSIONES.

Se ha obtenido un cuestionario validado que puede ser utilizado para realizar estudios sobre satisfacción laboral y desempeño docente, toda vez que la consistencia interna del cuestionario, usando el coeficiente alfa de Cronbach, es satisfactorio, con excepción de la dimensión denominada **Beneficios** del cuestionario de satisfacción laboral, cuyo $\alpha = 0.6$, las demás dimensiones reportaron un coeficiente superior a 0.75, por lo que, en términos generales se puede decir que el cuestionario es bueno. Respecto de la validez, también es satisfactoria después de haber realizado los ajustes a partir del criterio de expertos y la validez de cosntructo. Habrá que buscar corregir la dimensión **Beneficios**, toda vez que, aunque se aplicó un análisis factorial para seleccionar los ítems con mayor carga factorial, no se consiguió mejorar la consistencia interna del constructo.

Uno de los hallazgos principales, fue que existe sesgo es las respuestas debida a la escala y se intentó corregir haciendo un análisis sobre las puntuaciones obtenidas utilizando la respuesta media y la desviación típica. Como resultado de este análisis encontramos que 52% de los docentes tiene una satisfacción laboral baja o muy baja y el 48% refleja satisfacción laboral media y alta, aunque no existe diferencia significativa entre estos dos grupos. Respecto al desempeño docente, 45% de los docentes tiene un desempeño docente bajo y muy bajo, mientras que el 55% clasifica como medio y alto.

Cuando intentamos explicar el desempeño docente a partir de la satisfacción laboral, encontramos que éste puede ser explicado en un 39% a partir de la satisfacción laboral y el resto de variabilidad a partir de otros factores que no fueron considerados en este estudio. Para estudios futuros será necesario considerar, además de la satisfacción laboral, otros factores que nos ayuden a explicar el desempeño del docente.

Bibliografía

- Anaya. (2014). *Satisfacción Laboral del Profesorado*. España : Calidad de Revistas Científicas Españolas .
- Cantón Mayo, I., & Tardif, M. (2018). *Identidad profesional docente*. Madrid: Narcea ediciones S.A.
- Cantón Mayo, I., & Telléz Martínez, S. (2007). La satisfacción en el desempeño profesional de los docentes de educación infantil y primaria . Un estudio de caso. *Revista curriculum y formación de profesorado*, 15-25.
- Corbetta, P. (2007). *Metodología y técnicas de investigación social*. Madrid: MCGRAW HILL EDUCATION.
- Freitez, A. (1999). El rol de la educación en el marco de las teorías de la fecundidad: análisis de sus argumentos. *Temas de coyuntura*, 1-34.
- Hernández-Sampieri, R., & Fernández-Collado, C. &.-L. (2010). *Metodología de la Investigación*. México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.
- Martinez-Garrido, C. (2017). La Incidencia del Liderazgo y el Clima Escolar en la Satisfacción Laboral de los Docentes en América Latina.
- Mayo, I. C., & Martínez, S. T. (2016). La satisfacción laboral y profesional de los profesores. *Revista Lasallista de Investigación*, 1-34.
- Namakforoosh, M. N. (2007). *Metodología de la Investigación*. México: Limusa.
- Revista Educativa. (2017). aproximación al concepto de desempeño docente, una revisión conceptual sobre su delimitación.
- Sáenz, & Lorenzo. (2016). La satisfacción laboral y profesional de los profesores. *Revista Lasallista de investigación*, 214-226.
- Salluca Salluca, L. (2010). *Relación entre los niveles de satisfacción laboral y desempeño docente en instituciones educativas del Cerado Callao*. Lima, Perú: Universidad San Ignacio de Loyola.

CAPÍTULO 7

MODELOS OCULTOS DE MARKOV: APLICACIONES EN EL AMBIENTE Y EN LAS FINANZAS

María de Lourdes Sandoval Solís¹, Gladys Linares Fleites², Karla Mildred Cigarroa
Alonso³, Juan Badillo Iglecias⁴

¹Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México. ²Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México. ³Posgrado de Ciencias Ambientales. Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla. Edificio IC3 118, Ciudad Universitaria, Puebla, México. ⁴Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. Edificio CCO2 106, Ciudad Universitaria, Puebla, México.

RESUMEN.

A través de dos aplicaciones de los Modelos Ocultos de Markov se presenta la potencialidad de estos modelos para resolver problemas reales en dos diferentes ámbitos: cambio de uso de suelo y detección de desbalance financiero. En este trabajo, se presenta la distribución de los procesos de cambio cartografiados en la Barra San José, Chiapas, México, observándose que el proceso de cambio que predomina en el periodo de análisis en toda la zona de estudio es el incremento de los asentamientos humanos y el avance de la frontera agrícola, mientras que la deforestación de selvas, bosques y manglares se presenta en forma más acentuada. Por otra parte, ejemplificándose con la base de datos Berka, se presenta la detección de desbalance financiero a través de Modelos Ocultos de Markov, demostrándose numéricamente la convergencia de la matriz de transición y la relación del número de observaciones con la detección del desbalance financiero.

PALABRAS CLAVE: Cadenas de Márkov, Modelos Ocultos de Márkov, Cambio de Uso de Suelo, Detección de Desbalance Financiero.

ABSTRACT

Through two applications of the Hidden Markov Models, the potential of these models to solve real problems in two different areas is presented: change of land use and detection of financial imbalance. In this work, the distribution of the change processes mapped in Barra San José, Chiapas, Mexico is presented, observing that the process of change that predominates in the period of analysis throughout the study area is the increase in human settlements and the advance of

the agricultural frontier, while the deforestation of forests, forests and mangroves is more pronounced. On the other hand, exemplifying with the Berka database, the detection of financial imbalance is presented through hidden Markov models, demonstrating numerically the convergence of the transition matrix and the relationship of the number of observations with the detection of financial imbalance

KEY WORDS: Markov Chain, Hidden Markov Model, land use change analysis, financial imbalance detection.

1. INTRODUCCIÓN

Un Modelo Oculto de Markov (HMM, por sus siglas en inglés) es un proceso doblemente estocástico con un proceso estocástico subyacente que no es directamente observable (está "oculto") pero se puede observar solo a través de otro proceso estocástico que produce la secuencia de observaciones. Desde 1989, que Rabiner [7] presenta los HMM para el reconocimiento de voz, se han aplicado en diversos problemas. Aquí se presentan dos aplicaciones concretas: una es el cambio de uso de suelo en manglares en la zona de San José Chiapa, México y, otra, la detección del desbalance financiero.

Desde la década de los años 90 del pasado siglo, muchos investigadores han usado modelos de Markov o matrices de probabilidad de transición para comprender características de la dinámica territorial, como la diversidad, los controles de cambio, o la dependencia de escala. Los HMM son técnicas de modelado secuencial de datos, que actualmente se consideran una herramienta casi indispensable en el análisis del cambio de uso del suelo [3] y [9]. Por otra parte, en el campo de las finanzas, los HMM también permite modelar la información de las transacciones bancarias de un usuario

A continuación, se presenta, en los epígrafes 2 y 3, una breve exposición de los HMM y sus algoritmos computacionales. Posteriormente, en el epígrafe 4, se desarrolla la aplicación del cambio de uso de suelo en manglares en San José Chiapas y, en el epígrafe 5, se expone la aplicación a la detección del desbalance financiero. Para finalizar, se brindan las conclusiones y las referencias.

2. MODELOS OCULTOS DE MÁRKOV.

Un Modelo Oculto de Márkov (HMM) es un proceso estocástico doblemente embebido con un proceso aleatorio que no es observable (está oculto), pero que puede ser visto solamente a través de otro conjunto de procesos aleatorios que produzcan la secuencia de observaciones y satisface la siguiente propiedad: Propiedad de Márkov: Esta propiedad hace referencia a la falta de memoria de una proceso estocástico, esto es, el estado actual de un proceso únicamente depende de su estado anterior y no de la secuencia de eventos

que precedieron a éste último estado. Los HMM está definido por los siguientes componentes:

Q: Número de estados en el modelo, cada estado individual se etiqueta como q_1, q_2, \dots, q_N .

O: Secuencia de observaciones tomadas de un vocabulario $V = \{v_1, v_2, \dots, v_M\}$, denotadas como $O = (O_1, O_2, \dots, O_M)$.

A: la matriz de transición $A = \{a_{ij}\}$, donde cada a_{ij} representa la probabilidad del estado i al estado j :

$$a_{ij} = P(q_{t+1}=j | q_t = i).$$

B: Probabilidad de observación de cada símbolo del vocabulario, también llamada probabilidad de

emisión, $b_j(k) = P(O_t = v_k | q_t = j)$, es decir, la probabilidad de que la observación O_t sea generada por el estado j .

π : Distribución de probabilidad sobre el estado inicial $\pi = \{\pi_i\}$, definida como: $\pi_i = P(q_1=i)$, es decir, la probabilidad de iniciar en alguno de los estados de q .

Es evidente que una especificación completa de un HMM requiere la estimación de los parámetros del modelo, A, B y π por lo tanto un HMM se define como λ .

$$\lambda = (A, B, \pi) \tag{1}$$

Este modelo toma dos suposiciones importantes. La primera, al igual que en las cadenas de Markov de primer orden, es la suposición de Markov que establece que el estado actual únicamente depende de su predecesor, lo que puede ser expresado como:

$$\text{Suposición de Márkov } P(q_t | q_1, \dots, q_{t-1}) = P(q_t | q_{t-1}) \tag{2}$$

Como segunda suposición, la probabilidad de observación O_t depende únicamente del estado que la produce q_t y no de cualquier otro estado u observación, expresado matemáticamente como:

$$\text{Independencia de salida: } P(O_t | q_1 \dots q_t, \dots, q_T, o_1, \dots, o_t, \dots, o_T) = P(o_t | q_t) \tag{3}$$

3. ALGORITMOS DE LOS MODELOS OCULTOS DE MÁRKOV.

Los HMM, según Rabiner [7], se caracterizan por tres problemas esenciales, que son necesarios resolver para hacer que los HMM sean aplicables a problemas reales. Estos problemas son:

Problema 1 (Probabilidad): El primer problema consiste en la evaluación de una secuencia de observaciones O dado el modelo $\lambda = (A, B, \pi)$. La solución a este problema brinda la probabilidad de que la secuencia sea generada por ese modelo λ , es decir: $P(O|\lambda)$.

Problema 2 (Descodificación): Dada una secuencia de observaciones O y el modelo $\lambda = (A, B, \pi)$, es necesario determinar cuál es la secuencia de estados Q más probable que haya generado las observaciones O .

Problema 3 (Aprendizaje): Ajustar los parámetros A y B del modelo para maximizar las probabilidades de las observaciones $P(O|\lambda)$.

A continuación, se desarrollan aspectos fundamentales del problema.

Probabilidad

Para obtener la probabilidad total de la observación, se tiene que sumar sobre todas las posibles secuencias de Q estados ocultos, en general:

$$P(O|\lambda) = \sum_Q P(O, Q) = \sum_Q P(O|Q) P(Q) \quad (4)$$

Se nota que si se tiene un número de estados ocultos muy grande al igual que el tamaño de la cadena de observación es muy grande existe un problema computacional del orden exponencial $O(N^T)$ lo cual implica que el tiempo de cómputo es muy grande. Para atacar el problema de tiempo computacional se tendrá que hacer uso del algoritmo Forward que es del orden $O(N^2T)$, el cual también evita hacer cálculos repetidos

Algoritmo Forward

La probabilidad de estar en el estado j después de t observaciones, se define con la variable $\alpha_t(j)$

$$\alpha_t(j) = P(O_1, O_2 \dots O_t, q_t = j | \lambda) \quad (5)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\alpha_t(j) = \left(\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad (6)$$

Algoritmo Forward

Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

Recursividad:

$$\alpha_t(j) = \left(\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right) b_j(O_t) \quad 1 \leq j \leq N, 1 < t \leq T$$

Terminación:

$$P(O|\lambda) = \left(\sum_{i=1}^N \alpha_T(i) \right)$$

Algoritmo Backward

La probabilidad de las observaciones desde el tiempo t+1 hasta el final, y estar en el estado i al tiempo t, se define con la variable $\beta_t(i)$

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = i | \lambda) \quad (7)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con la suma de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (8)$$

Para más detalles de la demostración matemática de las recurrencias de los algoritmos consulte [4],[5] y [6].

Algoritmo Backward

Inicialización:

$$\beta_T(i) = 1 \quad 1 \leq i \leq N$$

Recursividad:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \beta_{t+1}(j) b_j(O_{t+1}) \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N$$

Terminación:

$$P(O|\lambda) = \sum_{j=1}^N \pi_j b_j(O_1) \beta_1(j)$$

Descodificación

El segundo problema del HMM es la descodificación, que consiste en obtener la secuencia de estados ocultos Q que generen las observaciones O, es decir, de entre todas las secuencias de estados que puedan generar determinada secuencia de observaciones, se desea encontrar aquella que sea más representativa de la observación O. Para ello, se tiene que encontrar cual es la secuencia de estados ocultos con mayor probabilidad. Se puede resolver este problema ejecutando el algoritmo Viterbi que es muy similar al algoritmo Forward.

La probabilidad de estar en el estado j después de t observaciones y pasar sobre la secuencia de estados más probables, q_0, q_1, \dots, q_{T-1} , se define con la variable $v_t(j)$, de la siguiente forma:

$$v_t(j) = \max_{q_0, q_1, \dots, q_{T-1}} P(q_0, q_1, \dots, q_{t-1}, O_1, O_2, \dots, O_t = j | \lambda) \quad (9)$$

Estas variables son guardadas en celdas, donde el valor de cada una se obtiene con el cálculo del valor máximo de las probabilidades de todos los caminos que llevan a dicha celda, de la siguiente forma:

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (10)$$

Ahora, se define también la variable denominada backtrace(bt) similar a la variable del algoritmo de Viterbi (v_t), en lugar de mantener la probabilidad más alta, esta variable mantiene el número del estado con dicha probabilidad.

$$bt_t(j) = \arg \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad (11)$$

El valor de cada celda se puede calcular mediante el siguiente algoritmo:

Algoritmo Viterbi.

Inicialización:

$$v_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

$$b_{t_1}(i) = 1 \quad 1 \leq i \leq N$$

Recursividad:

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad 1 \leq j \leq N, \quad 1 < t \leq T$$

$$bt_t(j) = \arg \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(O_t) \quad 1 \leq j \leq N, \quad 1 < t \leq T$$

Terminación:

$$P^* = \max_{i=1}^N v_T(i)$$

$$q_T^* = \arg \max_{i=1}^N v_T(i)$$

Para las demostraciones matemáticas detalladas de las relaciones utilizadas en los algoritmos revisar el informe de Millán (2019) [5]

Algoritmo Baum-Welch.

Este el problema crucial de los HMM y consiste en cómo aprenden los parámetros de un HMM, es decir, las matrices A y B. Por lo tanto, será necesario determinar las probabilidades que mejor se adaptan al modelo, a través de una secuencia de entrenamiento, Sin embargo, para un HMM, al tener estados ocultos, no se puede realizar directamente el cálculo de las transiciones, pues lo único que se conoce son las observaciones. Una forma de resolverlo es iniciar con una estimación para las probabilidades de transición y emisión e iterar para estimar mejores probabilidades, dicho algoritmo se conoce como Forward-Backward o Baum-Welch, basado en el algoritmo de Maximización de la Esperanza (EM, Expectation Maximization). Se inicia por estimar las probabilidades de transición \hat{a}_{ij} :

$$\hat{a}_{ij} = \frac{\text{estimación de transiciones del estado } i \text{ a } j}{\text{estimación de transiciones desde el estado } i} \quad (12)$$

Para estimar el numerador, se tienen que sumar las transiciones de i hacia j en cada tiempo t, la estimación de cada una de estas probabilidades en un tiempo t en particular se define como: La probabilidad de estar en el estado i en el tiempo t y en j en el tiempo t+1 dada la secuencia de observación y modelo, la denotaremos por $\xi_t(i, j)$, de esta manera:

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda) = \frac{P(q_t=i, q_{t+1}=j, O | \lambda)}{P(O | \lambda)} \quad (13)$$

Haciendo uso de las variables Forward y Backward llegamos al siguiente resultado:

$$\xi_t(i, j) = \frac{b_j(O_{t+1})\beta_{t+1}(j)a_{ij}\alpha_t(i)}{\sum_{i=1}^N \alpha_T(i)} \quad (14)$$

Este cálculo únicamente para un tiempo t en particular, así que para obtener la estimación de transiciones de i hacia j así como la estimación total de las transiciones de i a cualquier otro estado se tendrá que sumar sobre todos los tiempos, (hasta T-1, ya que el tiempo T no hay transición) de esta manera se tiene:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{k=1}^N \xi_t(i, k)} \quad (15)$$

Se procede de forma similar para la estimación de la probabilidad de emisión:

$$\hat{b}_j(k) = \frac{\text{estimación de estar en el estado } j \text{ y observar el símbolo } k}{\text{estimación de veces de estar en el estado } j} \quad (16)$$

Para estimar el numerador necesitamos conocer la probabilidad de estar en el estado j en el tiempo t y la denotaremos por $\gamma_t(j)$, formalmente como:

$$\gamma_t(j) = P(q_t = j | O, \lambda) = \frac{P(q_t=j, O | \lambda)}{P(O | \lambda)} \quad (17)$$

Nuevamente, usando las variables de Forward y Backward se llega a:

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{i=1}^N \alpha_T(i)} \quad (18)$$

De igual forma, este cálculo es únicamente para un tiempo t en particular donde se observa k, así, para obtener la estimación de la probabilidad de emisión de k, se tiene que sumar sobre todos los tiempos t en donde se observe el símbolo k, para la estimación de estar en el estado j, se suma sobre todos los tiempos, por lo tanto:

$$\hat{b}_j(k) = \frac{\sum_{t=1}^T O_{t=k} \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (19)$$

La demostración del algoritmo de Baum Welch la puede revisar en Richard (1998) [8].

4. APLICACIÓN 1: CAMBIO DE USO DE SUELOS EN MANGLARES DE 1978 A 2017

Un Modelo Oculto de Markov (HMM) se puede definir como una Cadena de Markov en un espacio producto. El análisis del Cambio de Uso de Suelo a través del método de cadenas de Márkov permite profundizar en proyecciones territoriales y realizar análisis probabilísticos y estadísticos de los problemas bajo estudio. Las cadenas de Márkov representan una clase de procesos estocásticos de gran interés para aplicaciones prácticas. En particular, las cadenas de Márkov de tiempo discreto (DTMC) permiten modelar las probabilidades de transición entre estados discretos mediante la ayuda de matrices. Para la aplicación del modelo de Márkov se han generado diversos paquetes de computación como lo refiere Bai, et al. (2017) [1].

El paquete Markovchain para R es un paquete específicamente dedicado al análisis DTMC, Spedicato (2017) [10] y [11]. Este tiene como objetivo lograr una mayor flexibilidad en el manejo de DTMC, que otras soluciones existentes no poseen, proporcionando clases S4 para Cadenas de Márkov homogéneas y no homogéneas. En aplicaciones de los HHM al estudio de cambio de cobertura y uso de suelos este paquete es de gran utilidad.

Distribución de los procesos de Cambio de Uso de Suelo

La distribución de los procesos de cambio cartografiados en la Barra San José, Chiapas se presenta en [8]. El proceso de cambio que predomina en el periodo de análisis en toda la zona de estudio es el incremento de los asentamientos humanos y el avance de la frontera agrícola, registrando un incremento de 5,941 ha en 39 años. Sin embargo, la deforestación de selvas, bosques y manglares se presentó en forma más acentuada, debido a que este proceso afectó el 49 % de la superficie

Magnitud de los procesos de Cambio de Uso de Suelo

La magnitud de los procesos de Cambio de Uso de Suelo en la Barra San José, Chiapas se muestra en la figura 1. En el mapa se observa que el proceso de cambio que predomina en el periodo de análisis es la deforestación de Selvas a Pastizales, Y la deforestación de Manglares a sitios de Asentamientos humanos.

Estos procesos de cambio también afectan las áreas forestales que se desarrollan en el interior de la reserva de la biosfera localizada en la Barra San José; sin embargo, los impactos más importantes se localizan en las periferias de los Asentamientos humanos.

Los procesos de Cambio aquí reportados se deben en gran medida a que la ganadería extensiva constituye en el sureste de México una alternativa con bajos costos y mejor beneficio económico para los campesinos y pequeños productores, ante la constante crisis de los precios de cultivos agroforestales con mayores beneficios ambientales como el café

y el cacao (*Theobroma cacao*); además del fuerte impacto de las olas de migrantes que recibe el estado de Chiapas.

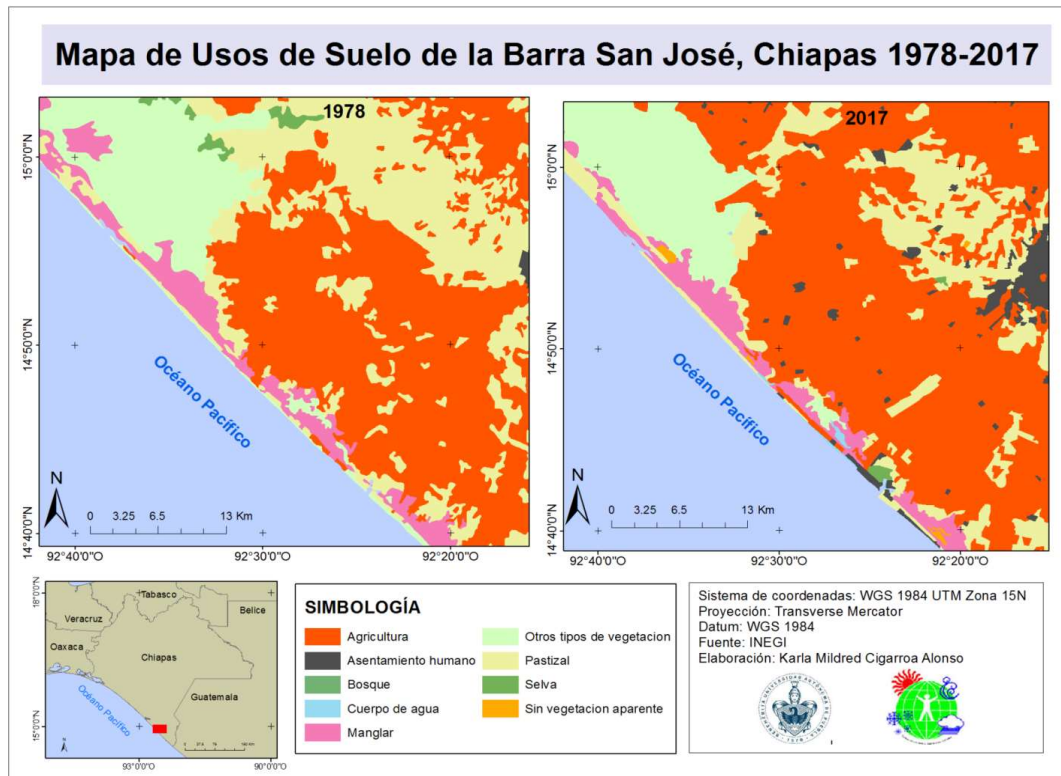


Figura 1. Mapa de Usos de Suelo y tipos de vegetación de la Barra San José, Chiapas para el periodo 1978-2017

Sistema de clasificación de Usos de Suelo

En la tabla 1 se relacionan las siglas utilizadas para cada una de las clases de Usos de suelo que se relacionan en las matrices de transición y probabilidad.

Tabla 1. Clases de las matrices de transición de Uso de Suelo y Probabilidad

Clase	Nombre	Siglas
1	Agricultura	AGR
2	Asentamiento humano	AH
3	Bosque	BO
4	Cuerpo de agua	CA
5	Manglar	M
6	Otros tipos de vegetación	O
7	Pastizal	PA
8	Selvas	SE
9	Sin vegetación aparente	SVA

A partir de las matrices de detección de cambio generadas de la sobre posición de los mapas 1978-2017 (Tabla 2) se estimaron los porcentajes de cambio en 1978 dominaban las Selvas y Pastizales inducidos o cultivados, esta condición cambio para 2017 aumentando la Agricultura y las Áreas sin vegetación aparente, indicativo de un alto grado de deforestación en la zona. Los Asentamientos humanos aumentaron drásticamente de 0.05 a 1.51% del total de la superficie. A lo largo de la franja de manglar se observaron trazos y delimitaciones de terrenos que fragmentan aún más a este ecosistema. Este patrón y su afectación pueden considerarse de gran magnitud, ya que la fragmentación del manglar afecta la integridad y funcionalidad ecológica, además de que limita aún más su regeneración.

Tabla 2. Matriz de transición de Uso de Suelo para el periodo 1978-2017

1978	2017									Total 1978
	AGR	A H	BO	C A	M	O	PA	SE	SV A	
Agricultura	9,816	27 7	482	6	22	18	1,027	191	5	11,844
Asentamiento humano	9	36	0	0	0	0	3	0	0	48
Bosque	673	10	4,20 5	0	0	15	400	964	0	6,267
Cuerpo de agua	4	4	0	32 5	59	68	32	2	16	510
Manglar	139	20	0	12 5	2,66 4	259	508	25	120	3,859
Otros tipos de vegetación	283	24	0	14	152	1,84 3	213	148	42	2,718
Pastizal	4,897	57 5	286	83	350	167	16,93 6	1,859	103	25,255
Selva	1,032	21	846	5	14	89	2,946	7,372	30	12,354
Sin vegetación aparente	0	5	0	9	9	0	5	2	32	62
Total 2017	16,85 2	97 0	5,81 9	56 7	3,26 9	2,45 9	22,07 0	10,56 2	347	62,916

La probabilidad de cambio para el periodo 1978-2017 se muestra en la Tabla 3. La categoría que tienen mayor probabilidad de permanecer en el tiempo es Asentamiento humano (0.97). Mientras que la clase que tuvo una menor probabilidad de permanecer en el mismo tipo de cobertura-uso, fue Sin vegetación aparente (0.56). El Manglar tuvo una probabilidad baja de mantenerse (0.69), registrando transiciones hacia la Agricultura, Cuerpos de agua, Otros tipos de vegetación y Pastizales.

Tabla 3. Matriz de probabilidad de transición durante 1978-2017

1978	2017									Total
	AGR	AH	BO	CA	M	O	PA	SE	SVA	
Agricultura	0.83	0.02	0.04	0.00	0.00	0.00	0.09	0.02	0.00	1.00
Asentamiento humano	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Bosque	0.11	0.00	0.67	0.00	0.00	0.00	0.06	0.15	0.00	1.00
Cuerpo de agua	0.00	0.00	0.00	0.66	0.12	0.13	0.06	0.00	0.03	1.00
Manglar	0.04	0.00	0.00	0.03	0.69	0.07	0.13	0.01	0.03	1.00
Otros tipos de vegetación	0.10	0.01	0.00	0.00	0.06	0.68	0.08	0.05	0.01	1.00
Pastizal	0.19	0.02	0.01	0.00	0.01	0.01	0.67	0.07	0.00	1.00
Selva	0.08	0.00	0.07	0.00	0.00	0.01	0.24	0.60	0.00	1.00
Sin vegetación aparente	0.00	0.09	0.00	0.11	0.17	0.00	0.04	0.04	0.56	1.00

En la Figura 2 se muestra la probabilidad de cambio de las clases entre el periodo 1978-2017. Se observa que todas las clases, a excepción de los Asentamiento humanos, indican probabilidades de transición hacia Pastizales. Es clara la dinámica de desplazo de los Bosques y Selvas hacia Agricultura y Pastizales. Se observa un intercambio importante entre el Manglar y los Cuerpos de agua, denotado por la proximidad entre las clases, marcando la fuerte dinámica entre los ecosistemas costeros de transición. Los sitios que ya no son productivos, por ser escasos en recursos forestales, ser parcelas de cultivos abandonadas o pertenecer a zonas con sobrepastoreo muestran altas transiciones a Asentamientos humanos, como reflejo de una evolución socio-cultural.

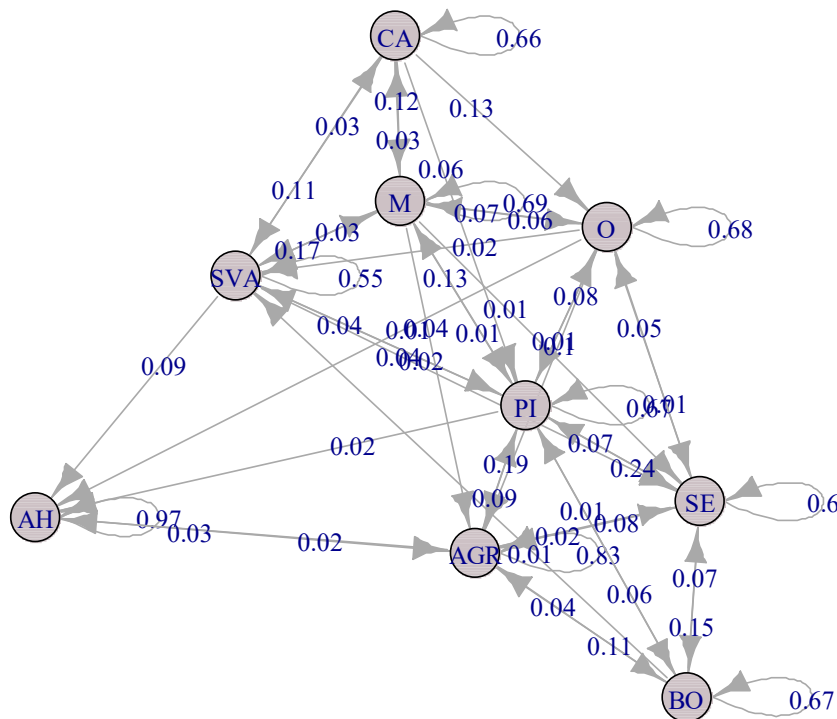


Figura 2. Diagrama de flujo de las probabilidades de transición de cada Uso de Suelo en el periodo 1978-2017.

Puede apreciarse que se ha realizado un análisis de la magnitud, dinámica y distribución de los procesos de Cambio de Uso de Suelo del manglar en los márgenes del vaso principal del sistema lagunar Barra San José, Chiapas, utilizando Sistemas de Información Geográfica y Cadenas de Markov en el periodo 1978 a 2017.

Se corroboran los factores que influyen en el Cambio de Uso de Suelo en esta zona, esto es, que la acelerada deforestación y la intensificación de actividades antrópicas, como la ganadería y la agricultura, están dificultando el buen manejo y la conservación de los manglares de la zona.

Se deben desarrollar buenas estrategias sobre el manejo de los manglares de la Barra San José, Chiapas, antes de implementar alguna política que pueda resultar en severas afectaciones en los remanentes de estos ecosistemas y su diversidad.

Debe destacarse que la importancia del estudio del Cambio en la Cobertura y Uso de Suelo está siendo ampliamente reconocida en la actualidad, planteándose que “una ciencia integrada del suelo” está emergiendo, uniendo las ciencias ambientales, humanas, los desarrollos tecnológicos en Sistemas de Información Geográfica junto a la modelación de procesos estocásticos, con el fin de resolver diferentes inquietudes sobre los Cambios y Uso de Suelo y los impactos que estos cambios tienen en la humanidad y en el medio ambiente.

Es necesario dar a conocer estos resultados a las diferentes autoridades encargadas de la planificación territorial, de manera que conozcan e implementen metodologías de planificación espacial para detectar y establecer las posibles modificaciones del crecimiento urbano que permitan reorientar y minimizar los impactos sobre los manglares de la Barra San José, Chiapas, bajo un contexto de sustentabilidad.

5. APLICACIÓN 2: DETECCIÓN DE DESBALANCE FINANCIEROS

Cuando se tiene una cuenta bancaria el banco quiere detectar cuando hay una transacción bancaria que desbalancea la cuenta, es decir, está en números rojos. Se usan Modelos Ocultos de Márkov para detectar transacciones bancarias que conducen a un desbalance financiero.

Información seleccionada de la base de datos Berka.

La información contenida en la base de datos Berka [2], contiene alrededor de un millón de transacciones bancarias. Se separó por el número de cuenta con sus respectivas transacciones y balance de cuenta. Posteriormente se seleccionaron dos cuentas bancarias que pudieran aportar diferentes puntos de vista sobre el comportamiento de los movimientos que se realizan. Para el experimento se tomó una cuenta con balance

positivo y una cuenta con algunos balances negativos. Las dos cuentas seleccionadas se muestran en la tabla las cuentas y el estado de balance.

Tabla 4. Cuentas bancarias de la Base de Datos Berka [2]

Número de cuenta	Tipo de balance	Cantidad obtenida del balance	Estado de balance
1069	Balance positivo	25220.50	Positivo
2335	Balance negativo cercano al neutro	2785.50	Negativo

Se muestra a continuación el comportamiento en las transacciones de las cuentas seleccionadas. Se enfoca en los montos que representa cada transacción y el balance que tiene la cuenta en cada momento de la transacción.

Cuenta 1069

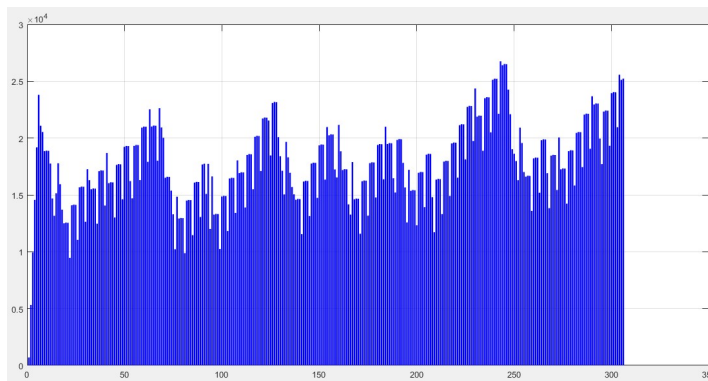


Figura 3. Comportamiento del balance positivo en la cuenta 1069

Cuenta 2335

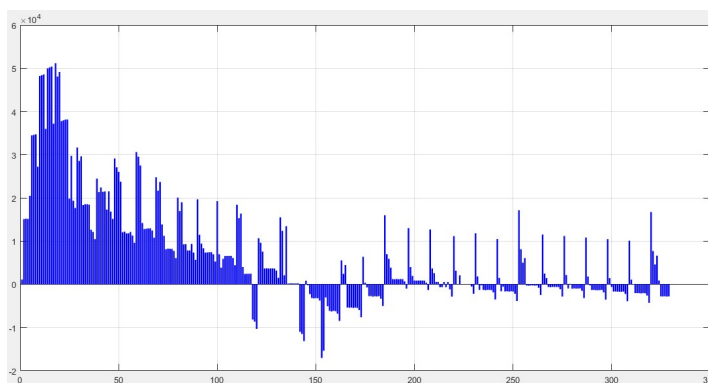


Figura 4 Comportamiento del balance negativo en la cuenta 2335

Determinación del Modelo Oculto de Márkov.

En el caso sobre el análisis de la Base de datos Berka, se usa como estados del modelo la información de las variables del campo Operation: "VYBER KARTOU" retiro de tarjeta de crédito, "VKLAD " crédito en efectivo, "PREVOD Z UCTU" Deposito de otro banco, "VYBER" retiro en efectivo, "PREVOD NA UCET " remesas a otro banco, "YYYYY" Datos sin identificación.

Para definir las observaciones del modelo se siguió lo propuesto en el artículo Srivastava (2008) [12], dividir las diferentes cantidades de las transacciones en tres particiones usando el algoritmo del vecino más cercano con tres centroides obteniendo la observación por estado en la forma {1,2,3}.

Se muestra en la figura 3 la determinación del HMM para cada cuenta bancaria.

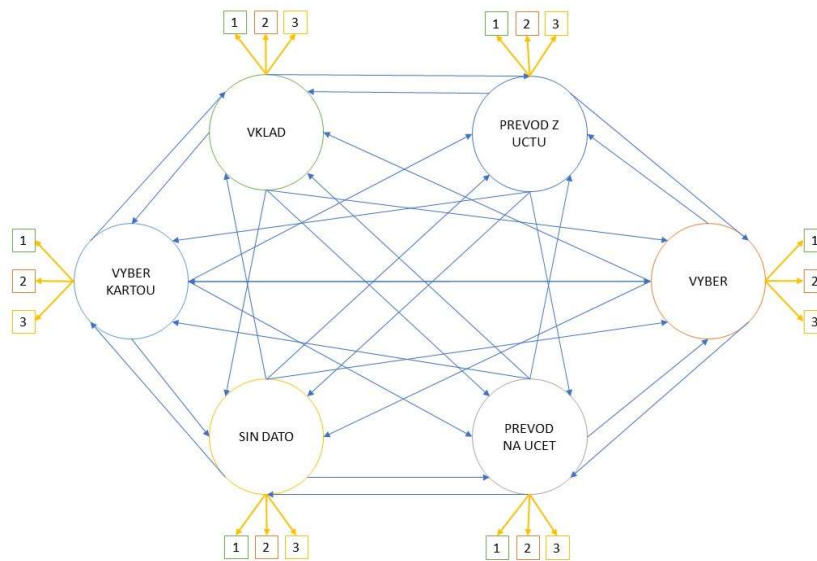


Figura 5. Modelo Oculto de Márkov para las cuentas bancarias de la base de datos Berka [2].

Convergencia de las matrices de transición.

Se analiza el comportamiento de las matrices A y B del HMM, se nota que entre mayor sea el tamaño de muestra para calcular estas matrices, cada término de la matriz de transición converge. A continuación, se muestran los valores de la transición del estado i al estado j conforme aumenta el número de observaciones. Se muestra los resultados para una matriz de transición de tres estados, el elemento de la matriz A_{11} es cero

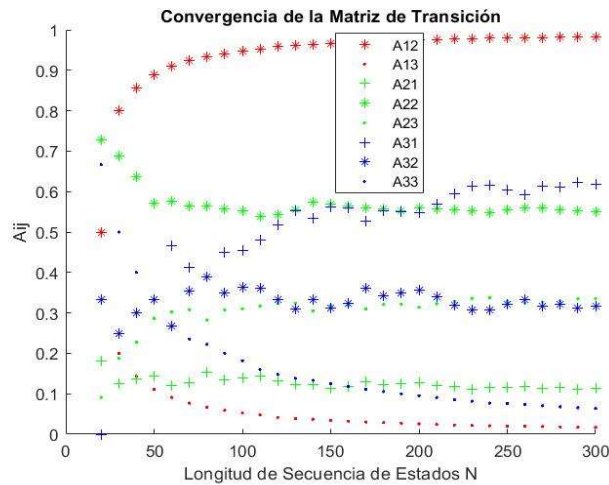


Figura 6. Convergencia de cada término de la matriz de transición

Determinación del Desbalance Financiero

Para realizar las pruebas sobre las cuentas bancarias se compara el logaritmo de la probabilidad de la secuencia de estados más probable para la secuencia de observaciones O_1, O_2, \dots, O_t con el logaritmo de la probabilidad de la secuencia de estados para la secuencia de observaciones

O_1, O_2, \dots, O_{t+1} que se obtienen al aplicar el algoritmo Viterbi . A continuación, se describe el procedimiento:

Paso uno: Se determina un número aleatorio, que en las secuencias de observaciones se considera como la

observación O_{t+1} que se quiere determinar si es un desbalance o no.

Paso dos: Para entrenar el HMM se calculan las matrices de emisión y transición, se usa la secuencia de

observaciones de O_1, \dots, O_t y sus respectivos estados.

Paso tres: Determinar la secuencia de observaciones inicial de tamaño N , O_1, \dots, O_N ,

siendo $O_N = O_t$.

Paso cuatro: Determinar la secuencia de estados más probable y comparar el número de estados dado en el

paso dos.

Paso cinco: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados obtenida en el paso 4.

Paso seis: Determinar la secuencia de observaciones final de tamaño $N+1$, O_1, \dots, O_{N+1} , siendo

$O_{N+1} = O_{t+1}$, que incluye la observación que se desea determinar si es desbalance financiero.

Paso siete: Determinar la secuencia de estados más probable para la secuencia de observaciones del paso 6

Paso ocho: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados obtenida en el paso 7.

Paso nueve: Se calcula Delta definida como la diferencia de Logaritmo de la probabilidad de la secuencia de

estados Final (Paso 8) menos Logaritmo de la probabilidad de la secuencia de estados Inicial (Paso 5).

Paso diez: Se calcula la Delta proporcional respecto a Logaritmo de la probabilidad de la secuencia de estados Inicial

A continuación, se presenta los resultados para cada cuenta bancaria.

Cuenta Bancaria 1069

En el caso de la cuenta 1069 no se tiene ninguna transacción bancaria en la cual se ocasione un balance negativo. Así que se elige aleatoriamente una observación O_{t+1} , las matrices de transición y emisión resultan:

Matriz de transición.

0	0	0	1	0	0
0	0.21052632	0.15789474	0.10526316	0.42105263	0.10526316
0	0	0.08888889	0.33333333	0	0.57777778
0	0.08333333	0.05	0.275	0.4	0.19166667
0.01639344	0.03278689	0.52459016	0.19672131	0.08196721	0.14754098
0	0.03389831	0	0.96610169	0	0

Matriz de emisión.

1	0	0
0.15789474	0.15789474	0.68421053
0	1	0
0.74166667	0.05	0.20833333
0.01639344	0.19672131	0.78688525
1	0	0

Paso tres: Determinar la secuencia de observaciones inicial de tamaño N , O_1, \dots, O_N ,

siendo $O_N = O_t$

Secuencia Observaciones Inicial

2	1	1	3	2	1	1	3	3	2	1	1	3	2	3	1	1	3	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia Estados Inicial

4	5	4	5	3	3	4	5	4	3	5	4	5	3	4	6	4	5	5	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso cuatro:

Secuencia de estados más probable obtenida

4	6	4	5	3	6	4	5	5	3	6	4	5	3	4	6	4	5	3	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Teniendo esta secuencia de estados generada por el algoritmo de viterbi se compara con la secuencia de estados de entrada, se obtiene 0.7000

Paso cinco: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados

Logaritmo de la probabilidad de la secuencia de estados Inicial: -18.6465

Paso seis: Secuencia Observaciones Final

2	1	1	3	2	1	1	3	3	2	1	1	3	2	3	1	1	3	2	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso ocho: Logaritmo de la probabilidad de la secuencia de estados más probable Final:
-19.1147

Paso nueve: Delta: -0.4682

Paso diez: Delta proporcional = 0.0251

Del paso tres al diez se realiza para diferentes tamaños de N . En cada una de las corridas se genera un número aleatorio diferente. En la siguiente tabla se reporta el promedio de cinco corridas para cada tamaño de la secuencia de observaciones inicial.

Tabla 5. Comportamiento de la Cuenta bancaria 1069 de la Base de Datos Berka [2], con balance positivo

Longitud de la Observación	Delta	Delta Proporcional
20	-0.56176	0.0354
17	-0.56176	0.04016
14	-0.56176	0.0544
11	-0.56176	0.05748
8	-0.56176	0.0871
5	-0.56288	0.14836

Cuenta Bancaria 2335

En el caso de la cuenta 2335 se encuentra un número considerable de transacciones que generan en algún punto un balance negativo, que después se recupera a positivo sin embargo para efectos del experimento se define el HMM para un elemento antes del primer balance negativo. Para este caso se obtienen las matrices de transición y emisión:

Matriz de transición.

0	0	0	0	1	0
0	0.11904762	0.0952381	0.23809524	0.35714286	0.19047619
0	0.18181818	0.13636364	0.31818182	0.13636364	0.22727273
0	0.08045977	0.05747126	0.2183908	0.14942529	0.49425287
0	0.2195122	0.09756098	0.18292683	0.45121951	0.04878049
0.0106383	0.07446809	0.0212766	0.38297872	0.13829787	0.37234043

Matriz de emisión.

0	0	1
0.47619048	0.45238095	0.07142857
0.95454545	0.04545455	0
0.71264368	0.10344828	0.18390805
0.75609756	0.09756098	0.14634146
1	0	0

Paso tres:

Secuencia Observaciones Inicial

1	1	2	2	1	1	1	1	1	1	1	1	2	1	1	3	1	1	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Secuencia de estados Inicial

4	3	5	3	3	2	4	6	4	6	5	5	5	5	2	4	4	6	4	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso cuatro: Secuencia de estados más probable obtenida

5	5	2	4	6	6	6	6	6	6	6	6	6	4	6	6	4	6	6	6	6
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Se compara la secuencia de estados generada por el algoritmo de viterbi con la secuencia de estados de entrada, en este caso se obtiene: 0.2500

Paso cinco: Calcular el Logaritmo de la Probabilidad total de la secuencia de estados más probable Inicial:

-12.5840

Paso seis: Secuencia Observaciones Final, que incluye la observación que se desea determinar si es desbalance financiero

1	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	3	1	1	1	1	3
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Paso ocho: Logaritmo de la probabilidad de la secuencia de estados Final: -14.9128

Paso nueve: Delta: -2.3288

Paso diez: Delta proporcional = 0.1851

Del paso tres al diez se realiza para diferentes tamaños de N. En la siguiente tabla se reporta Delta y la Delta proporcional para diferentes tamaños de la secuencia de observación inicial.

Tabla 6. Comportamiento de la Cuenta bancaria 2335 de la Base de Datos Berka [2], con balance negativo

Longitud	Delta	Delta Proporcional
20	-1.4774	0.1412
17	-1.4774	0.1726
14	-1.4774	0.238
11	-1.4773	0.2767
8	-1.477	0.3033
5	-1.4739	0.5127

Se observa en la Tabla 2, que cuando no hay desbalance financiero el valor absoluto de Delta es menor que uno y el porcentaje de variación es menor a 15%. Mientras que la Tabla 3 se muestra que cuando existe un desbalance financiero, el valor absoluto de la variable Delta es mayor que uno y el porcentaje de variación es mayor a 14%.

Los Modelos Ocultos de Márkov permiten modelar la información de las transacciones bancarias de un usuario.

Este modelo lo representan las matrices de Transición y Emisión. Se demuestra numéricamente, con la base de datos Berka [2], la convergencia de la matriz de transición cuando el tamaño de la secuencia de estados aumenta.

6. CONCLUSIONES

Se han presentado dos aplicaciones de los HMM. Una sobre el Cambio de Uso de Suelo, donde se utiliza Cadenas de Markov para modelar y resolver el problema y otra, para detectar el desbalance de cuentas bancaria.

En la primera, los resultados muestran que, en la Barra San José, Chiapas se conjugan dinámicas regionales y locales de cambios, las cuales deben ser tomadas en cuenta en el proceso de planeación y ejecución de políticas de conservación para lograr una mejor conservación y uso de los recursos. Estos estudios ayudan a conocer el tipo de manejo y uso que el hombre hace de la naturaleza en un territorio, y a partir de esto, se puede predecir la dirección del cambio y, en consecuencia, se pueden medir las implicaciones ambientales o sociales de estos cambios probables, sus repercusiones económicas y establecer mecanismos para controlar las causas de origen, contribuyendo a generar un sistema de monitoreo periódico.

En la segunda, también estos modelos permiten determinar, para una cuenta bancaria, si existe un desbalance financiero, comparando dos secuencias de tamaño N y $N+1$. Como trabajo futuro resultaría de gran interés aplicar este proceso a una base de datos de tarjetas de crédito, lo que permitiría detectar la existencia de fraudes.

REFERENCIAS

1. BAI, J., C. DEL CAMPO, L. R. KELLER. (2017): Modelos de cadenas de Márkov en la práctica: una revisión de opciones de software de bajo coste. *Revista Investigación Operacional*. Vol. 38, No. 1, pp 56-62.
2. BASE DE DATOS BERKA , recuperada de <http://lisp.vse.cz/>, consultado : marzo de 2019
3. CIGARROA-ALONSO K.M, LINARES-FLEITES G, SANDOVAL-SOLÍS M.L, VALERA-PÉREZ M.A. (2017). Análisis de Cambio de Uso de Suelo a través de Modelos Ocultos de Márkov. *Algoritmos, estrategias y modelos en las aplicaciones de métodos cuantitativos*, Editores Bouza Herrera C. N, García Rodríguez J. F, Santiago Moreno A, Rueda García M., RIDECA. Habana, Cuba. pp 166-177.

4. JURAFSKY, D., & MARTIN, J. (2000): *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd edition), pp. 122-139.
5. MILLÁN ÁNGELES, V. (2019): Modelos Ocultos de Márkov, Informe de Prácticas Profesionales, Benemérita Universidad Autónoma de Puebla.
6. PACHECO MARTÍNEZ, M. (2019): Redes Neuronales Recurrentes Aplicadas al Reconocimiento de Voz y Traducción Automática de español a japonés. Tesis de Licenciatura en Ingeniería en Ciencias Computacionales. Benemérita Universidad Autónoma de Puebla.
7. RABINER, LAWRENCE R. (1989): A tutorial on Hidden Márkov Model and Selected Applications in Speech Recognition. *Proceeding of the IEEE*. Vol. 77, No. 2, pp 257-286.
8. RICHARD D., SEAN R. E., ANDERS K., GRAEME M. (1998): Biological sequence analysis. The Pitt Building. Trumpington Street, Cambridge, United Kingdom: Cambridge University Press.
9. SANDOVAL-SOLÍS, M.L., LINARES-FLEITES, G., CIGARROA, K.M. Y VALERA, M.A. (2017): Cadenas de Márkov en la Práctica: Aplicación con Lenguaje R. *Desarrollo E Implementación De Las Ciencias Computacionales*. Benemérita Universidad Autónoma de Puebla, México. pp 9-17
10. SPEDICATO, G.A. (2016): Márkovchain: An R Package to Easily Handle Discrete Markov Chains. R package version 0.4.
11. SPEDICATO, G.A. (2017): Discrete Time Márkov Chains with R. The R Journal. IOP Web.
<https://journal.r-project.org/archive/2017/RJ-2017-036/index.html>
12. SRIVASTAVA, A., KUNDU, A., SURAL, S., & MAJUMDAR, A. (2008): Credit Card Fraud Detection Using Hidden Markov Model. *In Dependable and Secure Computing, IEEE Transactions on*, pp. 37-39.

CAPÍTULO 8

UNA REVISIÓN DE RECIENTES ENFOQUES EN LA IDENTIFICACIÓN DE CLÚSTERES EN EL ANÁLISIS DE GRANDES MASAS DE DATOS

Carlos N. Bouza-Herrera¹, Agustín Santiago-Moreno², Sira M. Allende-Alonso¹, José M. Sautto-Vallejo²

¹Facultad de Matemática y computación, Universidad de la Habana.

²Facultad de Matemáticas, Universidad Autónoma de Guerrero.

Abstract

This paper identifies some of the new developments in the identification of clusters when dealing with Big Data problems. We deal with the discussion of hierarchical clusters within the framework of Artificial Intelligence. Expectation-Maximization (EM) algorithms are discussed in this context. We also discuss how; optimization models allow determining the initial step when using EM algorithms.

Keywords: identification of clusters, Big Data, Artificial Intelligence, Expectation-Maximization algorithms

Resumen

En este trabajo identificamos los nuevos desarrollos en la identificación de clústeres al tratar con Grandes Bases de Datos. Se trata el problema de los clústeres jerárquicos y se discute esto en el marco de la Inteligencia Artificial. Los algoritmos del tipo Expectation-Maximization (EM) son discutidos en este contexto. También se discuten como los modelos de optimización permiten obtener deterministamente soluciones a partir de fijar un paso inicial al hacer uso de algoritmos EM.

Palabras Clave: identificación of clústeres, Big Data, Inteligencia Artificial, algoritmos Expectation-Maximization

1. INTRODUCCIÓN

En este trabajo hacemos una revisión de recientes resultados en la labor de identificar clústeres. En las diversas áreas donde se desarrollan estudios cuantitativos se trabaja con conjuntos de datos del tipo $\vec{x} = (x_1, \dots, x_n)$, donde $x_i \in \mathbb{R}^d, d \geq 1$. El interés de establecer patrones de comportamiento de los datos ha generado diversas disciplinas dentro de la estadística analítica. Un deseo común es formar clústeres con los objetos analizados usando la data asociada a cada uno de ellos. Usualmente se desea que no haya solapamiento entre los clústeres, aunque esto puede complejizar el uso de los métodos usualmente presentes en los softwares comerciales.

Una hipótesis comúnmente utilizada es que \vec{x} sigue una distribución normal d-variada

$$f(\vec{x}|\theta) = N(\vec{x}|\vec{\mu}, \underline{V}) = \frac{\exp\left\{-\frac{1}{2}D(\vec{x}|\vec{\mu}, \underline{V})\right\}}{(2\pi)^{\frac{d}{2}} |\underline{V}|^{\frac{1}{2}}},$$

donde:

1. $D(\vec{x}|\vec{\mu}, \underline{V}) = (\vec{x} - \vec{\mu})^T \underline{V}^{-1} (\vec{x} - \vec{\mu})$, distancia de Mahalanobis
2. $\vec{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$, es el vector de medias
3. $\underline{V} = \{\sigma_{ij}\}_{d \times d}$ es una matriz definida positiva (matriz de covarianzas).

Una familia de distribuciones que ha ido ganando en popularidad es la de la llamada del t-tipo. Esta contiene modelos que son superiores al normal pues describe mejor las colas y es más robusta ante observaciones extremas. Esto último es consecuencia de utilizar la distancia de Mahalanobis en el denominador. Su expresión es:

$$f(\vec{x}|\theta) = \frac{\Gamma\left(\frac{\nu+d}{2}\right) |\underline{V}|^{\frac{1}{2}}}{(v\pi)^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) \left(\frac{[1 + D(\vec{x}|\vec{\mu}, \underline{V})]}{\nu}\right)^{\frac{\nu+d}{2}}}, \quad \nu = \text{grados de libertad}$$

2. LAS MEZCLAS

Un marco muy general es el dado por considerar que tratamos con una población heterogénea y que la descripción de la distribución es hecha por la función de densidad multivariada

$$p(\vec{x}|\psi) = \prod_{j=1}^k \pi_j f(\vec{x}|\theta_j) = \prod_{j=1}^k \pi_j h_j \frac{D(\vec{x}|\vec{\mu}_j, \underline{V}_j)}{|\underline{V}_j|^{\frac{1}{2}}},$$

h_j es una función estrictamente positiva continua sobre \mathbb{R} y monótona decreciente en $(0, \infty)$

$f(\vec{x}|\theta_j)$ puede considerarse como la densidad en el clúster j -ésimo y π_j como el peso, tasa o importancia de este clúster.

Al suponer la existencia de $k \geq 2$ clústeres se hace conveniente considerar que se trabaja con la mezcla de densidades normales:

$$\sum_{1 \leq j \leq k} \pi_j N(\vec{\mu}_j, \underline{V}_j); \pi_j > 0, \sum_{1 \leq j \leq k} \pi_j = 1.$$

$\pi_j, \vec{\mu}_j$ y \underline{V}_j se asocian al clúster j -ésimo.

Bajo este modelo tenemos la función de verosimilitud para la data \vec{x}

$$f(\vec{x}; (\vec{\mu}_1, \dots, \vec{\mu}_k), (\underline{V}_1, \dots, \underline{V}_k)) = \prod_{1 \leq j \leq k} \sum_{1 \leq i \leq k} \pi_j N(x_i | \vec{\mu}_j, \underline{V}_j)$$

Note que en el caso particular $\vec{\mu}_j = x_j, \underline{V}_j = \varepsilon I_d$ con $\varepsilon \rightarrow 0$ esta función no tiene máximo. Así que este es un problema mal condicionado de optimización. Por lo que el paradigma usual no es aplicable, pues $f(\vec{x}; (\vec{\mu}_1, \dots, \vec{\mu}_k), (\underline{V}_1, \dots, \underline{V}_k))$ no está acotada, ver Ritter (2015). Cuando la data a analizar posee al menos $d+1$ puntos y el espacio generado es afín y se puede asegurar que:

1. Existe una vecindad abierta de $((\pi_1, \dots, \pi_k), (\vec{\mu}_1, \dots, \vec{\mu}_k), (\underline{V}_1, \dots, \underline{V}_k))$ tal que $f(\vec{x}; (\vec{\mu}_1, \dots, \vec{\mu}_k), (\underline{V}_1, \dots, \underline{V}_k))$ posee un máximo local T_n en ella.
2. La sucesión T_n es fuertemente consistente en

$$((\pi_1, \dots, \pi_k), (\vec{\mu}_1, \dots, \vec{\mu}_k), (\underline{V}_1, \dots, \underline{V}_k))$$

La solución de tal problema ha sido atacada por diversos autores usando el algoritmo EM, ver McLaschlan-Peel, (2000), usando optimización, vea Fritz et al. (2013) etc.

3. CONSTRUCCIÓN DE CLÚSTERES

La determinación del clúster es uno de los problemas de más amplio uso del arsenal de la estadística. En la práctica la Inteligencia Artificial utiliza muchas herramientas que implementan de una otra forma la construcción o detección de clústeres en una masa de datos multivariados. Las herramientas y la interpretación de ellos dependen de la naturaleza de la data. La data podemos clasificarla como simétrica, asimétrica cuando usamos una métrica. Gran parte de la literatura trata de los problemas en términos métricos. Métodos con una base matemática teórica, así como heurísticas conforman una literatura de miles de trabajos científicos.

Consideremos que deseamos determinar una partición en k clústeres $C = (C_1, \dots, C_k)$

Recientemente Gallegos-Ritter (2018) propusieron utilizar para construir clústeres el modelo

$$\begin{aligned} \log f \left(\left((\pi_1, \dots, \pi_k), (C_1, \dots, C_k), (\vec{\mu}_1, \dots, \vec{\mu}_k), (V_1, \dots, V_k) \right) \right) &= \log f(\vec{\pi}, C, M, V) \\ &= \sum_{1 \leq j \leq k} \sum_{i \in C_j} \log \left(\pi_j N(x_i | \vec{\mu}_j, V_j) \right) \end{aligned}$$

Cuando para todo j se cumple que $\|C_j\| = n_j \geq d + 1$ se tiene un problema de optimización con restricciones.

Note que si hacemos una adecuada clusterificación las medias de los clústeres deben ser diferentes. Es decir, debe ser rechazada la hipótesis nula

$$H_0: \vec{\mu}_1 = \dots = \vec{\mu}_k$$

En la docimasia de tal hipótesis aparecen métodos generales para diseñar adecuadamente la prueba.

Una hipótesis es la de homocedasticidad. Comúnmente se utiliza la llamada Λ -Wilks para homocedasticidad que es realizada usando el estadístico de prueba:

$$W = \frac{f_1(\vec{x} | \vec{\pi}, C, \hat{M}, \hat{V})}{f_0(\vec{x} | \vec{\pi}, C, \hat{M}, \hat{V})} = \left[\frac{\det(\hat{V})}{\det(\hat{V}^*)} \right]^{\frac{n}{2}}$$

donde

$$\hat{M} = (\bar{x}_1, \dots, \bar{x}_k), \bar{x}_j = \text{mean vector in } C_j$$

$$\hat{M} = \frac{1}{n} \sum_{j=1}^k n_j \bar{x}_j \quad n_j = \text{unidades en el clúster } j\text{-ésimo, } n = \text{tamaño de la muestra total}$$

$$\hat{V} = \text{matriz de dispersión de los datos,}$$

$$\hat{V}_j = \text{matriz de dispersión in } C_j$$

$$\hat{V} = \frac{1}{n} \sum_{j=1}^k n_j \hat{V}_j$$

W se distribuye de acuerdo a una potencia negativa de Λ con parámetros $d, n-k, k-1$.

En los 90's se propusieron métodos alternativos para paliar la existencia de algunas fuentes de heteroesdasticidad. Vea Banfield-Raftery (1993) y Celeux-Govaert (1995).

Ante restricciones cruzadas en los clústeres, cuando se trabaja con mezclas de normales, se puede re-parametrizar en términos de los valores propios usando

$$V_j = \lambda_j \underline{D}_j \underline{A}_j \underline{D}_j^T, j = 1, \dots, k;$$

$$\underline{D}_j = \text{matriz ortogonal de los vectores propios de } V_j$$

Estos vectores describen la orientación de la dispersión. \underline{A}_j es una matriz diagonal escalada, $|\underline{A}_j| = 1$, cuyos elementos son proporcionales a los valores propios de V_j en

orden decreciente dándole forma . λ_j esa una constante de proporcionalidad relacionada al volumen de los clústeres, $\lambda_j^{\frac{d}{2}} = |V_j|^{\frac{1}{2}}$.

Todos ellos se consideran parámetros independientes. Estos se restringen al mismo clúster o se permite que varíen entre ellos. Las posibles combinaciones de sus varios valores determinan 14 modelos posibles, ver, Garcia-Escudero et al (2018). El software MCLUST, ver Fraley et al. (2012) y Scurcca et al (2016).

Un análisis similar fue desarrollado por Andrews et al. (2011)

A partir de la propuesta de Rocci et al. (2016) de utilizar

$$\vec{x}^+ = \underline{A}\vec{x} + \vec{b}$$

nos lleva a la función de verosimilitud

$$L(\psi) = n \log |\underline{A}| + \sum_{i=1}^n \log \left(\sum_{j=1}^k \pi_j N(\vec{x}_i^+ | \vec{\mu}_j^+, \underline{V}_j^+) \right)$$

De ella se deriva que la clasificación basada en el máximo de las probabilidades a-posteriores es invariante ante el grupo de transformaciones lineales afines. En este caso la generalización de la restricción es

$$a \leq \lambda_i(\underline{V}_j \underline{B}^{-1}) \leq b; \underline{B} \text{ simétrica y definida positiva}$$

Esta matriz representa la información a-priori sobre la estructura de la matriz de covarianzas. Esta nueva restricción nos lleva a la equi-varianza del máximo de la función verosimilitud.

Es usual que se haga uso de algoritmos EM para hacer la estimación de máximo verosímil. Para modelos de clustering y clasificación basándose en mezclas de componentes elípticos.

4. CLUSTERING JERÁRQUICO

La determinación de clústeres jerárquicos es de mucha importancia en las diferentes aplicaciones. Las observaciones son conectadas determinando clústeres. Una partición es determinada y es analizado el gráfico correspondiente.

La red creada se denota $N_X = \{(X, A_X)\}$ donde X es un conjunto finito de nodos y $A_X: X \times X \rightarrow \mathbb{R}_+$. El método de clustering que se utilice busca una partición $P_X = \{B_1, \dots, B_J\}$, $\cup_{1 \leq j \leq J} B_j: B_j \cap B_t = \emptyset, j \neq t$ de X .

Una relación de equivalencia es identificada por \sim y es una relación binaria tal que si $x(1), x(2)$ y $x(3)$ pertenecen a X entonces :

$$x(i) \sim x(j); x(i) \sim x(j) \text{ si y solo si } x(j) \sim x(i); \text{ si } x(1) \sim x(2) \text{ y } x(2) \sim x(3) \\ \Rightarrow x(1) \sim x(3)$$

En una partición P_X que induce una relación de equivalencia, \sim_{P_X} , tenemos que $x(1) \sim_{P_X} x(2)$ si y solo si ambos pertenecen al mismo bloque.

Cada B_t es llamado bloque (clúster) y el conjunto potencia de X se denota como $X^\#$, conjunto de todos de los bloques pertenecen a X .

Contemporáneamente se hace importante no considerar solo la partición como dada por un algoritmo, método o data sino determinar familias de particiones que puedan ser indexadas en cierta forma. Esto permitirá desarrollar estudios que determinen el comportamiento de estos. Un acercamiento es hacer uso de propiedades de las particiones en los análisis. Pueden verse estas ideas iniciales ya en Jain-Dubes (1988) y Lance-Williams (1967), Carlsson-Mémoli (2010) probaron que la propuesta de conexión simple, ver Jain-Dubes (1988), es el único clustering jerárquico que satisface los axiomas siguientes:

Axioma del valor: en una red con dos nodos esos son unidos a una resolución igual al máximo de la dos disimilaridades envueltas.

Axioma de la transformación: el nivel de la resolución en que dos nodos de la red objetivo se vinculan no es mayor que el nivel en que lo hacen en la red original.

Si el método los satisface se dice que es admisible.

Diversas extensiones se han desarrollado de estos axiomas para otras clases de problemas, ver Carlsson et al. (2018).

Note que la salida de un clúster jerárquico no es una partición sino una colección imbricada de particiones : $D_X(\delta)$, donde $\delta \geq 0$ es un parámetro de resolución ue sub-indiza el dendograma D_X . Denotamos la existencia de conexión entre $x(i)$ y $x(t)$ (estar en el mismo bloque , clúster) como $x(i) \sim_{D_X} x(j)$. Un dendograma debe cumplir las condiciones

Condición de acotación: si $\delta = 0$ la partición determina clústeres $C_j = \{u_j\}, j = 1, \dots, n$. Si $\delta^* \gg$ (suficientemente grande) $D_X = \{x_1, \dots, x_n\}$.

Condición de jerarquía: Al hacer crecer δ los clústeres son agregados pero no pueden ser divididos .

Así que:

1. $D_X(0) = \{\{x\} | x \in X\}$, $D_X(\delta^*) = \{X\}$ para un cierto parámetro δ_0 todo $\delta^* \geq \delta_0$, si $\delta < \delta^*$
2. Para todo par $x(i), x(j)$ tales que $x(i) \sim_{D_X(\delta)} x(j) \implies x(i) \sim_{D_X(\delta^*)} x(j)$.

Puede modelarse la construcción de dendogramas al considerar que la construcción de clústeres jerárquicos es una función

$$\mathfrak{D}: N_X \rightarrow \mathfrak{D}; \mathfrak{D} = \text{espacio de los dendogramas}$$

Así que un dendograma es $D_X = \mathfrak{D}(X, A_X)$.

Para una red particular si se satisfacen los axiomas del valor y la transformación debemos construir la disimilitud simétrica

$$\forall x, x' \in X, \quad A_X^\dagger = (x, x') = \max\{A_X^\dagger = (x, x'), A_X^\dagger = (x', x)\}$$

Con ella el problema de hacer clústeres de datos simétricos se convierte en uno de linkeo simple, ver Carlsson et al. (2014, 2018). En este contexto se puede determinar un clustering recíproco.

Una cadena que une $x(1)$ y $x(2)$ es una secuencia de nodos $\{x(1) = x_0, \dots, x_t = x(2)\}$. Se denota tal cadena como $C\{x(1), x(2)\}$. Las cadenas pueden concatenarse. Si $C=C\{x(1), x(2)\}$ y $C'=C\{x(2), x(3)\}$ estas están concatenadas. Note que $C\{x(1), x(2)\} \neq C\{x(2), x(1)\}$.

En el marco teórico que se ha abierto hay la posibilidad de modelar el análisis de los dendogramas. Como todos sabemos su análisis va a depender de la subjetividad del analista. Además, al enfrentarnos a grandes masas de datos es mucho más difícil hacerlo. Carlsson et al. (2014, 2018) propusieron de uso de ultramétricas para dar una evaluación de los dendogramas.

Tomemos una función no negativa definida sobre el conjunto de nodos X . Para que $u_X: X \times X \rightarrow \mathbb{R}_+$ sea una métrica, debe satisfacer las condiciones:

Identidad : que $\forall x, x^* \in X, u_X(x, x^*) = 0$ si y solo si $x = x^*$.

Simetría: $\forall x, x^* \in X, u_X(x, x^*) = u_X(x^*, x)$.

Desigualdad triangular fuerte: $\forall x, x^*, x^{**} \in X, u_X(x, x^{**}) \leq \max\{u_X(x, x^*), u_X(x^*, x^{**})\}$.

Claramente el espacio ultra métrico es un caso particular del métrico.

Sea \mathcal{U} el espacio de las redes provisto con ultra métricas. En este marco de trabajo se puede establecer una estructura que preserve la biyección del mapeo entre dendogramas y ultra métricas. Carlsson-Mémoli (2010) probaron la validez de que si tenemos bien definidos los mapeos $\mathfrak{D}, \mathcal{B}$:

1. $\mathfrak{I}: \mathfrak{D} \rightarrow \mathfrak{U}$ tal que $\mathfrak{I}(D_X) = (X, u_X)$. Tomemos $u_X(x, x^*)$ como la menor resolución en la que se agrupan $x, x^*: u(x, x^*) = \min\{\delta \geq 0 | x \sim_{D_X} x^*\}$.
2. $\mathfrak{B}: \mathfrak{D} \rightarrow \mathfrak{U}$ para todo $\delta \geq 0$ definida como $x \sim_{u_X(\delta)} x^* \Leftrightarrow u(x, x^*) \leq \delta$.

Hagamos $D_X(\delta) = \{X \text{ mod } \sim_{u_X(\delta)}\}$, $\mathfrak{B}(X, u_X) = D_X$. Note que $\mathfrak{I} \circ \mathfrak{B}$ es la identidad en \mathfrak{U} y $\mathfrak{B} \circ \mathfrak{I}$ lo es en \mathfrak{D} .

Estos resultados soportan que los métodos jerárquicos descritos anteriormente inducen ultra métricas en X basadas en funciones de disimilitud. Así que $\mathcal{H}: \mathcal{N} \rightarrow \mathcal{U}$, donde $\mathcal{N} \supseteq \mathcal{U}$. Así dos métodos jerárquicos son equivalentes solo si $\mathcal{H}(N) = \mathcal{H}^*(N)$ para todo N de \mathcal{N} .

5. LAS GRANDES MASAS DE DATOS

Actualmente un problema que pervade las aplicaciones contemporáneas es la existencia de grandes bases de datos. Los repositorios como YouTube, Amazon, Google, Facebook etc. permiten determinar grandes grupos relativamente homogéneos de diferentes tamaños. Estos, además, exhiben, comúnmente, información de grandes dimensiones, al identificar los datos existentes asociados a cada unidad. El interés puede centrarse no en los grandes clústeres sino en los pequeños pues los decisores desean, en muchas ocasiones, analizar qué es lo que genera grupos muy diferenciados de las grandes tendencias. En otros, el interés está en el estudio de los grandes grupos. Esos son indicadores de atipicidad y en los grandes repositorios es común que sea difícil el identificar tales clústeres, pues en las grandes bases de datos se tiende a sobredimensionar los comportamientos medios. En todos los casos es de esperar que los clústeres sean de tamaños muy diferentes, son altamente desbalanceados. Esto es identificado en la literatura con el tratar con datos no balanceados. Es un reto extraer información de grandes bases de datos desconocidos para identificar grupos “interesantes”. La construcción de los clústeres en datos no –balanceados genera retos adicionales a los que clásicos métodos de clustering.

Los métodos usuales de clustering fallan en la identificación de pequeños grupos en grandes repositorios de datos. Al tratar con datos de alta dimensionalidad los métodos tratarán de identificar sub-espacios de menor dimensión para detectar tales clústeres. Esto plantea la necesidad de poseer alguna información a apriorística para fijar valores de parámetros que permitan orientar la búsqueda de sub-espacios de menor dimensión. Si los datos no están muy contaminados con ruidos van a ser exitosos métodos basados en modelos o usando funciones de densidad.

Kraccywz (2016) señaló como los métodos basados en centroides tienden a generar clústeres de igual tamaño, al mezclar los clústeres pequeños con los grandes. Este es denominado efecto de uniformización. Su propuesta fue un método híbrido donde se

combinen la concepción de los centroides con ideas asociadas al uso de funciones de densidad no paramétrica.

Una propuesta de algoritmo que parece como satisfactoria es la propuesta de Bordonivà et al. (2018). Este se centra en la identificación de clústeres pequeños en grandes masas de datos. Este se basa en el llamado factor local de outliers (local outliers factor, LOF). Vea Breunig et al. (2000). Este mide el grado de aberración de una observación en la masa de n datos. Sea una vecindad local de h -vecinos cercanos de una observación \vec{x}_i cuyo tamaño es $N_h(\vec{x}_i)$, y la densidad $g(\vec{x}_i|N_h(\vec{x}_i))$. Este factor es calculado mediante

$$LOF_h(\vec{x}_i) = \frac{1}{N_h(\vec{x}_i)} \sum_{\vec{x} \in N_h(\vec{x}_i)} \frac{g(\vec{x}|N_h(\vec{x}_i))}{g(\vec{x}_i|N_h(\vec{x}_i))}$$

Si la observación tiene una densidad similar a la de los vecinos $LOF_h(\vec{x}_i) \cong 1$. En el caso en que es mucho mayor que 1 esta observación es considerada un outlier. Esa medida no es seriamente afectada por el valor de h . Zimek et al (2012) probaron la efectividad de esta medida incluso al tratar con data de alta dimensionalidad.

En las implementaciones puede usarse LOF para eliminar outliers, cuando el interés está en los clústeres grandes.

El algoritmo propuesto por Brodinová et al. (2018), llamado IClust aparece como un software adecuado para identificar pequeños clústeres en una gran masa de datos de alta dimensionalidad.

En el contexto del tratamiento de grandes masas de datos nos encontramos con la existencia de bloques de variables (componentes). Es interés establecer conexiones entre las relaciones entre estos “bloques de variables”. Cada bloque es obtenido para cada observación. Es común asumir que las observaciones provienen de poblaciones que son homogéneas. Esto es cierto en áreas como los estudios basados en el monitoreo. Si este no es el caso los métodos estadísticos no poseen las propiedades que la teoría les asigna. Así que antes de aplicar los métodos se va a dividir la población en clústeres (poblaciones “homogéneas”). Un viejo enfoque es considerar que la estructura de los clústeres es identificable a partir de la llamada “clusterwise” regresión. Vea los primeros intentos de desarrollo de este enfoque en Spath (1979). En este enfoque se asume que la regresión incorpora la información de los grupos a través de un modelo particular de regresión. Así que se buscan particiones de las observaciones dentro de los clústeres que minimicen la suma de cuadrados, al considerar la estructura determinada al hacer la partición usando un método de clustering.

Partiendo de la variable independiente tenemos la matriz $Y_{n \times q}$ y k matrices con las variables independientes, que determinan los bloques: $X_j = (\vec{x}_{1j}, \dots, \vec{x}_{IJ(j)})$ y la definida sobre la población entera a $\underline{X} = (\underline{X}_1, \dots, \underline{X}_k)$, $J = \sum_{j=1}^k J(j)$. Los bloques pueden poseer pesos diferentes. Las matrices de trabajo poseen columnas estandarizadas,

por lo que la suma por columna es cero y la norma es uno. Así que es mejor notacionalmente considerar que trabajamos con

$$\underline{X}_{j*} = \frac{\underline{X}_j}{\sqrt{\frac{k}{n-1} \text{traza} \left((\underline{X}_j)^T \underline{X}_j \right)}}, \quad \underline{Y}_{j*} = \frac{\underline{Y}}{\sqrt{\frac{1}{n-1} \text{traza} \left((\underline{Y})^T \underline{Y} \right)}}$$

La proyección ortogonal $P(j)$ del espacio Euclidiano \mathbb{R}^n es dada por

$$P(j) = \underline{X}_j \left((\underline{X}_j)^T \underline{X}_j \right)^{-1} (\underline{X}_j)^T$$

La información relevante es expresada por la combinación lineal de las columnas de las matrices envueltas al determinarse los vectores de coeficientes (loadings) de las matrices descritas mediante las relaciones

$$\vec{u} = \underline{Y}\vec{v}; \quad \vec{t} = \underline{X}\vec{w}; \quad \vec{t}(j) = \underline{X}_j\vec{w}_j, j = 1, \dots, k,$$

En la población completa, el componente global de las variables independientes es expresada por \vec{t} y en los clústeres por $\vec{t}(j)$.

Los vectores de observación van a ser agrupados en H clústeres. Sea $C(h)$ el clúster h -ésimo, $n(h)$ el número de observaciones en este y $\underline{X}_{j(h)}, \underline{Y}_h$ las correspondientes matrices.

Los métodos de este enfoque buscan minimizar simultáneamente las sumas de cuadrados sobre los clústeres. En esto se asemejan a la política que justifica los métodos de las k -medias. La literatura en el tema evidencia esto. Vea como Spath (1979) y De Sarbo-Cron (1988) buscaron resolver un problema de máxima verosimilitud. Tomar en cuenta no solo las sumas de cuadrados sobre los clústeres sino también la interna fue estudiada por Vicari-Vichi (2013). El uso de mínimos cuadráticos parciales fue la propuesta de Vinzi et al. (2005) y de Preda-Saporta (2005). Sardedt (2008) hizo un análisis del uso del análisis de senderos. Los soportes computacionales de estos populares son FIMIX-PLS, REBUS-PS, PLS.IRSS etc. Ver Bougeard et al (2018). Estos últimos proponen dos nuevos algoritmos: CW. MBPS y CW. MBRAC.

6. EL PROBLEMA DE OPTIMIZACIÓN

La función de verosimilitud puede presentar varios máximos locales que son espurios y que son consecuencia de que el correspondiente clúster contiene pocos puntos, que están muy cerca o acaso que estos son pertenecientes a un espacio de dimensión menor. McLachlan-Peel (2000) señalan que estas soluciones espurias son caracterizadas como que una raíz del problema de máxima verosimilitud está cerca de la frontera del espacio paramétrico y que algunos de sus componentes sobredimensionan pequeños patrones aleatorios en la data en vez de reflejar una estructura de grupo.

Varias soluciones pueden ser recomendadas. Ninguna es completamente satisfactoria. Una es la de utilizar algún algoritmo estocástico, de optimización global. Su naturaleza es la que garantizará la convergencia. Por ejemplo, el Recocido Simulado posee tal

propiedad, aunque su velocidad de convergencia es lenta. Esto hace que su comparación con los algoritmos del tipo EM no le hagan absolutamente más recomendable.

Otro enfoque popular es monitorear los tamaños relativos de las proporciones de mezcla, McLachlan-Peel (2000) o evitar la degeneración introduciendo una verosimilitud acotada penalizada, que la evite en la clausura del espacio paramétrico, Ciuperca et al (2003). Este último utiliza un punto de vista Bayesiano.

6.1 Optimización con restricciones

Al buscar la maximización de la verosimilitud se plantea usualmente resolver el problema

$$\hat{\psi} = \text{ArgMax}(\vec{x}|\psi) = \prod_{j=1}^k \pi_j f(\vec{x}|\theta_j)$$

Un enfoque inicial es el debido a Hattaway (1985) que buscaba la solución para el caso de una mezcla de distribuciones Gaussianas univariadas

$$f(\vec{x}; (\mu_1, \dots, \mu_k), (\sigma_1^2, \dots, \sigma_k^2)) = \prod_{1 \leq j \leq k} \sum_{1 \leq i \leq k} \pi_j N(x_i | (\mu_1, \dots, \mu_k), (\sigma_1^2, \dots, \sigma_k^2))$$

El espacio paramétrico es

$$S(\psi) = \left\{ (\pi_1, \dots, \pi_k, \mu_1, \dots, \mu_k, \sigma_1^2, \dots, \sigma_k^2) \left| \sum_{j=1}^k \pi_j = 1, \pi_j > 0, \mu_j \in]-\infty, \infty[, \sigma_j > 0, j = 1, \dots, k \right. \right\}$$

Como ya hemos señalado el problema de máxima verosimilitud para mezclas de funciones elípticas nos lleva a un problema de optimización mal condicionado. Por ello los algoritmos usuales de optimización fallan y aparecen componentes degenerados. Sea la clausura

$$\bar{S}(\psi) = \{ \psi \in S(\psi) | \exists j_0 \in \{1, \dots, k\}, t \in \mathbb{N} \text{ tales que } \vec{\mu}_{j_0} = \vec{x}_t, |V_{j_0}| = 0 \}$$

el j_0 -ésimo elemento es denominado componente degenerado.

tomemos una constante positiva c y denotemos por el conjunto $S_c(\psi)$ tal que $c \leq \min_{k \neq j} \frac{\sigma_k^2}{\sigma_j^2}$. Utilizando esta restricción, si la muestra posee al menos $k+1$ puntos diferentes existe un optimizador global de la función de verosimilitud sobre $S_c(\psi)$.

En el caso multivariado la restricción es $c \leq \min_{k \neq j} \frac{|V_k|}{|V_j|}$. Vea una discusión sobre este problema en McLachlan-Peel (2000).

La restricción en términos de los valores propios de la propuesta de Hattaway (1985) se escribe como

$$\min_{k \neq j} \lambda \frac{|V_k|}{|V_j|} \geq c > 0$$

Usándola se puede obtener una formulación de la maximización de la verosimilitud global restringida. Aunque esta restricción es de fácil comprobación en las aplicaciones aun no existen algoritmos suficientemente simples para resolver el problema de optimización. Buscando un algoritmo adecuado es implementable partir de la propuesta de Ingrassia (2004). Considerando que se poseen dos matrices $d \times d$ simétricas y definidas positivas \underline{A} y \underline{B} se tiene que si:

$\lambda_{Max}(\underline{C}) = \text{el mayor valor propio de la matriz}$

$\lambda_{min}(\underline{C}) = \text{el menor valor propio de la matriz}$

Entonces

$$\lambda_{Max}\left(\frac{\underline{A}}{\underline{B}}\right) \leq \frac{\lambda_{Max}(\underline{A})}{\lambda_{Max}(\underline{B})}, \quad \lambda_{min}\left(\frac{\underline{A}}{\underline{B}}\right) \geq \frac{\lambda_{min}(\underline{A})}{\lambda_{min}(\underline{B})}$$

Se sigue que para $a, b > 0$ tales que $\frac{a}{b} \geq c$. Entonces $\forall i = 1, \dots, d$. y $j = 1, \dots, k$ se tiene que $a \leq \lambda_i V_j \leq b$

Este problema de optimización puede ser resuelto usando el paquete tclust de R, vean Fritz et al. (2013).

6.2 El algoritmo Expectation-Maximization (EM)

Los algoritmos del tipo EM son basados en un proceso iterativo que es muy poderoso y que lleva a obtener, deterministamente soluciones a partir de fijar un paso inicial. EM es un algoritmo iterativo que busca la maximización de la función de verosimilitud cuando se cuenta con información parcial. Fue propuesto por Dempster et al (1977). Una discusión amplia puede ser obtenida en McLachlan-Krisnan (2008). El problema más complicado en las aplicaciones es determinar un adecuado punto de partida. La literatura presenta muchos resultados que tratan de derivar estrategias exitosas para inicializar. Las propuestas buscan el evitar la convergencia del algoritmo a soluciones espurias.

Recordando que el j_0 -ésimo elemento de

$$\bar{S}(\psi) = \{\psi \in S(\psi) | \exists j_0 \in \{1, \dots, k\}, t \in \mathbb{N} \text{ tales que } \vec{\mu}_{j_0} = \vec{x}_t, |V_{j_0}| = 0\}$$

es denominado componente degenerado se tiene que la correspondiente solución es una degenerada.

El vector observado es considerado incompleto pues no se poseen los vectores paramétricos. El algoritmo genera secuencias de ψ . Comenzando por uno inicial ψ^0 se

genera un sucesión $\psi^{(r)}$ tal que la sucesión de las funciones de verosimilitud es no decreciente.

Si trabajamos con una mezcla de distribuciones normales en la iteración $r+1$ se computa

$$\tau_{ij}^+ = \frac{\pi_j^-}{\sum_{j=1}^k N(x_i | \vec{\mu}_j^-, \underline{V}_j^-)} N(x_i | \vec{\mu}_j^-, \underline{V}_j^-)$$

$\vec{\mu}_j^-, \underline{V}_j^-$ son los estimados realizados en la iteración r -ésima. El algoritmo se actualiza al calcular

$$\vec{\mu}_j^+ = \frac{\sum_{i=1}^n \tau_{ij}^+ \vec{x}_i}{\sum_{i=1}^n \tau_{ij}^+}, \quad \underline{V}_j^+ = \frac{\sum_{i=1}^n \tau_{ij}^+ (\vec{x}_i - \vec{\mu}_j^+) (\vec{x}_i - \vec{\mu}_j^+)^T}{\sum_{i=1}^n \tau_{ij}^+}$$

Bajo ciertas condiciones de regularidad:

ψ^r converge a un conjunto compacto de máximos locales de la función de verosimilitud.

Ingrassia-Rocci (2007) consideraron el caso en que se cumple $a \leq \lambda_i \underline{V}_j \leq b$ para todo i y j . Tomando $\underline{V}_j = \xi^2 \underline{W}_j$, $j=1, \dots, k$, tal que $\min_{ij} \lambda_i \underline{W}_j = 1$ y son mas débiles las restricciones

$$1 \leq \lambda_i \underline{W}_j \leq \frac{b}{a} \leq \frac{1}{c}$$

Estas restricciones son de fácil implementación usando el algoritmo EM. La r -ésima iteración del algoritmo es computada a partir del estimador de la matriz de covarianza $\underline{V}_j^{(r)}$ cuyos valores propios $\lambda_{ij}^{(r)} = \lambda_i \underline{V}_j^{(r)}$ son tales que satisfagan las restricciones al fijar $\lambda_{ij}^{(r)} = \min\left(b, \max\left(a, l_{ij}^{(r)}\right)\right)$, $l_{ij}^{(r)} = \text{actualizacon de } \lambda_i \underline{V}_j^{(r)} \text{ en el } M - \text{paso}$.

Rocci et al. (2016) generalizaron la restricción $a \leq \lambda_i \underline{V}_j \leq b$ para reforzar la equi-varianza para transformaciones lineales afines. Tomando $\underline{A}_{d \times d}$ no singular y $\vec{b} \in \mathbb{R}^d$. Tomando

$$\vec{x}^+ = \underline{A} \vec{x} + \vec{b}$$

podemos ver que $N(\vec{x} | \vec{\mu}_j, \underline{V}_j) = |\underline{A}| N(\vec{x}^+ | \vec{\mu}^+, \underline{V}^+)$; $\vec{\mu}^+ = \underline{A} \vec{\mu} + \vec{b}$, $\underline{V}^+ = \underline{A} \underline{V} \underline{A}^T$ la verosimilitud de los datos transformados está dada por

$$L(\psi) = n \log |\underline{A}| + \sum_{i=1}^n \log \left(\sum_{j=1}^k \pi_j N(\vec{x}^+ | \vec{\mu}_j^+, \underline{V}_j^+) \right)$$

Fritz et al (2013) establece que el óptimo de los valores propios en el espacio restringido es obtenido en una forma cerrada en cada paso del algoritmo EM. Esto supera los

resultados anteriores que solo obtenían aproximaciones como los de García- Escudero et al. (2008), Ingrassia-Rocci (2007) y Greselin-Ingrassia (2010).

Tomemos los estimadores $\{\psi^{(r)}\}_r$ generada por el algoritmo EM. En la estimación de los parámetros de una mezcla de normales univariadas, y denotamos

$$f_{ij}^{(r)} = \pi_j^{(r)} N(x_i | \mu_j^{(r)}, \sigma_j^{(r)2})$$

Si el componente j_0 está cerca de la unidad degenerada x_{i0} la densidad $f_{i0j_0}^{(r)}$ es alta y muy pequeñas para $f_{ij_0}^{(r)}$ en las otras unidades x_i . La norma de $\vec{v}_0 = \left(\frac{1}{f_{i0j_0}}, \{f_{ij_0}\}_{i \neq i_0} \right)$ es pequeña. En la iteración r denotemos $\vec{v}_0^{(r)}$. Entonces si $\|\vec{v}_0\|$ es suficientemente pequeña se contrae el mapeo generado por el algoritmo EM y hay una convergencia y su punto fijo es degeneración. Además, cerca de los valores menores de los valores propios decrecen con rapidez. Vea Biernacki-Chétien (2003) y Ingrassia-Rocci (2011) para las demostraciones. Estos últimos consideraron adecuado acotar los valores propios entre dos interacciones consecutivas para controlar la velocidad de la variación de los mayores y los menores valores propios en cada iteración. Esto será implementable a partir del algoritmo EM de Ingrassia-Rocci (2007).

References

Andrews J. L., McNicolls P.D.y Subedi S. (2011): Model based classification via mixtures of multivariate t-factor analyzers. Stat. Comp. 21, 361-373.

Banield J. D. y Raftery A. E. (1993): Model based Gaussian and non-Gaussian clustering. Biometrics, 49, 803-821.

Baudry J.P y Celeaux G. (2015): EM for mixtures. Stat. Comp. 22, 1021-1029.

Bienacki C., y Chretien S. (2003): Degeneracy in the maximum likelihood estimation of univariate Gaussian mixtures with EM. T. Prob. Letters. 61, 373-382.

Bougeard S., Abdi H., Saporta G. y Niang N. (2018): Clusterwise analysis of multiblock component methods. Adv. Data Anal Classi. 12, 285-313.

Breunig M.M., Kriegel H-P., Ng R.T. y Sander J. (2000): LOF: identifying density-based estimation. In Chen W. Et al. (eds) ACM SIGMOD Int Conf. On Management of Data, 93-104.

- Brodinová Š., Zaharieva, M., Filzmoser, P., Ortner Th. y Ch. Breiteneder (2018). Clustering of unbalanced high-dimensional media data. *Adv. Data Anal. Classif.* 12, 261-284.
- Carlsson G y Mémoli F. (2014): Hierarchical quasi-clustering methods for asymmetric networks. *Int. Conf. Mach. Learning* 32, 352-360.
- Carlsson G y Mémoli F. Ribeiro A. Segarra S. (2010): Characterization stability and convergence of hierarchical clustering methods. *J. Mach. Learning Res.* 11, 1425-1470.
- Carlsson G y Mémoli F. Ribeiro A. Segarra S. (2018): Hierarchical clustering methods of asymmetric networks. *Adv. Data Anal. Class.* 12, 65-105.
- Ciuperca G. Ridolfi A. y Idier J. (2003): Penalized maximum likelihood estimator for normal mixtures. *Scand. J. Stat.* 30, 45-59.
- Dempster A., Laird N. Y Rubin D. (1977): Maximum likelihood for incomplete data via the EM algorithm. *J. royal Stat. Soc. serie B.* 39, 1-38.
- DeSarbo W. y Cron W. (1988): A maximum likelihood methodology fo clusterwise linear regression. *J. Classif.* 5, 249-282.
- Fraley C. y Raftery A. E. (2006): MCLUST version 4 for R package for normal mixtures modeling model-based clustering, classification and density estimation Univ. Of Washington, Seattle.
- Fraley C., Raftery A. E., Murphy T y Scrucca L. (2012): (2006): MCLUST version 3: An R packge for normal mixtures modeling and model based clustering. TEchnical report
- Fritz H., Garcia-Escudero L. A. y Mayo-Iscar A. (2013): a fast algorithm for robust constrained clustering. *Comp. Sta. Data Anal.* 61- 124-136.
- Gallegos M. T. y Ritter G. (2018): Probabilistic clsutering via Pareto solutions and signifcne tests. *Adv. Data Anal. Calssif.* 12, 179-202.
- García-Escudero L.A., Gordaliza A., Greselin F., Ingrassia S. Y Moyo-Iscar A. (2018): Eigen values and constraints in mixture modeling: geometric and computational issues. *Adv. Data Anal. Clasif.* 12, 203-233.
- Greselin F. e Ingrassia S. (2010): Constrained monotone EM algorithms for mixtures of multivariate distributions. *Stat. Comp.* 20, 9-22.
- Hathaway R. J. (1985): A constrained formulation of máximum-likelihhod for normal mixtures. *Ann. Stat.* 13, 795-800.

- Ingrassia S. (2004): A likelihood based constrained algorithm for multivariate Gaussian mixture models. *Stat. Methods App.* 13, 151-166.
- Ingrassia S. Y Rocci R. (2007): Constrained monotone EM algorithm for finite mixtures of multivariate Gaussians. *Comp. Stat. Data Anal.* 51, 5339-5351.
- Ingrassia S. Y Rocci R. (2011): Degeneracy of the EM algorithm for the MLE of multivariate Gaussian mixtures and dynamic constraints. *Comp. Stat. Data Anal.* 55, 1715-1725.
- Jain A y Dubes R. C. (1988): *Algorithms for clustering Data*. Prentice Hall Inc. N. York.
- Krawczyk B. (2016): Learning from imbalanced data: challenges and future directions. *Prog. Artif. Intell* 21, 1-12.
- McLachlan G. J. y Krisnan. T., (2000): *The EM algorithm and extensions*. Wiley, N. York.
- Lance G. N. y Williams W. T. (1967): A general theory of classificatory sorting strategies 1: hierarchical systems. *Comput. J.*, 9, 373-389.
- McLachlan G. J. y Peel, D. (2000): *Finite mixture models*. Wiley, N. York.
- Preda C. y Saporta G. (2005): Clusterwise PLS regression on a stochastic process. *Comp. Stat. Data Anal.* 49, 99-108.
- Ritter G. (2015): Robust cluster analysis and variable selection. Monog. In *Sta. And Applied Probability*. Vol 137. Chapman & Hall/CRC, Boca de Raton.
- Rocci R., Gattone S. A. Y Di Mari R. (2016): A data driven equivariant approach to constrained Gaussian mixture modeling *Adv. Data Anal. Class.*
- Sardedt M. (2008): A review of recent aproches for capturing heterogeneity in partial least squares path modeling. *J. Model Management* 3, 140-161.
- Scurceca L., Fop M., Murphy T. B. Y Raftery A. (2016): MCLUST 5 clustering, classification and density estimation using Gaussian mixture models. *R Journal*, 8, 289-317.
- Spath (1979): Cluster wise linear regression. *Computing*, 22, 367-373.
- Vicari D. YyVichi M. (2013): Multivariate linear regression for heterogeneous data. *J. App. -sta* 14, 1209-1230.

Vinzi V. Lauro C. Amaato S. et al. (2005): PLS typological regression. In Vinci et al (eds) *New Developments in Classification and Data Analysis*, 133-140. Springer, Berlin

Zimek A., Schubert E. y Kriegel H. P. (2012): A survey of unsupervised outlier detection in high-dimensional numerical data. *Stat. Anal. Data Mining*. 5, 363-387.