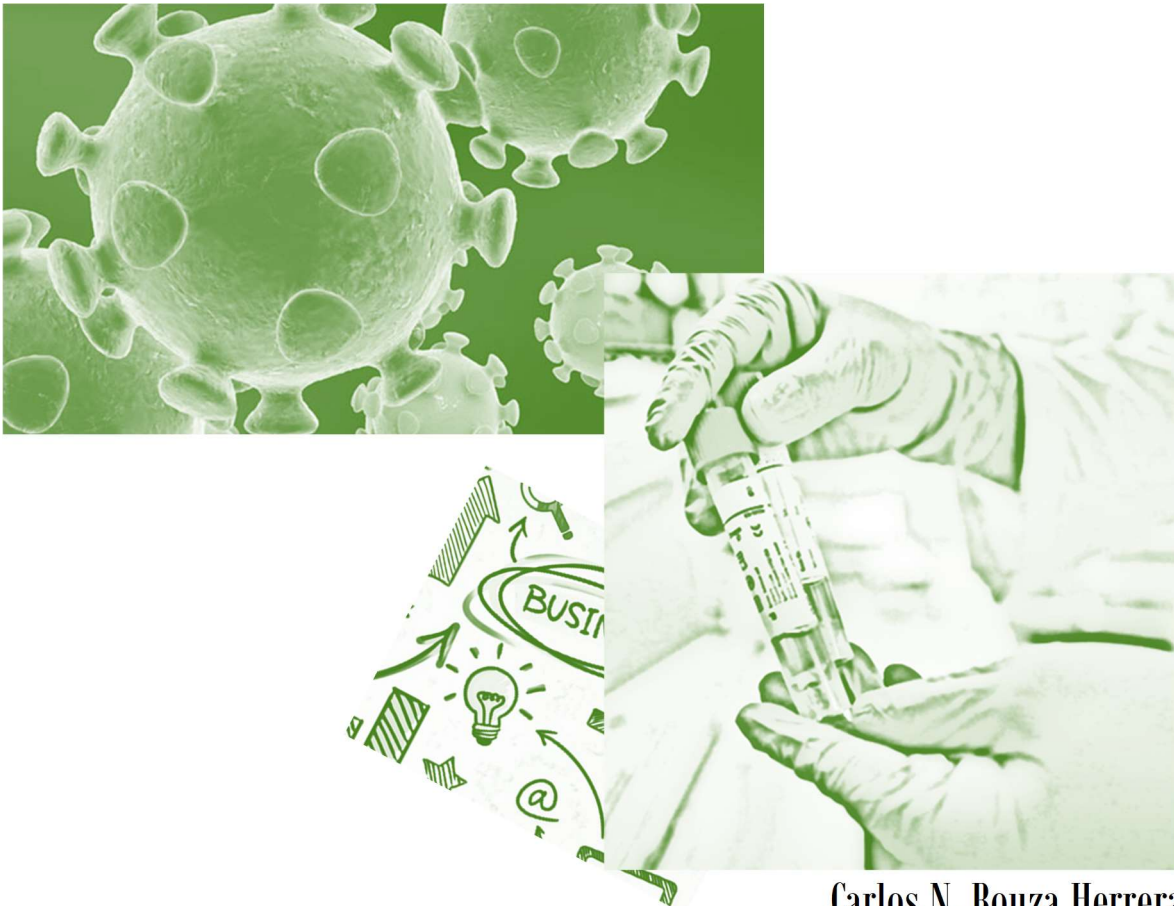


Desarrollo de nuevos modelos y métodos matemáticos para la toma de decisiones

Tomo 1



Carlos N. Bouza Herrera

Agustín Santiago Moreno

María del Mar Rueda García

José Félix García Rodríguez

ISBN:978-607-8755-15-8



9 786078 755158

Desarrollo de nuevos modelos y métodos matemáticos para la toma de decisiones

Tomo 1

Serie Métodos cuantitativos aplicados

Editores:

Carlos N. Bouza Herrera

Agustín Santiago Moreno

María del Mar Rueda García

José Félix García Rodríguez

Primera edición. Marzo del 2021.

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

ISBN: 978-607-8755-15-8

ISBN:978-607-8755-15-8



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, Brasil.

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.

Habana, Cuba, 6 marzo de 2021.

Índice de capítulos

CAPITULO	TÍTULO	AUTORES	PÁGINAS
1	ENTRENAMIENTO DE ESTUDIANTES DE MEDICINA EN LA DISPERSIÓN DE LA COVID-19 CON MODELO ESTOCÁSTICO	José A. Betancourt Bethencourt	1-7
2	EL MUESTREO DE GRAFOS EN LA MODELACIÓN DE EPIDEMIAS: EL COVID 19	Carlos N. Bouza, Sira M. Allende, Nilda E. Ponce	9-20
3	SOBRE EL TAMAÑO DE MUESTRA PARA UN ESTUDIO DE LA PREVALENCIA DE ENFERMEDADES CORONARIAS	Agustín Santiago Moreno, Carlos Bouza Herrera, José Maclovio Sautto Vallejo, Nurys Bárbara Armas Rojas y Yanela Yordanka	21-29
4	ESTIMACIÓN DEL BALANCE BENEFICIO/RIESGO EN ENSAYOS CLÍNICOS	Carmen Viada, Carlos Bouza, Martha Fors, Mabel Alvarez, Aliuska Frias, Lázara García, Yanela Santiesteban, Yuliannis Santiesteban, Mayelin Troche, Liana Martínez.	31-45
5	UNA REVISIÓN DE MÉTODOS PARA EL CLUSTERING ESPACIAL EN POBLACIONES CON ESTRUCTURA COMPLEJA: ALGORITMOS BASADOS EN LA CONECTIVIDAD Y EN LA FUNCIÓN DE DENSIDAD.	Sira M. Allende Alonso y Carlos N. Bouza Herrera	47-65
6	ANÁLISIS DE FACTORES DE RELACIONADOS CON LA EVOLUCIÓN CLÍNICA DE MUJERES CON DISFUNCIONES DEL SUELO PÉLVICO	R.S. Borges Sandrino, V. Sistachs Vega, E. Ramón Musibay y M. A. Díaz Martínez	67-79
7	CRIMES IN COLOMBIA (2005-2018): INFERRING CRIME RATE RELATIONSHIPS AMONG GEOGRAPHICAL STATES	Cesar Augusto Vargas-García, Jenny Paola Lis-Gutiérrez, Mercedes Gaitán-Angulo, Linda Carolina Henao, Melissa Lis-Gutiérrez, Henry Zapata Patarroyo	81-94
8	COLOMBIA EN EL COMERCIO DE CUERO, CALZADO Y MARROQUINERÍA CON CHILE: VENTAJAS COMPARATIVAS Y SUS DETERMINANTES	Jenny Paola Danna-Buitrago, Rémi Stellian, Carolina Henao, Ana Fernández, M. Andreina Moros Ochoa	95-115
9	GESTIÓN DEL CONOCIMIENTO COMO DETERMINANTE DE LA CAPACIDAD DE INNOVACIÓN Y VENTAJA COMPETITIVA DE LAS ORGANIZACIONES	José Félix García Rodríguez, Georgina Guadalupe Shriner Sierra, David Martínez Luis e Ignacio Caamal Cauich	117-127
10	FORMACIÓN DEL PRECIO DE EXPORTACIÓN DEL LIMÓN PERSA DEL MUNICIPIO DE MISANTLA, ESTADO DE VERACRUZ, MÉXICO	Ignacio Caamal Cauich, Verna Grisel Pat Fernández, José Félix García Rodríguez, José Antonio Ávila Dorantes y Alberto Pérez Fernández	129-138

11	EVOLUCIÓN DE LAS TIC, DESARROLLO DE LA INDUSTRIA 4.0 Y SU APLICACIÓN EN EL SECTOR TURÍSTICO	M. Andreina Moros Ochoa, Paula Viviana Robayo Acuña y Gilmer Yovani Castro Nieto	139-155
----	--	--	---------

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)
M. Andreina Moros Ochoa	Colegio de Estudios Superiores de Administración (CESA)
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
José A. Betancourt Bethencourt	Universidad de Ciencias Médicas Carlos J. Finlay, Cuba
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 Cauich	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
Nilda E. Ponce	Ministerio de Salud Pública Municipio Cotorro, La Habana, Cuba
Verna Gricel Pat Fernández	Universidad Autónoma Chapingo (UACH).
Gilmer Yovani Castro Nieto	Pontificia Universidad Javeriana (Bogotá, Colombia)

Prólogo

La innovación tecnológica-científica al incidir en el ámbito educativo, enfrenta nuevos retos y desafíos, sobre todo ante los cambios en el tema del conocimiento, cada vez más integrados inter y transdisciplinariamente. El reto que nos imponen el surgimiento de nuevas enfermedades, la aparición de nuevos enfoques en la economía, inseguridad, violencia, desarrollo organizacional y la evolución acelerada de las tecnologías de la información y comunicación, nos obliga a dar respuestas más dinámicas para hacer frente a estos fenómenos.

Retos así, llevan a las instituciones de educación superior a fortalecer sus políticas educativas mediante un plan de desarrollo institucional, además de integrar los cambios en la estructura organizacional para poner a punto el quehacer académico ante estos desafíos. Un punto neurálgico para hacer frente a todo ello es precisamente la capacidad de profesionalización de los principales protagonistas de la transformación de la educación: los docentes-investigadores.

Esta obra representa de una manera concisa, cómo las disciplinas aportan desde diferentes focos de atención resultados metodológicos aplicados al contexto social, que mediante modelos matemáticos generan nuevos conocimientos y oportunidades de solución orientados a la toma de decisiones.

En la naturaleza de los eventos de la vida diaria surgen acontecimientos que pueden ser medidos; tales acontecimientos traen consigo un razonamiento lógico que permite detectar variables que pueden ser sometidas a un proceso de medición que contribuya a lograr un proceso planificado. Además de lo anterior, la importancia de los modelos y métodos matemáticos radica en ser una herramienta sustancial para controlar el riesgo. La generosidad de estos modelos es que funciona como un eje transversal en todas las áreas del conocimiento, asumiendo un proceso sistémico, que, si bien resulta un poco complejo cuando se encuentran los elementos que se interrelacionan o interactúan entre sí, se logra una confirmación metrológica y el control continuo de los procesos de medición del acontecimiento del que trate.

Sin lugar a dudas, el contenido de esta obra dará al lector una visión extraordinaria de la transdisciplina como una herramienta de solución a los problemas sociales mediante modelos matemáticos en donde las diferentes áreas de conocimiento aportan una orientación investigativa que aborda el objeto de estudio desde distintos focos de atención.

*M.C. Margarito Radilla Romero
Jefe del área de capacidad y desempeño de la
Dirección General de Planeación de la UAGro*

Capítulo 1

pp 1-7

ENTRENAMIENTO DE ESTUDIANTES DE MEDICINA EN LA DISPERSIÓN DE LA COVID-19 CON MODELO ESTOCÁSTICO TRAINING MEDICINE STUDENTS IN COVID-19 SPREAD USING STOCHASTIC MODEL

José A. Betancourt Bethencourt*

* Universidad de Ciencias Médicas Carlos J. Finlay, Cuba

RESUMEN

En Cuba, a nivel nacional se formulan de manera disciplinaria excelentes modelos matemáticos, pero estos no están contemplados para su uso a nivel provincial ni hay superación de posgrado de estos temas en salud en la Universidad de Ciencias Médicas de Camagüey. Los estudiantes de medicina deben conocer mediante simulaciones el efecto de diferentes intervenciones durante la pandemia de COVID-19. En este trabajo se usó un modelo estocástico del programa R que les permitió a los estudiantes visualizar el efecto del uso inadecuado y adecuado de medidas de higiene (lavado de manos y de superficies y el uso del nasobuco) así como el efecto del distanciamiento social o excesiva interacción.

PALABRAS CLAVE: pandemia, COVID-19, simulación, programa R, modelos estocásticos, transdisciplina

ABSTRACT

There are in Cuba excellent mathematical models to analyze COVID-19, unfortunately these models are not include in medicine or professional teaching in the University of Medical Sciences of Camagüey. The medicine students should know by means of simulations the effect of different interventions during the pandemic of COVID-19. In this work a stochastic model of the program R was used that allowed the students to visualize the effect of the inadequate and appropriate use of measures of hygiene (washing of hands and of surfaces and the use protection for mouth and nose) as well as the effect of the social distancing or excessive interaction.

KEY WORDS: pandemics, COVID-19, simulation, R program, stochastic models, trans discipline

INTRODUCCIÓN

La dinámica actual conlleva a aplicar teorías transdisciplinarias, por tanto, las universidades tienen el reto de liderar el proceso de transformación, introduciendo cambios en los procesos formativos que permitan a los futuros profesionales en salud pública, encontrar en la transdisciplinariedad, una herramienta para fortalecer la investigación y el diálogo de saberes.(Castilla et al., 2018)

Dado el impacto de la COVID-19, se manifiestan cambios en el uso de la Inteligencia Artificial, nuevas maneras de aprendizaje a distancia, consultas médicas a distancia y otros, lo que demuestra que para lograr resiliencia hay que fortalecerse en estos métodos adaptativos que requieren enfoques integradores.(Nuno Martins, 2020)

La actual pandemia de COVID-19 es mayormente un problema social, esta situación no solo es de la medicina, sino que afecta a la sociedad de manera holística, por lo que al investigarla debe hacerse con visión integradora.(Moradian, 2020, Laupacis, 2020) que genere nuevos conocimientos, que faciliten la integración de saberes y que produzcan impacto social (Smetschka and Gaube, 2020)

Europa ha definido financiamientos para el desarrollo de nuevos métodos analíticos que permitan estimar parámetros claves de los datos de contactos y serológicos, y la creación de bases de datos adecuadas que se coleccionan de manera sistemática.(CORDIS, 2020). Muchos académicos consideran que se deben incrementar los métodos de adquisición de diversas habilidades on line (Sandhu and Wolf, 2020, Ferrel and Ryan, 2020)

En Cuba a nivel nacional se formulan de manera disciplinaria excelentes modelos matemáticos (Reinaldo et al., 2020), estos no están contemplados para su uso a nivel provincial ni hay superación de posgrado de estos temas en salud en la Universidad de Ciencias Médicas de Camagüey. Por lo que se prepara el presente entrenamiento para estudiantes de pregrado y de posgrado de la localidad.

MÉTODO

Se realizaron dos simulaciones con el paquete EpiModel (Jenness et al., 2016) presentes en el programa R (R_Core_Team., 2020). Se calculó el número reproductivo básico (R_0) con el paquete earlyR (Jombart et al., 2017). La población de la comunidad Bellavista en Camagüey, donde se hizo la simulación es de 4000 habitantes.

RESULTADOS

En la primera simulación (figura 1) el modelo contempla un índice de transmisibilidad del 5%, lo que denota fallas en las medidas de prevención en cuanto a la higiene (lavado de manos y de superficies y el uso del nasobuco) así como el efecto de poco distanciamiento social pues hay un promedio de 20 contactos diarios por persona.

Con esta primera simulación, a los 15 días de mantenerse la situación sin cambios se afectaría un 70 % de la población (2800 personas) con la sola entrada de un enfermo. Se comprueba que es un brote epidémico (figura 2) con un R_0 de 1.421

En la segunda simulación (figura 3) el modelo contempla un índice de transmisibilidad del 1%, lo que denota mejorasen las medidas de prevención en cuanto a la higiene (lavado de manos y de superficies y el uso del nasobuco) así como el efecto de mayor distanciamiento social pues se redujo el promedio a 10 contactos diarios por persona, ya aquí $R_0 < 1$ por lo que se evitó el brote epidémico. Solo se mantendrían casos endémicos perfectamente controlables por las autoridades de salud.

DISCUSIÓN

En un trabajo realizado en Camagüey se encontró que se realizaban por parte de las personas 18.8 contactos estrechos por día (Bethencourt Bethencourt et al., 2011) típico de la idiosincrasia del cubano que favorece en general la transmisión de enfermedades de transmisión aérea. En China se demostró que solamente con el distanciamiento social durante la pandemia fue suficiente para el control de la COVID-19. (Zhanget, 2020)

Con un modelo estocástico de transmisión estos autores demostraron que en la mayoría de los escenarios el rastreo de contactos y el aislamiento fueron suficientes para controlar la enfermedad en un término de tres meses. (Joel Hellewell et al., 2020)

Con un modelo estocástico de transmisión, estos autores demostraron que en la mayoría de los escenarios el rastreo de contactos y el aislamiento fueron suficientes para controlar la enfermedad en un término de tres meses. (Joel Hellewell et al., 2020)

Para el futuro un grupo de investigadores sugieren incluir en los modelos la transmisibilidad de individuos asintomáticos y considerar subpoblaciones relacionadas con la edad género, etc., introducir medidas preventivas en los modelos en esta COVID-19 y en virus futuros, incluir en los modelos algunos datos imprecisos usando ecuaciones diferenciales fuzzy y finalmente incluir las cargas virales. (Ndairou et al., 2020). Ya hay modelos que contemplan infecciones no detectadas(Ivorra et al., 2020)

En esta experiencia comprobaron la reducción del número reproductivo básico con la restricción de viajes.(Kucharski et al., 2020) , En una revisión realizada se encontró que la media de R_0 para COVID-19 está alrededor de 3.28 con una mediana de 2.79(Liu et al., 2020)

El hecho es que se necesita el estilo de educación transdisciplinaria para resolver diferentes problemas que se presentan en la vida real (Belcher et al., 2016, Méndez, 2015, Koch et al., 2015, Betancourt Bethencourt et al., 2014)

CONCLUSIONS

Se constata de manera visual por parte de los estudiantes el efecto de las intervenciones.

Es una herramienta analítica sustentable basada en programa libre.

Se demuestra la necesidad de integrar la epidemiología y la matemática.

ANEXOS

Figura 1. Posible escenario de dispersión en el que se mantienen sin modificar las medidas higiénicas y de distanciamiento social.

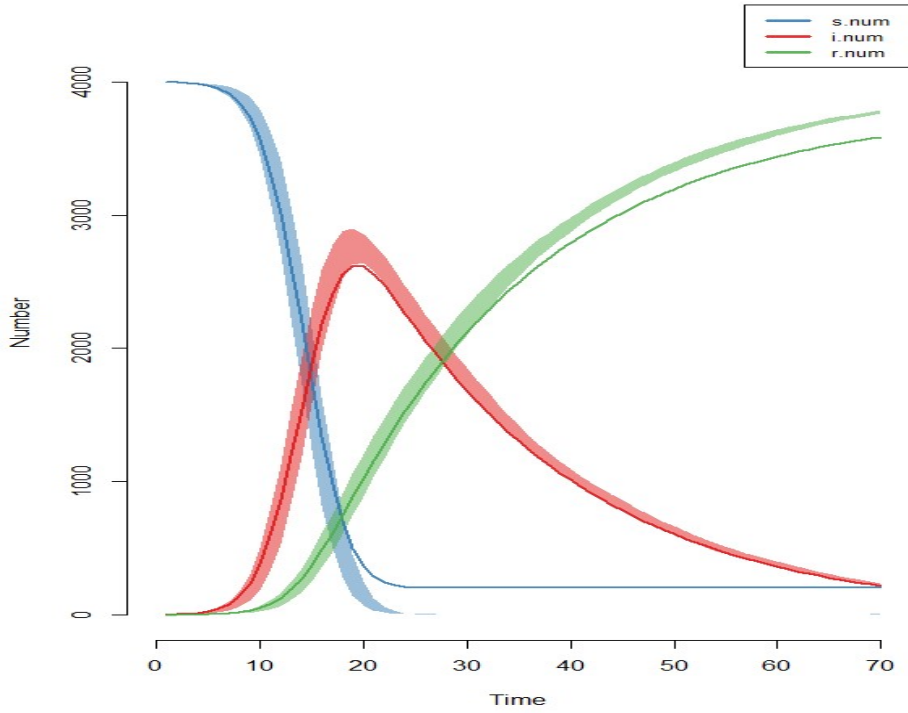


Figura 2. Estimación del número reproductivo básico en ese brote epidémico

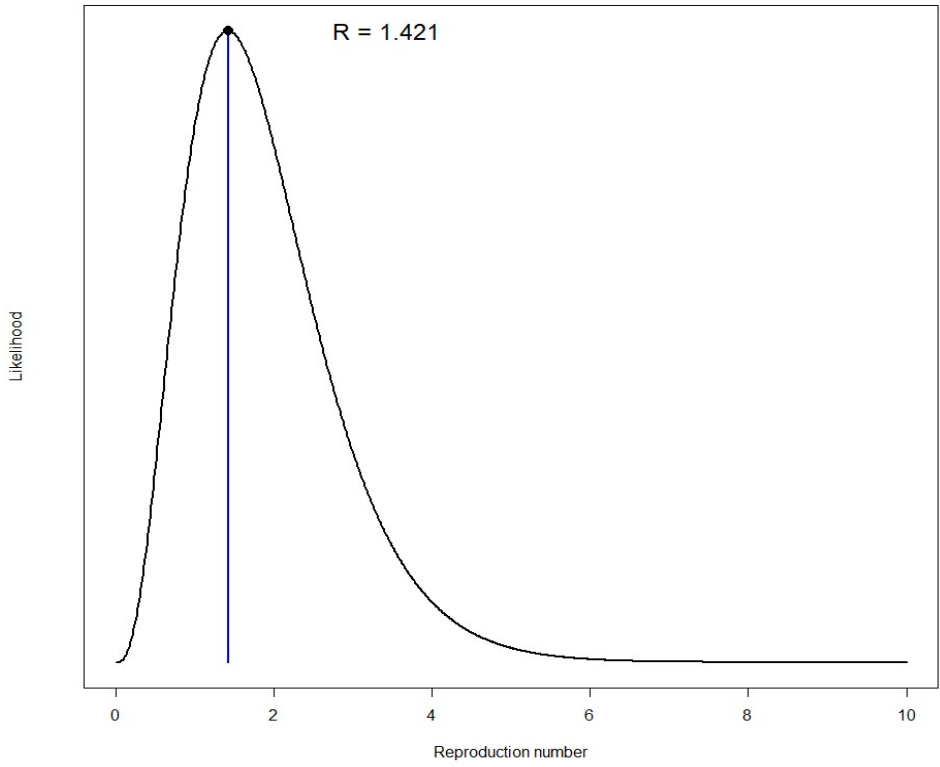
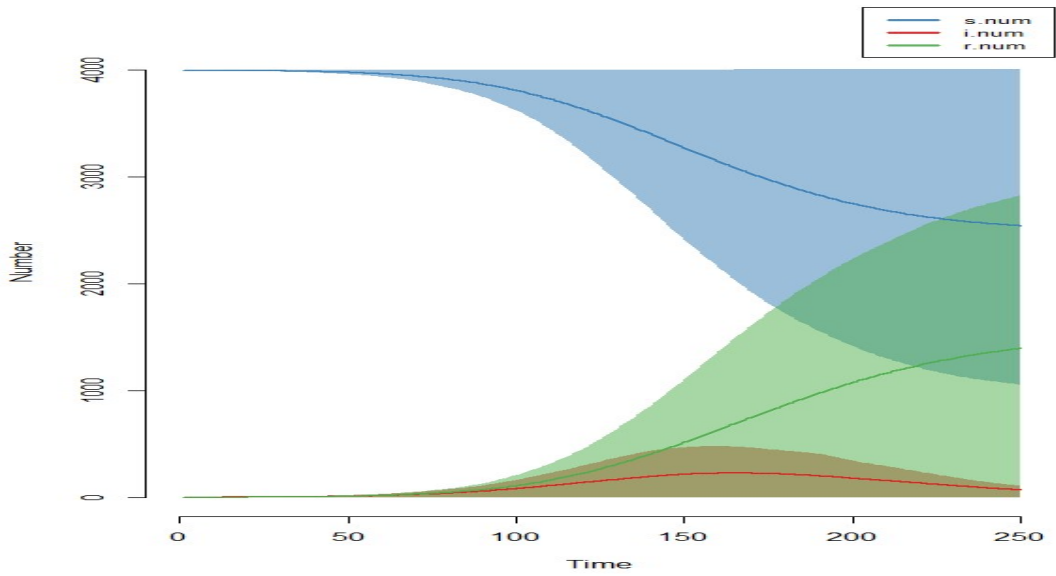


Figura 3. Posible escenario de dispersión en el que se modifican positivamente las medidas higiénicas y se eleva el distanciamiento social.



REFERENCIAS

- BELCHER, M., RASMUSSEN, K., KEMSHAW, M. & ZORNES, D. 2016. Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, 25, 1-17.
- BETANCOURT BETHENCOURT, J., MIRABAL NÁPOLES, M. & ACO FRANCOISE, L. 2014. Hacia la investigación transdisciplinaria mediante el aprendizaje cooperativo. *Revista Electrónica Educare*, [Online], 18. Available: <http://www.revistas.una.ac.cr/index.php/EDUCARE/issue/current>.
- BETANCOURT BETHENCOURT, J., GONZÁLEZ VAZQUEZ, R. & ACO FRANCOISE, L. 2011. Determinación de patrones de contactos para enfermedades de transmisión érea en comunidades de Camagüey. *Rev. Cubana Hig Epidemiol.* [Online], 49. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-30032011000100002.
- CASTILLA, M., M. G. & VILLADIEGO, E. 2018. Salud Pública: Un campo de confrontación del paradigma disciplinar y transdisciplinar. *Avances en Salud*, 2, 48 - 57.
- CORDIS. 2020. Translational and Transdisciplinary research in Modeling Infectious Diseases. Available: <https://cordis.europa.eu/project/id/682540>.
- FERREL, M. & RYAN, J. 2020. The Impact of COVID-19 on Medical Education. *Cureus Cureus*; DOI 10.7759/cureus.7492 [Online], 12. Available: <https://www.tandfonline.com/doi/pdf/10.1080/10872981.2020.1764740>.
- IVORRA, B., FERRÁNDEZ, M. R., VELA-PÉREZ, M. & RAMOS, A. M. 2020. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun Nonlinear Sci Numer Simulat* 88 (2020) 105303 [Online]. Available: <https://doi.org/10.1016/j.cnsns.2020.105303>.
- JENNESS, S., GOODREAU, S., WANG, L. & MORRIS, M. 2016. EpiModel: Mathematical Modeling of Infectious Disease. The Statnet Projec. Available: <http://CRAN.R-project.org/package=EpiModel>.
- JOEL HELLEWELL, ABBOTT, S., GIMMA, A., BOSSE, N. I., JARVIS, C. I., RUSSELL, T. W., MUNDAY, J. D., ADAM J KUCHARSKI, W JOHN EDMUNDS & EGGO†, R. M. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* [Online], 8. Available: [https://doi.org/10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7)
- JOMBART, T., CORI, A. & NOUVELLET, P. 2017. earlyR: Estimation of Transmissibility in the Early Stages of a Disease Outbreak. R package version 0.0.1. Available: <https://CRAN.R-project.org/package=earlyR>.
- KOCH, T., OLIVA, I., VALENCIA, S. & SALAZAR, J. P. 2015. Prospectiva Transdisciplinaria en Educación Superior: Construcción de Perfiles Vocacionales según Grados de Flexibilidad Disciplinar en Estudiantes de Último Año de Nivel Secundario. *Formación universitaria*, 8, 65-78.
- KUCHARSKI, A. J., RUSSELL, T. W., DIAMOND, C., YANG LIU, EDMUNDS, J., SEBASTIAN FUNK & EGGO, R. M. 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect* [Online], 20. Available: [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).
- LAUPACIS, A. 2020. Working together to contain and manage COVID-19. *CMAJ* 2020 March 30;192:E340-1. doi: 10.1503/cmaj.200428; [Online]. Available: <https://www.cmaj.ca/content/cmaj/192/13/E340.full.pdf>.
- LIU, Y., GAYLE, A. A., WILDER-SMITH, A. & ROCKLÖV, J. 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, doi: 10.1093/jtm/taaa021 [Online]. Available: <https://academic.oup.com/jtm/article-abstract/27/2/taaa021/5735319>
- MÉNDEZ, F. 2015. Transdiscipline and research in health: Science, society and decision making. *Transdisciplina y la investigación en salud: Ciencia, sociedad y toma de decisiones.*, 46, 126-132.
- MORADIAN, N. 2020. The urgent need for integrated science to fight COVID-19 pandemic and beyond. *J Transl Med* (2020) 18:205 [Online]. Available: <https://doi.org/10.1186/s12967-020-02364-2>.
- NDAÏROU, F., AREA, I., NIETO, J. & TORRES, D. M. 2020. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons and Fractals* [Online], 135. Available: <https://doi.org/10.1016/j.chaos.2020.109846>.
- NUNO MARTINS, P. 2020. A transdisciplinary perspective of the current crisis situation in the world. *International Journal of Multidisciplinary Research (IJMR)* [Online], 6. Available: <https://doi.org/10.36713/epra4404>.

- R_CORE_TEAM. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. [Online]. Viena, Austria. Available: <http://www.r-project.org>.
- REINALDO, O. F., ABEL PADRÓN PADILLA & TAMAYO, E. C. 2020. COVID-19 en Cuba: El pico se adelanta, ¿qué dicen los modelos matemáticos y cómo interpretarlos? Available: http://www.cubadebate.cu/especiales/2020/04/27/covid-19-en-cuba-el-pico-se-adelanta-que-dicen-los-modelos-matematicos-y-como-interpretarlos/#.XtUSqK_B-ic.
- SANDHU, P. & WOLF, M. D. 2020. The impact of COVID-19 on the undergraduate medical curriculum. *MEDICAL EDUCATION ONLINE* [Online], 25. Available: <https://doi.org/10.1080/10872981.2020.1764740>; <https://www.tandfonline.com/doi/pdf/10.1080/10872981.2020.1764740>.
- SMETSCHKA, B. & GAUBE, V. 2020. Co-creating formalized models: Participatory modelling as method and process in transdisciplinary research and its impact potentials. *Environmental Science and Policy* 103, 41–49.
- ZHANGET, J. 2020. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 10.1126/science.abb8001 (2020). [Online]. Available: <https://science.sciencemag.org/content/sci/early/2020/05/04/science.abb8001.full.pdf>.

Capítulo 2

pp 9-14

EL MUESTREO DE GRAFOS EN LA MODELACIÓN DE EPIDEMIAS: EL COVID

19

Carlos N. Bouza¹, Sira M. Allende¹, Nilda E. Ponce²

¹MATCOM, Universidad de La Habana. La Habana, Cuba

²Ministerio de Salud Pública Municipio Cotorro, La Habana, Cuba

RESUMEN:

Se presentan modelos que capturan la dinámica de los contagios entre los individuos en la población estudiada en una epidemia. Los métodos de muestreo discutidos toman en cuenta la naturaleza de los procesos de contactos para fijar los procedimientos de selección. Se desarrollan métodos para seleccionar una muestra usando una estructura de grafos.

PALABRAS CLAVE: métodos de muestreo, procedimientos de selección, estructura de grafos.

ABSTRACT:

Models that capture the dynamics of the contagions among individuals, in the studied population in an epidemic, are presented. The sampling methods are discussed, taking into account the nature of the contact processes, for fixing selection procedures. The use of a graph structure for sampling are developed.

Key words: sampling methods, selection procedures graph structure.

1. INTRODUCCIÓN

Está demostrado que difusión del Coronavirus COVID-19 se lleva a cabo fundamentalmente a través del contacto social. Para valorar las potencialidades de la difusión a gran escala del contagio, es necesario tratar de establecer la dinámica de este fenómeno, para buscar mitigar esta. Hacer modelos que describan la contaminación en forma adecuada es de importancia central para establecer políticas. Existen muchos trabajos sobre la influenza, ver Ferguson et al. (2020). Este nuevo virus tiene un comportamiento muy diferente de sus parientes, por ejemplo, la tasa de mortalidad respecto a la edad. Se recomienda a los interesados consultar Tang et al (2020), Kucharski et al (2020), Zhigljavsky et al. (2020).

La Teoría de Grafos provee herramientas para modelar con flexibilidad el fenómeno de la difusión del virus. Su uso permite analizar diferentes escenarios y caracterizar el impacto de la implementación de políticas de control del COVID-19. En este trabajo vamos a presentar elementos para el desarrollo de encuestas en poblaciones donde los contactos son modelados usando un enfoque basado en la Teoría de Grafos. Se describe como se enmarcan poblaciones tomando en cuenta la naturaleza de los procesos de contactos. Los modelos que se presentan capturan la dinámica de los contagios entre los individuos en la población estudiada. Los métodos de muestreo discutidos toman en cuenta la naturaleza de los procesos de contactos para fijar los procedimientos de selección. El objetivo de seleccionar una muestra usando una estructura de grafos es permitir construir

modelos decisionales confiables e interpretables para describir la epidemia y evaluar el impacto de diferentes políticas de control.

La necesidad de muestrear grafos ha sido históricamente fuente de diversos modelos teóricos particulares. Vean antecedentes de enfoques, modelos, métodos y experiencias en Birnbaum-Sirken (1965), Frank (1977, 1980a, 1980b, 2011), Sirken (2005) y Goodman (1961).

El grafo obtenido partiendo de muestras se permite condensar la información de grafo poblacional, permitiendo diseccionar mejor el fenómeno que se modela. El problema general del muestreo de grafos es obtener un grafo lo más parecido posible al poblacional. En este trabajo se presentan dos estimadores que siempre serán adecuados para la estimación de parámetros desconocidos, una vez determinados por el decisor. Se ilustra el comportamiento de estos considerando la estimación del total. Los estimadores presentados son muy robustos, en términos de las propiedades estadísticas más destacadas, respecto a cambios en el diseño muestral.

Debido a la poca información confiable existente y los problemas que presenta el manejo de la data de los repositorios hace que la data utilizable no permita más que hacer ilustraciones del comportamiento de los métodos presentados.

En la sección 2 se discute como el interés establecer la existencia de relaciones entre los individuos de la población lleva a usar como modelo un grafo y como en este contexto la teoría de la percolación permite estudiar la expansión de la enfermedad a largo plazo al examinar la mezcla de redes lo que ayuda en la descripción de lo que puede ocurrir en una epidemia. En la sección se presenta un marco general para la selección de la muestra y se discuten aspectos básicos de los diseños muestrales. Son discutidos aspectos particulares que aparecen al muestrear grafos y se discuten los basamentos de ciertos procedimientos de muestreo adecuados. Finalmente se trata del muestreo de grafos que admiten usar una representación bipartita respecto a la incidencia de contactos en una población. Diseños y procedimientos de muestreo bien conocidos son caracterizados en este contexto, lo que permite unificar el tratamiento de la selección de muestras en grafos.

2. ALGUNOS ELEMENTOS DE LA UTILIDAD DE GRAFOS EN LA MODELACIÓN DE EPIDEMIAS

La mayor parte de las epidemias comienzan a partir de algunos infestados los que transmiten la infección a otros miembros de la población. La experiencia del SARS ha generado diversos estudios pos-epidemia para establecer el origen de la infección. Los interesados pueden consultar Haydon et al. (2003); Riley et al. (2003).

Está documentado que en el brote de SARS que tomó lugar en el periodo 2002–2003 esta se extendió a una velocidad menor que lo esperado. Por su parte la epidemia del COVID19 que se enfrenta se ha expandido muy rápidamente.

Al considerar los grafos debe tenerse en cuenta la diferencia en la denominación de los mismos elementos. En epidemiología se identifican huéspedes y contactos que es lo que en la sociometría de denominan actores y relaciones. Estos no son sino los nodos y aristas que se definen en el contexto de la Teoría de Grafos (TG). Usando las herramientas de esta teoría podremos identificar y estudiar problemas como los asociados las vecindades de los individuos de interés, como los portadores, a partir de observar las aristas que les conectan con otros. El tamaño de ellas es lo que conoce como en TG como grados del nodo, el nivel de aglomeración que tienen se puede caracterizar cuantitativamente etc.

En las Ciencias Sociales el interés se centra generalmente en el estudio del fenómeno que genera las conexiones. Otras propiedades de la estructura generada (grafo) no les son importantes. Estos estudios son de carácter exploratorio y buscan describir cómo evolucionan las redes sociales. Vea los trabajos pioneros en esta área debidos a Leinhardt (1977), Scott (1991), Wasserman-Faust (1994) y Karlberg (1997). La naturaleza de las conexiones como reflexividad, simetría, transitividad permiten establecer patrones para la circulación de información y de la cohesión de grupos sociales. También son de su interés establecer liderazgos (dominancia de la red por algunos nodos). En estudios de la transmisión de infecciones el problema es raigalmente similar. Desde la epidemiología tendremos que un líder será un súper-contaminador, la circulación del contagio la virulencia de la infección etc.

Así que la experiencia de la modelación mediante grafos proveniente de las Ciencias Sociales provee de un marco de trabajo útil para la epidemiología. En los estudios sociales el objetivo es conocer la red, mientras que epidemiología lo que se busca es como se disemina la enfermedad. Construir un grafo permitiría tener una base para estudiar las dinámicas de la transmisión.

Veremos cómo elementales herramientas de la TG proveen de métodos cuantitativos poderosos que, provenientes de la descripción de redes, permiten caracterizar el comportamiento de epidemias.

La matriz de incidencias está definida como

$$\underline{A} = [a_{ij}]_{N \times N}; a_{ij} = \begin{cases} 1 & \text{si } i \text{ se relaciona con } j \\ 0 & \text{otro caso} \end{cases}$$

Esta en las Ciencias Sociales es conocida como ‘socio-matriz’. En epidemiología diremos que $a_{ij} = 1$ identifica que la conexión entre dos individuos y que i pasa la infección a j . Así \underline{A} resume el estado de las conexiones en un grupo de personas estudiadas. Por ejemplo, en el SIDA si “ i ” tiene el virus y tiene contacto sexual con “ j ” $a_{ij} = 1$ indica que hay un posible contagio, analizando los a_{jk} ’s podremos establecer la red de transmisión con raíz en “ i ”. Esta matriz brinda información por sí misma y operaciones sencillas dan medidas cuantitativas útiles. Por ejemplo

$$\bar{n} = \frac{1}{N} \sum_{i,j} a_{ij} = \frac{\|\underline{A}\|}{N} = \frac{1}{N} \text{traza}(\underline{A}^2)$$

es el número promedio de contactos por individuo en la población. La matriz \underline{A}^m contiene información sobre los caminos en una red de largo m . Otras funciones de ella brindan información sobre la transitividad o la medida de aglomeración. Tome, por ejemplo

$$\phi = \frac{\text{traza}(\underline{A}^3)}{\|\underline{A}^2\| - \text{traza}(\underline{A}^2)}$$

Esta es la razón del número de triángulos dentro de la red de los tríos conectados (triadas). Mayores valores de este implican que hay una mayor aglomeración de estas. Otro uso de la matriz de adyacencia es el de conectividad. Se dice que la red está conectada si se puede acceder a todo individuo (nodo) desde cualquier otro. En términos epidemiológicos esto equivaldrá a que la infección puede afectar a toda la población a partir de un solo infectado. Se identifica esto al observar que

$$\lim_{m \rightarrow \infty} (1 + A)^m = \sum_{\infty < m=1} A^m$$

no tiene términos iguales cero. Tener ceros en estas matrices identifican que la red tiene componentes separados.

No es superfluo apuntar que cada epidemia tiene características diferentes que hacen que cada infección tenga sus especificidades que hacen necesario diseñar una red para modelar cada una. Esto fija que se planteen nuevos problemas para hacer los protocolos de medición y almacenamiento de datos.

En los estudios epidemiológicos es de interés establecer la existencia de relaciones entre los individuos de la población en términos de la motivación de estos. Esto lleva a usar no solo moverse a través del grafo sino ponderar usando alguna variable que responda a la motivación. Al considerar que las conexiones deben ser ponderadas en términos de una variable aparecen nuevos problemas con la obtención y archivo de la data y los modelos son más complejos. Se requiere de un planeamiento riguroso de la toma de datos y su procesamiento desde las etapas iniciales de la epidemia. Brotes anteriores permiten establecer estos aspectos burocráticos al enfrentarse un rebrote. Recomendamos a los interesados ampliar en los antecedentes de tal enfoque en Grimmett (1989), Müller et al (2000).

La teoría de la percolación lleva a modelos de redes mezcladas que permiten estudiar la expansión de la enfermedad a largo plazo. En percolación se examina como se forman estructuras conectadas dentro de las redes obteniéndose lattices cuadrados (sitios vecinos se conectan aleatoriamente con una cierta probabilidad p). Si p es suficientemente grande es posible hallar un camino de un extremo del lattice a otro. Es claro que, si p es la probabilidad de transmisión, las aristas dentro del lattice pueden representar los eventos de transmisión. Vea Newman et al. (1999) Al obtener el tamaño de los clústeres conectados que genera un modelo de percolación se puede relacionar este con el tamaño que se espera de la epidemia en la red. Si los individuos se sitúan en los puntos del lattice con una probabilidad p obtenemos una representación de la epidemia en una población con susceptibles. La probabilidad p juega un papel significativo en la descripción de lo que puede ocurrir en una epidemia.

Consideremos una población de individuos $U = \{u_1, \dots, u_N\}$. Cada u_i es un elemento que identificamos con un nodo. Las conexiones entre los individuos determinan un conjunto de pares

$$a_{ij} = \begin{cases} 1 & \text{si } i \text{ y } j \text{ se conectan} \\ 0 & \text{en otro caso} \end{cases} \quad i, j \in U; \underline{A} = [a_{ij}]$$

$G = (U, A)$ es el grafo de la población. Cuando $a_{ij} = a_{ji}$ se dice que el grafo es no dirigido. Por ejemplo pertenecer al mismo núcleo familiar. Si puede ser que $a_{ij} \neq a_{ji}$ tenemos un grafo dirigido. Tal es el caso cuando evaluamos los contactos entre una persona infestada y otros de su entorno.

El interés de los epidemiólogos es medir diversas variables. En este contexto consideremos que esta variable está motivada por desear evaluar un evento, como es al desear evaluar el riesgo de muerte. Sin perder en generalidad identificaremos los términos: variable = motivo. Por ejemplo, el nivel de riesgo de una persona generará la necesidad de hacer una medición motivada por este interés. Tomemos $\chi = \chi(G)$ como el conjunto de todos los motivos en G . Una vez determinado un motivo $\chi(k) \in \chi = \chi(G)$ es de interés determinar los individuos que se asocian a este: M_k con cardinalidad $|M_k|$.

Ejemplos. Si G es no-dirigido

1. Nos interesa identificar tríos de individuos ($a_{ij} = a_{ih} = a_{jh}$), o sea que $\prod_{i \neq j \in M_k} a_{ij} = 1$, lo que define $\chi(k)$ y $|M_k| = 3$. Esto nos permite estudiar la problemática en una familia de 3 personas.

2. Deseamos estudiar cadenas. $\chi(k)$ es el motivo generado por el interés en los individuos conectados en G . Tomemos $t_1 \neq t_2 \in M_k$ donde, $a_{t_1 t_2} = 1$ o hay una sucesión de nodos $S(q) = j_1, \dots, j_q \in M_k \setminus \{t_1, t_2\}$ tal que $a_{t_1 j_1} a_{j_1 j_2} \dots a_{j_{q-1} j_q} a_{j_q t_2} = 1$, lo que determina nuestro motivo. En la epidemia estaríamos modelando el

medir variables de interés en una cadena que conecta posibles infestados. Es claro que $|M_k|$ no tiene que ser igual a $|M_h|$, si $h \neq k$.

3. UN MARCO GENERAL PARA LA SELECCIÓN DE LA MUESTRA

3.1. Algunos elementos sobre los Diseños muestrales

Como vimos trabajaremos con una población $U = \{u_1, \dots, u_N\}$. Todo subconjunto $s \subset U$ es llamado muestra de la población U . Sea $A = 2^U$ el conjunto de los subconjuntos de U , y d una medida de probabilidad definida sobre A . d es llamado diseño de muestreo. El diseño muestral puede ser implementado mediante un procedimiento mediante el cual se extrae un conjunto de unidades. Usando la muestra elegidos podemos obtener una generalización a toda la población usando los resultados obtenidos.

Todo diseño muestral comienza con la búsqueda de información que ayude a la identificación de las características de la población bajo estudio. Además, salvo en poblaciones muy pequeñas y accesibles nunca se observan todas las unidades de la población. La teoría estadística sugiere tomar muestras que asignen una adecuada probabilidad a cada individuo de ser elegido. En la práctica no siempre es fácil elegir muestras de la población objetivo y en realidad solo las podemos seleccionar de un marco existente (listas, mapas, etc.). Cuando definimos un diseño de muestreo le asignamos una probabilidad $d(s)$ a cada muestra de la población y una vez definida podemos calcular las probabilidades de inclusión de un elemento de la población

$$\pi_i = \sum_{\{s \subset \{1, \dots, N\} | s \ni i\}} d(s)$$

y de aparezcan dos simultáneamente

$$\pi_{ii^*} = \sum_{\{s \subset \{1, \dots, N\} | s \ni \{i, i^*\}\}} d(s)$$

Similarmente podemos definir la probabilidad de inclusión de un conjunto de k individuos, al tener plenamente definido el diseño muestral $d(s)$, mediante

$$\pi_{u^1, \dots, u^k} = \sum_{\{s \subset U | s \ni \{u^1, \dots, u^k\}\}} d(s)$$

Un problema usual es tener que estimar el total de la variable Y , por ejemplo, el número total de contagiados que desarrollan complicaciones renales. Una solución es el uso del estimador Horvitz-Thompson del total poblacional

$$t_{Y(HT)} = \sum_{i \in S} \frac{Y_i}{\pi_i}$$

su error de muestreo es

$$V(t_{Y(HT)}) = \sum_{i=1}^N \frac{(1 - \pi_i)}{\pi_i} Y_i^2 + \sum_{i=1}^N \sum_{j \neq i}^{N-1} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} Y_i Y_j$$

Este estimador permite caracterizar muchos problemas. Si usamos muestreo simple aleatorio sin reemplazo estamos usando el diseño

$$d(s) = \binom{N}{n}^{-1} \text{ si } |s| = n, d(s) = 0 \text{ en otro caso}$$

y

$$\pi_i = \pi_j = \frac{n}{N}, \pi_{ij} = \frac{n(n-1)}{N(N-1)}$$

Para cada problema asociado a la estimación de un parámetro podremos identificar varios estimadores bien conocidos o desarrollarlos. Vea libros de textos como Arnab (2017), Wayne (2009).

3.2. Muestreando grafos

Utilizando un diseño muestral seleccionamos una muestra de nodos s_0 (muestra inicial) la probabilidad de inclusión de un individuo es $\pi_i = \Pr(i \in s_0) > 0$. A partir de s_0 aplicamos un método de muestreo a partir de fijar un procedimiento de observación (OP), a partir de las aristas que inciden en los nodos en ella.

El número de aristas (conexiones) en la muestra es $A_{s_0} \subseteq A$. Para la muestra observada podemos denotar la muestra de las aristas $s_{oe} \subseteq U \times U$. Así que toda arista observada pertenece a s_{oe} especificada la matriz de adyacencias A_{s_0} , que es observada a partir de fijar un OP que implementa un cierto diseño muestral d .

En el estudio del COVID19, como en casi todas las epidemias similares, es de interés conocer los individuos que inciden en la arista (i,j) . Sea este conjunto

$$INC(a_{ij}) = \{u_t | t \in s_0 \text{ e incide en } (i,j)\}$$

de este podemos determinar $INC(A_{s_0}) = \cup_{a_{ij} \in A_{s_0}}$ que es el conjunto de nodos incidentes a A_{s_0} . A partir de estos razonamientos tenemos la posibilidad de usar una herramienta para determinar los individuos portadores que inciden en la conexión entre los individuos i -ésimo y j -ésimo y los que inciden a todas las aristas (conexiones) observadas.

Fijemos algunas estructuras del grafo muestreado.

Habíamos fijado que al seleccionar la muestra obtuvimos $G_s = \{U_s, A_s\}$ y al usar un OP la muestra final es $s_0 \cup INC(A_{s_0})$. O sea que serán evaluados no solo los individuos obtenidos en la muestra sino también los conectados con ellos. De ahí que, a partir de considerar los motivos, para todo

$$\chi(k) \in \chi_s = \chi(G_s) \text{ si y solo si } M_k \times M_k \subseteq s_{oe}.$$

La inclusión de $\chi(k) \in \chi_s$ implica que $M_k \in U_s$. Por otra parte $\chi(k) \in \chi_s$ no es implicado por que $M_k \in U_s$.

La trascendencia de estos hechos, en el contexto del estudio, deberán ser tomados en cuenta por los epidemiólogos a la hora de hacer sus generalizaciones a partir de la muestra observada.

Note que en muchas ocasiones el objetivo de seleccionar la muestra es establecer, comprender y/o identificar algunos patrones del comportamiento de la epidemia del COVID19 para ser usada que serían útiles para el manejo de la misma en otros brotes.

4. DISEÑOS PARA MUESTREAR POBLACIONES CON REPRESENTACIÓN BIPARTITA DE LA INCIDENCIA (BIGS)

4.1. Representación bipartita de la incidencia de una población

Este es un modelo en el que se representa el muestreo de grafos de poblaciones finitas cuando el interés está en el comportamiento bipartito. Mediante esta representación se pueden estudiar diversas estructuras donde los individuos son de naturaleza diferente. Su uso permite desarrollar estudios de conexiones no-convencionales. La presentación es dada por utilizar el conjunto

$$B = \{F, \chi(G); H\}$$

para representar G y un conjunto de motivos $\chi(G)$. H es el conjunto de aristas, pero estas solo existen entre los elementos de F y $\chi(G)$, pero no internamente. Las aristas entre los elementos de esos conjuntos tienen una existencia posible tal que $i \in F$ y $\chi(k) \in \chi = \chi(G)$ si y solo si $\chi(k) \in \chi_s$ cada vez que $i \in s_0$.

Tal es el caso en que $u_i = \text{tratamiento } i$ y $u_j = \text{paciente } j$. Cuando medimos a_{ij} estamos estableciendo que el tratamiento i -ésimo es recibido por el paciente j -ésimo, y a_{ji} si el paciente i -ésimo recibe el tratamiento j -ésimo. En el contexto epidemiológico podemos tener que F es un conjunto de hospitales, tratamientos etc. y $\chi(G)$ el conjunto de personas hospitalizados, de pacientes bajo un cierto tratamiento (o varios), etc. Así, al considerar el par paciente-hospital

$$h_{it} = \begin{cases} 1 & \text{si el paciente } t \text{ es asignado al hospital } i \\ 0 & \text{en otro caso} \end{cases}$$

además, si pasa por varios hospitales

$$\sum_{i \in F} h_{it} > 1$$

Sea $B[\chi(k)]$ el conjunto de los predecesores de $\chi(k)$

$$h_{ik} = \begin{cases} 1 & \text{si } i \in B[\chi(k)] \\ 0 & \text{en otro caso} \end{cases}$$

Así que el motivo $\chi(k)$ se observa si en la muestra está al menos un hospital de $B[\chi(k)]$ está en s_0 .

Es interés de los epidemiólogos el usar una cierta variable Y . Para ello tomamos cada motivo $k \in \chi(k)$ y asociarle a Y : Y_k . Por ejemplo, esta puede ser el número de días de ingreso, el riesgo de muerte, el costo de su tratamiento etc. Se desearía estimar algún parámetro como el total, la media etc. El total de Y es

$$T_Y = \sum_{k \in \chi(k)} Y_k$$

De la teoría del muestreo aparece como recomendable estimarle usando

$$T_{Y_{s_0}} = \frac{1}{n} \sum_{k \in \chi(G_{s_0})} \frac{Y_k}{\pi_k} = \frac{1}{n} \sum_{k \in G} \frac{Y_k}{\pi_k} \delta_k; \quad \delta_k = \text{número de veces en que se observó } k; \quad \delta_k \sim \text{Bin}(n, P_k),$$

$$T_{Y_{s_0}} = \sum_{i \in s_0} \frac{Y_i}{\pi_i} = \sum_{i \in G} \frac{Y_i}{\pi_i} \delta_i; \quad \delta_i = \begin{cases} 1 & \text{si } i \in s_0 \\ 0 & \text{en otro caso} \end{cases}; \quad \delta_i \sim B(\pi_i)$$

Estos no son necesariamente los mejores, pero son una solución es función de la estimabilidad. Note que al usar el muestreo simple aleatorio estos no son sino la función de la media muestral pues $\frac{1}{N} = P_k; \frac{n}{N} = \pi_i$

4.2. Algunos Diseños muestrales alternativos para estudios del Covid 19

4.2.1. Muestreo Indirecto (Indirect sampling)

Podemos ver que los conocidos modelos de Birnbaum and Sirken (1965) y Lavallée (2007), por ejemplo, caen en esta familia de diseños. La selección de los pacientes no es hecha sino en forma indirecta al seleccionar la muestra de F, que puede ser los tratamientos a que se someten las personas en proceso de curación, el hospital al que fueron conducidos, el consultorio del médico de la familia etc. Al considerar los estudios epidemiológicos tenemos que, en vez de hacer la selección del conjunto de personas directamente, lo hacemos indirectamente a través de hacer las selecciones en F e identificar los elementos a evaluar. Es más amable y barato no depender de un marco muestral donde se listen las personas y usar el marco de las entidades a que adscriben, cuyo control se tiene, como en los casos ejemplificados anteriormente

En el marco teórico del muestreo BIG ($B = \{F, \chi(G); H\}$) tomemos que

$$i \in F, \quad u_j \in \chi(G)$$

entonces la adyacencia es evaluada mediante;

$$\delta_{ij} = \begin{cases} 1 & \text{si } (ij) \in H \text{ (} h_{ij} = 1 \text{) si se llega a } j \text{ a través de algún } i \in s_0 \\ 0 & \text{en otro caso} \end{cases}$$

Sea el diseño para seleccionar la muestra inicial $d(s_0)$.

La probabilidad de inclusión de un individuo es:

$$\pi_j = 1 - P\left(\bigcap_{i \in B_j} i \notin s_0\right); B_j = \{i \in F | \delta_{ij} = 1\}$$

$P\left(\bigcap_{i \in B_j} i \notin s_0\right) = \bar{\pi}_{B_j}$ es la probabilidad de que ningún antecesor de j se encuentre en s_0 . Fíjese que al estudiar los contagios esta es la probabilidad de que si j tiene el virus no fue contagiado por nadie en la muestra.

La probabilidad de observar un par de personas está dada por

$$\pi_{jt} = 1 - \left(\bar{\pi}_{B_j} + \bar{\pi}_{B_t} - \bar{\pi}_{B_j \cup B_t}\right)$$

Esta va a ser de importancia para establecer si un par de personas pueden ser observadas en la muestra, por ejemplo, una pareja de trabajadores. Establecer otras probabilidades de inclusión conjunta pueden ser determinadas igualmente, de ser necesario.

4.2.2. Muestreo de Redes (Network sampling)

Sirken (2005) ejemplifica este modelo al muestrear núcleos familiares y estudiar parejas. Tal es el caso al tomar los esposos. Estas pueden ser reportados por más de un núcleo. Por ejemplo, el hombre habita con la familia de la madre en una dirección diferente que la de la esposa que lo hace con la suya. El epidemiólogo puede estar

interesado en estudiar si ambos conyugues poseen el mismo nivel de riesgo de ser contaminados. Este es una forma de hacer muestreo indirecto, pero tiene peculiaridades asociadas al OP instrumenta la selección de unidades.

El marco de muestreo de los núcleos es F , controlados por oficinas estatales y políticas, y de F se selecciona una muestra inicial s_0 utilizando un diseño muestral d , implementado por un OP. En el ejemplo que hemos usado para ilustrar, un matrimonio estará en la muestra si se observa al menos uno de los núcleos en los que poseen direcciones. La red de pareja es el motivo de interés.

Para cada $i \in F$ and $j \in \chi(k)$ se tiene que $(ij) \in H$ si al menos una pareja de M_j está el núcleo i . La probabilidad de inclusión de los motivos sigue siendo

$$\pi_j = 1 - P\left(\bigcap_{i \in B_j} i \notin s_0\right); B_j = \{i \in F | \delta_{ij} = 1\}; P\left(\bigcap_{i \in B_j} i \notin s_0\right) = \bar{\pi}_{B_j}$$

$$\pi_{jt} = 1 - \left(\bar{\pi}_{B_j} + \bar{\pi}_{B_t} - \bar{\pi}_{B_j \cup B_t}\right)$$

4.2.3. Muestreo de Clústeres Adaptativos (Adaptive cluster sampling , ACS)

Para dar una idea del modelo consideremos que tenemos una malla espacial, por ejemplo, en un municipio las manzanas. F es el conjunto de las mallas. El interés el epidemiólogo es medir una variable Y_i , por ejemplo, los habitantes en el área con alto riesgo de muerte de contraer el virus. El epidemiólogo fija un umbral con valor T y selecciona una muestra inicial s_0 . Cuando la malla pertenece a ella y $Y_i > T$ se muestrean las mallas que tienen fronteras con ella. El proceso termina, cuando no se añaden más mallas. Esto permitiría tener un método efectivo para estimar el número de personas con alto riesgo, su media por malla en un área, etc. En el ejemplo tendríamos una estimación para el municipio.

El clúster adaptativo es el conjunto de las mallas contiguas y

$$h_{ij} = \begin{cases} 1 & \text{si } i \in s_0 \text{ y } j \in \chi(j) \\ 0 & \text{en otro caso} \end{cases}$$

4.2.4. Muestreo de una Bola de Nieve de Q-etapas (Q-stage snowball sampling , Q-SBS)

Este método fue inicialmente propuesto por Goodman (1961) y su nombre e inspirado en la idea de una bola de nieve que rueda colina a abajo. El procedimiento genera un grafo dirigido en el que cada nodo posee solo una arista de salida. Este fue modelado dentro de la Teoría de Grafos en los trabajos de Frank (1977) y Frank - Snijders (1994) al considerar una etapa. Esto ha sido extendido para $Q \geq 1$. Se requiere identificar los predecesores de todos los motivos observados. Partiendo de un grafo no-dirigidos $G = (U, A)$ donde

$$U = \{u_1, \dots, u_N\}, u_i \text{ malla } i - \text{ésima}$$

$$(ij) \in A \text{ si y solo si } i \text{ y } j \text{ son vecinos y } Y_q > T, q = i, j$$

$$u_i \text{ tal que } Y_i > T \text{ es un nodo aislado en } G$$

En este procedimiento $F = U$ y se selecciona una muestra inicial s_0 , cuyos elementos son denominados semillas. La propagación de la muestra se lleva cabo a partir de aquellos nodos no aislados en G .

En la primera etapa la muestra está compuesta por los nodos $s_1 = s_0 \cup \alpha(s_0)$. Las nuevas semillas son los $u_i \in s_1$ y $s_1 \setminus s_0$ es la primera ola de muestras de la “bola de nieve”, y así sucesivamente hasta la etapa Q. Todas las redes de mallas seleccionadas son tratadas como motivos de interés. Estas tendrán al menos un predecesor en s_0 .

Note que este método es un procedimiento que satisface las intenciones de detectar el desarrollo de una epidemia partir de seleccionar una muestra de mallas. Por ejemplo, si tomamos una malla en un municipio y aplicamos este OP tendríamos un cubrimiento del mismo. Si incluimos otros municipios algunas de las semillas de estos pueden ser nodos de otros lo que daría una visión general del desarrollo de la epidemia del COVID en un ámbito más amplio.

Veamos algunas particularidades.

Si tomamos $\chi(k) \in \chi_s = \chi(G)$ y $i \neq j \in M_k$ podemos determinar el campo más corto de i a j en $\chi(G)$ dado un cierto OP. Sea su longitud v_{ij} y que la incidencia es reciproca. Partiendo de $i \in s_0 \cap M_k$ para observar todos sus nodos se requieren $Q = \text{Max}_{j \in M_k} v_{ij}$ etapas. Para cualquier $\chi(k) \in \chi_s = \chi(G)$, $i \in M_k$, la distancia $\Delta_{i(k)}$ entre la observación i y $\chi(k)$ es el número mínimo de etapas necesarias para observar $\chi(k) \in \chi_s = \chi(G)$ comenzando el SBS por la semilla i .

En la Teoría de Grafos se tiene unos resultados que juegan un papel importante en caracterizar la epidemia del COVID19

1. $\forall \chi(k) \in \chi(G)$, $i \in M_k$ si los nodos de M_k están conectados.
2. $\Delta_{i(k)} = \begin{cases} \text{Max}_{j \in M_k} \text{si } | \text{ArgMax}_{j \in M_k} v_{ij} | = 1 \\ 1 + \text{Max}_{j \in M_k \setminus i} v_{ij} \end{cases}$;
3. Si $\exists t \neq i$ que no es conectado en $\chi(G)$,
 $\Delta_{i(k)} = 1 + \text{Max}_{j \in M_k \setminus i} v_{ij}$; $M_k \setminus i = \{t \in M_k \text{ que se conectan con } i\}$
4. Si $\exists \chi(k) \in \chi(G)$, y un par $i \neq j \in M_k$ tal que $\Delta_{i(k)} = \Delta_{j(k)} = \infty$ no hay una representación factible de $\chi(G)$.

Veamos el caso en que el motivo esta fuera de M_k . En tales casos es necesario una estructura adicional. Tomando el grafo $G_{\bar{u}} = \{U_{\bar{u}}, A_{\bar{u}}\}$; donde $\bar{u} = \{u_1, \dots, u_w | u_r \in \chi(G); r = 1, \dots, w\}$ (un súper nodo) es el conjunto de los nodos puede expresarse mediante $A = A_1 \cup A_2 \cup A_3$ donde

$$A_1 = A \cap \{U_{\bar{u}} \times U_{\bar{u}}\}; A_2 = A \cap \{\bar{U}_{\bar{u}} \times \bar{U}_{\bar{u}}\}; A_3 = A \cap \text{donde } \bar{U}_{\bar{u}} = U \setminus U_{\bar{u}}$$

Eliminemos las aristas en A_1 y mantengamos todas las que están A_3 y no sean incidentes a los nodos en $U_{\bar{u}}$. Analizando A_3 reemplacemos las aristas $\{(ij) : i \in U_{\bar{u}}, (ij) \in A\}$ por una sola arista \bar{u} en $A_{\bar{u}}$, y procedemos similarmente con las aristas $\{(ji) : i \in U_h, (ji) \in A\}$ al sustituirles por $(\bar{u}j)$ en $A_{\bar{u}}$. Este el grafo transformado $G_{\bar{u}} = \{U_{\bar{u}}, A_{\bar{u}}\}$ es el determinado por el súper nodo.

$\forall \chi(k) \in \chi(G)$, y $i \notin M_k$ se tiene que $\Delta_{i(k)} = \infty$ si $v_{ij} = \infty$ para todo $j \in M_k$, así que no se puede llegar a $\chi(k)$ desde i . En otro caso podemos particionar los nodos de M_k usando v_{ij} para cada $j \in M_k$. Los nodos en M_k cuya distancia a i es u pertenecen a $U_u = \{j | j \in M_k, v_{ij} = u\}$. Dando un paso más observamos $U_u \times U_u$ partiendo de U_u y G_h es transformado a partir de G saliendo del híper nodo, \bar{u} si $U_h \neq \emptyset$, y $M_{ku} = \{\bar{u}\} \cup \{M_k \setminus U_u\}$ $M_k \setminus h = \{h\} \cup (M_k \setminus U_h)$. En este caso se tiene que

4. $\forall \chi(k) \in \chi(G), y i \in M_k$ se tienen que $\Delta_{i(k)} = \min_{u \geq 1} (u + \Delta_{\ddot{u}(k)})$ donde $\Delta_{\ddot{u}(k)}$ es la distancia existente entre el hiper nodo \ddot{u} y el motivo M_{ku} en G.

5. Conclusiones

Este trabajo sugiere como abordar estudios basados en grafos para seleccionar muestras para estudios epidemiológicos. Este ha sido motivado por los estudios que se desarrollan para evaluar aspectos de la pandemia del Covid19, tomando como referencia la problemática presentada en un municipio de La Habana, Cuba.

REFERENCIAS

- [1] Arnab R. (2017): Survey sampling theory and applications, Academic Press, N. York.
- [2] Birnbaum, Z. and M. Sirken (1965). Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates. Vital and Health Statistics, PHS Publication No. 1000-Series 2, No. 11. National Center for Health Statistics, Washington, D. C.: U. S. Government Printing Office.
- [3] Frank, O. (1977). Estimation of graph totals. Scandinavian Journal of Statistics 4, 81–89.
- [4] Frank, O. (1980a). Estimation of the number of vertices of different degrees in a graph. Journal of Statistical Planning and Inference 4, 45–50.
- [5] Frank, O. (1980b). Sampling and inference in a population graph. International Statistical, Review 48, 33–41.
- [6] Frank, O. (2011). Survey sampling in networks. The SAGE Handbook of Social Network Analysis.
- [7] Frank, O. and T. Snijders (1994). Estimating the size of hidden populations using snowball sampling. Journal of Official Statistics 10, 53–67.
- [8] Ferguson M.N. et al. (2020): Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>. (consultado Mayo 2020)
- [9] Fuller W. A. (2009): Sampling Statistics, Wiley, Chichester.
- [10] Goodman, L. (1961). Snowball sampling. The Annals of Mathematical Statistics 32, 148–170.
- [11] Kucharski A., Russell T., Diamond C., Liu Y. (2020): Analysis and projections of transmission dynamics of nCoV in Wuhan. **CMMID repository**, 2.
- [12] Lavallée, P. (2007). Indirect Sampling. New York, USA: Springer Science and Business Media.
- [13] Pham, H. (2020): On Estimating the Number of Deaths Related to Covid-19, Mathematics 2020, 8, 655; doi:10.3390/math8050655
- [14] Sirken, M. (2005). Network sampling. Encyclopedia of Biostatistics. John Wiley and Sons, Ltd., Online, DOI: 10.1002/0470011815.b2a16043.
- [15] Thompson, S. (1990). Adaptive cluster sampling. Journal of the American Statistical Association 85, 1050–1059.
- [16] Association 85, 1050–1059.
- [17] Tang B., Bragazzi N.L., Li Q., Tang S., Xiao Y, Wu J. (2020): An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). **Infectious Disease Modelling**, 5, 248255.
- [18] Thompson, S. (1991). Adaptive cluster sampling: Designs with primary and secondary units. Biometrics 47, 1103–1115.
- [19] Thompson, S. (2012). Sampling (3rd ed.). New York: Wiley.
- [20] Zhang, L.-C. and M. Patone (2017). Graph sampling. Metron 75, 277–299. Bipartite incidence graph sampling
- [21] Zhang L-Ch and M. Öguz-Alper (2020): Bipartite incidence graph sampling, arXiv: 2003.09467v1[stat.ME] 20 March 2020 (consultado Mayo 15, 2020).
- [22] Zhigljavsky, A., et al. (2020): Generic probabilistic modelling and non-homogeneity issues for the UK epidemic of COVID-19. **arXiv:2004.01991v1[stat.AP]4Apr2020**. (Consultado mayo 2020).

Capítulo 3

pp 21-29

SOBRE EL TAMAÑO DE MUESTRA PARA UN ESTUDIO DE LA PREVALENCIA DE ENFERMEDADES CORONARIAS

Agustín Santiago Moreno¹, Carlos Bouza Herrera², José Maclovio Sautto Vallejo¹,
Nurys Bárbara Armas Rojas³ y Yanela Yordanka³

¹Universidad Autónoma de Guerrero, México

²Universidad de la Habana, Cuba

³Instituto de Cardiología, Cuba

ABSTRACT

Cardiovascular diseases are the main death cause of adults in the world. México and Cuba have a similar behavior, as the rest of Latin America countries, due to the high prevalence of it. Which should be the sample size in studies of this disease is to be responded usually by statisticians. Here we propose a methodology for dealing with this problem. It has been developed for solving a real-life problem appearing in a medical research.

Key words: cardiovascular risk, sample size, sampling, odds, Poisson

RESUMEN

La enfermedad cardiovascular causa la mayor parte de las muertes de adultos en el mundo. México y Cuba tienen un comportamiento similar al resto de los países Latinoamericanos, debido a su alta prevalencia. Usualmente los estadísticos deben determinar el tamaño de la muestra al enfrentarse con estudios de esta enfermedad. Aquí proponemos una metodología para tratar este problema. Fue desarrollada para dar solución a un problema de la vida real que apareció en una investigación médica.

Palabras clave: Riesgo cardiovascular, tamaño de la muestra, muestreo, odds, Poisson

1. INTRODUCCIÓN.

Las enfermedades cardiovasculares son la principal causa de muerte en adultos en el mundo (Molinero, 2003), (Ruiz, Segura, & Agustí, 2012), (OMS, 2015). El riesgo cardiovascular se define como la probabilidad de muerte cardiovascular de una persona en un periodo determinado (10 años) (Gobierno Federal, SALUD, SEDENA Y SEMAR), (Alcocer, Lozada, Fanghänel, Sánchez-Reyes, & Campos-Franco, 2011). Existen diversos tipos de enfermedades cardiovasculares, entre ellas podemos mencionar la hipertensión arterial, enfermedad arterial coronaria, enfermedad valvular cardíaca, accidente cerebrovascular (trombosis o derrame cerebral) y fiebre reumática o enfermedad cardíaca reumática (DMEDICINA.com, Salud y bienestar, 2015). Cada año, aproximadamente 37 millones de personas en el mundo sufren un evento cardiovascular (ataque al corazón o una enfermedad vascular cerebral) (Alejandro, 2008), (López, 2013) y alrededor de 17.5 millones de personas mueren por dichas causas (OMS, 2015). Con frecuencia las enfermedades cardiovasculares resultan de la interacción de diferentes factores de Riesgo Cardiovascular, en donde se incluyen hipertensión, niveles elevados de lípidos en la sangre, diabetes, obesidad y tabaquismo (Ruiz, Segura, & Agustí, 2012). Por tanto, los principales factores de riesgo cardiovascular son: Presión arterial alta o hipertensión arterial (HTA), colesterol elevado o lipoproteína de baja densidad (LDL), las lipoproteínas de alta densidad (HDL, del inglés High density lipoprotein), diabetes, obesidad y sobrepeso, tabaquismo, inactividad física, género, herencia y la edad (Alcocer, Lozada, Fanghänel, Sánchez-Reyes, & Campos-Franco, 2011).

Según afirmación de la OMS, la mayoría de las enfermedades cardiovasculares se pueden prevenir cuando el tratamiento se enfoca de manera global a todos los factores de riesgo que presenta una persona, por ejemplo, tratar la hipertensión arterial al mismo tiempo que la elevación del colesterol; el tener alta la presión arterial o el colesterol por separado representa un riesgo para la persona, pero si están juntos el riesgo se incrementa significativamente. Por lo que la Organización Mundial del Corazón (WHO, 2013), señala de acuerdo a los estudios realizados, que el tratamiento de los factores de riesgo se sigue enfocando en el manejo por separado de cada uno de ellos, en lugar de ser enfocado a la atención del riesgo global del individuo (WHO, 2003).

La literatura médica divide los factores de riesgo cardiovascular en dos categorías: principales y contribuyentes. Los principales factores de riesgo son aquellos cuyo efecto de aumentar el riesgo cardiovascular ha sido comprobado. Los factores contribuyentes son aquellos que los médicos piensan que pueden dar lugar a un mayor riesgo cardiovascular pero cuyo papel exacto no ha sido definido aún (Farmacias ABC, 2006). Los principales factores de riesgo cardiovascular son: Presión arterial alta o hipertensión arterial (T), colesterol elevado o lipoproteína de baja densidad (C), las lipoproteínas de alta densidad (H), diabetes (D), obesidad (OB) y sobrepeso (Sp), tabaquismo (Tab), inactividad física (IF), sexo (Hombre, Mujer), herencia (Her) y la edad (E) (Alcocer, Lozada, Fanghänel, Sánchez-Reyes & Campos-Franco, 2011).

A más factores de riesgo, mayor las probabilidades de padecer una enfermedad del corazón. Algunos factores de riesgo pueden cambiarse, tratarse o modificarse y otros no. Pero el control del mayor número posible de factores de riesgo, mediante cambios en el estilo de vida y/o medicamentos, puede reducir el riesgo cardiovascular (Farmacias ABC, 2006).

Los factores contribuyentes de riesgo cardiovascular son: el estrés, hormonas sexuales, anticonceptivos orales y el alcohol, aunque para este último los estudios demuestran que el riesgo cardiovascular es menor en las personas que beben cantidades moderadas de alcohol que en las personas que no beben. Según los expertos, el consumo moderado es un promedio de una o dos bebidas por día para los hombres y de una bebida por día para las mujeres (Texas Heart Institute, 2015)

Según los datos de la ENSANUT 2012, estamos ante un verdadero problema de salud que día a día va creciendo, y que tanto las autoridades de salud como los usuarios de los servicios, debemos tomar conciencia de la realidad y hacer una verdadera medicina preventiva para enfrentar el problema (ENSANUT, 2012).

En estas notas nos interesa el desarrollo de un método sencillo para determinar tamaños de muestras en estudios que se restringen a la descripción de un segmento de la población.

2. CARACTERÍSTICAS DE LOS ESTUDIOS.

Se considera la realización de un estudio transversal, no experimental y explicativo. Una muestra debe ser tomada. Aunque existen gran número de trabajos relativos al estudio de los riesgos de la enfermedad cardiovascular, el conocido como estudio de Framingham constituye un pilar básico, y en diferentes formas es ampliamente utilizado para la toma de decisiones terapéuticas en base a la estimación de riesgo proporcionada por el modelo al introducir las características de riesgo del paciente concreto (Rodríguez, 2014).

En 2010 se condujo un estudio sobre los factores de riesgo cardiovascular en siete ciudades (Barquisimeto en Venezuela, Bogotá en Colombia, Buenos Aires en la Argentina, Lima en Perú, México DF en México, Quito en Ecuador y Santiago de Chile en Chile) denominado CARMELA (Cardiovascular Risk Factor Multiple Evaluation in Latin America). Los hallazgos de este estudio no fueron más allá de un estudio descriptivo de la población estudiada y la clasificación de la población en riesgo cardiovascular se realizó utilizando los puntajes de riesgo de Framingham.

El muestreo es comúnmente mencionado como una técnica principal al tratar de validar criterios sobre actividades, validez de decisiones etc. Este es frecuentemente mencionado en la literatura, conferencias etc. para validar los datos colectados y las generalizaciones que se hacen en base a la muestra observada.

Independientemente de eso hay aún cierta renuencia en los estudios médicos de adoptar e integrar métodos de muestreo a su investigación y actividad diaria o científica. Esta renuencia es debida en muchas ocasiones a la imposibilidad o dificultad de determinar adecuadamente un método de muestreo suficientemente adaptado a los objetivos que tiene su investigación. Eso es debido al no haber recibido un entrenamiento en las técnicas que les permitieran entender cómo operan y acudir a un especialista en muestreo antes de acometer la recolección de los datos.

Este reporte explora la naturaleza de algunos conceptos de la teoría y se enfoca en el problema de determinar el tamaño de muestra para el problema específico planteado en una investigación médica. Este es “Enfermedades coronarias: prevalencia en pacientes con insuficiencia arterial periférica”. Dadas sus características los resultados que discutiremos permitirán que se enfrenten similares investigaciones.

Motivado por la necesidad de buscar el tamaño de muestra en este estudio sobre la prevalencia de personas asintomáticos cardiovasculares, sin manifestaciones clínicas en corazón ni cerebro, que presentan dolores en las extremidades inferiores. Se reporta en la literatura que la prevalencia se encuentra entre 25% y 50%. La pregunta científica es si se puede determinar la arterioesclerosis clínica en corazón y cerebro en personas que acuden al médico manifestando dolores en las piernas.

En este reporte presentamos la base teórica que sustenta el proponer que el tamaño de muestra tenga un valor mínimo de 165.

3. EL PROBLEMA DEL MUESTREO

El análisis de una muestra permite inferir conclusiones susceptibles de generalización a la población de estudio con cierto grado de certeza. El total de observaciones en las cuales se está interesado, sea su número finito o infinito, constituye lo que se llama una población. La muestra es una pequeña parte de la población estudiada. Ésta debe caracterizarse por ser representativa de la población.

Una muestra puede ser de dos tipos:

a) No probabilística

En la muestra no probabilística la selección de las unidades de análisis depende de las características, criterios personales, etc. del investigador por lo que no son muy confiables en una investigación con fines científicos. Este tipo de muestra adolece de fundamentación probabilística, es decir, no se tiene la seguridad de que cada unidad muestral integre a la población total en el proceso de selección de la muestra. Entre los muestreos no probabilísticos están los procedimientos de muestreo intencional y accidental:

Muestreo Intencional: Procedimiento que permite seleccionar los casos característicos de la población limitando la muestra a estos casos. Se utiliza en situaciones en las que la población es muy variable y consecuentemente la muestra es muy pequeña.

Muestreo Accidental: Consiste en tomar casos hasta que se completa el número de unidades de análisis que indica el tamaño de muestra deseado.

Estos procedimientos son considerados útiles en el desarrollo de análisis cualitativos. Esto es, cuando el objetivo principal es resumir, analizar e interpretar la información obtenida mediante métodos cualitativos (entrevistas, análisis documental, etc.)

Las principales técnicas usadas son las de Categorización (Los datos se revisan y se reducen a unidades llamadas categorías), Análisis de contenido, permiten la descripción de la información a través de variables pre definidas en el estudio. Esto permite cuantificar datos cualitativos.

b) **Probabilística.**

El muestreo probabilístico permite conocer la probabilidad que cada unidad de análisis tiene de ser integrada a la muestra mediante la selección al azar.

Este tipo de muestreo comprende una gran cantidad de procedimientos de muestreo. Ante cada problema particular es necesario usar uno de los existentes en la literatura o desarrollar uno especial. Los más conocidos son simple o al azar, estratificado, sistemático y por conglomerados.

Muestreo Simple: Este genera una muestra aleatoria simple que resulta de aplicar procedimiento probabilístico que garantiza que todas las muestras posibles de un determinado tamaño tengan la misma probabilidad de ser elegidas. Esta definición refleja que la probabilidad de selección de la unidad de cada análisis es independiente de la probabilidad que tienen el resto de unidades de análisis que integran una población. Esto significa que tiene implícita la condición de equiprobabilidad.

4. EL PROBLEMA DEL TAMAÑO DE LA MUESTRA EN ESTUDIOS MÉDICOS SOBRE PREVALENCIA.

Algunos conceptos clave para determinar el tamaño de una muestra son:

1. Intervalo de Confianza: Algunos lo identifican con el “margen de error”. Este indica la precisión que soporta el estimado muestral utilizado. Provee límites superiores e inferiores para el parámetro que estima a partir de una probabilidad de error prefijado α .
2. α es llamado nivel de confianza, usualmente se usa $\alpha = 0.10$; 0.05 o 0.01 .
3. Tamaño de muestra n : es determinado para un cierto α . Este se relaciona con el tamaño N de la población
4. Porcentaje o Prevalencia: Es el porcentaje de elementos de la población que poseen una cierta propiedad. (por ejemplo, porcentaje de pacientes con la característica de interés). Si es desconocido se usa como cota máxima $p = 0.50$. Usando este valor se garantiza obtener más exactitud en los estimados.

Cuando se conoce la distribución de la característica es posible diseñar planes de muestreo por variables que tengan riesgos especificados.

La determinación de n y del rango de aceptación en un plan de muestreo se hace de acuerdo con el riesgo que están dispuestos a correr de cometer los errores tipo I y tipo II. Como se sabe, un plan de muestreo debe especificar el tamaño de la muestra, al cual llamamos n , y el rango de aceptación para el promedio o porcentaje de la muestra.

Para determinar el rango de aceptación se debe tener en cuenta que la variable:

$$Z = (\bar{x} - \mu) / [s / \sqrt{n}]$$

Z: rango aceptación

\bar{x} : media experimental

μ : media teórica

s: desviación estándar muestral

n: cantidad a analizar

Se asume una distribución normal de las variables.

En el cálculo del tamaño de la muestra se asume que esta es obtenida usando muestreo simple aleatorio. En tales casos para calcular el tamaño de una muestra hay que tomar en cuenta tres factores:

- El **porcentaje de confianza** con el cual se quiere generalizar los datos desde la muestra hacia la población total.
- El **porcentaje de error** que se pretende aceptar al momento de hacer la generalización.
- El **nivel de variabilidad** que se calcula para comprobar la hipótesis.
- La **confianza** o el **porcentaje de confianza** es el porcentaje de seguridad que existe para generalizar los resultados obtenidos.

Para evitar un costo muy alto para los estudios, o debido a que en ocasiones llega a ser prácticamente imposible el estudio de todos los casos, entonces se busca un porcentaje de confianza menor. Comúnmente en las investigaciones sociales se busca un 95%.

El porcentaje de error equivale a elegir una probabilidad de aceptar una hipótesis que sea falsa como si fuera verdadera, o la inversa: rechazar a hipótesis verdadera por considerarla falsa.

Al igual que en el caso de la confianza, si se quiere eliminar el riesgo del error y considerarlo como 0%, entonces la muestra es del mismo tamaño que la población, por lo que conviene correr un cierto riesgo de equivocarse. Comúnmente se aceptan entre el 4% y el 6% como error, tomando en cuenta de que no son complementarios la confianza y el error.

Debe fijarse la probabilidad (o porcentaje) con el que se aceptaría y se rechazaría la hipótesis que se quiere investigar, usando información sobre en alguna investigación anterior o en un ensayo previo a la investigación actual. El porcentaje con que se aceptó tal hipótesis se denomina variabilidad positiva y se denota por p , y el porcentaje con el que se rechazó se la hipótesis es la variabilidad negativa, denotada por $q=1-p$.

En el caso de variables cualitativas, que toman solo valores cero o uno, hay que considerar que p y q son complementarios, es decir, que su suma es igual a la unidad: $p + q = 1$. Además, cuando se habla de la máxima variabilidad, en el caso de no existir antecedentes sobre la investigación (no hay otras o no se pudo aplicar una prueba previa), entonces los valores de variabilidad a tomar son $p = q = 0.5$.

Una vez que se han determinado estos tres factores, entonces se puede calcular el tamaño de la muestra
En el caso de que no se conozca con precisión el tamaño de la población

$$n = \frac{z^2 pq}{E^2}$$

Donde:

- n es el tamaño de la muestra;
- z es el nivel de confianza;
- p es la variabilidad positiva;
- q es la variabilidad negativa;
- E es la precisión o error

La variabilidad y el error se pueden expresar por medio de porcentajes, hay que convertir todos esos valores a proporciones en el caso necesario.

El nivel de confianza no es ni un porcentaje, ni la proporción que le correspondería, a pesar de que se expresa en términos de porcentajes. El nivel de confianza se obtiene a partir de la distribución normal estándar, pues la proporción correspondiente al porcentaje de confianza es la integral de la curva normal que se toma como la confianza, y la intención es buscar el valor z de la variable aleatoria que corresponda a tal área.

En el caso de que se conozca el tamaño de la población:

$$n = \frac{z^2 pqN}{NE^2 + z^2 pq}$$

Donde:

- n es el tamaño de la muestra;
- z es el nivel de confianza;
- p es la variabilidad positiva;
- q es la variabilidad negativa;
- N es el tamaño de la población;
- E es la precisión o el error.

Al usar muestro no probabilístico. El tamaño de la muestra muchas veces es determinado también usando el buen juicio del investigador. Un ejemplo en medicina es proveído por Gordon et al. (2014). No realizó un piloteo específico. Los tamaños de muestra se movieron entre 37% y 100 de la población obteniendo respuestas de entre 21 y 89% de los entrevistados a las diferentes preguntas.

5. UNAS REGLAS DE ÚTILES.

En estudios médicos es necesario distinguir entre el desarrollo de estudios aleatorizados u observacionales. Al usar un muestreo por conveniencia es necesario hacer algunos cambios respecto a la teoría basada en la estadística paramétrica. Es usual que esta muestra se considere no representativas. Una interpretación teórica sostiene que no es posible hacer una generalización en base a tales muestras sobre lo que sucede en la población. Esto puede ser cierto o no pues dependerá de cuan adecuada sea la relación real entre muestra y población.

Al hacer un estudio observacional de cohorte, diseñado con cuidado y maestría se puede confiar en los datos obtenidos como representativos de la población. Es tarea de los especialistas en estadística guíen al investigador a ser cautos y objetivos para hace generalizaciones.

Tomaremos el evento.

A=no tener manifestaciones clínicas en corazón ni cerebro pero que manifiesta tener dolor en la pierna
La variable de interés es

$$Y_i = \begin{cases} 1, & i \in A \\ 0 & \text{en otro caso} \end{cases}$$

Los individuos en la población son clasificados a priori en la tabla de Contingencia 2x2

	Se entrevista	No se entrevista
Pertenece a A	π_{11}	π_{21}
No pertenece a A	π_{12}	π_{22}
	π_1	π_2

O sea que

- π_{11} =frecuencia relativa de pacientes que poseen la característica y se entrevistan
- π_{21} =frecuencia relativa de pacientes que poseen la característica y no se entrevistan
- π_{12} =frecuencia relativa de pacientes que no poseen la característica y se entrevistan
- π_{22} =frecuencia relativa de pacientes que no poseen la característica y no se entrevistan
- $\pi_1 = \pi_{11} + \pi_{12} = P(\text{se entrevistan})$
- $\pi_2 = \pi_{21} + \pi_{22} = P(\text{no se entrevistan})$
- $\pi_A = \pi_{11} + \pi_{21}$

Al tomar una muestra tendremos en la muestra el número de pacientes con la característica

$$T_Y = \sum_{i=1}^n Y_i$$

podemos considerar que su distribución es una Binomial: $B(n, \pi_A)$. Se sabe que esta tiene como límite una distribución Poisson si π_A es pequeña y n es grande.

La muestra genera la tabla

Pertenece a A	T_Y
No pertenece a A	$n - T_Y$
	n

En el estudio de esta prevalencia podemos aceptar la aproximación Poisson. Es clara esta acepción pues es raro observar una persona con la característica en una población de tamaño medio. Piense que Cuba tiene una población de 12 millones de personas, La Habana más de 2 millones etc. Por ello podemos considerar que T_Y sigue una distribución de Poisson con $\lambda = n\pi_A$. En los problemas que se enfrentan en investigaciones de este tipo es que los pacientes que tienen la característica de interés son pocos respecto a la población completa de personas, incluso en investigaciones relativamente pequeñas. Piénsese en entrevistar la población de los pacientes que acuden en un hospital con trastornos circulatorios en las piernas. Estos deben ser algunos millares al año.

El problema que enfrentamos es el estudio de la prevalencia de trastornos cardiovasculares no detectados no es sino una prueba de hipótesis. Esta es establecer si es grande el número de los pacientes que acuden con malestares en las piernas que son asintomáticos de una dolencia cardiológica. El interés del investigador es establecer la validez de su hipótesis de que

$$\theta_0 = E(T_Y) > \theta_1 = E\left(\sum (1 - Y_i)\right)$$

Note que :

- θ_0 es el número de asintomáticos cardiovasculares con manifestaciones clínicas en corazón ni cerebro con dolor en las piernas
- θ_1 es el numero de asintomáticos cardiovasculares y sin manifestaciones clínicas en corazón ni cerebro

Así que nos enfrentamos a un problema de prueba de hipótesis sobre la igualdad de los parámetros de dos variables con distribución Poisson. Regresando a la tabla de contingencia tenemos que podemos tomar en consideración el valor del odd:

$$\pi_0 = \frac{\pi_{21}}{\pi_{11}}$$

y considerar que trabajamos con una Poisson con parámetro $\theta = nR\pi_0$. Para ella la estabilización de la varianza $\sigma^2 = p(1 - p)$ se obtiene cuando tenemos $p = 0.5$. Así que siguiendo a Van Belle (2008) el tamaño de la muestra en tal caso, un estudio observacional será, para $\alpha = 0.05$:

$$n = \frac{1}{p(1 - p)(\sqrt{\theta_0} - \sqrt{\theta_1})^2} = \frac{4}{\pi_0(\sqrt{R} - 1)^2}$$

Esto garantiza una potencia

$$Pot = 1 - \beta = 0.80.$$

R es el riesgo relativo que fijará el investigador y debe ser mayor que uno. Es usual utilizar $R=3$. En el caso bajo investigación el mínimo de las prevalencias fijado por los médicos fue $\pi_0 \cong 0.33$. Por tanto,

recomendaríamos para el estudio observacional que analizamos, con $R=0.3$, que un tamaño de la muestra adecuado es

$$n = \frac{4}{0.33(\sqrt{R} - 1)^2} \cong 65$$

Por tanto, el estudio puede llevarse con mínimo de 65 personas. Incrementarlo puede ser motivado por necesidades de los experimentos.

6. CONCLUSIONES

Sin duda, la validez científica de los estudios en ciencias de la salud, es de fundamental importancia para el diagnóstico, tratamiento, control y cuidados en las enfermedades que aquejan a la humanidad. Con el presente trabajo pretendemos contribuir modestamente a ese objetivo. La propuesta está orientada a proporcionar elementos de juicio suficientes para la toma de decisiones, al momento de seleccionar una muestra de pacientes que participarán en un estudio médico. Ese es el objetivo de las fórmulas básicas de muestreo y las reglas básicas enunciadas. Esperamos que, con su utilización, se mejoren los reportes de investigación en estudios médicos.

REFERENCIAS

- [1] Adam Lee Gordon, Adrian Blundell, Jugdeep K. Dhesi, Calum Forrester-Paton, Jayne Forrester-Paton, Hannah K. Mitchell, Nicola Bracewell, Jocelyn Mjojo, Tahir Masud, John R. F. Gladman (2014): UK medical teaching about ageing is improving but there is still work to be done: the second national survey of undergraduate teaching in ageing and geriatric medicine. *Age and ageing* 2014; 43: 293–297
- [2] Alcocer, L. A., Lozada, O., Fanghanel, G., Sánchez-Reyes, L., & Campos-Franco, E. (2011): Estratificación del riesgo cardiovascular global. Comparación de los métodos Framingham y SCORE en población mexicana del estudio PRIT. *Cirugía y Cirujanos*, Vol. 79, No. 2, pp. 168-174.
- [3] Campos-Nonato, I., Hernández-Barrera, L., Rojas-Martínez, R., Pedroza-Tobías, A., Medina-García, C., & Barquera, S. (2013): Hipertensión arterial: prevalencia, diagnóstico oportuno, control y tendencias en adultos mexicanos. *Salud Pública Méx* 2013, 55,144-150.
- [4] J. Castellanos Vázquez; A. Santiago Moreno; C. Bouza Herrera; J. M. Sautto Vallejo (2019): Valoración De Riesgo Cardiovascular Mediante Modelos De Clasificación. *Inv. Oper.* 39, 81-88
- [5] Chen, Z., Bai, Z., Sinha, B.K., 2004. Ranked Set Sampling: Theory and Applications. *Lectures Notes in Statistics*, 176. Springer, N. York.
- [6] DMEDICINA.com, Salud y bienestar. (9 de octubre de 2015). *Factores de riesgo cardiovascular*. Recuperado el 2 de Julio de 2016, de <http://www.dmedicina.com/enfermedades/enfermedades-vasculares-y-del-corazon/factores-de-riesgo-cardiovascular.html>
- [7] Doug Stewart Daegis (2011): Application of Simple Random Sampling 1 (SRS) in eDiscovery. Fourth DESI Workshop on Setting Standards for Electronically Stored Information in Discovery Proceedings.
- [8] FARMACIAS ABC. (2006): *Riesgo cardiovascular, la importancia de la prevención*. Recuperado el 06 de Julio de 2016, de <http://www.farmacias-abc.com.ar/contenido/Index624.asp>
- [9] Molinero, L. M. (2003): Modelos de riesgo cardiovascular. Estudio de Framingham. Proyecto SCORE. Asociación de la Sociedad Española de Hipertensión. Liga Española para la Lucha contra la Hipertensión Arterial.
- [10] OMS. (2015): Enfermedades cardiovasculares. Recuperado el 2 de Julio de 2016, de Organización Mundial de la Salud: <http://www.who.int/mediacentre/factsheets/fs317/es/>
- [11] QUO. (2012): El 30% de mexicanos son hipertensos. Recuperado el 2 de Julio de 2016, de <http://quo.mx/noticias/2012/05/02/el-30-de-mexicanos-son-hipertensos>

- [12] Ruiz, E., Segura, L., & Agusti, R. (2012): Uso del Score de Framingham como Indicador de los Factores de Riesgo de las Enfermedades Cardiovasculares en la población Peruana. *Revista Peruana de Cardiología*, Vol. XXXVII, N° 3.
- [13] Texas Heart Institute. (2015): Factores de riesgo cardiovascular. Recuperado el 6 de Julio de 2016, de http://www.texasheart.org/HIC/Topics_Esp/HSmart/riskspan.cfm
- [14] Van Belle Gerald (2008): *Statistical rules of thumbs*. Wiley Series in Prob. And Stat. Wiley, N. Jersey.
- [15] WHO. (2003): *The World Health Report 2003*. Recuperado el 2 de Julio de 2016, de World Health Organization: http://www.who.int/whr/2003/en/whr03_en.pdf?ua=1

Capítulo 4

pp 31-45

ESTIMACIÓN DEL BALANCE BENEFICIO/RIESGO EN ENSAYOS CLÍNICOS

Carmen Viada¹, Carlos Bouza², Martha Fors³, Mabel Alvarez¹, Aliuska Frias¹, Lázara García¹, Yanela Santiesteban¹, Yuliannis Santiesteban¹, Mayelin Troche¹, Liana Martínez¹.

¹Centro de Inmunología Molecular, Habana, Cuba.

²Facultad de Matemática y Computación, Universidad de la Habana, Cuba.

Universidad del País Vasco, País Vasco, España.

³Universidad de las Américas, Quito, Ecuador.

Resumen

La evaluación de riesgos y beneficios es uno de los requisitos fundamentales en la revisión ética de la investigación con participantes humanos. Como resultado, los investigadores deben evaluar e intentar minimizar todos los riesgos, previsibles involucrados en la investigación propuesta, y los miembros de los comités de ética en investigación deben evaluar y hacer un balance de los riesgos y beneficios potenciales implicados en cada propuesta de investigación, como parte de sus obligaciones éticas respecto de los protocolos de investigación. Sin embargo, la literatura actual proporciona escasas guías sobre los detalles específicos de cómo debe ocurrir este proceso de equilibrio. En consecuencia, este artículo ofrece algunos detalles del proceso para equilibrar los riesgos y beneficios en la investigación biomédica y les recuerda a los miembros de los comités de ética de investigación su responsabilidad de proteger a aquellos que son vulnerables a la explotación en proyectos de investigación.

Palabras clave: Sujetos de investigación-Protocolos-Ética, Experimentación humana, Comité de Revisión y Ética, Proyectos de investigación, Vulnerabilidad en salud.

Abstract

The evaluation of risks and benefits is one the main requisites in investigation ethics with humans' participants. As a result, investigators should consider and look for minimizing all the risks, involved and previsible, in the proposed research, and the members of the Committees of Ethics must evaluate and make a balance of potential de risks and benefits involved in each proposed investigation as part of the ethical duties with respect to the research protocols. Nevertheless, nowadays literature gives scarce guidance on the specific details on how the process must behave for obtaining equilibrium between risks and benefits. Consequently, this article gives some details on equilibrating risks and benefits in biomedical investigation and points out to the members of the ethical committees of their responsibility of protecting vulnerable persons in the development of research projects.

Key Words: Subjects of Ethical-Investigation-Protocols, Experimentation with humans, Ethical Revision Committee, Research Projects, Vulnerability in health

1. Introducción

Toda investigación biomédica con participantes humanos debe ser revisada a priori por un comité reconocido de ética en investigación.¹ Quien presida el comité determinará si la propuesta de investigación requiere una revisión completa por parte de todos los miembros del mismo, si califica para una revisión ágil de parte de algunos miembros, o si la propuesta de investigación entra en la categoría de exención de revisión ética. A partir

de la sumisión de una propuesta ante el Comité de Revisión y Ética (CRE), sus miembros deben realizar una evaluación adecuada de ésta, asegurándose de que se cumplen los siete requisitos éticos para la investigación con sujetos humanos.² Estos requisitos incluyen una determinación del posible valor de la investigación, de la validez científica de la metodología propuesta, un proceso de selección razonable de los participantes, un balance favorable de los riesgos previsible y de los posibles beneficios del trabajo de investigación, y una evaluación del bienestar de los participantes en la investigación propuesta. Este último concepto incluye asegurar que todos los participantes serán respetados a lo largo de todo el proceso de investigación, que su información personal será protegida de manera efectiva (confidencialidad) y que el proceso de consentimiento informado cumple con las normas aceptadas internacionalmente.³

La revisión de la literatura publicada reveló que se ha discutido intensamente acerca de los riesgos y los beneficios de la investigación, sin embargo, no se encontró ninguna publicación que pudiera orientar a los miembros del comité de ética y a los propios investigadores respecto de la determinación detallada del riesgo, sus niveles y sus posibles categorías y, específicamente, cómo obtener un balance de estos riesgos y beneficios potenciales (ya sea para los individuos, las comunidades o para un país o sociedad). En consecuencia, este artículo busca proporcionar una orientación detallada sobre algunas de estas cuestiones, en un contexto de posible explotación en ciertos trabajos de investigación.

2. Riesgo de la investigación

El riesgo en investigación es la probabilidad de daño, pérdida, lesión u otras consecuencias adversas que pudieran ocurrirle a alguien como resultado de su participación en un estudio de investigación⁴. Habitualmente, cuando se menciona algún riesgo de daño, el daño físico instantáneamente se viene a la mente y, usualmente, los sujetos sólo piensan en el riesgo de lesiones físicas a su persona. Sin embargo, el concepto de daño tiene más de una dimensión. Por lo tanto, los investigadores, los miembros de los comités de ética y los posibles participantes de los proyectos de investigación deben estar conscientes de que los daños pueden tener lugar durante los proyectos e incluyen aspectos tales como daño físico, moral, psicológico, social, jurídico y financiero.

El riesgo de daños físicos incluye el riesgo de enfermedad, lesión, dolor y otros males que están asociados con el bienestar físico (por ejemplo: lesión durante procedimientos médicos invasivos o posibles efectos secundarios de una droga de la investigación).⁴ El concepto de daño moral comienza con el respeto a la persona y a su bienestar, con un reconocimiento y un respeto específico por la dignidad y la integridad corporal de la persona.⁵ La cuestión que aquí subyace es que todos los seres humanos tienen valor moral y, por la naturaleza de dicho valor, debe respetárselos, así como a su dignidad y a su integridad corporal. Como resultado, las personas deberían tener un control total sobre sus cuerpos, su información personal y su tejido corporal. Así, si se pretendiera utilizar su información o su tejido corporal para una investigación (por ejemplo, tejido extraído durante una cirugía de rutina o de emergencia), incluso de forma anónima, esa información o tejido corporal le pertenecen a la persona y, por lo tanto, su permiso debe ser obtenido por adelantado. No hacerlo es causarle daño moral a su dignidad y a su integridad.

El riesgo de sufrir daños psicológicos incluye la posibilidad de producir estados negativos o alterados en la conducta, incluyendo ansiedad, depresión, culpa, sentimientos de inutilidad, enojo o miedo.⁴

Estos pueden tener lugar si los participantes de la investigación precisan recordar eventos penosos o aprender sobre la posibilidad genética de desarrollar una enfermedad incurable, o si los participantes se sienten amenazados o estresados como resultado de su participación.

El riesgo de daños sociales implica la posibilidad de que se produzca una distribución en las redes sociales de los participantes (familia, amigos, conocidos, comunidades cívicas y religiosas), o una alteración en sus relaciones con otros, y puede implicar estigmatización, vergüenza o pérdida de respeto. Puede tener lugar, por ejemplo, si se diera una revelación inadvertida de información sensible o embarazosa (por ejemplo, portar VIH o padecimientos mentales). Un riesgo de daño jurídico incluye el riesgo de descubrimiento y enjuiciamiento

por una conducta delictiva, por ejemplo, si se revelase información sobre el consumo abusivo de sustancias ilegales a la policía.⁴

Los riesgos de daño financiero o económico incluyen la posibilidad de incurrir o de imponer cargas financieras, penurias o costos financieros directos o indirectos a los participantes en razón de su implicación en el proyecto de investigación. Esto puede ocurrir, por ejemplo, si el diagnóstico de VIH de una persona es revelado inadvertidamente durante el proceso de investigación, dando como resultado que el participante afectado pierda su empleo o el acceso a los beneficios de un seguro.

3. Índices de riesgo

Para llevar a cabo el análisis de riesgo se utilizan diferentes herramientas de cálculo (hay múltiples aplicaciones informáticas) con las cuales se dimensiona la probabilidad de los riesgos obteniéndose el índice de riesgo que es una función de la gravedad del fenómeno (gR) y de la probabilidad (pR) de que se produzca dicho fenómeno. De esta forma, un fenómeno que sea muy frecuente con consecuencias muy graves hace lógicamente que el riesgo sea inaceptable. Y por el contrario una probabilidad muy remota, aunque la gravedad sea muy elevada permite que el riesgo sea aceptable. (Tabla 1)

$$IR = f(gR \Delta pR)$$

Tabla 1. Probabilidad de los Riesgos

Probabilidad	Fatal (Muerte o lesión grave irreversible)	Significativa (Lesión grave reversible)	Marginal (Complicaciones menores)	Irrelevante (Simples molestias)
Frecuente	1	3	7	13
Probable	2	5	9	16
Ocasional	4	6	11	18
Remota	8	10	14	19
Improbable	12	15	17	20

Esta es la base del análisis de riesgos, pero, como se comentaba previamente, no ha tenido demasiada repercusión en el ámbito farmacéutico.

Tipos de enfoque del análisis de riesgo

Se han planteado enfoques tanto cualitativos como cuantitativos a la hora de materializar este tema. Varios programas están en experimentación o ya en activo con un enfoque cualitativo como son PrOACT-URL, BRAT (Benefit-RiskActionTeam), CMR CASS, FDA BRF (The US Benefit-Risk Framework). Estos se dedican a conferir transparencia y legitimidad a las decisiones de la autoridad sanitaria. Los enfoques cuantitativos son principios de cálculo para la evaluación de riesgos de una forma numérica basándose en diferentes modelizaciones.

Niveles de riesgo

“Riesgo” es una palabra que expresa “probabilidades”, y cuando se utiliza en relación a la investigación, está expresando la probabilidad de que ocurra un daño durante la misma. En este sentido, los riesgos en la investigación biomédica pueden clasificarse en riesgo mínimo, bajo, medio y alto.⁴ (Tabla 2)

Riesgo mínimo

Este es, de hecho, el menor riesgo posible. No hay ninguna categoría “sin riesgo”, ya que una vez que los pacientes están involucrados en la investigación, como mínimo pierden tiempo personal que de otro modo

podrían estar utilizando en su vida personal, sus actividades o su trabajo. La participación en la investigación y la incertidumbre sobre los resultados también conlleva al menos un grado mínimo de aprehensión para los participantes, lo cual tiene consecuencias psicológicas. Por lo tanto, las propuestas de investigación nunca deben afirmar que “no hay riesgo”.

Los procedimientos que pueden clasificarse como de riesgo mínimo incluyen interrogar, observar y medir a los participantes en un área o materia que no sea polémica, siempre que los procedimientos se lleven a cabo de una manera sensible y que se haya dado el consentimiento.⁶ En la investigación biomédica, esta categoría incluye también la recolección de una muestra simple de orina y el uso de sangre de una muestra que ya se había extraído como parte del manejo rutinario de casos de atención médica. La mayoría de los proyectos de investigación en ciencias sociales y de la cultura, generalmente se clasifican como de riesgo mínimo, ya que por lo general implican un riesgo muy bajo de daño físico.⁷ Sin embargo, debido al posible riesgo de daño psicológico significativo que se puede producir en algunas investigaciones, como las que exponen a los participantes a estímulos muy potentes (por ejemplo, estudios que contienen violencia o pornografía), o que pueden causar daño mental a largo plazo (por ejemplo, depresión o insomnio) más allá de los riesgos de la vida normal; a esos proyectos podría corresponderles una clasificación en otras categorías, dependiendo de las circunstancias particulares.⁸ Algunos proyectos de investigación en ciencias sociales pueden incluso ser clasificados como de alto riesgo si la investigación conlleva un riesgo significativo para la seguridad de los participantes (por ejemplo, violencia doméstica con riesgo de muerte).⁸

Bajo riesgo

En esta categoría de la investigación biomédica se insertan los procedimientos cuyo empleo durante el proceso investigativo puede causar dolor breve o sensibilidad, o pequeños moretones y posibles pequeñas cicatrices.⁶ Deberíamos tener en cuenta que cuando los niños participan de una investigación, estos siempre son aprehensivos, por lo que cualquier uso de agujas durante el proceso, ya sea para inyecciones, venopunción o algo relacionado, automáticamente se incluye en una clasificación de bajo riesgo en lugar de riesgo mínimo.

Riesgo medio: Esta categoría describe a los procedimientos usados durante la investigación que involucran riesgos que son intermediarios entre los riesgos “bajo” y “alto”, y tales juicios son mejor efectuados por un Comité de Revisión y Ética. Nótese que la investigación biomédica que implica, por ejemplo, una o dos imágenes radiográficas (rayos X) tomadas a partir del abdomen de una mujer embarazada significaría algo más que un bajo riesgo para el feto, pero aun así la investigación puede ser permitida por parte del comité de ética si existiese un mayor beneficio proporcional que se pueda obtener para la mujer embarazada en cuestión a partir del resultado del estudio.

Riesgo alto: En la investigación biomédica esta categoría de riesgo suele implicar investigaciones que utilizan procedimientos como la biopsia pulmonar o hepática, la punción arterial o el cateterismo cardíaco, procedimientos que normalmente no pueden justificarse únicamente con fines investigativos. Tales operaciones sólo deben llevarse a cabo cuando la investigación se combina con un diagnóstico y un tratamiento que se destinan a beneficiar a los participantes interesados de un modo particular. Los ensayos clínicos que utilizan agentes de tratamiento experimental también se clasifican como de alto riesgo.⁶

Tres dimensiones clave⁹

- Gravedad
 - De las RAM
 - De la enfermedad (tratada y no tratada)
- Duración
 - De las RAM
 - De la enfermedad
 - De los efectos beneficiosos

- Frecuencia (incidencia)
 - De las RAM
 - De la enfermedad
 - Del tratamiento exitoso

Tabla 2. Descripción de las dimensiones clave

	Alto	Medio	Bajo
Gravedad	Mortal	Incapacitante	Incómodo
Duración	Permanente	Persistente	Temporal
Incidencia	Común	Infrecuente	Raro

4. Beneficios de la investigación

“Beneficio” es una palabra que expresa un hecho o estado de cosas y, por lo tanto, cuando este término se utiliza durante las etapas de planificación o de revisión ética en la investigación, las personas deberían utilizar, en realidad, el término “Beneficio posible” o “beneficios potenciales”, ya que es una anticipación de los resultados y no un hecho. También debemos señalar que los beneficios podrían presentarse en más de un terreno. Pueden tener lugar sólo para el individuo (beneficios concretos para los participantes del estudio), o puede resultar en un Beneficio para la sociedad (cuando la investigación es vista como un bien social, en Beneficio de las generaciones actuales y futuras). El estudio no debería ser aprobado para proceder a menos que se indique algún Beneficio anticipado o potencial, especificando si ese Beneficio será un nuevo conocimiento o una confirmación de los resultados de otras investigaciones dirigidas previamente.

El Consejo de Organizaciones Internacionales de Ciencias Médicas (CIOMS) publica las Directrices Internacionales del CIOMS³, las cuales establecen una variedad de intervenciones que pueden estar implicadas en la investigación, algunas de las cuales pueden tener un Beneficio terapéutico directo, mientras que otras pueden incluir intervenciones no benéficas, como cuando el estudio se realiza únicamente para responder a una pregunta de investigación. Por lo tanto, los posibles beneficios incluyen la obtención de conocimiento, comprensión y entendimiento, así como posibles ganancias en habilidad, prestigio o experiencia para los investigadores o la institución de investigación. Los participantes individuales pueden beneficiarse físicamente, a través del bienestar social directo o a través de las instituciones que sostienen su bienestar.

La investigación también puede ser beneficiosa para una comunidad local.³ En la investigación biomédica esto incluye mejoras en los servicios de salud; mejores medios de vida para los miembros de la comunidad; una reducción de la carga de pacientes en el sistema de atención de la salud; soluciones a los problemas de la atención en salud; aliviar los crecientes costos del cuidado de la salud y la creación de una industria de conocimiento intensiva sobre el cuidado de la salud.⁴

La evaluación de riesgos y beneficios

Tanto la “probabilidad” de los riesgos como la “magnitud” de los posibles daños deben ser consideradas en esta etapa. Esto significa la probabilidad o posibilidad de que ocurra un daño y la magnitud del mismo, incluidas sus consecuencias.⁴ Por lo tanto, al realizar esta evaluación, debemos preguntar: ¿Qué tan probable es que ocurra este daño particular (probabilidad)? Después de responder a ese interrogante, debemos preguntarnos: ¿Qué tan grande magnitud) podría ser este daño, si ocurriera? ¿Qué efecto puede tener? Y luego, tendríamos que comparar los riesgos “predecibles” con los beneficios “previsibles”.

Para que las propuestas de investigación cumplan con este requisito ético, debe existir una relación riesgo/Beneficio favorable. En otras palabras, los riesgos deben ser inferiores a los beneficios previstos, o los beneficios esperados deben superar a los riesgos. Así, los beneficios que puedan lograrse para el individuo o para la sociedad deben superar los riesgos potenciales para los participantes de la investigación. Lo que es más importante, cuanto más probables o más graves sean los riesgos potenciales, mayor debería ser la magnitud de los beneficios potenciales.⁴

Al llevar a cabo esta determinación, los comités de ética en investigación deben ser conscientes de que algunas investigaciones biomédicas pueden ofrecer beneficios directos a los participantes, tal es el caso de ciertas personas que padecen una enfermedad particular que puede ser susceptible al tratamiento que se está probando. En este escenario, los participantes pueden estar dispuestos a aceptar los riesgos o efectos secundarios del tratamiento que, probablemente, hubieran sido considerados inaceptables por parte de otras personas que no padecen esa enfermedad específica. En consecuencia, la evaluación de riesgos y beneficios requiere un ejercicio de buen juicio y debe basarse en la información disponible en el protocolo de investigación, así como en una apreciación completa del contexto.⁴ Todas estas determinaciones deben ser transparentes y justificables.

A la par de este proceso, debe considerarse si la investigación biomédica propuesta continuará proporcionando el tratamiento probado a los participantes cuando el proyecto haya terminado, particularmente cuando los servicios de salud locales no proporcionan tales tratamientos. Esta es una cuestión importante de justicia para los participantes de investigaciones en salud en los países del Sur (países de ingresos bajos y medios), y se torna aún más crucial cuando los participantes están cerca del final de la vida y su estado de salud es incurable. Los comités de ética en investigación deben deliberar profundamente sobre estos temas al evaluar dichas propuestas.

Pasos en el balance beneficios/riesgos

Para ejecutar el balance beneficio/riesgo se evaluarán por un lado los beneficios y por otro los riesgos.

Evaluación de beneficios:

Viene determinada por

- Características de la condición: prevalencia, gravedad, historia natural, etc.
- Objetivo del tratamiento: prevención, curación, mantenimiento, etc.
- Eficacia y efectividad: basadas en las variables de resultado y en medidas características de la epidemiología (Riesgo Relativo ($RR=le/lo$), Reducción Relativa del Riesgo ($RRR=[le-lo]/lo$), Reducción Absoluta del Riesgo ($RAR=[le-lo]$), Odds Ratio ($OR=a*d/b*c$), Número Necesario a Tratar ($NNT=1/RAR$), número-necesario-para-dañar ($NND = 1 / AAR$), Aumento Absoluto del Riesgo (AAR)), donde le es el riesgo o incidencia de curación en el grupo de expuestos a la intervención y lo es el riesgo o incidencia de curación en el grupo de control o no expuestos.

Tabla 3. Tabla estándar de 2x2.

	Casos	No Casos	
Expuestos	a	b	Total de expuestos (mi)
No expuestos	c	d	Total de no expuestos (mo)
	Total de casos (ni)	Total de no casos (no)	Población total (n)

Evaluación de riesgos:

Habrà que considerar:

- Gravedad y probabilidad de que ocurra el efecto no deseado.
- Grupos vulnerables.
- Seguridad de las alternativas.
- Riesgos “paralelos” que no tienen por qué ser intrínsecos del medicamento, como pueden ser los de una prueba diagnóstica necesaria para la administración de un fármaco.
- Impacto económico e intangible, como puede ser el coste del descrédito de una empresa debido a la aparición de una alerta por efectos adversos de un medicamento. En este caso puede que la empresa ante el temor de las pérdidas por este aspecto o por temor a reclamaciones jurídicas decida voluntariamente dejar de comercializar un medicamento.

Tras la evaluación de cada una de las partes, se procederá a la evaluación del balance en sí. Esta evaluación se hará sobre condiciones de uso real, comparando con cualquier alternativa de tratamiento existente. Surge la duda de si esta evaluación debe hacerse desde una perspectiva individual (como pensará el clínico) o desde una perspectiva social (opción defendida por los decisores de las autoridades sanitarias). La valoración es cualitativa, no hay una ratio umbral que determine lo que es aceptable en cuanto al riesgo, pero a pesar de esto y de ser una decisión altamente subjetiva, está dotada del “sentido común” en el que ha de inspirarse.

Métodos antiguos

- Número-necesario-para-tratar (NNT) y número-necesario-para-dañar (NND) ^{10, 11}
 - Efecto beneficioso único, efecto adverso único
 - Principalmente útil para un único EC aleatorizado
- “Regla de Tres” ⁹
 - Tres dimensiones (gravedad, frecuencia, duración)
 - Enfermedad, efecto del fármaco en la enfermedad, efectos adversos
 - Tres efectos adversos más graves y más frecuentes

Métodos más complejos

- Basado en análisis de decisiones ^{12,13}
 - En su mayor parte basado en criterios: definir la jerarquía de criterios relevantes; evaluar la importancia relativa de estos criterios; y puntuar a los fármacos diferentes en los criterios
- Basado en modelos estadísticos
 - Modelos “Global” ¹⁴
 - Simulaciones basadas en Pacientes ¹⁵

Ningún método ha conseguido aún un uso generalizado ni en el ámbito académico ni el en ámbito regulador. Según el Informe CIOMS IV, 1998 “Es un aspecto frustrante de la evaluación beneficio-riesgo que no se haya definido y probado un algoritmo o un método de medida que combine los datos de beneficio y de riesgo, y que pueda permitir hacer comparaciones cuantitativas sencillas de las diferentes opciones de tratamiento, que por turnos puedan facilitar la toma de decisiones.” Por otra parte, las notificaciones espontáneas tienen un impacto sustancial. ^{16,17}

Estimación bayesiana del Balance Beneficio-Riesgo

Como medida de Balance Beneficio-Riesgo se estimaron los Factores de Bayes (FB):

$$\text{Factor de Bayes} = FB = \frac{\pi(\text{beneficio} | x) / p(\text{beneficio})}{\pi(\text{riesgo} | x) / p(\text{riesgo})}$$

Las decisiones se pueden ubicar según el siguiente criterio:

- | | |
|------------------------------------|---|
| Si $FB \geq 1$: | Evidencias a favor del beneficio. |
| Si $1 > FB \geq 10^{-1/2}$: | Evidencia mínima contra el beneficio. |
| Si $10^{-1/2} > FB \geq 10^{-1}$: | Evidencia sustancial contra el beneficio. |
| Si $10^{-1} > FB \geq 10^{-2}$: | Fuerte evidencia contra el beneficio. |
| Si $10^{-2} > FB$: | Decisiva evidencia contra el beneficio. |

Las estimaciones se realizan utilizando el sistema EpiDat Versión 3.0, en la opción Método→Análisis bayesiano→Dos poblaciones→Valoración de hipótesis.

Esta evaluación lleva a calificar un medicamento de la siguiente forma en base a su riesgo:

- **Riesgo aceptable** en las condiciones de uso autorizadas: será suficiente con incluir advertencias en la ficha técnica y en el prospecto.
- **Riesgo aceptable en ciertas condiciones:** habrá que definir las restricciones y controlar la prescripción y la dispensación.
- **Riesgo inaceptable:** se procederá a la retirada del mercado.

Para arribar a un juicio sobre la aceptabilidad ética de los riesgos en la investigación propuesta, es esencial que los miembros de los comités de ética identifiquen todos los riesgos y quiénes pueden verse afectados.⁴ Esta evaluación incluirá los posibles efectos sobre los participantes, pero puede implicar también la evaluación del riesgo para otros sujetos que existen por fuera del ámbito de la investigación (por ejemplo, en proyectos de investigación genómica, cuyos resultados pueden afectar a miembros de la familia de los participantes). El comité debe evaluar la probabilidad y la magnitud de los riesgos, así como la medida en que éstos pueden ser minimizados. Los miembros del comité deben entonces identificar los beneficios potenciales que se podrían esperar razonablemente, y a quiénes se agregarían probablemente estos beneficios, ya sea a los participantes y/o a otros. El Comité de Revisión y Ética debe, entonces, ejercer su juicio sobre el asunto y efectuar un balance de las dos caras de la cuestión.

Algunas formas de minimizar riesgos

Los riesgos deben ser minimizados utilizando procedimientos que sean consistentes con el diseño de investigación. Esto variará de acuerdo a la metodología empleada, sin embargo, ésta debería ser una meta prioritaria durante las distintas etapas del diseño de la investigación. Así, por ejemplo, al evaluar el diseño de un estudio, deberíamos tener en cuenta que un tamaño inadecuado de la muestra puede no conducir a resultados significativos, afectando la magnitud del Beneficio posible. En este caso, se corre el riesgo de desperdiciar el tiempo de los participantes, además de exponerlos a investigaciones cuyo resultado puede no ser aceptado por la comunidad científica en general.

De este modo, el proyecto debe reunir a un equipo de investigación con experiencia suficiente para llevar a cabo con éxito el trabajo propuesto. Esto sería particularmente importante para la investigación biomédica que involucra procedimientos invasivos, en orden a minimizar los daños físicos. Dicho proyecto, además, debería incluir personal capacitado que pueda responder a posibles emergencias. También debería contar con un plan de monitoreo de la seguridad, que incluya la protección y la confidencialidad de los datos recogidos.⁴

5. La explotación en la investigación biomédica

La salud y la investigación biomédica procuran beneficiar a todos los miembros de la sociedad, pero debe prestarse mayor atención a aquellos trabajos que involucran a personas que pueden ser consideradas vulnerables o privadas de sus derechos.¹⁸ La vulnerabilidad indica un modo de estar expuesto, susceptible de ser dañado física, psíquica o emocionalmente.^{19, 20}

Así, las personas pueden ser clasificadas como “vulnerables” si tienen una capacidad reducida para protegerse y proteger sus asuntos personales.²¹ Es posible que no tengan suficiente educación o inteligencia, poder político, recursos, fortaleza u otros atributos que sean necesarios para proteger sus propios intereses. Las Directrices Internacionales del CIOMS establecen específicamente que se debe proporcionar una justificación especial para invitar a personas vulnerables a participar de la investigación y, si son seleccionadas, debe haber una protección adecuada de sus derechos y de su bienestar.³

La investigación en salud también debe intentar beneficiar y priorizar a aquellos cuya salud es relativamente inferior debido a su posición precaria en los estratos socioeconómicos más bajos de la sociedad.¹⁸ Los sistemas socioeconómicos prevaletentes en muchos países de ingresos bajos y medios son moldeados y guiados, en gran parte, por el capitalismo y por las inclinaciones capitalistas de quienes están en el poder en esas sociedades. En muchos de esos sistemas, el principal interés del sector privado – y muchas veces el único – es maximizar

los beneficios para sí y para sus accionistas. Corresponde, por lo tanto, al Estado y a la sociedad civil procurar amortiguar los efectos plenos de las políticas económicas que no buscan beneficiar ni garantizar el bienestar de las personas de los estratos socioeconómicos más bajos. En estas circunstancias, la investigación biomédica debe centrarse en cómo mejorar las intervenciones prácticas atendiendo a esos grupos, en orden a beneficiar su salud y su condición socioeconómica. De este modo, debería darse prioridad a este tipo de investigaciones.¹⁸

La cuestión ética más importante en situaciones de vulnerabilidad y privación de derechos es la de la justicia, la cual exige una distribución equitativa de los beneficios y de las cargas de la investigación. Ningún grupo debe soportar cargas excesivas mientras otros reciben beneficios sin compartir por igual esas cargas. Existen personas o grupos específicos que tradicionalmente son considerados vulnerables, ya que tienen una capacidad mental o una libertad limitadas para consentir o rechazar su participación en la investigación.³

Este grupo incluye a niños y personas con trastornos mentales o del comportamiento que los hacen incapaces de proporcionar el consentimiento informado. También puede considerarse que algunas personas con buena capacidad mental están en situación de vulnerabilidad, como por ejemplo cuando se encuentran en posiciones inferiores en un grupo o sociedad jerárquica, o son institucionalizadas en situaciones que limitan su autonomía y sus decisiones. Los ejemplos de esto también incluyen a los estudiantes que participan el cuidado de la salud o la formación en investigación; los empleados de las empresas farmacéuticas; los miembros de las fuerzas nacionales de seguridad de un país; o las personas mayores en hogares de ancianos.

A pesar de todo lo anterior, la investigación biomédica a veces debe incluir a estas personas, ya que ellas y su enfermedad específica pueden verse beneficiadas con el resultado de la investigación en cuestión (por ejemplo, las mujeres embarazadas que pueden obtener un Beneficio a partir del estudio de determinados medicamentos o vacunas). Entonces, dependiendo de las circunstancias, las personas vulnerables no deberían ser excluidas de la investigación, sino que deberían tomarse precauciones adicionales para protegerlas y cuidar su bienestar.²¹ Asimismo, debido a que estas personas son vulnerables, deben ser protegidas de la explotación.

La explotación se refiere a un estado de cosas en el que una persona o un grupo de personas en posiciones de poder e influencia aprovechan injustamente la vulnerabilidad de otra persona o de un grupo de personas para beneficiarse a sí mismas.²²⁻²⁴ La explotación puede tener lugar a nivel individual o puede involucrar a un grupo, a una comunidad o a una población.²⁵ Cuando consideramos la posible explotación en investigación, debemos ser conscientes de que muchos países del Sur (países de ingresos bajos y medios) tienen un número relativamente mayor de personas no educadas, desahucadas y marginadas que los países del Norte, por lo que se debe hacer un mayor énfasis en el principio ético de la justicia más que en la noción norteamericana de autonomía.²⁵

La autonomía de una persona siempre debe ser respetada y siempre debe obtenerse y documentarse su consentimiento informado, pero cuando los comités de ética están evaluando una investigación de múltiples centros que involucra a países del Norte y a países de bajos y medios ingresos, deben hacer un énfasis mayor en la cuestión de la justicia, asegurando al mismo tiempo la documentación del consentimiento informado. En otras palabras, los comités de ética tienen la responsabilidad de evaluar las propuestas de investigación para determinar si lo que los patrocinadores e investigadores potenciales están proponiendo hacer es justo para los participantes y para las comunidades de las que estos provienen. Esto debería ocurrir independientemente de que estas personas pudieran dar o no su consentimiento informado.

La justicia también requiere que los comités de ética aseguren que la investigación biomédica refleja las necesidades de salud de la comunidad particular que estará involucrada en el estudio.³ En consecuencia, los miembros del comité de ética deben preguntarse: 1) En la investigación propuesta, ¿se explota a los participantes de investigación? 2) ¿Existe un equilibrio justo entre los beneficios, las cargas y los riesgos para estos participantes? 3) ¿Su comunidad está siendo sobre-investigada? 4) Si no hay beneficios directos para los participantes, ¿su comunidad local o su sociedad se beneficiarán con la investigación?

La asimetría actual e histórica de los recursos y de la influencia entre los países del Norte y los países del Sur han hecho que muchas personas en los países del Sur sean “vulnerables” y, por tanto, más susceptibles de explotación.²⁵ Los investigadores, los patrocinadores, los gobiernos de los países anfitriones, así como los

comités locales de ética en investigación, deben ser sensibles a todas estas cuestiones y procurar proteger a todas las personas de la posible explotación en la investigación biomédica.

Ejemplo 1

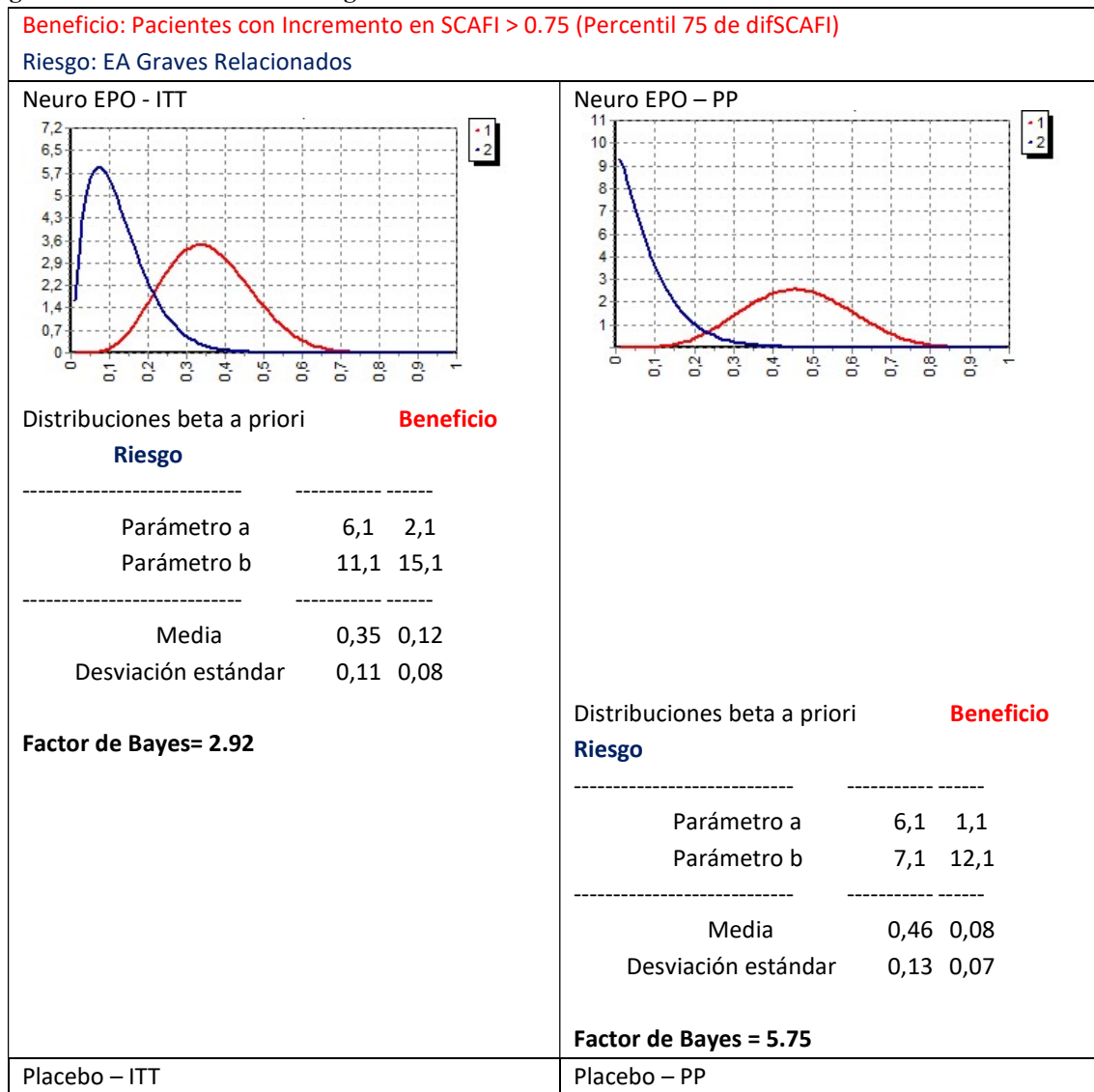
Se consideró como:

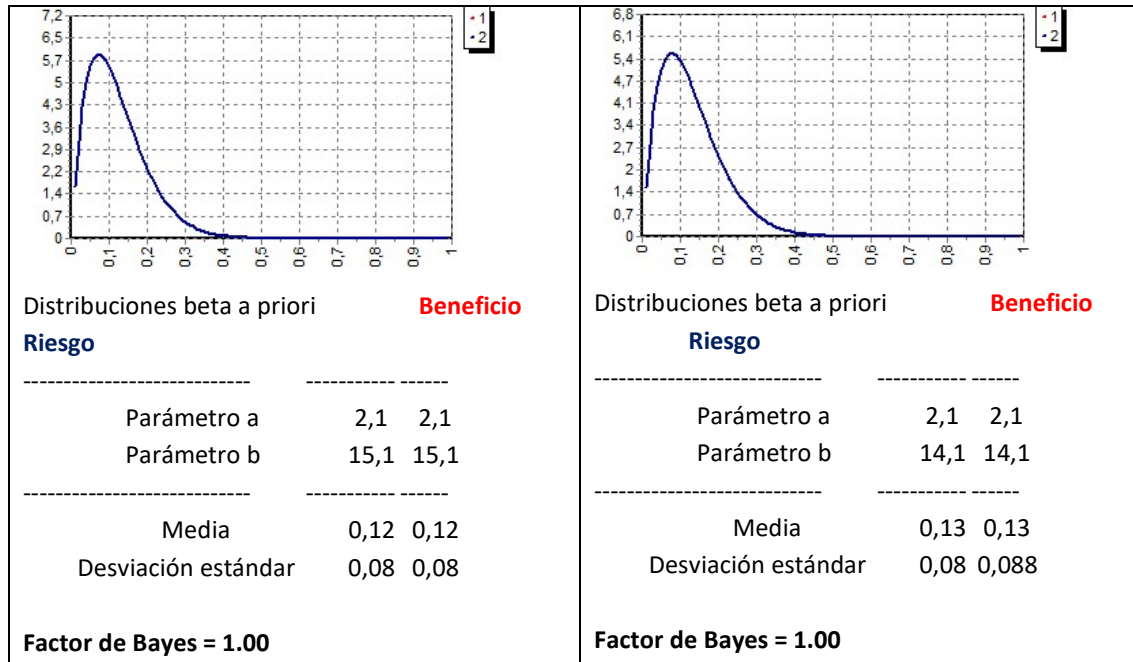
Beneficio = Pacientes con Incremento en SCAFI >0.75

Riesgo = Eventos adversos graves relacionados con el tratamiento.

En el escenario de Beneficio considerado se estima el Factor de Bayes superior a 1, indicando mayor evidencia de beneficio que riesgo, sólo en el grupo Placebo para el análisis por intención de tratar (ITT) y análisis por protocolo (PP) en el escenario de Beneficio con Índice Funcional de Ataxias Espinocerebelosa (SCAFI>0.75) el Factor de Bayes es igual a 1 indicando similar beneficio que riesgo. (Figura)

Figura 1. Análisis Beneficio-Riesgo





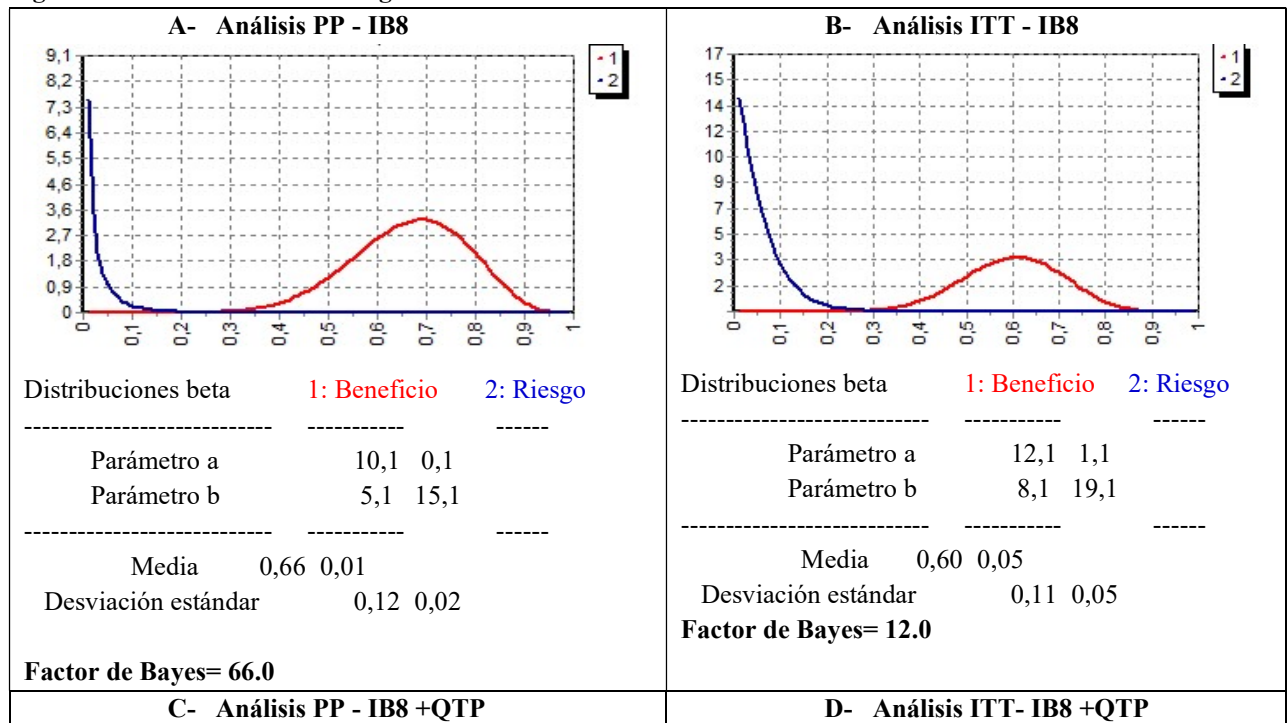
Ejemplo 2

Se consideró como:

Beneficio: Respuesta objetiva.

Riesgo: Evento adverso grave relacionado con el tratamiento

Figura 2. Análisis Beneficio-Riesgo



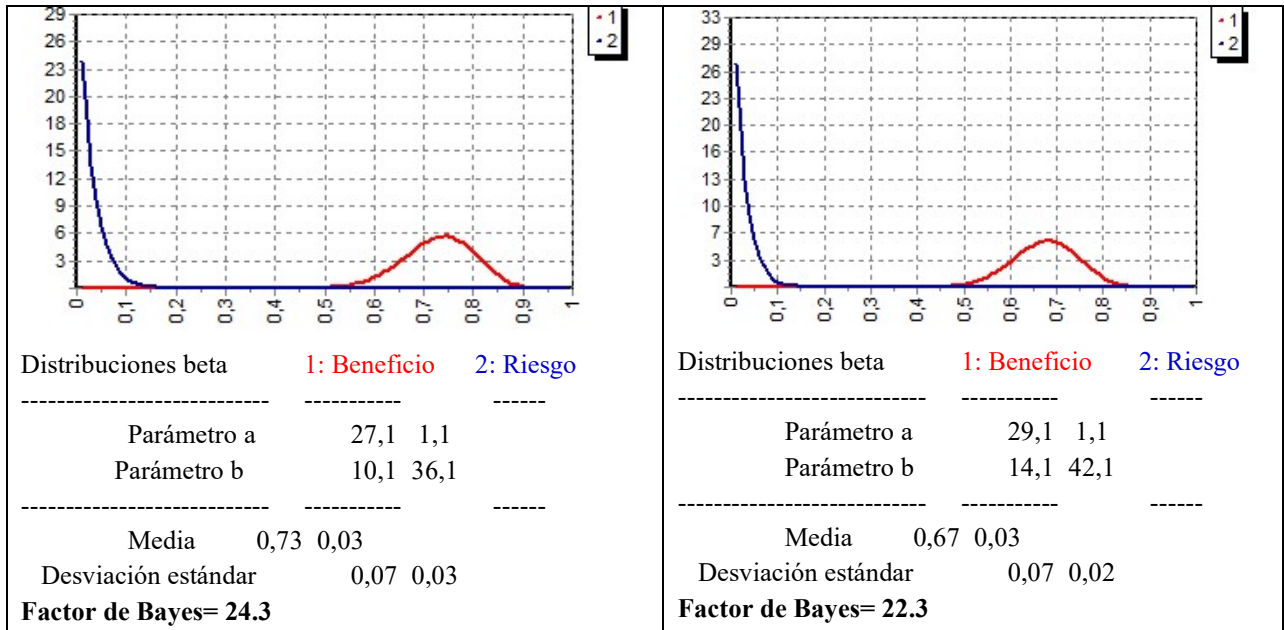


Tabla 4. Factores de Bayes derivados del Análisis Beneficio-Riesgo

	PP		ITT	
	1B8	1B8+QT	1B8	1B8+QT
Beneficio	10/15	27/37	12/20	29/43
EAG relacionado	0/15	1/37	1/20	1/43
Factor de Bayes	66.0	24.3	12.0	22.3

En la Figura 2 y Tabla 4 se resumen los diferentes escenarios considerados para el análisis Beneficio-Riesgo. Teniendo en cuenta que la serie de casos estudiados, se considera que existe un sub-registro de la información de seguridad, el escenario más honesto lo constituye el Riesgo definido como la ocurrencia de algún Evento Adverso, caso en el cual se concluye que el Beneficio de recibir el tratamiento es 24.3 veces superior (en el análisis PP) y 22.3 veces superior (en el análisis ITT) que el Riesgo de presentar algún evento adverso grave relacionado.

6. Discusión

Los comités de ética en investigación deben evaluar el nivel de riesgo para los participantes, establecer los posibles beneficios de la investigación, sopesar y hacer un balance de estos dos potenciales previsible para determinar si las propuestas particulares deberían ser aprobadas o desaprobadas. Los participantes de la investigación deben estar protegidos contra los riesgos excesivos, por lo que debe existir un marco sistemático para evaluar los riesgos, en lugar de que los investigadores y los miembros de los comités de ética dependan de sus propios juicios intuitivos.²⁶

Las directrices internacionales sobre ética establecen que la investigación clínica sólo es ética cuando el riesgo para los participantes es razonable.^{27, 28} Para cada propuesta, los miembros del comité de ética deben determinar si podría surgir más de una categoría de riesgo en la investigación en análisis (por ejemplo, riesgo de daño físico, daño psicológico, daño social, daño económico o daño legal). Luego, deben determinar el nivel de riesgo (mínimo, bajo, medio o alto) para cada uno de los posibles tipos de categorías identificados en la investigación propuesta. Deben enumerarse o delinearse en un gráfico cada una de las categorías identificadas con su respectivo nivel de riesgo proporcional.

Los miembros del comité deben tratar de determinar entonces los posibles beneficios, si los hay, de la investigación propuesta. Debe tenerse especial cuidado porque los beneficios sociales pueden ser a veces inciertos.²⁹ Estos beneficios pueden aparecer establecidos en la propuesta de investigación particular pero los miembros del comité deben llegar a tal determinación por sí mismos. Se debe documentar si los beneficios potenciales de la investigación se obtendrían para los participantes o para la sociedad, y si ese Beneficio potencial se dará en la modalidad de nuevos conocimientos o de confirmación de una investigación anterior. Debe señalarse si los participantes pueden beneficiarse directamente o indirectamente mediante posibles mejoras en las instituciones que velan por su bienestar.

Los miembros del comité de ética en investigación también deben intentar identificar si algún Beneficio podría llegar a las comunidades locales de las cuales proceden los participantes, lo cual podría tomar la forma de mejoras en los servicios de salud locales que beneficiarían la vida y el apoyo de otros miembros de la comunidad. También debe elaborarse la documentación de todos los beneficios potenciales para la sociedad en general, y esto puede generar mayor conocimiento e ideas que pueden producir nuevas investigaciones en campos relacionados.

Considerando todo lo anterior, los miembros del comité deben entonces proceder a evaluar la probabilidad de que tengan lugar los riesgos identificados y la posible magnitud de los efectos si dichos riesgos se materializaran. Una determinación similar debe realizarse para los posibles beneficios derivados del proyecto de investigación.³⁰ Todos los análisis de riesgo-Beneficio deben incluir un juicio cuantitativo de todos los riesgos percibidos, cuáles riesgos pueden considerarse aceptables y cuáles son los beneficios percibidos.³¹ Por lo tanto, los riesgos previsibles y los beneficios razonablemente previstos del trabajo de investigación deberían ser comparados analíticamente por escrito o visualmente en un gráfico, centrándose en determinar si los beneficios previstos son tan grandes en su magnitud y alcance como para compensar los diversos riesgos predecibles de que se produzcan daños, con su posible magnitud, en caso de que tuvieran lugar. Los comités de ética en investigación pueden, entonces, proceder a aprobar esas propuestas con la preponderancia de que la probabilidad y la magnitud de los beneficios potenciales derivados de la investigación superarán la probabilidad y la magnitud de los daños si tuvieran lugar, una vez que se hayan cumplido todos los otros requisitos éticos para la investigación biomédica.

7. Conclusiones

En cuanto a las limitaciones metodológicas de este tipo de análisis pueden destacarse las diferencias en la forma de estudiar los beneficios y los riesgos lo que lleva a un sesgo en la obtención de resultados. Por otra parte, sobre todo la evaluación de la eficacia se realiza en estudios de grupo y habría que plantearse cómo se distribuyen los resultados favorables en esos grupos, cuando no son aplicables análisis de subgrupos o modelos de análisis de varianza. Otra limitación a tener en cuenta son las diferencias en las variables de resultado como por ejemplo un resultado de eficacia que puede ser la morbi-mortalidad frente a un resultado de seguridad como la aparición de efectos adversos. Todo esto hace que sean difícilmente aplicables las aproximaciones cuantitativas que anteriormente se planteaban, siendo la solución más práctica emplear una aproximación de tipo cualitativo, pero con sentido clínico.

Los investigadores, al preparar los documentos de consentimiento informado para la inscripción de los participantes y los protocolos para la revisión ética, deberían intentar explicitar todos los riesgos previsibles asociados a sus proyectos de investigación.³²⁻³⁶ Todos los potenciales beneficios identificados en el documento deben referirse claramente a los participantes individuales, a sus sistemas de salud (o algo similar), a las personas de su comunidad local o a la sociedad en general. Las consideraciones pertinentes a este proceso de equilibrio deben tener peso ético.

Los comités de ética también deben considerar las cuestiones asociadas a la justicia y asegurarse de que la investigación refleja las necesidades de salud de la población a ser estudiada. Será responsabilidad de los comités de ética garantizar que las comunidades dentro de su jurisdicción particular no sean sobre-investigadas, y no deberán aprobar investigaciones donde los participantes puedan ser explotados. Particularmente, en los

países de ingresos bajos y medios del Sur, las investigaciones propuestas deberían beneficiar directamente a los participantes, a su comunidad local o a la sociedad en su conjunto.

Referencias

- [1] Aarons D. Research ethics. *West Indian Med J.* 1995; 44(4):115-8.
- [2] Aarons DE. Issues in bioethics: teaching research ethics. *West Indian Med J.* 2003; 52(2):145-50.
- [3] Caribbean Public Health Agency. Research ethics committee. CARPHA. Trinidad & Tobago; 2014. p. 11.
- [4] Aarons D. Ethical issues surrounding body integrity and research. *West Indian Med J.*
- [5] 2014; 63(5):399-400b
- [6] Aarons DE. Research: an ethical answer in addressing our people's health problems and inequities. *West Indian Med J.* 2015;64 (Suppl 2); 97-100.
- [7] Aarons DE. Family health, public health, and vulnerability in research: a Caribbean perspective. *Revista Redbioética/Unesco.* 2016a; 7(14).
- [8] Aarons DE. Risk/benefits balance and exploitation in biomedical research. *Annals of the 6th International Congress of RedBioética/Unesco.* Costa Rica: Ministry of Health; 2016b.
- [9] Arora SK, Shah D, Chaturvedi S, Gupta P. Defining and measuring vulnerability in young people. *Indian J Community Med.* 2015 [acceso 22 fev 2017]; 40(3):193-7. DOI: 10.4103/0970-0218.158868
- [10] Clarke A, Deeks JJ, Shakir SAW. An assessment of the publicly disseminated evidence of safety used in decisions to withdraw medicinal products from the UK and US markets. *Drug Safety,* 2006; 29(2): 175-181.
- [11] Council for International Organizations of Medical Sciences. International guidelines for biomedical research involving human subjects. Geneva: CIOMS; 2016 [acceso 20 fev 2017].
- [12] Disponible: <http://bit.ly/2p7Mwar>
- [13] Dal-Ré R, Rid A, Emanuel E, Wendler D. The potential exploitation of research participants in high income countries who lack access to health care. *Br J Clin Pharmacol.* 2016 [acceso 22 fev 2017]; 81(5):857-64. DOI: 10.1111/bcp12879
- [14] Edwards IR, Wiholm B-E, Martinez C. Concepts in Risk-Benefit Assessment. *Drug Safety,* 1996; 15(1): 1-7.
- [15] Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *Jama.* 2000; 283(20):2701-11.
- [16] Felli JC, Noel RA, Cavazzoni, PA. A Multiattribute Model for Evaluating the Benefit-Risk Profiles of Treatment Alternatives. *Medical Decision Making,* 2009; 29(1): 104-115.
- [17] Fischhof B, Slovic P, Lichtenstein S, Read S, Combs B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences.* 1978; 9(2):127-52.
- [18] Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine,* 1988; 318(26): 1728-1733.
- [19] Mancini G, Schulzer M. Reporting Risks and Benefits of Therapy by Use of the Concepts of Unqualified Success and Unmitigated Failure. *Circulation,* 1999; 99: 377-383.
- [20] Miller FG, Jofe S. Limits to research risks. *J Med Ethics.* 2009; 35(7):397-9.
- [21] Mussen F, Salek S, Walker, S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiology and Drug Safety,* 2007; 16(S1): S2-S15.
- [22] National Advisory Board on Research Ethics. Ethical principles of research in the humanities and social and behavioral sciences and proposals for ethical review. The National Advisory Board on Research Ethics' Proposals. Helsinki; 2009. p. 3.

- [23] Olivier P, Montastruc JL. The nature of the scientific evidence leading to drug withdrawals for pharmacovigilance reasons in France. *Pharmacoepidemiology and Drug Safety*, 2006; 15(11): 808-812
- [24] Prat B, Lwin KM, Zion D, Nosten F, Lof B, Cheah PY. Exploitation and community engagement: can advisory boards successfully assume a role minimizing exploitation in international research? *Dev World Bioeth*. 2015 [acceso 22 fev 2017]; 15(1):18-26. DOI: 10.1111/dewb.12031
- [25] Snyder J. Exploitations and their complications: the necessity of identifying the multiple forms of exploitation in pharmaceutical trials. *Bioethics*. 2012 [acceso 22 fev 2017]; 26(5):251-8. DOI: 10.1111/j.1467-8519.2010.01847
- [26] Sutton A, Cooper N, Abrams K, Lambert P, Jones D. A Bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty. *Journal of Clinical Epidemiology*, 2005; 58(1): 26-40.
- [27] van Staa TP, Cooper C, Barlow D, Leufkens HG. Menopause. 2008 Mar-Apr;15(2):374-81. Individualizing the risks and benefits of postmenopausal hormone therapy. *Menopause*, 2008; 15(2): 374-381.
- [28] Straehle C. Vulnerability, health agency and capability to health. *Bioethics*. 2016; 30(1):34-40. DOI: 10.1111/bioe.12221
- [29] Renn O. Three decades of risk research: accomplishments and new challenges. *J Risk Res*. 1998 [acceso 2 mar 2017]; 1(1):49-71. Disponible: <http://dx.doi.org/10.1080/136698798377321>
- [30] Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. *Jama*. 2010; 304(13):1472-9.
- [31] The National Committee for Research Ethics in the Social Sciences and the Humanities. Guidelines for Research Ethics in the Social Sciences, Law and the Humanities. Norway; 2006. p. 12.
- [32] Weijer C, Miller PB. Refutng the net risks test: a response to Wendler and Miller's 'Assessing research risks systematically.' *J Med Ethics*. 2007; 33(8):481-6.
- [33] Weijer C, Miller PB. When are research risks reasonable in relation to anticipated benefits? *Nat Med*. 2004; 10(6):570-3.
- [34] Wendler D, Miller FG. Assessing research risks systematically: the net risks test. *J Med Ethics*. 2007; 33(8):481-86.
- [35] Working CIOMS Group IV. Benefit-Risk Balance for Marketed Drugs: Evaluating Safety Signals. CIOMS: Geneva, 1998.

UNA REVISIÓN DE MÉTODOS PARA EL CLUSTERING ESPACIAL EN POBLACIONES CON ESTRUCTURA COMPLEJA: ALGORITMOS BASADOS EN LA CONECTIVIDAD Y EN LA FUNCIÓN DE DENSIDAD.

Sira M. Allende Alonso¹ y Carlos N. Bouza Herrera¹

¹Universidad de La Habana

ABSTRACT

In these notes we present a critical review of clustering methods based on connectivity and density functions. They should be used when dealing with complex populations, where metric criteria do not describe the grouping of objects adequately. A discussion on their performance is developed.

KEYWORDS: Connectivity, Density, DBSCAN, DENCLUE, OPTICS, Kernel.

RESUMEN

En estas notas se hace una revisión crítica de métodos basados en la conectividad y en la función de densidad. Estos deben ser usados al enfrentarse con poblaciones complejas, en las que criterios del tipo métrico no describen adecuadamente la forma en que se agrupan los objetos. Una discusión sobre su desempeño es llevada a cabo.

PALABRAS CLAVE: Conectividad, Densidad, DBSCAN, DENCLUE, OPTICS, Kernel.

1. INTRODUCCIÓN.

Es común la necesidad de hacer estudios sobre la forma en que objetos se agrupan en el espacio. El estudio de clustering espacial está asociado a complejidades particulares. En ellos es requerido frecuentemente hacer estudios sobre el tejido social. La distancia física no describe como los individuos, familias y otros conglomerados humanos, diseminados en el espacio físico dado por una ciudad, un país etc. se relacionan por su similitud. Los estudios, al considerar que cada objeto está asociado a múltiples variables llevan a enfrentar problemas del tipo Big Data, pues el número de variables va a ser muy grande. Los descriptores de los individuos son obtenibles de diversas fuentes: datos censales, tarjetas de crédito, redes sociales etc. Así al considerar que los individuos no están agrupados espacialmente el concepto de similitud deja de ser naturalmente describible por alguna función de distancia. Es necesario usar otro tipo de funciones para hacer la descripción de cómo se agrupan. Así nos enfrentamos a la necesidad de determinar clústeres utilizando métodos que no dependan de distancias, métricas etc.

Un enfoque que permite superar las limitaciones de criterios del tipo métrico es utilizar criterios basados en la conectividad y la densidad.

En este trabajo se presenta este enfoque. Se discuten varios métodos de común uso.

En la próxima sección se hace una discusión sobre métodos basados en la conectividad y la función de densidad, estableciéndose como estos se dividen en paramétricos, mezcla, no-paramétricos e histograma y como estos permiten modelar la determinación de clústeres a partir de modelos. La sección 2 presenta el algoritmo DBSCAN (Density Based Spatial Clustering of Applications with Noise). Este fue propuesto por Ester, Kriegel

et al. (1996) y fue el primero de los desarrollados usando la noción de densidad de un clúster. Esta es seguida por la presentación de DENCLUE (DENSITY-based CLUstERING) el que hace uso de funciones de influencia, las que provienen de conceptos de la estadística robusta. El método propuesto por Ankerst, Breunig, Kriegel, and Sander es discutido en la siguiente sección seguida por presentaciones similares para los métodos del tipo jerárquico basado en densidades y el uso de Kérneles.

Una descripción de los algoritmos es desarrollada y se hacen comparaciones entre métodos basados en la densidad y con otros métodos no basados en ella.

2. Métodos Basados en la conectividad y la función de densidad

Estos métodos asumen que los clústeres son regiones de alta densidad en el espacio de la data dada por

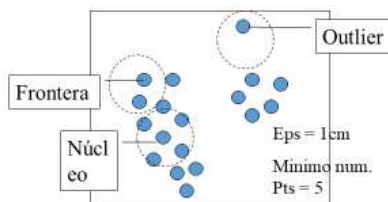
$$\vec{x}_t \in X \subset \mathbb{R}^d, d \in \mathbb{N}, t = 1, \dots, N$$

Tenemos como enfoques para construirlos en este contexto:

- Paramétricos
- Mezcla
- No-paramétricos
- Histograma

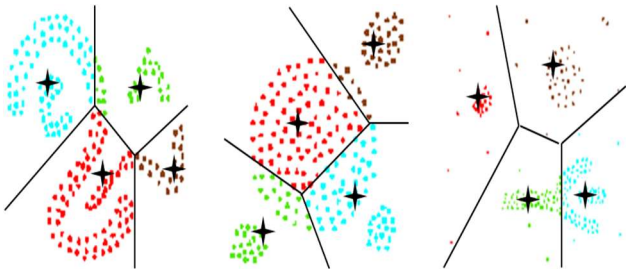
Los algoritmos basados en densidad son usados para construir clústeres en regiones densas de objetos en el espacio, donde se encuentran los datos que están separados por regiones de baja densidad. Los elementos aislados representan ruidos. Este tipo de método es muy útil para filtrar ruidos y obtener clústeres de diversas formas. Requiere de muchos puntos y su complejidad es $O(n^2)$. En el clustering basado en tales criterios (local cluster criterion) cada clúster de puntos es más denso que los que están fuera de él

Vea en la figura siguiente un ejemplo de la detección de puntos outliers, frontera y núcleo (core).

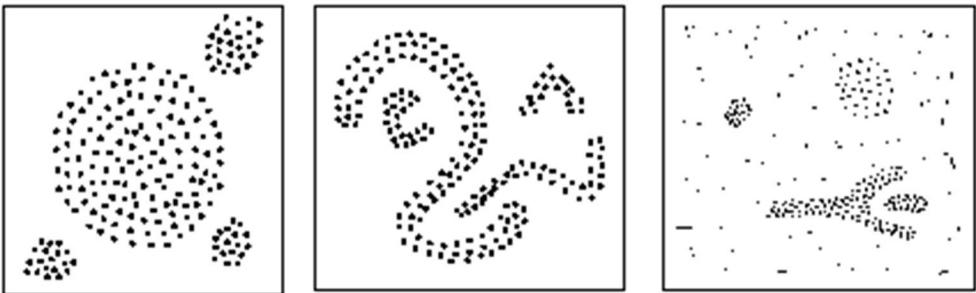


Detección de puntos extremos

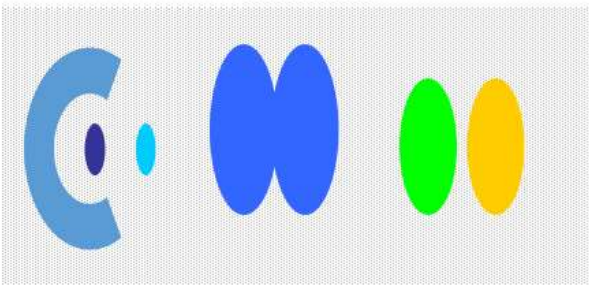
En las figuras próximas se muestran ejemplos de agrupamiento basados en densidad.



Resultados con un algoritmo K-medoides : $K=4$



Resultados con un algoritmo basado en la densidad



Uso de la densidad par construir clústeres

Los métodos típicos en este contexto son DBSACN, OPTICS, DENCLUE. Estos se basan en la conectividad local y funciones de densidad. Cada clúster posee una mayor densidad de puntos dentro de él que la que hay fuera de este. Su comportamiento es notablemente mejor en el descubrimiento de clústeres de forma arbitraria, en el tratamiento de los ruidos. Esto es, los clústeres son regiones densas en el espacio de la data, separados por regiones de baja densidad. Formalicemos las ideas básicas usando la intuición.

Para cada punto de un clúster la densidad local de los objetos a su alrededor debe exceder un punto umbral.

El conjunto de puntos formado, estará espacialmente conectado. La densidad local de un punto “p” es determinado por dos parámetros:

- ϵ = Máximo radio de la vecindad.
- $MinPts$ = Número Mínimo de puntos en una Eps-vecindad del punto p.

Fijemos algunas definiciones necesarias.

1. La ϵ -vecindad está formada por los objetos en un radio ϵ con centro en: $p = N_\epsilon(p) = \{q \in D \mid dist(p,q) \leq \epsilon\}$
2. Un punto q es un punto-core (punto núcleo) con $(\epsilon, MinPts)$ si $|N_\epsilon(q)| \geq MinPts$
3. El punto p es directamente densidad-alcanzable si existe una cadena de puntos $p_1, \dots, p_n, p_1 = q, p_n = p$ tales que p_i es directamente densidad-alcanzable desde p_{i+1} .

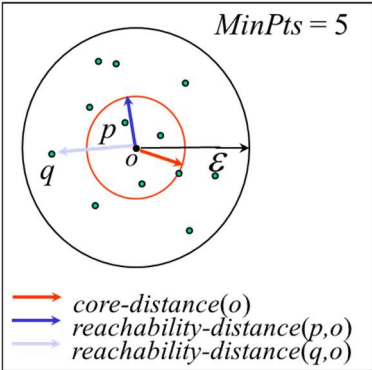


Ilustración del significado de varios puntos

Otra forma de plantear esto es decir que un objeto q será directamente densidad-alcanzable desde p si este es un objeto-core y q está en su ϵ -vecindad.

4. Se dice que un punto p esta densidad-conectado con otro q para $\epsilon, (MinPts)$ si existe un punto o p y q son densidad-alcanzables desde para $(\epsilon, MinPts)$.
5. La ϵ -vecindad es de alta densidad de p si contiene al menos $MinPts$ objetos

Ejemplo Sea $MinPts = 4$. La vecindad de q aparece en la siguiente figura.

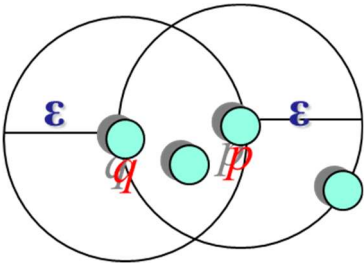


Ilustración del significado de una vecindad de q

Ejemplo. Sean $MinPts = 3$, $Eps = \text{radio de los círculos}$. En este caso M, P, O y R son core-objetos pues cada uno es una Eps -vecindad que contiene al menos 3 puntos. Vea la figura siguiente para ilustrar el comportamiento de estos puntos

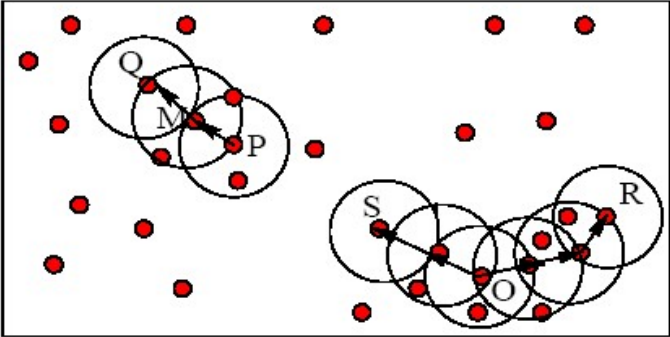


Ilustración de objetos en un ejemplo

Vea un ejemplo en el que q es directamente densidad-alcanzable desde p y p es indirectamente densidad-alcanzable desde q

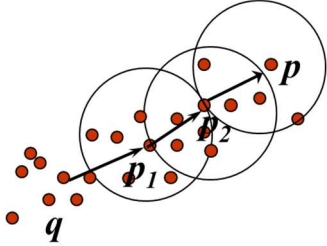


Ilustración del significado de densidad-alcanzable e indirectamente densidad-alcanzable

Note que la densidad-alcanzabilidad es asimétrica y no es suficiente para describir los clústeres y la densidad-conectividad es simétrica.

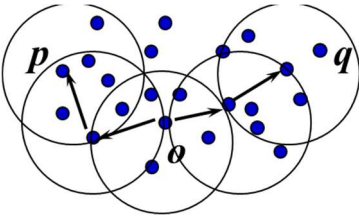


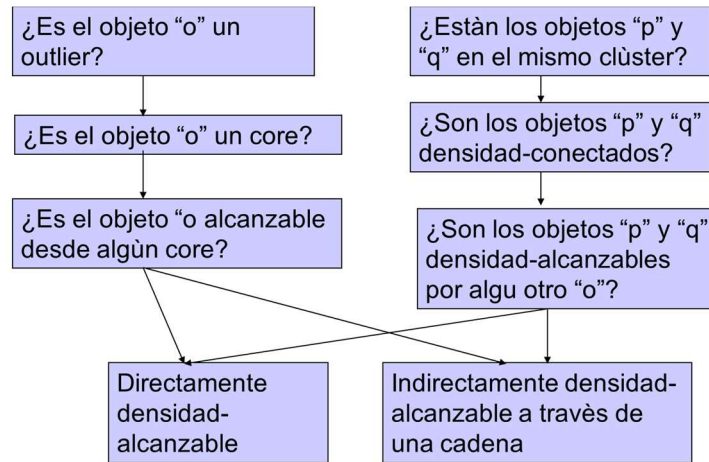
Ilustración de conectividad

Ahora podemos dar una definición formal de Clúster:

Dada una data D , los parámetros ϵ y $MinPts$ un clúster C es un subconjunto de objetos que satisfacen:

1. *Conectividad*: p, q pertenecen a C si están densidad-conectados.
2. *Maximalidad*: si $p \in C$ y q es densidad-alcanzable desde p , entonces $q \in C$. (Eso evita la redundancia).

Podemos dar una conceptualización del proceso de Clustering



Proceso de construir clústeres en el contexto dado por el uso de densidades

3. DBSCAN (Density Based Spatial Clustering of Applications with Noise)

Ese es el primer método desarrollado para hacer clústeres a partir de densidades. Fue propuesto por Ester, Kriegel et al. (1996). Se basa en la noción de densidad de un **clúster**:

Un clúster es un conjunto maximal de puntos densidad-conectados.

El método puede descubrir clúster de formas arbitrarias en el espacio de las datas en presencia de ruidos.

El algoritmo GDBSCAN (Generalized Density Based Spatial Clustering of Applications with Noise) busca formas densas en zonas geográficas. Se reconoce un clúster en una zona al evaluar si se rebasa cierto límite prefijado (threshold) de densidad. El valor de la densidad es determinado a partir de una relación entre el número de objetos y el área que ocupan en una zona perteneciente a un gran y único clúster. Los clústeres van a ser objetos o cuerpos geométricos, con sus características propias. Por ejemplo, su baricentro, el área total, la densidad, etc.

Todos los objetos pueden compararse entre sí con el objetivo de establecer relaciones de distancia entre los centros de los clústeres, si hay un solapamiento entre regiones, etc. Al considerar esto se pueden hacer fusiones y reorganizar los distintos grupos. Se realiza un proceso iterativo, y para cada objeto de la base de datos se establece si pertenece o no a alguno de los clústeres construidos, y la distancia a la que se encuentra de todos ellos. Si el objeto no se asigna a ningún grupo, se considera es un ruido y se pasa a analizar el siguiente elemento. Así se hace hasta completar un analizar todos los objetos. Si la condición de parada no se cumple y se repite cambiando las condiciones iniciales. El resultado final será un mapa de densidades, en el que las zonas con mayor densidad de objetos se agrupan para formar clústeres-geométricos. Estos tienen sus propias características que determinan su forma, densidad y situación en el espacio de coordenadas.

densidad-alcanzable

Este funciona como sigue:

- Arbitrariamente seleccionar un punto p
- Coleccionar todos los puntos densidad-alcanzable desde p para un Eps y un $MinPts$.
 - Si p es un punto- core se forma un clúster.
 - Si p es un punto frontera no hay puntos densidad-alcanzable desde p y revisita otro punto de la data.
 - Si p es un punto de ruido eliminar
- Continuar hasta que se procesen todos los puntos

Vea una ilustración en la figura siguiente.

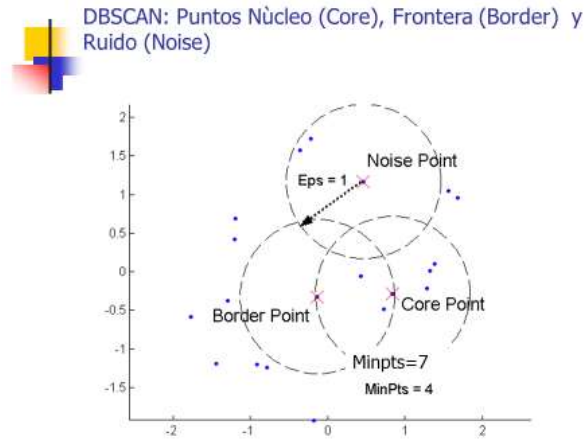


Ilustración de los puntos

Vea un algoritmo simplificado

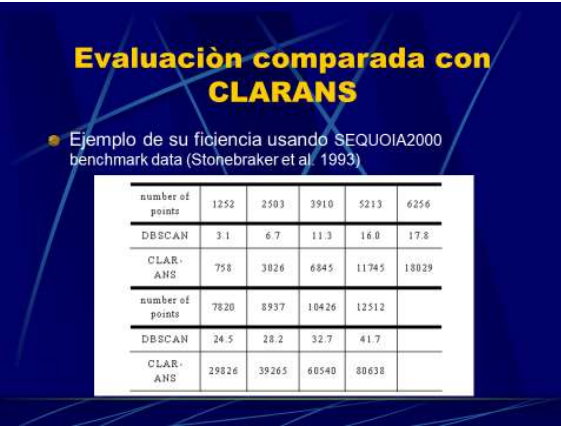
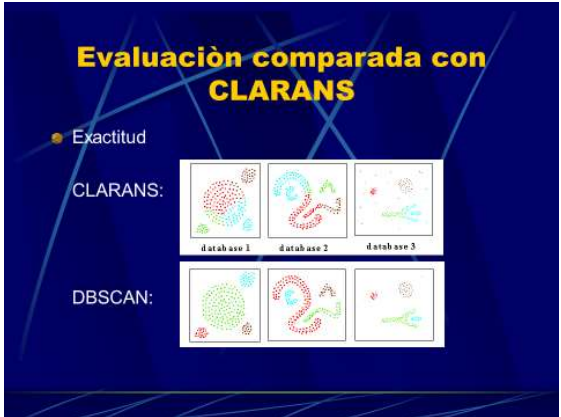
```

current_cluster_label ← 1
for all core points do
    if the core point has no cluster label then
        current_cluster_label ← current_cluster_label + 1
        Label the current core point with cluster label current_cluster_label
    end if
    for all points in the Eps-neighborhood, except  $i^{th}$  the point itself do
        if the point does not have a cluster label then
            Label the point with cluster label current_cluster_label
        end if
    end for
end for

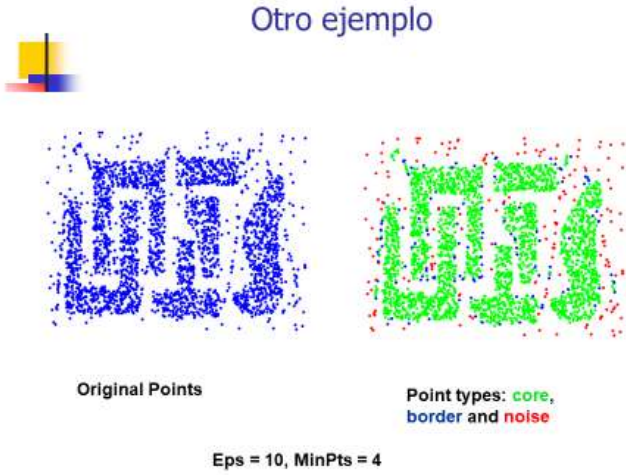
```

DBSCAN es sensible a los valores que se fijan de los parámetros. Generalmente toma en su procesamiento un tiempo $O(n \log n)$.

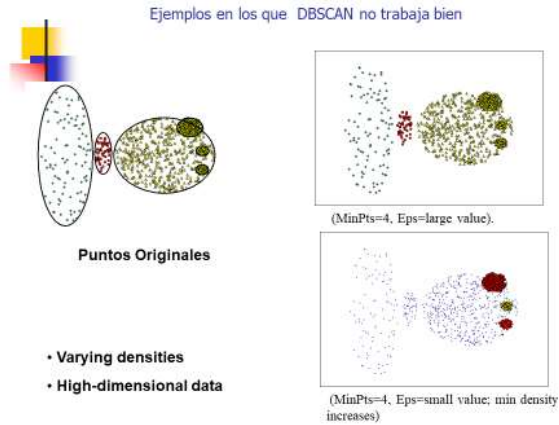
Podemos ver cómo se comporta respecto a CLARANS en las figuras siguientes.



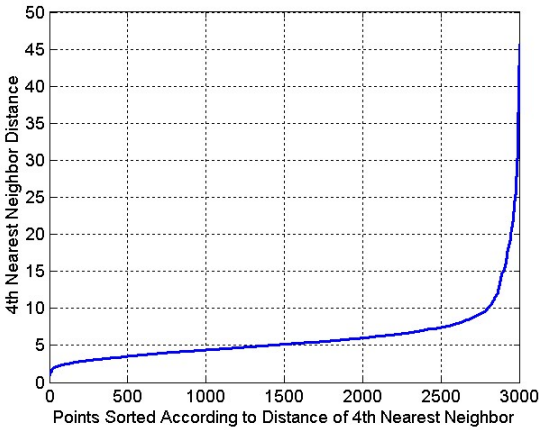
Como se ve DBSCAN es más eficiente que CLARANS por un factor de más de 100 en términos de eficiencia usando la data SEQUOIA 2000 benchmark.



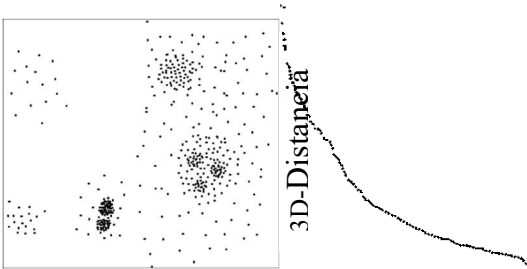
Este método no trabaja bien cuando las densidades son variables o los datos son de alta dimensionalidad. Tal es el caso ilustrado a continuación.



Podemos usar una Heurística para determinar EPS y MinPts. La idea es que los puntos de un clúster y sus K vecinos más cercanos están más o menos a misma distancia. Los puntos ruidosos van a estar muy distantes de estos. Así que plotamos la distancia de cada punto a sus vecinos. Si usamos $K=4$ y $Eps=10$ tendremos el gráfico siguiente:



Esto no siempre funciona adecuadamente. Veamos un ejemplo problemático.



4. DENCLUE (DENsity-based CLUstEring)

Este método utiliza las funciones de influencia para catalogar la influencia que cada objeto ejerce sobre los elementos cercanos. Estas funciones de influencia son muy similares a las funciones de activación $g(t, r)$ que

se usan en las redes neuronales. La densidad se computa como la suma de todas las funciones de influencia de todos los objetos de la base de datos Z.

$$densidad = \sum_{(t,r) \in Z} g(t,r)$$

Los clústeres se determinan al detectar los atractores, o máximos locales de la densidad. Este algoritmo de agrupamiento es robusto, permitiendo manejar datos ruidosos o erróneos.

Así, superado cierto valor umbral de distancia entre objetos, la salida del programa cambia de un estado a otro. Usualmente se reporta si el estado es inactivo (0) y otro activo (1). Se modela la densidad a partir de la noción de influencia desarrollada en la Estadística Robusta. Cada objeto ejerce influencia en su vecindad. La influencia decrece con la distancia. Por ejemplo, si consideramos cada objeto entran en consideración el radio, la cercanía y la magnitud del ruido. La influencia se representa por una función matemática, la influencia de y sobre x para una constante σ es

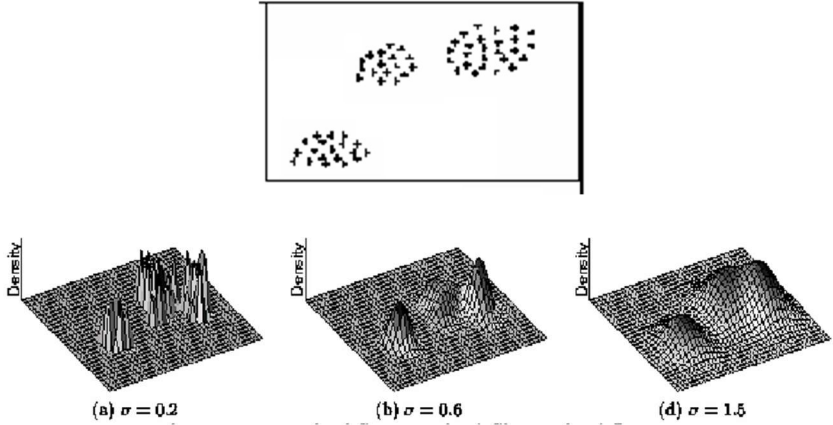
$$f_{square}^y(x) = 0, \text{ si } dist(x,y) > \sigma, \text{ (1 en otro caso)}$$

El valor umbral es una función de activación que es usualmente la gaussiana o la sigmoideal.

Veamos el caso Gaussiano.

$$f_{Gaussian}^D(x) = \sum_{i=1}^N e^{-\frac{d(x,x_i)^2}{2\sigma^2}}$$

Veamos un ejemplo para varios valores de σ en una misma data.



Función de activación Gaussiana bivariada para $\sigma=0,2 \ 0,6 \ 1,5$

El gradiente de una Gaussiana es determinado como sigue. La densidad es

$$f_{Gaussian}(x,y) = e^{-\frac{d(x,y)^2}{2\sigma^2}}$$

y la pendiente (gradiente) está dada por

$$\nabla f_{Gaussian}^D(x,x_i) = \sum_{i=1}^N (x_i - x) \cdot e^{-\frac{d(x,x_i)^2}{2\sigma^2}}$$

Los Clústeres pueden ser determinados matemáticamente identificando los llamados atractores de la densidad. Estos son localmente maximales sobre todas las funciones de densidad. Veamos unos atractores en la figura siguiente.

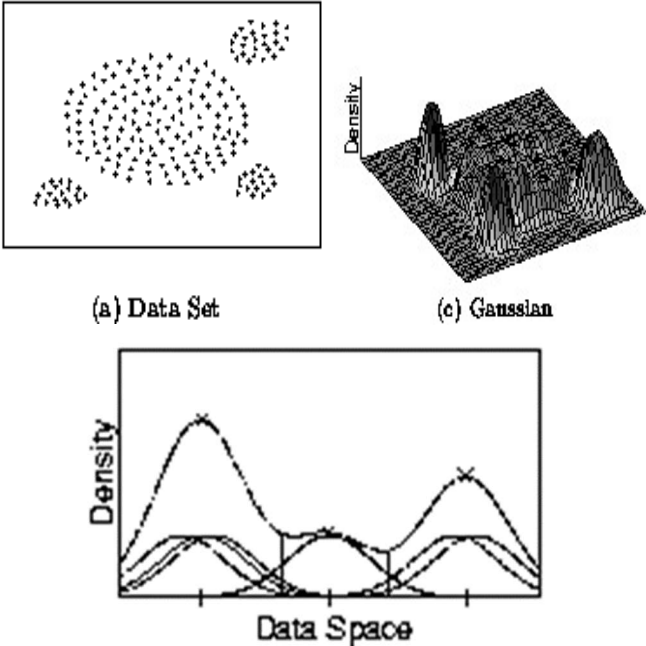


Ilustración de atractores Gaussianos y la densidad

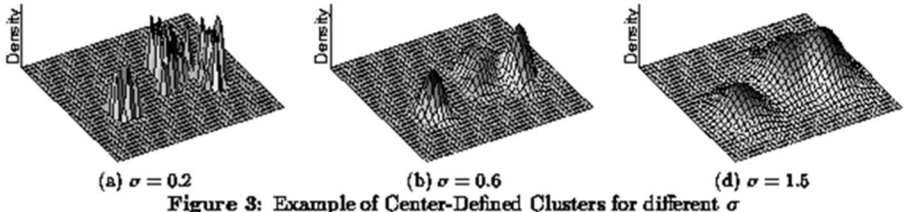


Figure 3: Example of Center-Defined Clusters for different σ

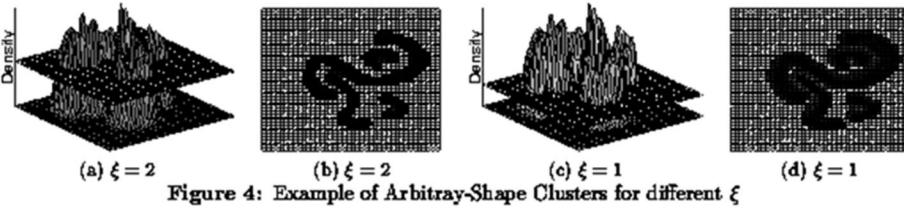


Figure 4: Example of Arbitrary-Shape Clusters for different ξ

Un grupo de clústeres son definidos por sus centros y los puntos se comunican mediante un sendero P . Se evalúa para cada objeto x sobre P la densidad $(x) \geq \xi$. El espacio se divide en rejillas de tamaño 2σ . Solo se consideran rejillas con muchos objetos. Para cada objeto se calcula su atractor usando un método del tipo “hill

climbing”. A partir de la estimación de la densidad diferenciamos y obtenemos la solución como sigue:

$$\begin{aligned}\nabla \hat{p}(\vec{x}) &= \frac{1}{h^{d+2}N} \sum_{t=1}^N K\left(\frac{\vec{x}-\vec{x}_t}{h}\right) \cdot (\vec{x}_t - \vec{x}) = 0 \\ \Rightarrow \vec{x} &= \frac{\sum_{t=1}^N K\left(\frac{\vec{x}-\vec{x}_t}{h}\right) \vec{x}_t}{\sum_{t=1}^N K\left(\frac{\vec{x}-\vec{x}_t}{h}\right)}\end{aligned}$$

Esto no es una solución, pero es usada para comenzar las iteraciones. Así el próximo paso es calcular

$$\vec{x}^{(l+1)} = \frac{\sum_{t=1}^N K\left(\frac{\vec{x}^{(l)}-\vec{x}_t}{h}\right) \vec{x}_t}{\sum_{t=1}^N K\left(\frac{\vec{x}^{(l)}-\vec{x}_t}{h}\right)}$$

Los atractores formarán la base para conformar todos los clústeres. El modelador debe introducir algunas innovaciones para evitar calcular los atractores para todos los puntos. Los Clústeres serán determinados por los máximos locales de la estimación de la densidad.

Este método tiene problemas relacionados con el tamaño del paso lo que afecta la eficiencia. Debe evitarse tener pasos pequeños innecesarios. Puede no ser efectivo por no converger a un óptimo local. Vea el siguiente ejemplo.

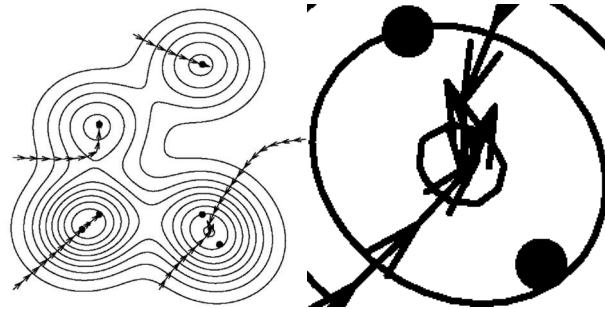


Ilustración de una baja eficiencia por no converger

Una variante mejorada es DENCLUE 2.0 Hill Climbing. Este es más eficiente ajustando automáticamente el tamaño del paso sin incrementar los costos. Posee un fundamento sólido desde el punto de vista matemático. También trabaja muy bien para datos con gran cantidad de ruido. Así que describen muy bien clústeres de formas arbitrarias en el caso de una dimensión alta. Es significativamente más rápido que DBSCAN, al que supera por un factor que llega a ser hasta del 45%. En su contra tiene el depender de un mayor número de parámetros.

Este método converge efectivamente lo que ha sido probado matemáticamente. El método hace uso de una nueva heurística del tipo Hill climbing dentro de la versión inicial de DENCLUE. Permite aplicar métodos EM para incrementar la aceleración como sigue.

$$\text{E-step } Q^{k+1} = \operatorname{argmax}_Q F(Q^k, \theta^k)$$

$$\text{M-step } \theta^{k+1} = \operatorname{argmax}_\theta F(Q^{k+1}, \theta^k)$$

Se asume que la data observada proviene de un cierto modelo. Se formulan los parámetros del modelo y hace uso de este para predecir los datos faltantes en el paso “expectation”. Con la data faltante y la observada se halla los que se consideren mejores valores de los parámetros en el paso “maximization” usando el método de Máxima Verosimilitud para estimar. Se desarrollan las iteraciones necesarias hasta obtener la convergencia.

Vea una comparación de este con la versión 1.0

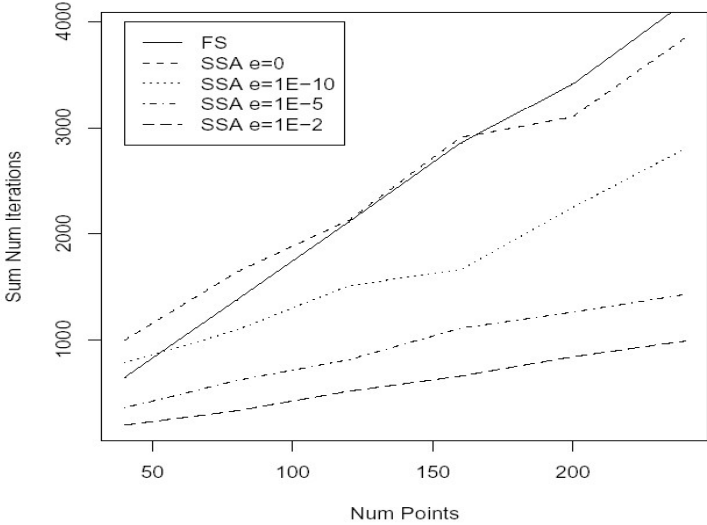
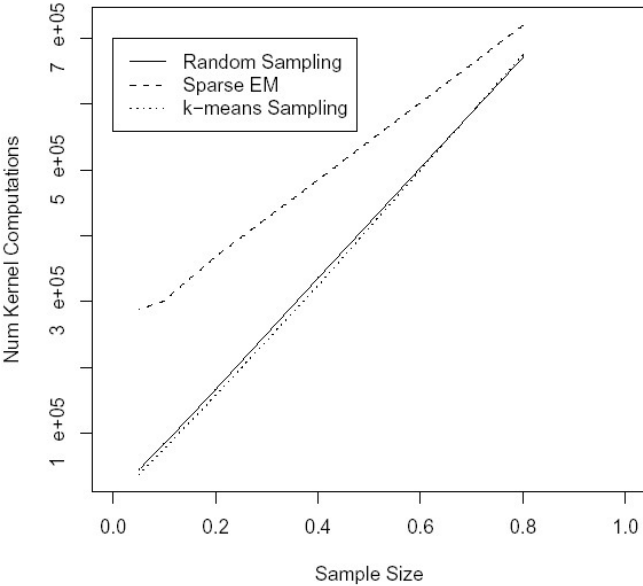


Ilustración con 16-dim. Data artificial

Una comparación de la aceleración de respecto a algunos métodos se puede ver en la figura siguiente.



Calidad en términos de la velocidad de los métodos del clustering DENCLUE 1.0 (NMI), 2.0 (SSA) y k-Medias con data real.

	k-means	SSA	Random Sampling	Sparse EM	k-means Sampling
iris	0.69±0.10	0.72	0.8: 0.66±0.05	0.8: 0.68±0.06	0.8: 0.67±0.06
			0.4: 0.63±0.05	0.4: 0.60±0.06	0.4: 0.65±0.07
			0.2: 0.63±0.06	0.2: 0.50±0.04	0.2: 0.64±0.07
ecoli	0.56±0.05	0.67	0.8: 0.65±0.02	0.8: 0.66±0.00	0.8: 0.65±0.02
			0.4: 0.62±0.06	0.4: 0.61±0.00	0.4: 0.65±0.04
			0.2: 0.59±0.06	0.2: 0.40±0.00	0.2: 0.65±0.03
wine	0.82±0.14	0.80	0.8: 0.71±0.06	0.8: 0.72±0.07	0.8: 0.70±0.11
			0.4: 0.63±0.10	0.4: 0.63±0.00	0.4: 0.70±0.05
			0.2: 0.55±0.15	0.2: 0.41±0.00	0.2: 0.58±0.21

Evaluación numérica de la calidad en términos de la velocidad de los métodos del clustering DENCLUE 1.0 (NMI), 2.0 (SSA) y k-Medias con varas datas reales.

5. OPTICS – Ankerst, Breunig, Kriegel, and Sander (SIGMOD'99)

Este método no produce clústeres explícitamente. Este genera una ordenación de los objetos representado los clústeres en base a una estructura de la densidad. La información que determina sobre los clústeres-ordenados es equivalente a los que se determinan usando criterios basados en la densidad correspondientes a una amplia gama de parámetros. Es Bueno para trabajar tanto automáticamente como interactivamente para hacer un análisis de clustering que incluya hallar elementos intrínsecos de la estructura del clustering. puede ser representado gráficamente o usar técnicas de visualización. Al ordenar los puntos a partir de la más corta distancia de alcanzabilidad se garantiza que los clústeres mayores sean construidos primeros para un valor constante de MinPts, y que para incrementar la densidad se requiere disminuir ϵ .

Los resultados son fácilmente analizados y no dependerán de la dimensión de la data. Por otra parte, es relativamente insensible a la fijación de los parámetros.

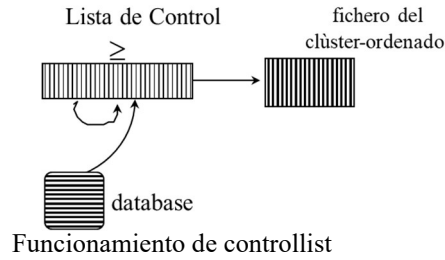
Es básico controlar la estructura lo que hace partir del procedimiento “controlList”. Este memoriza la menor distancia de alcanzabilidad (“distancia de un salto a ese punto”). Se visita cada punto y siempre se ejecuta el menor salto. La salida es una ordenación de los puntos, determinan los puntos-core, la distancia de alcanzabilidad. Vea un algoritmo simple.

```

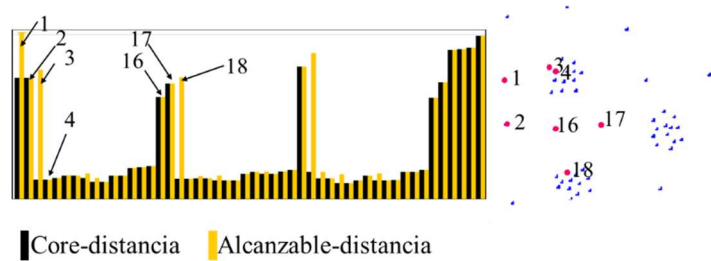
foreach  $o \in$  Database
  // initially,  $o$ .processed = false for all objects  $o$ 
  if  $o$ .processed = false;
    insert ( $o$ , “?”) into ControlList;
  while ControlList is not empty
    select first element ( $o$ ,  $r\_dist$ ) from ControlList;
    retrieve  $N_g(o)$  and determine  $c\_dist = core\_distance(o)$ ;
    set  $o$ .processed = true;
    write ( $o$ ,  $r\_dist$ ,  $c\_dist$ ) to file;
    if  $o$  is a core object at any distance  $\leq \epsilon$ 
      foreach  $p \in N_g(o)$  not yet processed;
        determine  $r\_dist_p = reachability\_distance(p, o)$ ;
        if ( $p, \_$ )  $\notin$  ControlList
          insert ( $p$ ,  $r\_dist_p$ ) in ControlList;
        else if ( $p$ ,  $old\_r\_dist$ )  $\in$  ControlList and  $r\_dist_p < old\_r\_dist$ 
          update ( $p$ ,  $r\_dist_p$ ) in ControlList;

```

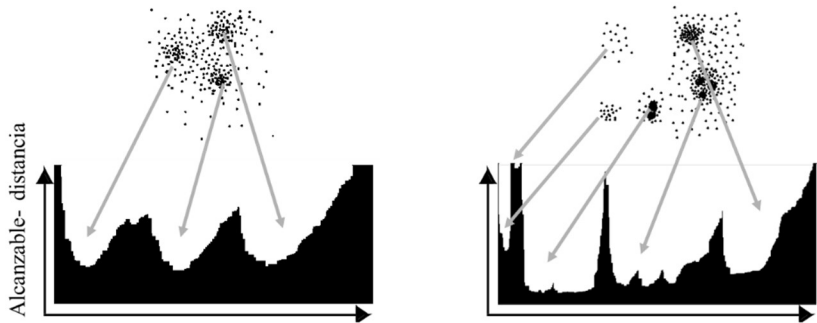
Vea gráficamente esto.



Veamos la construcción de *clústeres desarrollador usando una densidad plana* para $\epsilon^* \leq \epsilon$ y $\square MinPts$ se comienza con un objeto o mientras $c-dist(o) \leq \epsilon^*$ y $r-dist(o) > \epsilon^*$. Se continúa mientras $r-dist(o) \leq \epsilon^*$. Su comportamiento es aproximadamente, en términos de tiempo de cómputo runtime (DBSCAN ($\epsilon, MinPts$)), $O(n^2)$ sin soporte de un índice espacial (el peor caso es $O(n^2)$). Si se usa un índice espacial basado en un árbol él es $O(n \epsilon \log(n))$. Veamos un ejemplo:

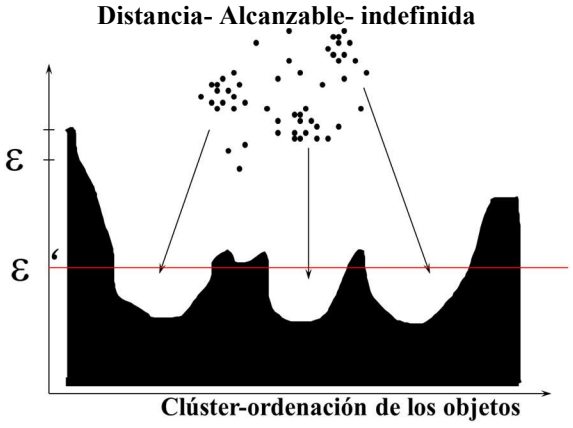
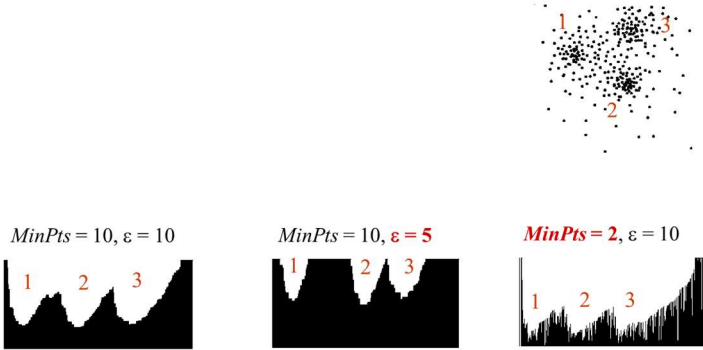


Un ploteo de la alcanzabilidad



Una ilustración de la alcanzabilidad en términos de distancia

Veamos unos ejemplos que ilustran el comportamiento de OPTICS.



Una comparación de este método con DBSCAN es ilustrado en lo que sigue.

DBSCAN VS OPTICS

	DBSCAN	OPTICS
Densidad	Valor Booleano (alto/bajo)	Valor Numérico (distancia -core)
Densidad-conectado	Valor Booleano (sí/no)	Valor Numérico (distancia-alcanzabilidad)
Estrategia de búsqueda	aleatorio	goloso

Vea un ejemplo en el que OPTICS funciona bien

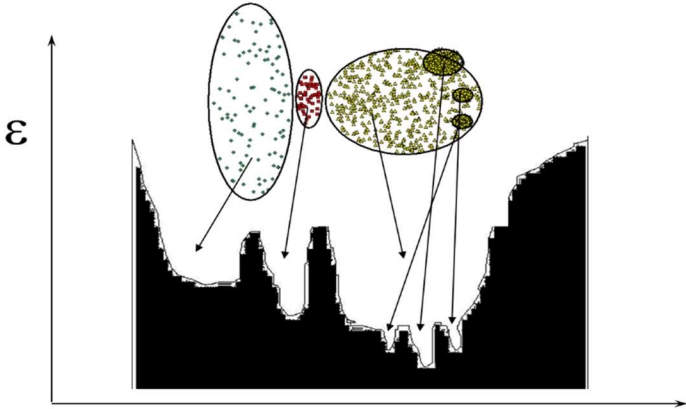


Ilustración de un ejemplo en el que OPTICS funciona bien

Veamos ahora un ejemplo en el que OPTICS no funciona bien.

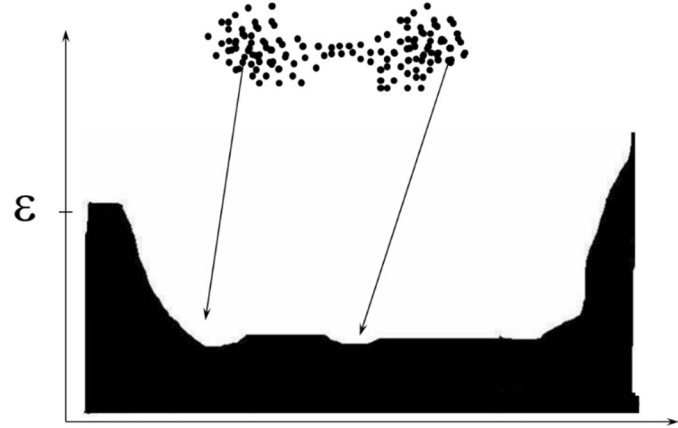
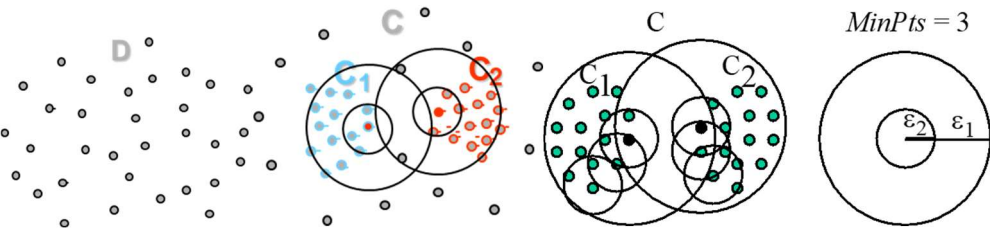


Ilustración de un ejemplo en el que OPTICS no funciona bien

6. CLUSTERING JERÁRQUICO BASADO EN DENSIDADES.

Los clústeres más densos están completamente contenidos en los menos densos. Vea en la siguiente figura los objetos procesados en la “derecha” se ordenan se mantienen records de la densidad del punto en su vecindad.



7. ESTIMACIÓN USANDO UN KERNEL

Estos métodos modelan la función de influencia para un punto de la data usando una función de kernel. La densidad es normalizada utilizando la suma de todos los Kérneles. Los Kérneles dependen de la determinación del llamado parámetro de suavizamiento “h”. El más usado es el llamado Kernel Gaussiano. Su fórmula y la de la función de densidad son

$$K(\vec{u}) = (2\pi)^{-\frac{d}{2}} \cdot \exp\left[-\frac{\vec{u}^2}{2}\right]$$

$$\hat{p}(\vec{x}) = \frac{1}{(Nh^d)} \sum_{t=1}^N K\left(\frac{\vec{x} - \vec{x}_t}{h}\right)$$

Vea una representación gráfica en la figura siguiente.



REFERENCIAS

- [1] Ankerst, M., M. Breunig, H.-P. Kriegel, and J. Sander (1999). Optics: Ordering points to identify the clustering structure, SIGMOD'99.
- [2] Aggarwal, C. C., C. Procopiuc, J. Wolf, P. S. Yu, and J.-S. Park. (1999). Fast algorithms for projected clustering. *SIGMOD'99*
- [3] Arabie, P., L. J. Hubert, and G. De Soete. (1996). Clustering and Classification. World Scientific, 1996
- [4] Boomija, M.C.A., M.Phil. (2008). Comparison of partition-based clustering algorithms journal of Computer Applications, Vol – 1, No.4, Oct – Dec 2008
- [5] Breunig, M., H.-P. Kriegel, R. Ng, J. Sander. (2000). LOF: Identifying Density-Based Local Outliers. SIGMOD 2000.
- [6] Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. (1996). A density-based algorithm for discovering clusters in large spatial databases. KDD'96.
- [7] Ester, M., H.-P. Kriegel, and X. Xu. (1995). Knowledge discovery in large spatial databases: Focusing techniques for efficient class identification. SSD'95.
- [8] Fraley C. and A. E. Raftery. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. American Stat. Assoc.*, 97:611–631.
- [9] Gandhi, G. and R. Srivastava. (2014). Analysis and implementation of modified k-medoids algorithm to increase scalability and efficiency for large dataset. *International Journal of Research in Engineering and Technology*. 3, Available @ <http://www.ijret.org>
- [10] Goel, A. A (2014). Study of Different Partitioning Clustering Technique. *IJSRD.*, 2
- [11] Guha, S., R. Rastogi, and K. Shim. (1998). Cure: An efficient clustering algorithm for large databases. SIGMOD'98.
- [12] Kriegel H, Schubert E, Zimek A (2017). The (black) art of runtime evaluation: Are we comparing algorithms or implementations? *Know. Inf. Syst.* 52,341–378.
- [13] Luxburg.A. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416.
- [14] Mathur, N., M. Tiwari, S. Khandelwal. (2015). Performance Analysis of Different Clustering Algorithm. *Journal of Computer Engineering* 17, 25-29.
- [15] Madeira S. C. and A. L. Oliveira. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 1.
- [16] McLachlan G. J. and K. E. Bksford. (1988). Mixture Models: Inference and Applications to Clustering. John Wiley & Sons.

- [17] Mirkin B. (1998). Mathematical classification and clustering. *J. of Global Optimization*, 12:105–108.
- [18] Nagpal, A., A. Jatain, D. Gaur. (2013). Review based on Data Clustering Algorithms. *Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013)*
- [19] Raymond T. Ng and Jiawei Han CLARANS: A Method for Clustering Objects for Spatial Data Mining *IEEE Transactions on knowledge and data engineering*, 14, 1003-1016
- [20] Schikuta E. (1996). Grid clustering: An efficient hierarchical clustering method for very large data sets. *Proc. 1996 Int. Conf. on Pattern Recognition*.
- [21] Schubert E, Koos A, Emrich T, Züfle A, Schmid KA, Zimek A (2015) A framework for clustering uncertain data. *PVLDB8* 12:1976–1979
- [22] Schubert E, Hess S, Morik K (2018). The relationship of DBSCAN to matrix factorization and spectral clustering. In: *LWDA, CEUR Workshop Proceedings*, 21, 330–334
- [23] Schubert E. and P. J. Rousseeuw. (2019). Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. *arXiv:1810.05691v4 [cs.LG]*.
- [24] Schubert E, Sander, Ester M, Kriegel H, Xu X (2016). DBSCAN revisited, r revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst.* 42, 1–19.
- [25] Schubert E, Zimek A (2019). ELKI: A large open-source library for data analysis - ELKI release 0.7.5" Heidelberg". *arXiv:1902.03616*
- [26] Song H, Lee J, Han W (2017). PAMAE: parallel k-medoids clustering with high accuracy and efficiency. In: *KDD*, 1087–1096.
- [27] Tung A. K. H., J. Hou, and J. Han. (2001). Spatial clustering in the presence of obstacles. *ICDE'01*.
- [28] Vijayarani, S. and Ms. P. Jothi. (2013). Comparative Analysis of Clustering Algorithms for Outlier Detection in Data Streams. *International Journal of Engineering Sciences & Research Technology Comparative*, 2.
- [29] Wagstaff K., C. Cardie, S. Rogers, and S. Schrödl. (2001). Constrained k-means clustering with background knowledge. *ICML'01*
- [30] Wang, H., W. Wang, J. Yang, and P. S. Yu. (2002). Clustering by pattern similarity in large data sets. *SIGMOD'02*
- [31] Xu, X., N. Yuruk, Z. Feng, and T. A. J. Schweiger. (2007). SCAN: A structural clustering algorithm for networks. *KDD'07*
- [32] Yu, H., J. Yang, and J. Han (2003). Classifying large data sets using SVM with hierarchical clusters. *KDD'03*.

Capítulo 6

pp 67-79

ANÁLISIS DE FACTORES DE RELACIONADOS CON LA EVOLUCIÓN CLÍNICA DE MUJERES CON DISFUNCIONES DEL SUELO PÉLVICO.

R.S. Borges Sandrino¹, V. Sistachs Vega², E. Ramón Musibay¹ y M. A. Díaz Martínez³

¹Facultad de Ciencias Médicas “Finlay-Albarran” La Habana, Cuba.

²Facultad de Matemática y Computación, Universidad de La Habana. La Habana, Cuba.

³CEMAT “Centro de Estudios Matemáticos”, CUJAE “Ciudad Universitaria José Antonio Echeverría”. La Habana, Cuba.

ABSTRACT

Pelvic floor dysfunctions (PFD) have a high prevalence and incidence in Cuba constituted by the uterine prolapses and the prolapse of the vaginal dome, the cystoceles and others, in spite of they have been analyzed through descriptive analysis, however this method does not explain the existent relations between the variables they fall in the clinical evolution. Statistical modelation by means of the multinomial regression permits a better characterization of the population in studies, facilitate the identification of the variables related with clinical evolution and predicts with an acceptable precision, as well as with a bigger efficiency and consistency.

KEYWORDS: Pelvic floor dysfunctions, statistical modelation, multinomial regression.

RESUMEN

Las disfunciones del suelo pélvico (DSP) tienen una elevada prevalencia e incidencia en Cuba estando constituidas por los prolapsos uterinos, de la cúpula vaginal, los cistoceles, y otros. A pesar de ello constituyen una de las afecciones estudiada solo a través de análisis descriptivo, sin embargo, este método no explica las relaciones existentes entre las variables que inciden en la evolución clínica de estas enfermas. La modelación estadística mediante el empleo de regresión logística multinomial permite una mejor caracterización de la población en estudio, posibilita la identificación de las variables relacionadas con evolución clínica y predice la misma con una precisión aceptable, así como con una mayor eficiencia y consistencia.

1. INTRODUCCIÓN

Las disfunciones del suelo pélvico (DSP) tienen una elevada incidencia en la población femenina estando constituidas por los prolapsos uterinos, de la cúpula vaginal, los enteroceles, los cistoceles, los rectoceles y la incontinencia urinaria o fecal. Al menos del 46% al 50% de la población femenina mundial por encima de los 50 años presenta algún tipo de disfunción y que una de cada cinco tiene más de una [1]. En Cuba la prevalencia en la mujer entre los 50 y 70 años es del 40% con un porcentaje significativo de cirugías [1,2]. En los Estados Unidos de Norteamérica, alrededor de 10 millones de adultos, sufren algunas de las disfunciones antes mencionadas, con un costo anual de atención médica especializada de 10 mil millones de dólares [2]. En Europa, según reportes del Grupo Español de SP, se estima que alrededor del 45% de las mujeres pudieran estar afectadas, mientras que en África el 50% de las mujeres con partos fisiológicos institucionales presentan al menos dos de las variedades antes mencionadas, lo cual evidencia que las DSP constituyen un problema mundial de salud con grandes limitaciones socio laborales y deterioro importante de las condiciones higiénicas [3].

A pesar de su incidencia en Cuba todavía son insuficientes las investigaciones relacionadas con el tema si tenemos en cuenta el grado de afectación socio psicológico con que conviven las mujeres en etapa laboral que las padecen. Múltiples factores han sido señalados como generadores o potencializadores de esta enfermedad sin embargo no son mencionados en todas las series y no han sido totalmente estudiados en la población cubana.

Por ello se hace necesario analizar las variables sociodemográficas y clínico quirúrgicas relacionadas con su desarrollo hecho que pudiera representar una valiosa herramienta en las manos de los galenos dedicados al tratamiento de las DSP.

Los objetivos de la siguiente investigación fueron la caracterización de mujeres en etapa laboral con DSP y determinar los factores de riesgo relacionados con la génesis de la enfermedad 3 Seleccionar a través de una regresión logística el mejor modelo predictivo. Para lograr estos objetivos en el epígrafe 2 se plantea el diseño dividido en análisis exploratorio y confirmatorio. En este último se conceptualiza sobre el análisis de Regresión Logística Binaria (RLB), así como los criterios para evaluar la capacidad predictiva y los criterios para la selección de modelos. En el epígrafe 3 se presentan los resultados y la discusión de los mismos.

2. DISEÑO DEL ESTUDIO Y PROCEDIMIENTO ESTADÍSTICO

2.1 Diseño y contexto del estudio

Se realizó un estudio prospectivo longitudinal con mujeres que presentaban disfunciones del suelo pélvico en el periodo comprendido entre enero del 2013 hasta diciembre del 2017. Se incluyeron consecutivamente 249 mujeres que aceptaron participar en el estudio bajo principio de consentimiento informado. Se excluyeron aquellas con ausencias a consulta y no seguimiento mínimo previsto, los trastornos neurológicos o tratamiento con relajantes musculares y la negación de participación.

2.1.1 Procedimientos estadísticos

Estuvo dividido en dos partes en la primera se realizó el análisis exploratorio de las variables explicativas se utilizaron medidas de resumen para describir los datos, de manera que las variables cualitativas se muestran en proporciones y porcentajes las variables cuantitativas se representan como mediana y 25% – 75% rango intercuantil (RIQ). La comparación entre el grupo de pacientes con evolución clínica buena, regular y mala se realizó mediante la prueba de chi cuadrado de Pearson (χ^2) o el test exacto de Fisher para las variables cualitativas, según fue apropiado; en las variables cuantitativas se empleó para la correlación entre dos variables el coeficiente de correlación de Pearson.

Para conocer la variable independiente con mayor influencia sobre la variable dependiente por categoría se aplicó la regresión logística multinomial. Las curvas de COR permitieron conocer el modelo con mejor poder de discriminación (área bajo la curva).

2.2.3 Análisis de regresión logística multinomial

La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica [4]. Las variables independientes pueden ser tanto continuas (covariables) como categóricas (factores), este fue el método seleccionado para identificar los factores de riesgo relacionados con la evolución clínica de las enfermas.

El objetivo de este tipo de regresión es encontrar el mejor modelo que se ajuste a un conjunto de datos y describir la relación entre la variable respuesta politómica y un conjunto de variables independientes seleccionando una de las tres categorías como referencia [4].

Se considera una variable aleatoria dependiente Y categórica nominal politómica con Soporte (Y)= {1,2,3} y con probabilidades $P_1=p(Y=1)$, $P_2=p(Y=2)$ y $P_3=p(Y=3)=1-p_1-p_2$.

Si se quiere analizar el efecto que ejercen dos variables explicativas continuas X_1 , X_2 sobre las probabilidades p_1 y p_2 que caracterizan a la variable Y. Se redefine a la variable Y mediante un vector (Y_1, Y_2) construido de la siguiente forma:

$$(Y_1, Y_2) = \begin{cases} (0,1) \text{ si } Y = 1 \\ (0,1) \text{ si } Y = 2 \\ (0,0) \text{ si } Y = 3 \end{cases}$$

Las variables Y_1 e Y_2 tienen una distribución de Bernoulli con $E(Y_1) = p_1$ y $E(Y_2) = p_2$, al igual que la variable dependiente en una regresión logística binaria clásica. Obviamente estas dos variables no son independientes ya que $Cov(Y_1, Y_2) = -p_1 p_2$.

Al cociente p_1/p_3 se le denomina ‘odds’ de la categoría 1 respecto de la categoría 3 y se le representa por $O_1(X_1, X_2) = O_1$. Este aspecto se repite para O_2 . De este modo puede observarse fácilmente que la razón de cambio en O_1 cuando X_1 se incrementa en una unidad manteniéndose constante, que recibe el nombre de ‘odds-ratio’ de la categoría 1 respecto de la variable X_1 y se representa por $OR_1(X_1)$ (idem. para $OR_1(X_2)$, $OR_2(X_1)$ y $OR_2(X_2)$).

Otra formulación alternativa, y quizás más conocida, se obtiene tomando logaritmos en ambas ecuaciones del modelo:

$$\ln\left(\frac{P_1}{P_3}\right) = Z_1 = \beta_{01} + \beta_{11} \cdot X_1 + \beta_{21} \cdot X_2$$

$$\ln\left(\frac{P_2}{P_3}\right) = Z_2 = \beta_{02} + \beta_{11} \cdot X_1 + \beta_{22} \cdot X_2$$

donde las expresiones del miembro izquierdo se denominan ‘logits’ (al igual que en la regresión logística binaria) y los parámetros representan las tasas de cambio en los ‘logits’ cuando una de las variables explicativas se incrementa en una unidad manteniéndose constante la otra [5].

Se deben incluir todas aquellas variables que se consideren clínicamente importantes para el modelo, independientemente de si se ha demostrado o no significación estadística en un análisis univariado o multivariado previo, ya que puede conducir a dejar de incluir en el modelo covariables con una débil asociación a la variable dependiente en solitario pero que podrían demostrar ser fuertes predictores de la misma al tomarlas en conjunto con el resto de covariables.

Con estos criterios debemos de conseguir obtener un modelo que sea lo más reducido posible que explique los datos (principio de parsimonia), y que además sea clínicamente congruente e interpretable. Posiblemente un mayor número de variables en el modelo implicaría mayores errores estándar [5].

Para ello se puede recurrir a métodos de selección paso a paso, bien mediante inclusión, hacia adelante, o por eliminación, hacia atrás, o a la selección de variables por mejores subconjuntos de covariables. Estos métodos se encuentran implementados en la mayoría de los paquetes estadísticos [5].

Para la estimación de los coeficientes del modelo y de sus errores estándar se utiliza la estimación por máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente y proporcionados por los datos de la muestra.

Para contrastar la bondad del ajuste global del modelo cuando el número de observaciones en cada combinación de valores de las variables explicativas es grande se utiliza el estadístico chi-cuadrado de Pearson y el estadístico de Wilks de razón de verosimilitud [6].

El estadístico chi-cuadrado de Pearson de bondad de ajuste a un modelo de regresión logística multinomial (M) de la forma anterior viene dado por:

$$X^2(M) = \sum_{q=1}^Q \sum_{j=1}^k \frac{(y_{j/q} - d_q \hat{\beta}_{j/q})^2}{d_q \hat{\beta}_{j/q}}$$

Este estadístico tiene distribución asintótica chi-cuadrado con grados de libertad obtenidos como la diferencia entre el número de parámetros y el número de parámetros independientes en el modelo.

El estadístico de Wilks también llamado devianza para el contraste de bondad de ajuste del modelo de regresión logística multinomial M se obtiene como menos dos veces el logaritmo del cociente entre el supremo de la verosimilitud bajo la hipótesis nula y el supremo de la verosimilitud en la población. A partir de esta expresión operando se obtiene la expresión de este estadístico que viene dada por:

$$G^2(M) = 2 \left[\sum_{q=1}^Q \sum_{j=1}^k y_{j/q} \ln \left(\frac{y_{j/q}}{\hat{m}_{j/q}} \right) \right]$$

se rechaza la hipótesis nula para ambos cuando $p\text{-valor} \leq \alpha$.

La calidad del ajuste en la RLM se mide mediante coeficientes de determinación conocidos como Pseudo- R^2 . El primero se basa en la función auxiliar Λ utilizada en el ajuste, se conoce como pseudo- R^2 de Mc-Fadden y viene dado por: $RMF = 1 - \Lambda_0$.

Su rango teórico de valores es $0 \leq R^2_{MF} \leq 1$, pero muy raramente su valor se aproxima a 1. Suele considerarse una buena calidad del ajuste cuando $0.2 \leq R^2_{MF} \leq 0.4$ y excelente para valores superiores [7].

Otros autores prefieren coeficientes basados directamente en la verosimilitud L, y no en la función auxiliar Λ . El más conocido es el pseudo- R^2 de Cox-Snell o el de Nagelkerke, su rango de valores es $0 \leq R^2_N \leq 1$ por lo que puede interpretarse del mismo modo que el coeficiente de determinación de la regresión lineal clásico [8].

Para determinar la capacidad de discriminación del modelo de asignar el resultado correcto los sujetos se utilizan las curvas de las características operativas del receptor (COR). El área bajo la curva de COR es la medida de discriminación que se utiliza con mayor frecuencia en modelos de error normal. Una curva COR enfrenta en un sistema de ejes la sensibilidad en el eje y, complementado con la especificidad en el eje x. El procedimiento consiste en determinar las correspondientes tablas de clasificación de puntos de corte de P ($Y=1|X$) y determinar a partir de ellas las correspondientes sensibilidades y especificidades que ofrece mayor sensibilidad y especificidad, se recurre al análisis gráfico de la curva, seleccionando como punto de corte aquél que corresponde con el punto de inflexión de la curva COR. Las mejores curvas serán aquellas con un área más próxima a la unidad. Como regla general, un área de 0,5 implica ausencia de discriminación; entre 0,7 y 0,79 es una discriminación aceptable; entre 0,8 y 0,89 es excelente; 0,9 o superior es una discriminación excepcional [7,8].

Todo el análisis estadístico se realizó mediante el programa para R versión 3.5.3, utilizándose como interfaz de desarrollo integrado el RStudio versión 1.2.1. Los gráficos presentados fueron creados utilizando la librería ggplot2 del mismo programa perteneciente a su tercera generación [10].

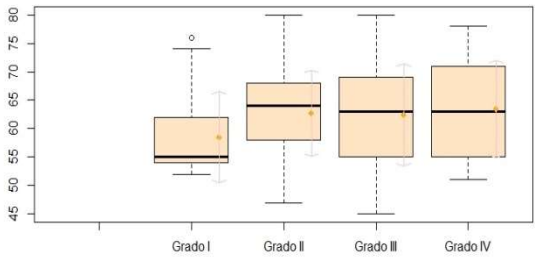
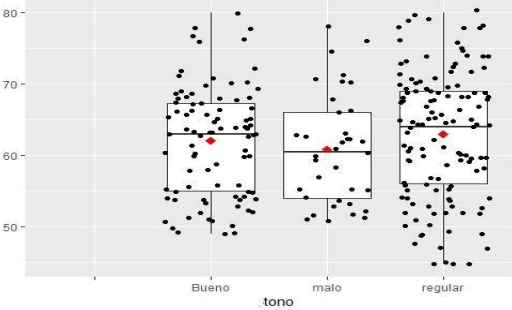
3. RESULTADOS Y DISCUSIÓN

3.1 Análisis exploratorio

Se estudiaron 249 pacientes con disfunciones del suelo pélvico atendidas en el Hospital General Docente “Carlos J Finlay”, la edad tuvo una mediana de 62,0 años (RIQ 55,0 – 69,0 años). La mayoría de los enfermos tenía comorbilidades crónicas siendo las más frecuentes la hipertensión arterial en 44 pacientes (17,6) y el asma en 18(7,2), 159 mujeres no presentaron enfermedades asociadas.

Tabla 1. Caracterización de pacientes con disfunciones del suelo pélvico.

Variable	Frecuencia	
Edad, años [mediana (RIQ)]	63,0(55,0-69,0)	
menos 50	18 (7,1)	
51 a 60	87 (34,4)	
61 a 70	103 (40,7)	
más 71	41 (16,2)	
Comorbilidades[mediana(RIQ)]	Frecuencia	
Asma	18 (7,2)	
Cardiopatía	13 (5,2)	
Diabetes Mellitus	15 (6)	
Hipertensión	44 (17,6)	
no	159 (63,6)	

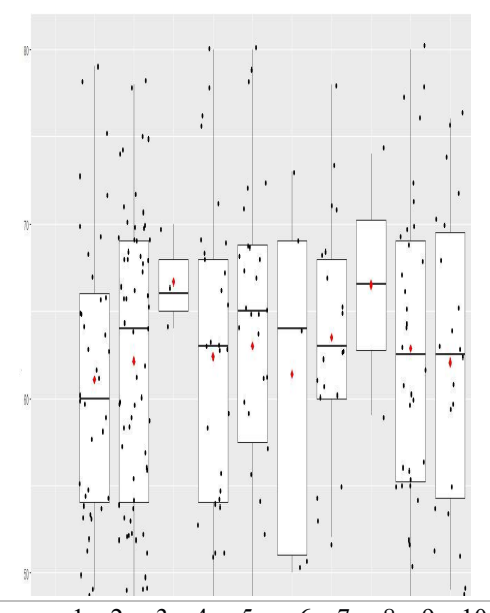
<p>Grado n(%)</p> <p>Grado I 14(5,6)</p> <p>Grado II 88(35,2)</p> <p>Grado III 129(51,6)</p> <p>Grado IV 18(7,2)</p>		
<p>Tono muscular n(%)</p> <p>Bueno 88(35,2)</p> <p>Regular 36(14,4)</p> <p>Malo 125(50,0)</p>		

El grado III afectó a 129 enfermas (51,6%). En cuanto al tono muscular perineal se detectó que 125 enfermas presentaban tono muscular malo (50%) (Tabla 1). El cistocele con incontinencia urinaria predominó entre los tipos de disfunción en 65 enfermas (26,10%) lo cual coincide con la literatura revisada donde suele ser la DSP de mayor incidencia (Tabla 2).

Las variables residuo, capacidad y flujo reflejan la funcionalidad del suelo pélvico por lo que su variación debe ser siempre explorada en relación con la edad, en la mayoría de las pacientes la capacidad vesical estuvo por encima de los 400 ml (media de $485 \pm 71,04$ y varianza de 50,47) sin cambios en relación con la edad, aspecto que se repite al analizar el flujo, sin embargo en el análisis del residuo urinario se puede observar la disminución del mismo al aumentar la edad, ninguna de las tres variables resultaron ser significativas en el análisis estadístico (Figura 1).

Tabla 2 Disfunciones del suelo pélvico

Tipo de disfunción		Frecuencia	Porcentaje
1	Cistocele	39	15,6
2	Cistocele con IU	65	26,0
3	Incontinencia fecal	3	1,2
4	Incontinencia Urinaria	29	11,6
5	Prolaputer	30	12,0
6	Prolp cúpula	5	2,0
7	Rectocele	20	8,0
8	Rectocele con IF	2	,8
9	Rectocistocele con IU	34	13,6
10	Retocistocele	22	8,8
Total		250	100,0



Un total de 208 pacientes no habían recibido tratamiento por este tipo de enfermedad (83,50%) mientras que 14 (5,6%) lo hicieron dos veces, sin embargo, el 8% de la serie había sido operado 6 o más veces.

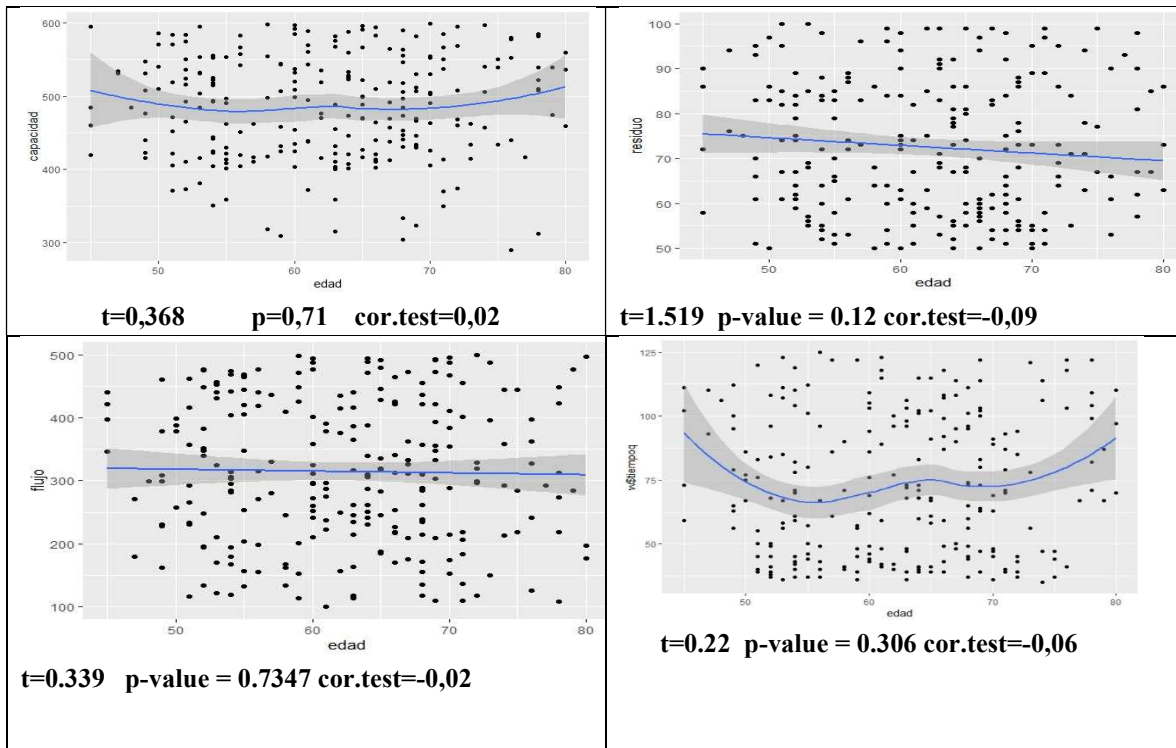


Figura 1. Relación de la edad con el residuo urinario el flujo la capacidad y el tiempo quirúrgico

La figura 2 muestra la relación entre el residuo urinario y el flujo en función de los grados, la cual fue positiva y significativa estadísticamente ($p = 0.01876$, $r = 0,74$)

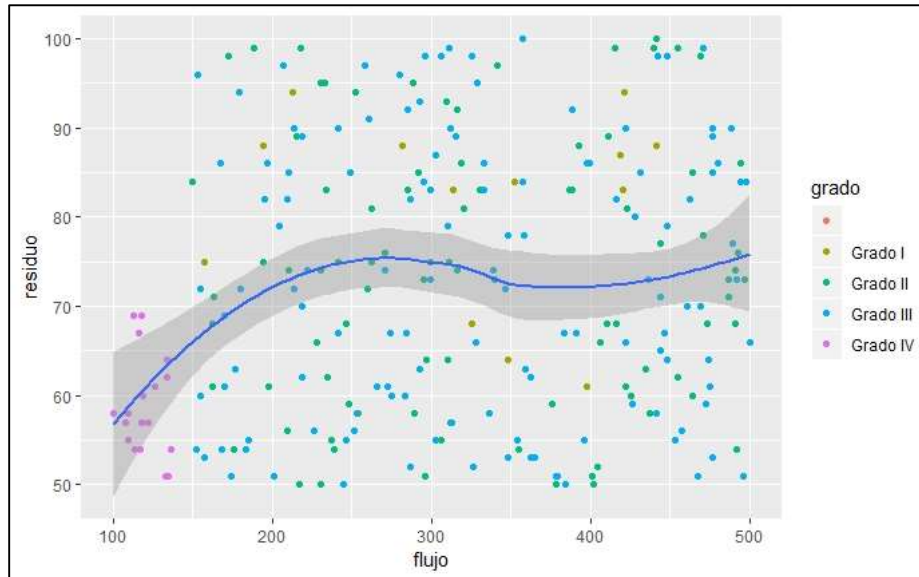


Figura 2. Relación entre residuo y flujo en función de los grados de disfunción.

Tabla 3. Distribución según operaciones previas y pacientes complicadas

Variables	Frecuencia
Operaciones previas n(%)	n(250)
0	208(83,50)
1	8(3,2)
2	14(5,6)
3	7(2,8)
4	6(2,4)
5	4(1,6)
6	2(8)
Pacientes complicadas n(%)	
No	234(93,6)
SI	15(6)
Evolución clínica n(%)	
Buena	201(80,72)
Regular	32(13,25)
Mala	16(6,02)
Convalecencia n(%)	
30 días	19(7,6)
31 a 60 días	84(16,4)
61 a 79 días	88(35,2)
Más de 80 días	58(23,2)

Se detectaron 15 pacientes complicadas después de aplicado el tratamiento (6.02%). La escala de Visik permitió conocer la evolución clínica de las enfermas [1-3], el 80,72% (201 pacientes) tuvo una evaluación clínica buena, sin embargo, en 32 (13,25) enfermas la evolución fue regular y en 16 mala (6,02%). La mayoría de las enfermas tuvieron entre los 61 y 79 días de convalecencia después del tratamiento y solo el 7,6% se recuperó en los primeros 30 días.

Regresión Logística multinomial

Para el análisis de RLM se incluyeron las variables originales según su significación clínica y por su posible efecto sobre la evolución clínica estas fueron edad, residuo, flujo capacidad, grado, tono, complicaciones, y operaciones previas.

Tabla 4 Regresión logística multinomial

calidad ^a		B	EE	Wald	gl	Sig.	Exp(B)
regular	Intersección	9,613	6,603	2,119	1	,045	
	Edad	,012	,024	6,785	1	,016	1,018
	Capacidad	,017	,007	6,245	1	,012	1,017
	Flujo	-,002	,004	5,245	1	,010	,998
	Residuo	-,013	,031	6,183	1	,029	,987
	Grado	,015	,008	6,356	1	,011	1,010
	Tono	,011	,005	5,335	1	,013	1,015
	Operaciones previas	,009	,020	,209	1	,647	,009
	Convalecencia	,067	,031	4,646	1	,073	,070
mala	Intersección	17,440	7,042	6,133	1	,013	
	Edad	,025	,009	8,830	1	,009	1,033
	Capacidad	,021	,008	7,830	1	,005	1,021
	Flujo	,006	,004	1,675	1	,006	1,006
	Residuo	,023	,035	2,450	1	,003	1,024
	Grado	,018	,007	6,630	1	,005	1,002
	Tono	,020	,007	7,530	1	,004	1,031
	Operaciones previas	,040	,022	3,293	1	,070	,041
	Convalecencia	,033	,035	,894	1	,344	,034
buena	Intersección	-15,027	7,219	4,333	1	,037	
	Edad	,012	,004	3,191	1	,022	1,004
	Capacidad	,014	,006	5,191	1	,023	1,014
	Flujo	-,001	,003	5,104	1	,010	,999
	Residuo	,001	,027	4,002	1	,024	1,001
	Grado	,015	,007	5,391	1	,025	1,016
	Tono	,013	,004	5,100	1	,021	1,013
	Operaciones previas	-,012	,018	,412	1	,521	,988
	Convalecencia	-,001	,025	,003	1	,957	,999

El análisis en este sentido debe hacerse según la categoría de evolución clínica teniendo la categoría mala como referencia en la cual se puede plantear al evaluar el ajuste del modelo que las variables edad, capacidad, flujo, residuo, grado y tono el estadístico de Wald fue significativo aspecto que indica la conservación en el modelo de dichas variables. Estos resultados se obtuvieron también en el resto de las categorías por lo cual el análisis a partir de esta etapa de la investigación se realizó sin las variables operaciones previas y convalecencia, aspecto que convierte el modelo inicial en uno depurado.

En la tabla 4 también se observan los odds ratios del modelo, las variables edad, capacidad, grado y tono presentan valores de 1 o superior por lo que pudiera plantearse que constituyen factores de riesgo para la evolución clínica de las enfermas en las tres categorías, sin ser despreciable las cifras alcanzadas por las variables flujo y residuo urinario.

La variable que tuvo un mayor efecto sobre la evolución clínica fue la edad para las enfermas para las categoría regular y mala, sin embargo, en la categoría buena la que más influyo fue el grado de prolapso.

La razón de verosimilitud es otra prueba empleada para medir el ajuste de un modelo de RLM. Este estadístico sigue una distribución χ^2 con p grados de libertad. Si este estadístico alcanza significación estadística, indica un buen ajuste. [9-11] (tabla 5). Todas las variables analizadas alcanzaron significación estadística en el modelo depurado, lo que indica que aportan al modelo y pueden permanecer en el mismo además de revelar un buen ajuste, aspecto que coincide con investigadores como Armstrong.

Tabla 5 Pruebas de la razón de verosimilitud

Efecto	Modelo Depurado			
	L. Verosimilitud	Chi-cuadrado	gl	Sig.
Intersección	371,547	83,062	4	,000
Edad	308,630	20,144	4	,000
Grado	310,330	21,845	4	,000
Tono	310,345	21,860	4	,000
Flujo	313,330	22,845	4	,000
Residuo	315,345	24,860	4	,000
Capacidad	322,009	35,523	4	,000

Los estadísticos chi-cuadrado de Pearson y la devianza resultaron útiles en la comparación de ambos modelos (inicial y depurado) en cuanto a la cual de los dos tiene mejor bondad de ajuste de manera general. El modelo depurado es el que mejor ajuste muestra al tener la valores devianza y el chi cuadrado de Pearson más bajos (tabla 6).

La devianza del modelo es una medida del grado de diferencia entre las frecuencias observadas y predichas, nos puede orientar durante la etapa de selección del modelo final [12,13].

Tabla 6. Bondad de ajuste del modelo

Modelos	Estadísticos	Chi-cuadrado	gl	Sig.
Inicial	Pearson	814,813	864	,883
	Desvianza	320,525	864	1,000
Depurado	Pearson	668,319	856	1,000
	Desvianza	263,024	856	1,000

En la tabla 7 se observa la calidad del ajuste del modelo utilizando los coeficientes Pseudo- R^2 de Mc-Fadden, de Cox-Snell y de Nagelkerke. El cálculo de éstos depende del valor de las devianzas del modelo inicial y del modelo depurado con sólo la constante [14,15]. Los tres coeficientes pseudo R^2 están por encima de 0,5 (Mc-Fadden 0,56 y Cox y Snell 0,66) destacándose el de Nagelkerke que alcanza 0,77 por lo que se puede afirmar que la calidad de ajuste del modelo es buena.

Tabla 3.13 Coeficientes Pseudo R^2 .

Pseudo R^2		Valores
Modelo Inicial	Mc-Fadden	0,239
	Nagelkerke	0,312
	Cox y Snell	0,156
Modelo Depurado	Mc-Fadden	0,561
	Nagelkerke	0,774
	Cox y Snell	0,663

Al comparar ambos modelos, se observó que el modelo depurado de RLM se ajustó mejor que el modelo inicial. Las características de las curvas COR y de las áreas bajo las mismas fueron mejor para el modelo depurado (área bajo la curva 0,85) que para la inicial área bajo la curva 0,67) (figura 3). La curva COR provee una información estadística y visual que ha facilitado su uso corriente en los estudios biomédicos [16,17]. Es un método muy útil para comparar la capacidad de discriminación de los modelos predictivos de manera general [18-20].

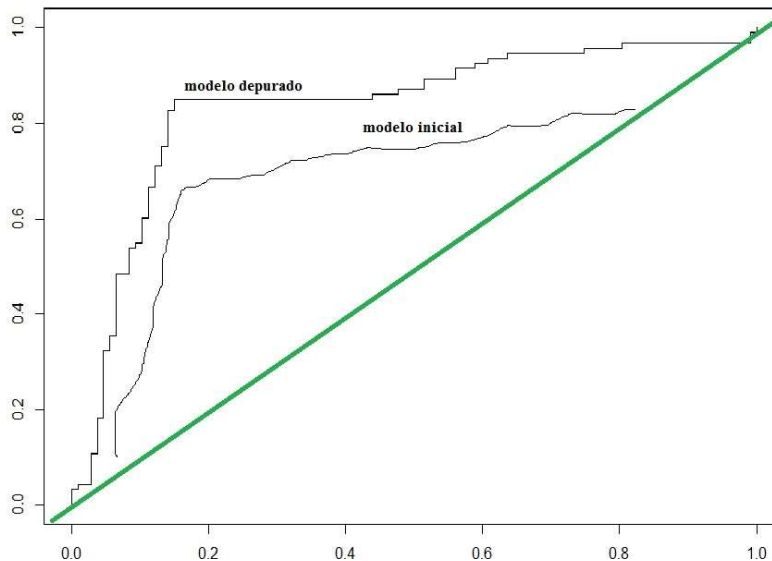


Figura 3 Discriminación de ambos modelos según curvas de COR

4. CONCLUSIONES

Los resultados obtenidos permiten aseverar que la regresión logística multinomial resulta útil en el análisis de variables cualitativas con más de dos categorías, permite identificar las variables que determinan o condicionan los resultados dentro de la variable dependiente tal como evolución clínica (buena, mala o regular). La selección automática de modelos permitió identificar y optar por el modelo que más se ajusta a los datos siguiendo el principio de parsimonia. El software estadístico R es una opción ventajosa no solo por la generación gráfica que ofrece sino porque permite automatizar este tipo de regresión por lo que se recomienda incrementar su uso en este tipo de análisis biomédicos donde es frecuente este tipo de variables.

REFERENCIAS BIBLIOGRÁFICAS

- [1] De Mello Portella P, Feldner PC Jr, Da Conceição JC, Castro RA, Sartori MG, Girão MJ. Prevalence of and quality of life related to anal incontinence in women with urinary incontinence and pelvic organ prolapsed. *Eur J Obstet Gynecol Reprod Biol.* 2017 Feb; 160 (2):228–31.
- [2] Núñez L. Natalia, Saxton M. Alejandro, Venegas de la Paz. Mauricio. Prevalencia de la incontinencia urinaria. *Kinesiología* abril-mayo 2016; 32(1): 17-26.
- [3] Deng DY. Urinary incontinence in women. *Med Clin North Am.* 2015; 95:101-109.
- [4] Hair J. F., Anderson R. E., Tatham R. L., Black W. C., *Análisis Multivariante*, PRENTIE HALL IBERIA, Madrid, 1999.
- [5] Anderson T.E., *A Semigraphical Method for the Analysis of Complex Problems.* *Technometrics* 2, 2015: 387-91.
- [6] Silvia Stela F. *Análisis exploratorio de datos espaciales. EstadísticaEspacial. Resúmenes. Revista Geográfica Digital. IGUNNE. [Internet]. 2016 [citado 2017 23 nov];11(21):aprox.40p.* Disponible en: <http://hum.unne.edu.ar/revistas/geoweb>
- [7] Pando V., *Regresión logística multinomial.* *Actas de la Reunión de Modelización Forestal. Cuad. Soc. Esp. Cien. For.* 18: 000-000 (2015). Disponible en researchgate.net/publication/40836593
- [8] Harrell, Jr., F. E. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis.* Springer-Verlag, New York, 2015.
- [9] Armstrong, BG, and M. Sloan. *Ordinal models for epidemiologic data.* *Am J Epidemiol* 2016; 129:191–204.
- [10] R Development Core Team (2005d), *Writing R Extensions*, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>, ISBN 3-900051-11-9.
- [11] Ananth, CV; Kleinbaum, DG. *Regression models for ordinal responses: A review of methods and applications.* *International Journal of Epidemiology.* 2017; 26: 1323-1333.
- [12] Hosmer, D.; Lemeshow, S. y Sturdivant, R. *Applied Logistic Regression.* (3rd edition), John Wiley, Canada, 2013.
- [13] Remy, N. J., *A note on the general definition of coefficient of determination.* *Biometrika*, 2015, 78:3, 691-692.
- [14] Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with R.* S, Springer, 4th edition, URL <http://www.stats.ox.ac.uk/pub/MASS4/>, ISBN 0-387-954570.

- [15] R Development Core Team (2005d), Writing R Extensions, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>, ISBN 3-900051-11-9.
- [16] Tutorial 13: Regresión Logística. <http://www.postdata-statistics.com>. Consultado en enero, 2018.
- [17] Arriaza, A.J, Fernández, F, López, M.A, Muñoz, M, Pérez, S & Sánchez, A (2018). Estadística Básica con R y R- commander. Servicio de publicaciones de la Universidad de Cádiz. <http://knuth.uca.es/ebrcmdr>.
- [18] Chambers, J. M. and Hastie, T. J. (1992), Statistical Models in S, London,UK: Chapman & Hall.
- [19] Wilkinson, L., The Grammar of Graphics (2nd ed.). Statistics and Computing, New York: Springer. 2015 (14,18)
- [20] Díaz, L. Selección de Modelos en Regresión Logística Binaria bajo el paradigma Bayesiano. Tesis de doctorado UH, Cuba, 2016.
- [21] Beltrán C. Aplicación del análisis de regresión logística multinomial en la clasificación de textos académicos: Biometría, Filosofía y Lingüística informática. Infotur 2016; 5:53-64.

**CRIMES IN COLOMBIA (2005-2018): INFERRING CRIME RATE
RELATIONSHIPS AMONG GEOGRAPHICAL STATES**

Cesar Augusto Vargas-García¹, Jenny Paola Lis-Gutiérrez², Mercedes Gaitán-Angulo², Linda Carolina Henaor², Melissa Lis-Gutiérrez³, Henry Zapata Patarroyo⁴

¹Agrosavia. Corporación Colombiana de investigación Agropecuaria.

²Fundación Universitaria Konrad Lorenz.

³Corporación Universitaria del Meta.

⁴Asociación Patria Mía.

Abstract

The purpose of this article is to identify which states had the greatest effect on the behavior of other spatial units with respect to the alleged criminal acts known to the Attorney General's Office since the entry into force of Law 906 of 2004 and Law 1098 of 2006, after the implementation of the Oral Accusatory Criminal System (SPOA) in Colombia between 2005 and 2018?

We used the time-lagged ordered lasso algorithm (Nguyen & Braun, 2018) and applied to crime data to infer interaction networks between crime events happened on states in Colombia. Our results provide further insights into interactions leading to coordinate crime rates among the states studied.

Resumen

El propósito de este artículo es identificar ¿cuáles son los estados que tuvieron mayor incidencia en el comportamiento de otras unidades espaciales con respecto a los presuntos hechos delictivos conocidos por la Fiscalía General de la Nación desde la entrada en vigencia de la Ley 906 de 2004 y la Ley 1098 de 2006, luego de la implementación del Sistema Penal Oral Acusatorio (SPOA) en Colombia entre 2005 y 2018?

Se utilizó el algoritmo time-lagged ordered lasso (Nguyen & Braun, 2018) y se aplicó a los datos sobre delincuencia para inferir redes de interacción entre los eventos delictivos ocurridos en los departamentos de Colombia. Los hallazgos proporcionaron una mayor comprensión de las interacciones que conducen a la coordinación de las tasas de criminalidad entre los departamentos estudiados.

Palabras Claves: Crimen, delincuencia, Colombia, Sistema Penal Oral Acusatorio, time-lagged ordered lasso

Introduction

According to data from the Organización Panamericana de la Salud (2019) half of all deaths of young people between 10 and 24 years in the Americas (totaling 230,000) are due to homicide (24%), followed by deaths in transit (20%) and suicide (7%). Nine out of ten homicide deaths, four out of five road traffic deaths and three out of four suicides are male.

In Colombia, the work of Vargas-García et al. (2018) made it possible to establish that the alleged criminal acts known to the Attorney General's Office since the entry into force of Law 906 of 2004 and Law 1098 of 2006, after the implementation of the Oral Accusatory Criminal System (SPOA), in Colombia, between 2005 and 2018, follows an Integrated Seasonal Autoregressive and Moving Average ARIMA (0,1,1)(2,0,0)₁₂. Although

¹ cavargas@agrosavia.co, jenny.lis@konradlorenz.edu.co, mercedes.gaitana@konradlorenz.edu.co, lindac.henaor@konradlorenz.edu.co, melissa.lis@unimeta.edu.co, edzapata7@poligran.edu.co

the study suggests a significant change in crime rates, the fact that the standard error obtained in the regression coefficient is greater than the estimated value of the regression coefficient makes it impossible to determine whether this change is an increase or a reduction in crime rates.

Taking into consideration this previous work, the purpose of the article is to identify which departments had the greatest effect on the behavior of other spatial units with respect to the alleged criminal acts known to the Attorney General's Office since the entry into force of Law 906 of 2004 and Law 1098 of 2006, after the implementation of the Accusatory Oral Criminal System (SPOA), in Colombia, between 2005 and 2018? An adaptation of Phixer's algorithm was used (Singh et al., 2012).

1. Literature Review

Criminology was born out of the need to understand crime. Studies on the criminal issue have diversified in recent decades in Latin America, research perspectives have evolved and developed in conjunction and progressively with other disciplines (Castro Valdebenito, 2018), but there is little research on the composition of crime or the combination of two or more types of crime, within small area units (Brantingham, 2016; Schreck, McGloin, & Kirk, 2009).

It has also been shown that intra-urban spatial patterns of many types of crime are positively correlated at the scale of small areas (Andresen, 2011) and many types of crime have been explained using a set of ecological theories, including social disorganization, routine activity and theories of the criminal pattern (Andresen, 2006; Kinney, Brantingham, Wuschke, Kirk and Brantingham, 2008; Roncek and Maier, 1991).

In more recent studies on crime analysis we find the works of Anselin (2019), who applies a modified Geary index to crime data in France. Fuente Flores & Sánchez Salinas (2017), through the construction of a spatial regression (RS) model of the estimated error with the technique of maximum likelihood, show that the use of non-residential land, public transport transfer stations and the percentage of male population from 18 to 24 years increase significantly the opportunities to be the object of a criminal act.

Quick, Li, and Brunton-Smith (2018) apply a set of Bayesian multivariate spatial models to analyze burglary, vehicle theft, and violent crime. The residual variability of each type of crime is divided into shared and specific components in order to control the effects of population density, deprivation, residential instability and ethnic heterogeneity. Shared components account for correlations between types of crime and identify general patterns of crime shared across multiple crimes.

Also applying spatial econometrics is the work of Ahmar & Aidid (2018), who use autoregressive spatial methods (SAR) and spatial error models (SEM) to analyse crime in South Sulawesi, Indonesia. The independent variables used are population density, number of poor people, GDP per capita, unemployment and the human development index (HDI).

For their part, Lizarzaburu-Bolaños, Burneo, Florez-Alata, & Noriega (2019), indicate the introduction of business administration sciences to understand the criminal matter of economic motivation. This paper addresses the crime of economic motivation as an incorrect business model. In the case of Colombia, some statistical studies are presented, for example, for the spatial distribution of military victims of the armed conflict (Escobedo, Lis-Gutiérrez, & Aguilera, 2020) and extra-judicial executions (Lis-Gutiérrez, & Aguilera, 2020a, 2020b).

2. Method

This section describes the data used, the design of the study, the procedure and the indicators and models used for the analysis.

2.1. The data

This paper uses as its primary source the records of criminal news in the Oral Criminal Accusatory System (SPOA) (Fiscalía General de la Nación, 2019) for alleged criminal acts known to the Attorney General's Office since the entry into force of Law 906 of 2004 and Law 1098 of 2006 (Attorney General's Office, 2018), between 2005 and 2018. The total number of records analyzed corresponded to 4.680.711.

Table 1. Sources and units of data used

Variable	Variable values
Year	Year between 2005 and 2018
Department	Name of political-administrative unit
Municipality	Name of political-administrative unit
Status	Active / Inactive
Stage	Execution of sentence, inquiry, investigation, trial, indictment, early termination
Law	Law 906, 1098 and Juvenile Jurisdiction
Number of crimes	Number

Source: own elaboration based on Fiscalía General de la Nación (2019).

2.2. Study design and procedure

This study has a quantitative approach with correlational scope. The design is non-experimental panel longitudinal cut, being the agents the states of Colombia and the period January 2005 to December 2018.

We assume a set of n states of a country whose crime rates in a given year are represented by random variables X_1, X_2, \dots, X_n . These random variables take values in the positive integers and represent the crime counts of corresponding states of a country. Ignoring self-edges, a network of n states contains $n(n - 1)$ edges that are directed from one state to another. An edge from state i to k signifies a causal effect of X_i on X_k that is mediated directly or through unknown factors not measured in a given year. Note that a state can effect its own, or other states' crime rates indirectly through intermediate states, creating feedforward and feedback loops in the network.

Most approaches for inferring interactions rely on mutual information or Bayesian methods that impose restrictions on the inferred network. For instance, Bayesian network approaches assume that there are no cycles (or feedbacks) between the nodes (states) (Wang & Huang, 2014). Moreover, techniques based on mutual information assume interactions to be undirected, and hence provide no information about causality (Wang & Huang, 2014).

Alternatively, since crime rates is rather a dynamic process, time-course data can be used to infer causality. However, temporal data tends to exhibit high autocorrelation and is usually only gathered for a few time points and factors (in our case, geographical states). In addition, many temporal methods typically assume that the dynamics of a given factor studied at a time point depends on the remaining factors only the immediately

preceding time point, while other methods include additional time points but do not impose any constraints to account for their temporal distance.

The time-lagged Ordered Lasso is a recently published regularization method that uses monotonicity constraints on the coefficients of a linear model to reflect the relative importance of the model features and has natural applications to time-lagged regression. Instead of assuming no correlation between the crime rates of the states through the years sampled, let $x_i(t)$ be a time series that records the crime rates of the state i through the time

$$\begin{aligned} \min_{\{w_{j,i,k}\}} \quad & \frac{1}{2} \sum_{t=1}^T \left(x_i(t) - \sum_{j=1}^p \sum_{k=1}^{l_{max}} w_{j,i,k} x_j(t-k\Delta t) \right)^2 \\ & + \lambda \sum_{j=1}^p \sum_{k=1}^{l_{max}} |w_{j,i,k}| \\ \text{subject to} \quad & |w_{j,i,1}| \geq |w_{j,i,2}| \geq \dots \geq |w_{j,i,l_{max}}|, \end{aligned}$$

(months, years, etc.). Then the following problem using the time-lagged Ordered Lasso can be formulated: where the monotonicity constraint of the time correlation is given by $|w_{j,i,1}| \geq \dots |w_{j,i,l_{max}}|$, and l_{max} is the maximum lagged variable. $w_{j,i,1} \dots w_{j,i,l_{max}}$ are the weights of the connections in the causality network, i.e., the crime rates in time 1 of state j influences state i in the current time. Note that given there are n states, then we should solve n of this problems.

The Jenks algorithm was used for spatial representation (Jenks, 1963; Khan, 2012; North, 2009; Lis-Gutiérrez et al., 2019a, 2019b). Its formula corresponds to the following:

$$J = C - \sum_{1 \leq j \leq (k-1)} dist(c_{j+1}, c_j) \quad (1)$$

As the equation shows, the Jenks algorithm not only looks for the minimum distance between the data points and the cluster centers to which they belong, it also finds the maximum difference between the groups.

3. Results

We applied the time-lagged Ordered Lasso to crime data to uncover interaction networks between crime events happened on states in Colombia. Our results provide further insights into interactions leading to coordinate crime rates among the states studied. This made it possible to identify influences received and caused.

3.1. Influences received

Table 2 and Graph 1 show that the departments with the most interactions (influences received) were: Santander and Valle del Cauca with 24 each, followed by Guaviare and Vaupés with 23 interactions each, and Sucre 22 (Graph 2). Those with fewer interactions were: Cundinamarca (14), Tolima (15) and La Guajira (15).

Table 2. Influences received by department

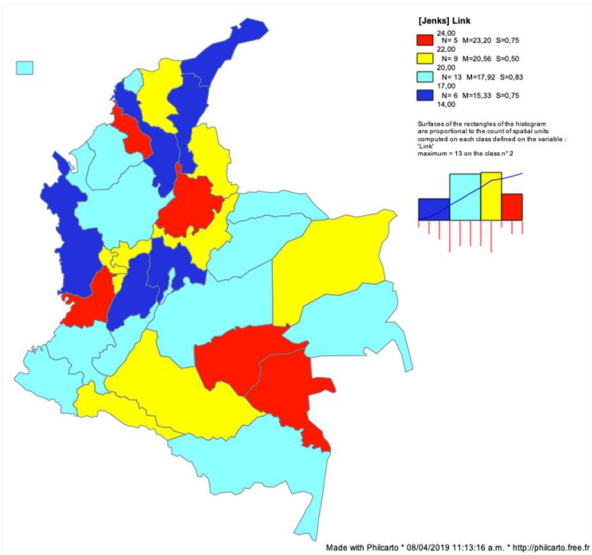
ID	Department	Number of links	Group according to Jenks algorithm
68	Santander	24	Rojo
76	Valle del Cauca	24	Rojo
95	Guaviare	23	Rojo
97	Vaupés	23	Rojo
70	Sucre	22	Rojo

ID	Department	Number of links	Group according to Jenks algorithm
15	Boyacá	21	Amarillo
17	Caldas	21	Amarillo
47	Magdalena	21	Amarillo
54	Norte de Santander	21	Amarillo
99	Vichada	21	Amarillo
18	Caquetá	20	Amarillo
86	Putumayo	20	Amarillo
63	Quindío	20	Amarillo
66	Risaralda	20	Amarillo
91	Amazonas	19	Azul claro
8	Atlántico	19	Azul claro
23	Córdoba	19	Azul claro
41	Huila	19	Azul claro
5	Antioquia	18	Azul claro
81	Arauca	18	Azul claro
11	BOGOTÁ, D. C.	18	Azul claro
94	Guainía	18	Azul claro
88	Archipiélago de San Andrés, Providencia y Santa Catalina	17	Azul claro
85	Casanare	17	Azul claro
19	Cauca	17	Azul claro
50	Meta	17	Azul claro
52	Nariño	17	Azul claro
13	Bolívar	16	Azul oscuro
20	Cesar	16	Azul oscuro
27	Chocó	16	Azul oscuro
44	La Guajira	15	Azul oscuro
73	Tolima	15	Azul oscuro
25	Cundinamarca	14	Azul oscuro

Source: own elaboration.

Graph 1 shows the distribution of the number of links according to Jenks algorithm, organized in 4 groups, according to Table 2.

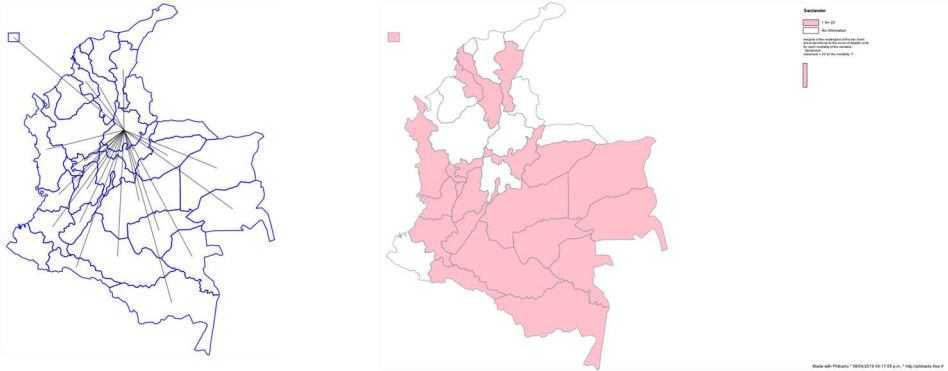
Graph 1. Distribution of the number of influences received according to Jenks algorithm.



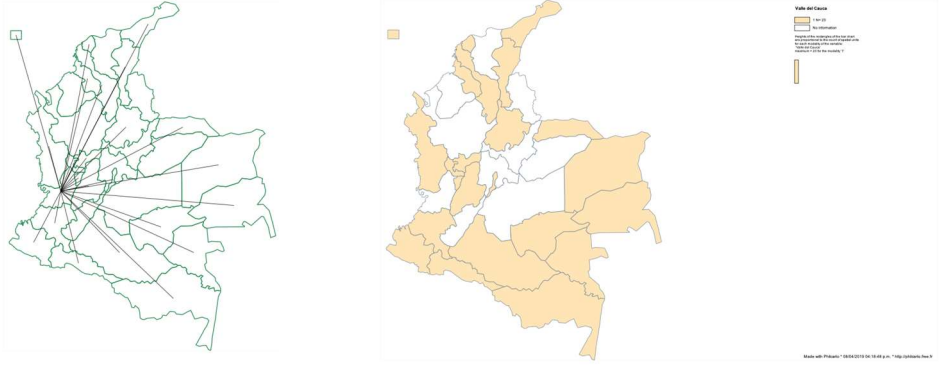
Source: own elaboration, using Philcarto (Waniez, 2020).

Graph 2. Departments with the most influences received

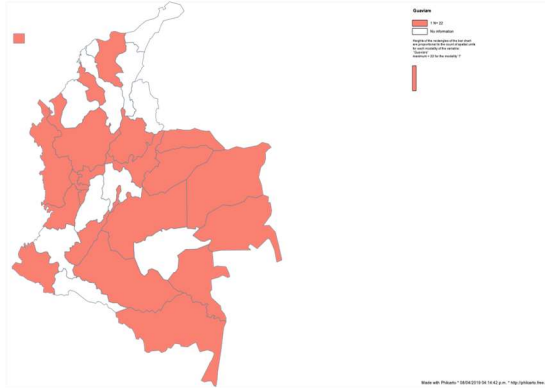
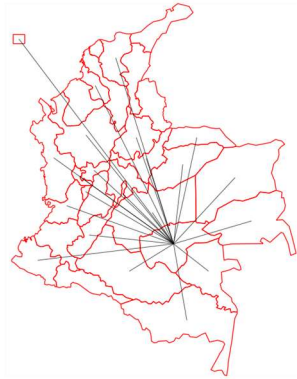
S
a
n
t
a
n
d
e
r



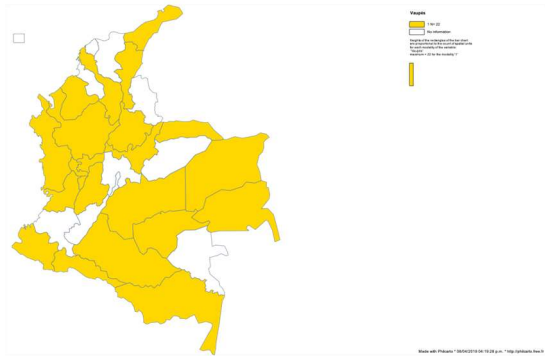
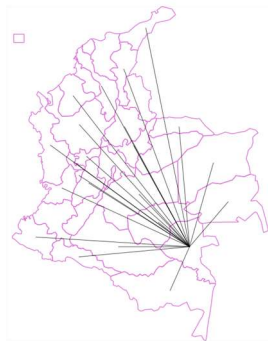
V
a
l
l
e
d
e
l
C
a
u
c
a



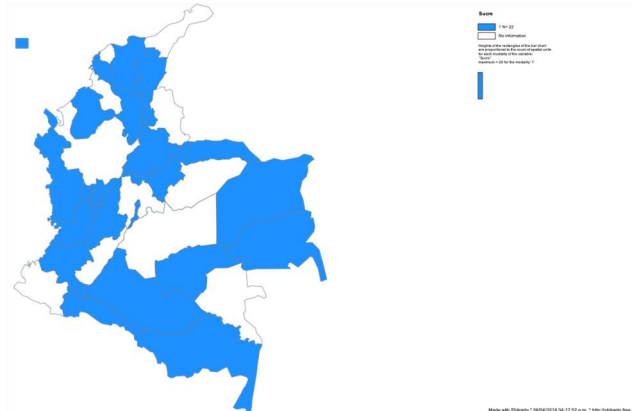
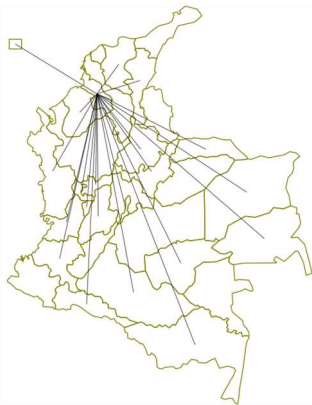
G
u
a
v
i
a
r
e



V
a
u
p
é
s



S
u
c
r
e



Source: own elaboration, using Philcarto (Waniez, 2020).

3.2. Influences caused

Table 3 and Graph 3 show that the departments with the most interactions (influences caused) were: Vichada (32), Quindío (30), Chocó and Risaralda (29), Caldas and Guainía (28) (Graph 3). Those with fewer interactions were: Magdalena (7), Norte de Santander (5) and Atlántico (4).

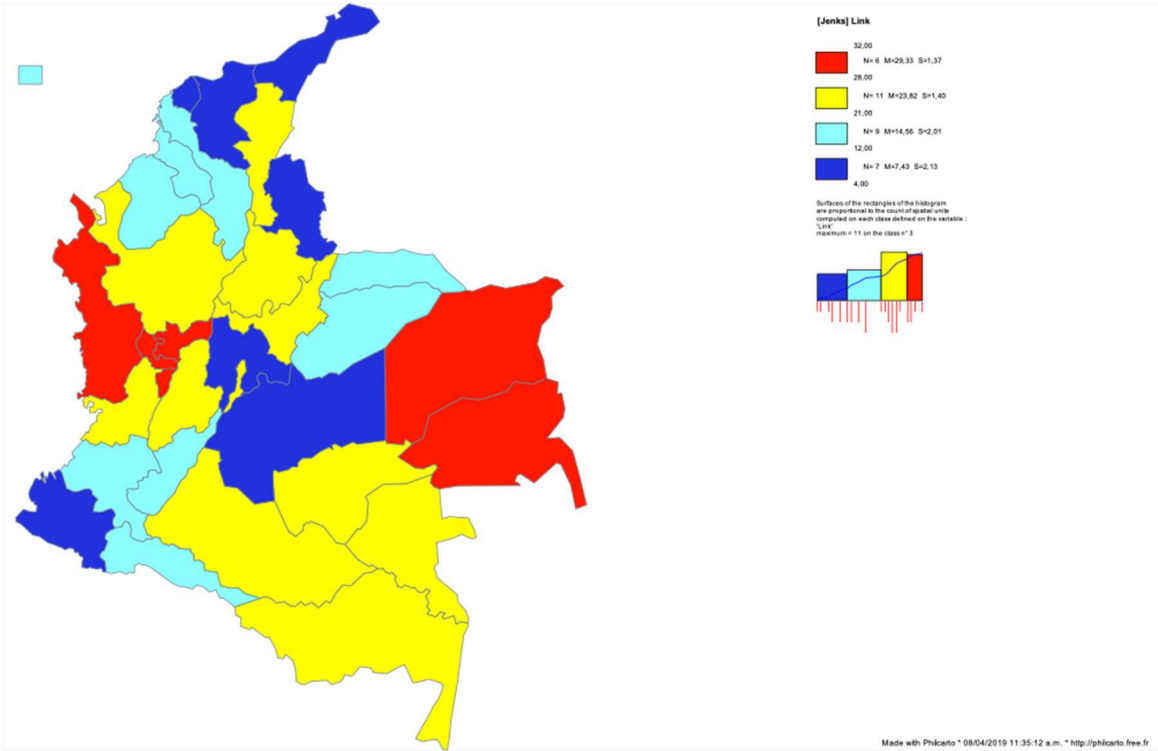
Table 3. Influences caused by department

ID	Department	Number of links	Group according to Jenks algorithm
99	Vichada	32	Rojo
63	Quindío	30	Rojo

ID	Department	Number of links	Group according to Jenks algorithm
27	Chocó	29	Rojo
66	Risaralda	29	Rojo
17	Caldas	28	Rojo
94	Guainía	28	Rojo
18	Caquetá	26	Amarillo
11	Bogotá, D. C.	25	Amarillo
20	Cesar	25	Amarillo
76	Valle del Cauca	25	Amarillo
91	Amazonas	24	Amarillo
15	Boyacá	24	Amarillo
95	Guaviare	24	Amarillo
68	Santander	23	Amarillo
97	Vaupés	23	Amarillo
5	Antioquia	22	Amarillo
73	Tolima	21	Amarillo
13	Bolívar	17	Azul claro
19	Cauca	17	Azul claro
86	Putumayo	17	Azul claro
81	Arauca	15	Azul claro
23	Córdoba	15	Azul claro
85	Casanare	13	Azul claro
70	Sucre	13	Azul claro
88	Archipiélago de San Andrés, Providencia y Santa Catalina	12	Azul claro
41	Huila	12	Azul claro
44	La Guajira	10	Azul oscuro
50	Meta	10	Azul oscuro
25	Cundinamarca	8	Azul oscuro
52	Nariño	8	Azul oscuro
47	Magdalena	7	Azul oscuro
54	Norte de Santander	5	Azul oscuro
8	Atlántico	4	Azul oscuro

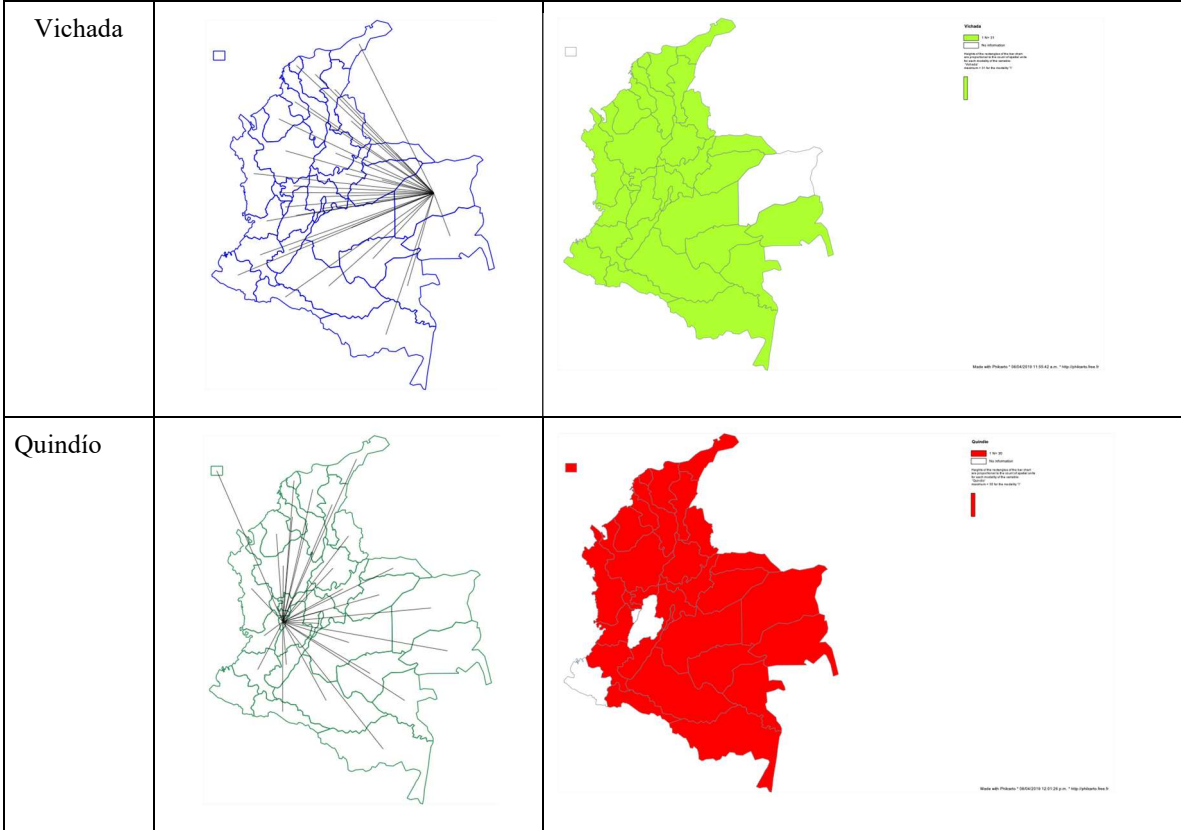
Fuente: elaboración propia.

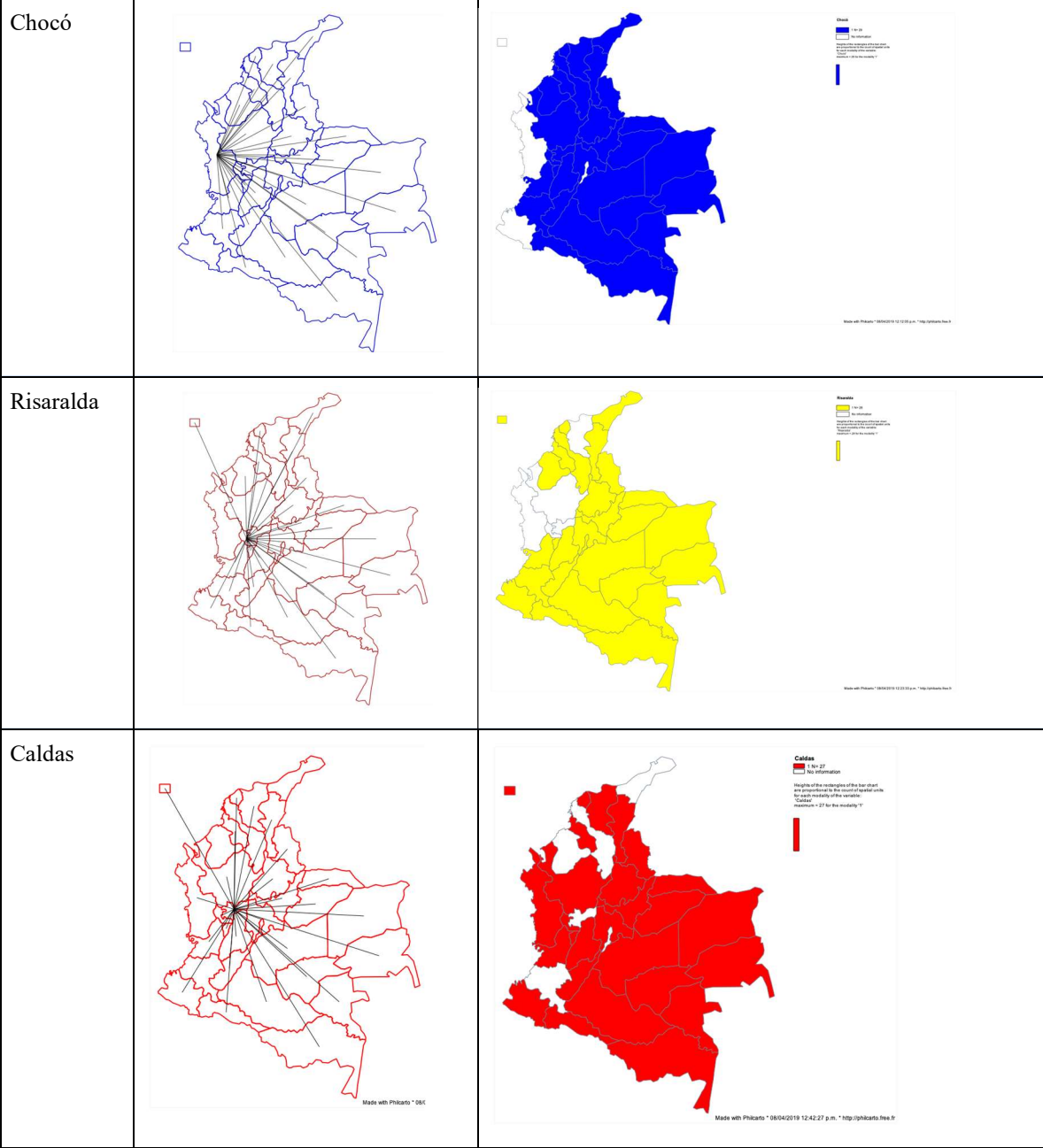
Graph 3. Distribution of the number of influences caused according to Jenks algorithm.

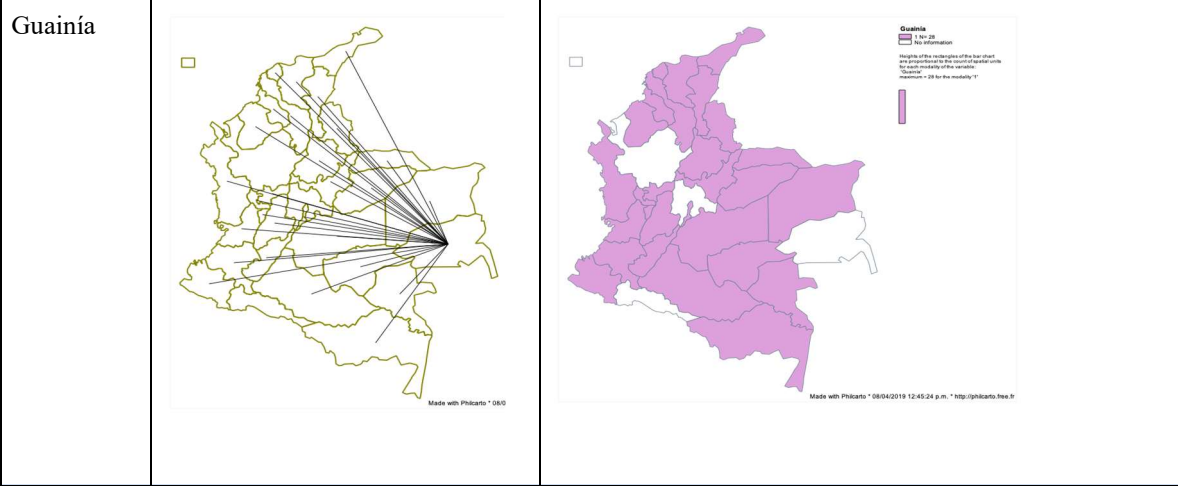


Source: own elaboration, using Philcarto (Waniez, 2020).

Graph 4. Departments with the most influence received

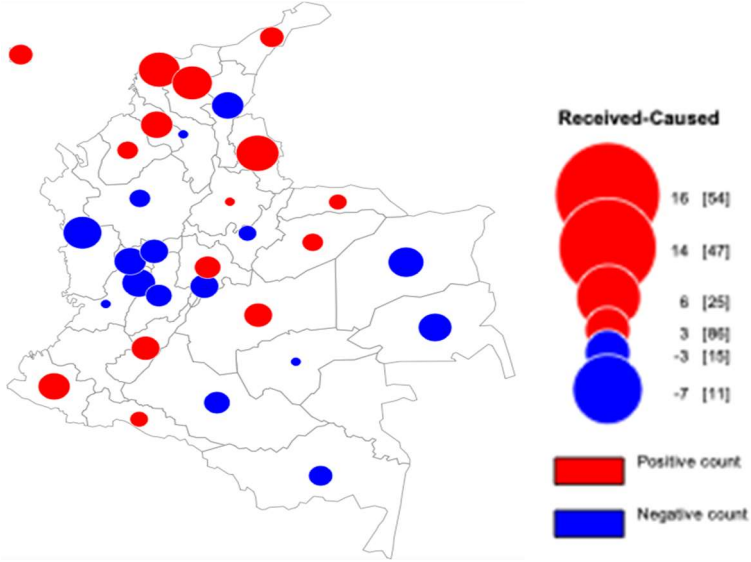




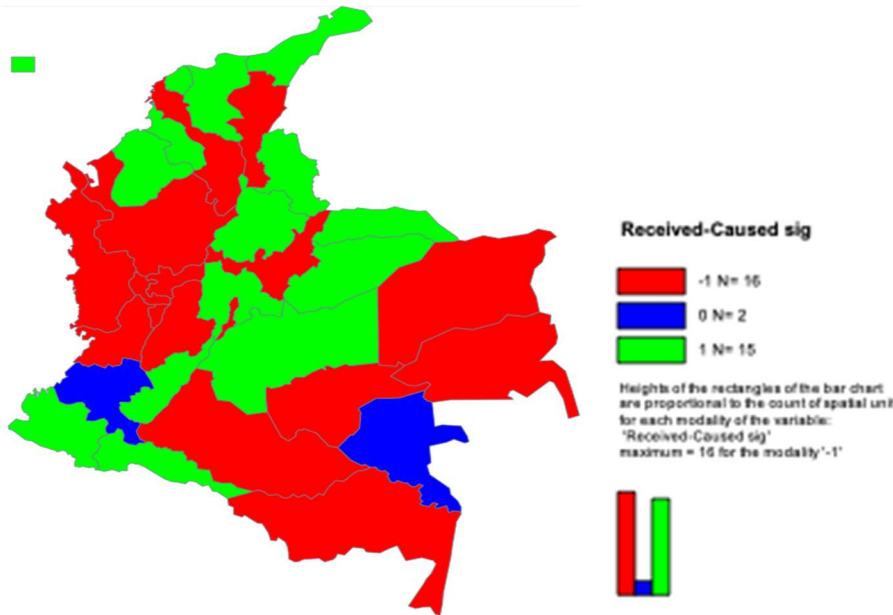


Source: own elaboration, using Philcarto (Waniez, 2020).

Graph 5 shows the difference between the influences received and those caused. Those departments with red dots (positive balance) imply that the effects on their spatial unit are greater than those caused. Conversely, those departments with blue dots (negative balance) imply that the effects on their spatial unit are less than those caused.



Graph 5. Difference between received and caused influences



Source: own elaboration, using Philcarto (Waniez, 2020).

Also, in the second map, the blue departments (Cauca and Vaupés) present a null effect, since the number of influences received and caused in each spatial unit is equal (17 and 23, respectively).

4. Discussion and conclusions

Using The time-lagged Ordered Lasso, it was possible to identify the spatial units that influenced other departments regarding the alleged criminal acts known to the Attorney General's Office since the entry into force of Law 906 of 2004 and Law 1098 of 2006, after the implementation of the Accusatory Oral Criminal System (SPOA).

It was found that all departments are simultaneously generators and receivers of conditions for the flourishing of the crime. Likewise, it was established that most of the departments cause more effects on other spatial units than those received (which generate more crimes): Amazonas, Caquetá, Guaviare, Guainía, Vichada, Bogotá, Boyacá, Chocó, Valle del Cauca, Antioquia, Caldas, Bolívar, Cesar, Tolima, Risaralda and Quindío.

Cauca and Vaupés are in balance in terms of the relationship between incidents or effects caused and received. These two departments share the following characteristics: presence of drug trafficking activities, indigenous people, high levels of corruption, presence of residual Organized Armed Groups, arms trafficking, and social inequality.

This study presented some limitations, such as (i) the joining of the different databases provided by the Fiscalía General de la Nación (one per year), (ii) a sample of 18 months after the signing of the peace process. Options for future work include: (i) disaggregate the different types of crimes and verify whether structural changes are maintained for all of them; (ii) apply spatial analysis techniques to identify spatial patterns (Vargas-García, 2020); (iii) apply supervised learning algorithms to predict the behavior of the most frequent crimes in the national territory; (iv) carry out analysis at the municipal level.

Referencias

- [1] Ahmar, A. S., & Aidid, M. K. (2018). Crime Modeling using Spatial Regression Approach. In *Journal of Physics: Conference Series*, 954(1), p. 012013). Retrieved from <http://iopscience.iop.org/article/10.1088/1742-6596/954/1/012013/meta>
- [2] Andresen, M. A. (2006). A spatial analysis of crime in Vancouver, British Columbia: A synthesis of social disorganization and routine activity theory. *The Canadian Geographer/Le Géographe canadien*, 50(4), 487-502.
- [3] Andresen, M. A. (2011). Estimating the probability of local crime clusters: The impact of immediate spatial neighbors. *Journal of Criminal Justice*, 39(5), 394-404.
- [4] Anselin, L. (2019). A local indicator of multivariate spatial association: extending Geary's C. *Geographical Analysis*, 51(2), 133-150. Retrieved from http://priede.bf.lu.lv/ftp/pub/TIS/datu_analiize/GeoDa/LA_multivariateGeary1.pdf
- [5] Brantingham, P. J. (2016). Crime diversity. *Criminology*, 54(4), 553-586.
- [6] Castro Valdebenito, H. J. (2018). Marco Antonio León León. Construyendo un sujeto criminal. Criminología, criminalidad y sociedad en Chile. Siglos XIX y XX. Santiago: Centro de Investigaciones Diego Barros Arana-DIBAM-Editorial Universitaria, 2016. *Historia Caribe*, 13(32), 219-222.
- [7] Escobedo, R., Lis-Gutiérrez, J. P., & Aguilera, D. (2020) Víctimas Militares en Colombia, 2004-2017: Un análisis Estadístico. En: A. Ciro y G. Barbosa. *Víctimas del Conflicto armado Colombiano* (pp 305-348). Bogotá, Colombia: Universidad Externado.
- [8] Fiscalía General de la Nación (2019). Estadísticas [Base de datos]. Bogotá: Fiscalía General de la Nación.
- [9] Fuentes Flores, C. M., & Sánchez Salinas, O. A. (2017). La distribución espacial del robo a transeúntes y el contexto socioeconómico en tres delegaciones de la Ciudad de México. *Elementos para una política de seguridad pública. Gestión y política pública*, 26(2), 417-451.
- [10] Jenks, G. F. (1963). Generalization in statistical mapping. *Annals of the Association of American Geographers*, 53(1), 15-26.
- [11] Khan, F. (2012). An initial seed selection algorithm for k-means clustering of georeferenced data to improve replicability of cluster assignments for mapping application. *Applied Soft Computing*, 12(11), 3698-3700. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1604/1604.04893.pdf>.
- [12] Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. G., & Brantingham, P. J. (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built environment*, 34(1), 62-74.
- [13] Lis-Gutierrez, J. P & Aguilera-Hernandez, D. (2020a). *Minería de datos aplicada a las ejecuciones extrajudiciales en Colombia*. En A. Ciro y G. Barbosa. *Garantía de no repetición: una contribución a la justicia transicional*. Bogotá: Universidad Externado de Colombia.
- [14] Lis-Gutierrez, J. P & Aguilera-Hernandez, D. (2020b). *Análisis estadístico de las ejecuciones extrajudiciales en Colombia*. En A. Ciro y G. Barbosa. *Garantía de no repetición: una contribución a la justicia transicional*. Bogotá: Universidad Externado de Colombia.
- [15] Lis-Gutiérrez, J. P., Zerda Sarmiento, A., Balaguera, M.I. Gaitán-Angulo, M., Lis-Gutiérrez, M. (2019a). *Uso del sistema de propiedad industrial para signos distintivos en Colombia: un análisis departamental (2000-2016)*. En: G.A. Campos Avendaño, M.A. Castaño Hernández, M. Gaitán Ángulo y V. Sánchez Mendoza (Compiladores). *Diálogos sobre investigación: avances científicos Konrad Lorenz (195-2014)*. Bogotá: Konrad Lorenz Editores. Disponible en: https://blogs.konradlorenz.edu.co/files/book_interactivo_dialogos_en_investigacion.pdf
- [16] Lis-Gutiérrez, J.P., Rincón, J.C., Gaitán-Angulo, M., Cubillos, J. Vargas, C. (2019b) Hechos Victimizantes en Colombia; antes, durante y después de la firma del Acuerdo de paz en la Habana. Konrad Lorenz (Ed) *Diálogos sobre Investigación: Avances Científicos Konrad Lorenz* (pp. 27-48). Bogotá, Colombia: Konrad Lorenz.
- [17] Lizaraburu-Bolaños, E. R., Burneo, K., Flórez-Alata, H. G., & Noriega, L. (2019). Visión empresarial como nuevo aporte a la teoría beckeriana del crimen en el Perú. *Revista Criminalidad*, 61(1), 29-50.
- [18] Nguyen, P., & Braun, R. (2018). Time-lagged Ordered Lasso for network inference. *BMC bioinformatics*, 19(1), 545.
- [19] North, M. A. (2009). A method for implementing a statistically significant number of data classes in the Jenks algorithm. In *Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 35-38). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/5358673/>.

- [20] Organización Panamericana de la Salud (OPS). (2019). The Health of Adolescents and Youth in the Americas. Implementation of the Regional Strategy and Plan of Action on Adolescent and Youth Health 2010-2018. Washington, D.C.: OPS. Retrieve from: <http://iris.paho.org/xmlui/handle/123456789/49545>
- [21] Quick, M., Li, G., & Brunton-Smith, I. (2018). Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, 58, 22-32. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0047235218301971>
- [22] Roncek, D. W., & Maier, P. A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of “hot spots”. *Criminology*, 29(4), 725-753.
- [23] Schreck, C. J., McGloin, J. M., & Kirk, D. S. (2009). On the origins of the violent neighborhood: A study of the nature and predictors of crime-type differentiation across Chicago neighborhoods. *Justice Quarterly*, 26(4), 771-794.
- [24] Singh, N. K., Ahsen, M. E., Mankala, S., Kim, H. S., White, M. A., & Vidyasagar, M. (2012). Reverse engineering gene interaction networks using the phi-mixing coefficient. *arXiv preprint arXiv:1208.4066*.
- [25] Vargas-García, C.A., Lis-Gutiérrez, J.P., Gaitán-Angulo, M., Balaguera, M.I. y Zapata Patarroyo, H. (2018). Evolución de la criminalidad en Colombia (2005-2018). En: J.P. Lis-Gutiérrez, C. Henao y L.E. Malagón-Castro. Técnicas de análisis cuantitativo aplicadas a las ciencias contables y económicas (240-260). Villavicencio: Unimeta.
- [26] Vargas-García, C., Lis-Gutiérrez, J. P., Gaitán-Angulo, M., & Lis-Gutiérrez, M. (2020, July). Parasite-Guest Infection Modeling: Social Science Applications. In *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*) (pp. 597-603). Springer, Cham. DOI https://doi.org/10.1007/978-3-030-53956-6_55
- [27] Wang, Y. R., & Huang, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *Journal of theoretical biology*, 362, 53-61.
- [28] Waniez, P. (2020). *Philcarto* [software]. Available at <http://philcarto.free.fr>.

Capítulo 8

pp 95-115

COLOMBIA EN EL COMERCIO DE CUERO, CALZADO Y MARROQUINERÍA CON CHILE: VENTAJAS COMPARATIVAS Y SUS DETERMINANTES

Jenny Paola Danna-Buitrago¹, Rémi Stellan², Carolina Henao³, Ana Fernández⁴,
M. Andreina Moros Ochoa⁵

Resumen

Este trabajo analiza el posicionamiento de Colombia frente a Chile en el comercio de productos de cuero, calzado y marroquinería (CCM) entre ambos países. Después de señalar diferentes mecanismos de liberalización, se muestra que las exportaciones colombianas de CCM hacia Chile son impulsadas por el calzado. Asimismo, se constata que las importaciones colombianas de CCM desde Chile decrecen desde el 2012, pero siguen siendo impulsadas por el cuero manufacturado. Luego, a partir de análisis en términos de ventajas comparativas reveladas se establece que el calzado colombiano tiende a consolidar sus ventajas comparativas frente a sus homólogos producidos en Chile, mientras que el cuero manufacturado tiende a consolidar sus desventajas comparativas. Además, se explica por qué las ventajas comparativas son reveladas a partir del indicador de contribución al saldo comercial, y se aplica un modelo probit que permite explorar los determinantes de tales ventajas. El trabajo termina con una discusión sobre los aportes del Programa de Transformación Productiva (PTP) al desempeño de los productos colombianos CCM en Chile, con énfasis en el calzado y el cuero manufacturado, lo que se traduce en un factor para el crecimiento económico colombiano.

Palabras claves: Cuero, calzado y marroquinería; Colombia; Chile; Comercio; Ventajas Comparativas Reveladas; Probit; Programa de Transformación Productiva.

Colombia and trade in Leather, footwear and leather goods with Chile: comparative advantages and their analysis

Abstract

This paper analyzes the position of Colombia in the trade of leather, footwear and leather goods (LFL) with Chile. First, different trade liberalization mechanisms are presented. In this regard, Colombian LFL exports to Chile are driven by footwear. In addition, Colombian LFL imports from Chile have been decreasing since 2012, but continue to be driven by manufactured leather. An analysis in terms of revealed comparative advantages suggests that Colombian footwear tends to consolidate its comparative advantages in the Colombia-Chile trade area, while manufactured leather tends to consolidate its comparative disadvantages. Comparative advantages are measured by the contribution-to-the-trade-balance index, and a probit model studies the determinants of comparative advantages. The paper ends discussing the influence of the Colombian Productive Transformation Plan (PTP) on the performance of Colombian LFL products in Chile with emphasis on footwear and manufactured leather, which contributes to Colombian economic growth.

Keywords: Leather, footwear and leather goods; Colombia; Chile; International Trade; Comparative Advantages Revealed; Probit; Productive Transformation Plan.

¹ Fundación Universitaria Los Libertadores (Colombia). Contacto: jenny.danna@libertadores.edu.co

² Profesor Facultad de Ciencias Económicas y Administrativas, Pontificia Universidad Javeriana (Colombia) e integrante del grupo de investigación Riesgos Financieros y Métodos de Valoración de Empresas-RISVAL de la misma institución. Contacto: rstellian@javeriana.edu.co.

³ Profesor investigador Fundación Universitaria Konrad Lorenz (Colombia). Contacto: lindac.henaor@konradlorenz.edu.co

⁴ Fundación Universitaria Los Libertadores (Colombia). Contacto: observatorio.tlc@libertadores.edu.co

⁵ Profesora Titular del Colegio de Estudios Superiores de Administración CESA y miembro del Grupo de Estudios de Administración del CESA (Colombia). Contacto: andreina.moros@cesa.edu.co

INTRODUCCIÓN

Desde hace más de dos décadas, los diferentes gobiernos colombianos han multiplicado la participación del país en acuerdos de integración comercial para apoyar una estrategia de crecimiento económico, expresada en los distintos Planes Nacionales de Desarrollo (PND) desde 1990, a través de los intercambios internacionales⁶. Esto, debido a que las exportaciones son consideradas como una oportunidad de vender la producción nacional en el extranjero, y las importaciones como la manera de conseguir activos y otros bienes de capital que mejoran la productividad y la calidad de la oferta exportable, creando así un círculo virtuoso que sostiene la producción, el empleo y la estabilidad macroeconómica (Stellian & Danna-Buitrago, 2017a). Para ello, la Ley 7 de 1991 plantea en su segundo artículo los siguientes principios: i) “Impulsar la internacionalización de la economía colombiana para lograr un ritmo creciente y sostenido de desarrollo”; ii) “Promover y fomentar el comercio exterior de bienes, tecnología, servicios y en particular, las exportaciones”; y III) “Estimular los procesos de integración y los acuerdos comerciales bilaterales y multilaterales que amplíen y faciliten las transacciones externas del país”.

Chile hace parte de los países hacia los cuales Colombia ha dirigido sus esfuerzos de integración económica internacional, incluso antes de la Ley 7 de 1991. En efecto, ambos países son miembros fundadores de la Asociación Latinoamericana de Integración (ALADI), creada en 1980 e implementada en 1981, cuyo propósito es la conformación de un mercado común latinoamericano. Desde entonces, el comercio se ha ido liberalizando y los dos países cumplen actualmente con las normas establecidas en el último Acuerdo de Libre Comercio (ALC) firmado en 2006.

Sin embargo, Colombia y Chile presentan un perfil exportador relativamente similar, orientado en gran medida por los recursos mineros y energéticos. Por lo tanto, la liberalización del comercio entre ambos países puede ponerse en tela de juicio por la falta de complementariedad económica (Lassudrie-Duchêne, 1971), lo que podría resultar en un débil dinamismo de los intercambios entre bilaterales (Duarte, González & Montoya, 2015). Al respecto, la política pública colombiana ha dedicado recursos para diversificar la oferta exportable del país, con el fin de que sus exportaciones sean menos dependientes de los recursos mineros y energéticos, principalmente mediante el Programa de Transformación Productiva (PTP) creado en 2009; un programa que apunta a mejorar la competitividad de las empresas colombianas pertenecientes a varias ramas consideradas como estratégicas debido a su potencial exportador⁷. Esto resulta relevante para Colombia, pues podría permitirle modificar su esquema de inserción internacional, mediante la reducción de exportaciones de materias primas a favor de productos de mayor valor agregado, lo que a su vez le permitiría reducir las importaciones de dichos productos (Torres, 2013).

Entre las ramas priorizadas por el PTP se encuentran por ejemplo los productos cosméticos y de aseo, los plásticos y la pintura, la química básica, los productos farmacéuticos, el cacao y sus derivados, la carne bovina y los productos lácteos⁸. Asimismo, el PPT contempla los productos de Cuero, Calzado y Marroquinería (en adelante productos CCM), gracias a los cuales Colombia podría convertirse en “líder de producción y exportaciones en la región y referente internacional en la creación de moda en manufacturas de cuero, basado en una industria a la vanguardia tecnológica, de alto valor agregado y con marcas posicionadas internacionalmente”⁹. Además, el PTP señala que el potencial exportador de Colombia para los productos en mención “es alto gracias a los tratados de libre comercio firmados por Colombia y a la creatividad, diferenciación y calidad en el diseño de los productos nacionales”¹⁰.

Al respecto, varios autores que estudian el potencial exportador de los productos colombianos CCM, indican que, es posible ganar partes de mercado en diferentes países y llegar a ser líder de producción y exportaciones, pero que para lograrlo es imperativo seguir innovando a fin de reforzar la competitividad y disponer de ventajas comparativas (Velásquez & Corrales, 2013; Forero, 2014; Arango, Betancourt & Martínez, 2015). En esta línea, el Ministerio colombiano de Comercio, Industria y Turismo señala que es posible lograr tal cometido en 20 mercados extranjeros que ofrecen oportunidades significativas de venta para dichos productos¹¹, entre los

6 Para los Planes Nacionales de Desarrollo de Colombia, véase: www.dnp.gov.co/DNPN/Plan-Nacional-de-Desarrollo/Paginas/Planes-de-Desarrollo-antiores.aspx

7 Para más información sobre el PTP, véase: www.colombiaproductiva.com/ptp-sectores

8 Para los demás productos priorizados por el PTP, véase www.ptp.com.co/

9 www.ptp.com.co/contenido/categoria.aspx?catID=782

10 www.ptp.com.co/ptp-sectores/historico/cuero-calzado-marroquineria

11 www.portafolio.co/economia/los-mercados-identificados-para-el-cuero-y-el-calzado-513770

cuales se encuentra Chile; Pero, al igual que los autores arriba mencionados, insiste en la necesidad de innovar para adquirir y consolidar ventajas comparativas que permitan incursionar y mantenerse en tales mercados. Finalmente, tanto el Ministerio como el PPT señalan la importancia de dinamizar las exportaciones de los productos CCM hacia Chile debido a las oportunidades que ofrece dicho mercado, indicando que esto impactaría positivamente en las actividades de más de 30.000 empresas relacionadas, desde los insumos hasta la comercialización.

Pese a la relevancia del tema por el potencial de los productos CCM en términos de generación de empleos e ingresos en Colombia por la cantidad de empresas relacionadas en todo el proceso, son pocos los trabajos que se interesan en el posicionamiento, ya sea del calzado, del cuero o de la marroquinería de origen colombiano, desde el punto de vista de las ventajas comparativas. Por una parte, García y Maldonado (2013) miden las ventajas comparativas del calzado de cuero colombiano frente a Estados Unidos y el mundo en distintos periodos entre 1980-2008. Por otra parte, Villegas y Zapata (2007) analizan la competitividad del cuero y del calzado, mediante diferentes indicadores entre 1999 y 2005, y pretenden hallar las ventajas comparativas de Colombia respecto a la Comunidad Andina¹². Sin embargo, ninguno de los trabajos se interesa en las relaciones comerciales entre Colombia y Chile, en las ventajas comparativas de los productos colombianos CCM frente a Chile, ni tampoco incluyen en sus cálculos todas las categorías de productos que conforman la rama CCM. Además, es posible afirmar que dichos trabajos carecen de un marco teórico robusto desde el cual se indique lo que entienden por ventajas comparativas, y los factores que inciden en su conformación. Lo mismo puede decirse sobre la escogencia de los indicadores pues no presentan justificación.

Teniendo en cuenta lo anterior, surgen una serie de cuestionamientos que guían el desarrollo de esta investigación: ¿Cómo ha sido la liberalización comercial de los productos CCM entre Colombia y Chile?, ¿Cómo ha sido históricamente el comercio de productos CCM entre Colombia y Chile?, ¿Existe un único producto que determine el comportamiento de las exportaciones colombianas del grupo CCM hacia Chile?, ¿Cómo determinar si los productos colombianos CCM disponen de ventajas comparativas frente a Chile?, ¿Que método debería utilizarse para determinarlo y por qué?, ¿Qué productos colombianos del grupo CCM disponen de (des)ventajas comparativas, siendo este un factor determinante del saldo comercial registrado entre los dos países?, ¿Cuáles son los determinantes de tales ventajas en caso de existir?. En este orden de ideas, el objetivo de esta investigación no es otro que determinar el posicionamiento de los productos colombianos CCM frente a sus homólogos chilenos, mediante el método de ventajas comparativas reveladas, así como los principales determinantes de tales ventajas.

El capítulo se divide en 4 secciones. En la primera, se hace una revisión de la literatura y una serie de análisis que permiten señalar cómo el comercio bilateral de productos CCM entre Colombia y Chile ha sido liberalizado, cómo ha sido el posicionamiento histórico de tales productos frente a sus homólogos chilenos y qué productos CCM han impulsado las exportaciones y/o las importaciones de Colombia hacia/desde Chile. En la segunda sección, se presenta la metodología, es decir, el método de medición de las ventajas comparativas elegido para el desarrollo de esta investigación, así como la justificación de tal elección. También se presenta el protocolo econométrico para la estimación de un modelo probit que permite determinar los factores que influyen en la conformación de las ventajas comparativas de los productos CCM. En la tercera sección, se presentan y discuten los resultados de la medición de las Ventajas Comparativas Reveladas (VCR) de los productos CCM de Colombia frente a Chile. Además, se presentan los resultados obtenidos de la estimación del modelo probit para determinar los factores que influyen en la conformación de las ventajas comparativas de los productos CCM fabricados en Colombia frente a Chile. En la cuarta sección, se presentan una discusión del papel del PTP frente a la evolución de los intercambios y de las ventajas comparativas de Colombia en la zona de intercambios Colombia-Chile en cuanto a los productos CCM. Finalmente, se enuncian las conclusiones de la investigación.

Así las cosas, es posible afirmar que este trabajo llena varios vacíos en la literatura: *i)* Se interesa en el conjunto de productos CCM; *ii)* Realiza mediciones para un periodo extendido, abordando más años respecto a los trabajos existente; *iii)* Utiliza un indicador de Ventajas Comparativas Reveladas conocido por su superioridad teórica y empírica (a diferencia de los indicadores utilizados en los trabajos arriba mencionados); *iv)* Se interesa

¹² Para una revisión de la literatura sobre los trabajos que se interesan en la competitividad de los productos colombianos de Calzado, Cuero y Marroquinería a partir de diferentes métodos, véase Páez, Jiménez & Danna-Buitrago (2018).

en las ventajas comparativas de los productos CCM de Colombia frente a Chile, lo que no se ha efectuado en ningún otro trabajo, y por último, v) Evidencia los determinantes de dichas ventajas en el mercado chileno.

1. EL COMERCIO BILATERAL DE PRODUCTOS CCM ENTRE COLOMBIA Y CHILE: REVISIÓN DE LA LITERATURA Y ANÁLISIS

Esta sección se divide en dos partes. Primero, se presentan la manera mediante la cual el comercio entre Colombia y Chile ha sido liberalizado, haciendo énfasis en los términos de liberalización del comercio de los productos CCM (1.1). Segundo, se señala cómo ha sido el comercio de tales productos entre los dos países y se determina qué productos CCM han sido los más relevantes en la dinamización de dicho comercio (1.2).

TÉRMINOS DE LIBERALIZACIÓN DEL COMERCIO DE PRODUCTOS CCM ENTRE COLOMBIA Y CHILE

El comercio bilateral Colombia-Chile de productos CCM ya ha sido completamente liberalizado. Esta situación toma su origen en 1980, año durante el cual Chile y Colombia se convierten en miembros fundadores de la Asociación Latinoamericana de Integración (ALADI), cuyo propósito es la conformación de un mercado común latinoamericano. Entre los diferentes mecanismos que apoyan este proyecto, los miembros de la ALADI tienen la posibilidad de firmar acuerdos “de alcance parcial” en los cuales participan solo una parte de sus miembros (artículos del 6 al 9 del Tratado de Montevideo, el cual crea la ALADI). Este tipo de acuerdo pretende favorecer futuros esfuerzos de liberalización a nivel de la ALADI para la totalidad de sus miembros, es decir, favorecer futuros acuerdos que se denominan “de alcance regional”.

Precisamente, Chile y Colombia utilizaron el mecanismo de los acuerdos “de alcance parcial” para liberalizar su comercio transfronterizo. En abril de 1983, los dos países firmaron el acuerdo de alcance parcial referenciado “R14”, el cual aclara las reglas de origen, y reduce o elimina los aranceles aduaneros para una lista de productos: aproximadamente 80 productos para Chile y 60 para Colombia. Entre estos productos, Chile autorizó por dos años la libre importación de cueros artificiales o regenerados procedentes de Colombia con un arancel del 5%, contra el 20% para otros países. Sin embargo, salvo esta excepción y pese a varios protocolos adicionales implementados hasta 1992 para ampliar la lista de productos sujetos a preferencias, los productos CCM no fueron tenidos en cuenta en los esfuerzos de liberalización contemplados en el acuerdo R14.

La liberalización del comercio de productos CCM entre Colombia y Chile empezó realmente a partir de 1995 debido a dos factores:

1. El primer factor es la adhesión de ambos países a la Organización Mundial del Comercio (OMC), creada en 1995. Bajo el subsecuente nuevo régimen comercial multilateral, Chile se comprometió a aplicar derechos aduaneros del 25% (en promedio) para los productos CCM procedentes de otros países miembros de la OMC, incluso Colombia. Igualmente, frente a los otros países miembros de la OMC, incluso Chile, Colombia se comprometió a aplicar derechos aduaneros del 8,238% (en promedio) para las pieles sin curtir, del 18,855% para las manufacturas de cuero, y del 19,268% para el calzado¹³.
2. El segundo factor es la sustitución del acuerdo R14 por el acuerdo CE24, firmado en diciembre de 1993 y complementado por seis protocolos adicionales hasta 1997. Además de aportar nuevas aclaraciones en cuanto a las reglas de origen, este acuerdo estipuló la eliminación total de los aranceles aduaneros a más tardar en 1999, salvo una lista de excepciones para las cuales la eliminación debió efectuarse en 5 años, entre 2007 y 2012 (véase el anexo 3 del acuerdo). De la misma manera, el CE24 mantiene aranceles aduaneros en el marco del Sistema Andino de Franjas de Precios (SAFP) para algunos productos agropecuarios¹⁴. Todos los productos CCM hicieron parte del proceso de desgravación general del acuerdo CE24, por lo que a partir de 1997 no se aplicaba ningún arancel aduanero para el comercio de dichos

¹³ Datos obtenidos de *WTO Tariff Analysis Online (TAO)*, <http://tao.wto.org/>.

¹⁴ Según el Sistema Andino de Franjas de Precios (SAFP), se aumenta el arancel aduanero si el precio internacional está por debajo de un nivel piso, y se reduce dicho arancel, incluso hasta cero, si el precio ha superado un nivel techo. Como ejemplos de productos que benefician del SAFP, es posible mencionar: la carne, los despojos de cerdo y pollo, los lácteos, el maíz y la harina de maíz, el arroz, el sorgo, los aceites y grasas, los azúcares, las tortas y demás residuos de la extracción del aceite.

productos entre Colombia y Chile. Así pues, la ausencia de arancel aduanero, obtenido por la negociación bilateral Colombia-Chile, superó todos los futuros esfuerzos de liberalización multilateral a través de la OMC.

Además, Chile y Colombia firmaron en 2006 tres nuevos protocolos adicionales al acuerdo CE24, después de los seis protocolos adicionales arriba mencionados, con el objetivo de profundizar la liberalización comercial entre ambos países¹⁵. El noveno y último protocolo adicional crea el Acuerdo de Libre Comercio (ALC), cuyo propósito es “establecer una zona de libre comercio” (artículo 1.1). En este sentido y basándose en el acuerdo CE24, el ALC formalizó nuevas disposiciones respecto a diversos temas como el trato nacional, las normas sanitarias y fitosanitarias, los impuestos y los subsidios a la exportación, al igual que consideraciones sobre el comercio electrónico, la transparencia y el medioambiente, entre otros.

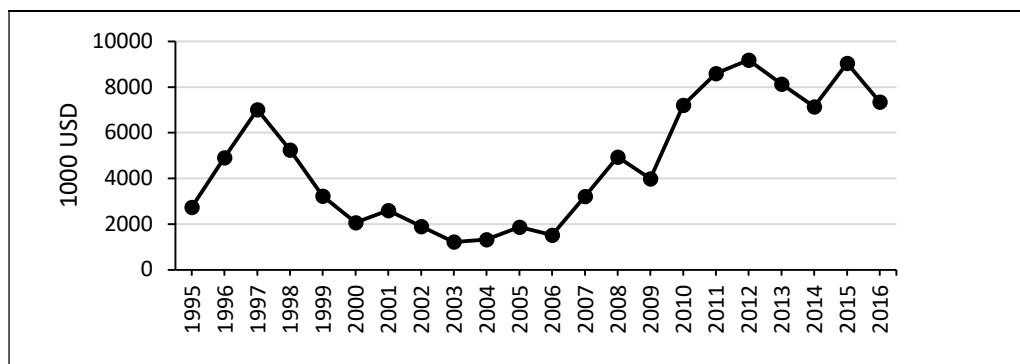
El ALC reitera que Chile y Colombia deben avanzar hacia la expansión y diversificación del comercio recíprocamente (art. 1.2 literal b), y hacia la eliminación de los obstáculos al comercio (literal c). Los productos CCM en particular ha resultado beneficiados de estas nuevas disposiciones.

COLOMBIA EN EL COMERCIO DE PRODUCTOS CCM CON CHILE

Tal como se muestra en la gráfica 1, el comercio de productos CCM, medido por la suma de las exportaciones y las importaciones de estos productos entre Colombia y Chile desde 1995, se ha caracterizado por tres fases:

1. **Fase inicial, 1995-1997:** El comercio de productos CCM crece, pasando de 2700 miles de dólares en 1995 a 7000 miles de dólares en 1997 después de que ambos países se convierten en miembros de OMC.
2. **Fase intermedia, 1997-2006:** Paradójicamente, la profundización de la liberalización mediante el acuerdo CE24 se acompaña de una contracción del comercio, pasando de 7000 miles de dólares a 1500 miles de dólares. Incluso, el comercio se ubicó por debajo de su nivel de 1995. En tal sentido, diversos factores exógenos al acuerdo CE24 impidieron la dinamización de los intercambios de productos CCM entre Colombia y Chile.
3. **Fase final, 2006-2016:** Con la entrada en vigor del ALC, el comercio logra una tendencia casi continua de crecimiento, pasando de 1500 miles de dólares en 2006, a más de 7000 miles de dólares en 2016, lo que equivale a un crecimiento del 17.06% anual, en un contexto en el que el comercio *total* ha crecido a un ritmo menor durante el mismo lapso de tiempo, a saber, el 6,59% anualmente (según los datos proporcionados por UNCTADstat). Este crecimiento resulta en un comercio de productos CCM entre Colombia y Chile que incluso desde 2010 supera su nivel de 1997, lo que sugiere que el ALC ha podido influir en la dinamización de dicho comercio.

Gráfica 1. El Cuero, Calzado y Marroquinería en el comercio bilateral Chile-Colombia (miles de dólares corrientes), 1995-2016.



Fuente: Elaboración propia con los datos proporcionados por UNCTADstat.

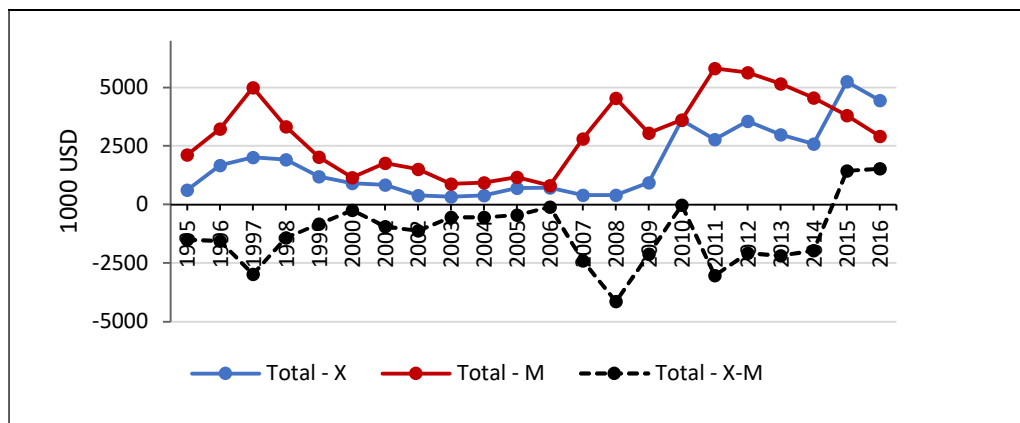
Dentro de esta evolución del comercio bilateral Colombia-Chile de productos CCM entre 1995 y 2016, ¿cómo se posiciona Colombia? Para dar respuesta este cuestionamiento y según la información registrada en la gráfica 2a y 2b, se pueden identificar dos fases:

¹⁵ Incluso, los protocolos adicionales introdujeron cantidades intercambiables sin arancel aduanero (“cuotas”) para los productos agropecuarios, reduciendo el impacto del SAFP.

1. **1995-2006, es decir, entre los primeros años de la OMC y la implementación del acuerdo CE24:** Las importaciones colombianas de productos CCM desde Chile siempre han superado las exportaciones. Esto significa que Colombia ganó menos oportunidades de venta en el mercado chileno (exportaciones) comparado con las oportunidades de venta perdidas en el mercado nacional (importaciones) para los productos CCM. Esto sucedió pese a que las importaciones tendieron a decrecer, después de un pico registrado en 1997.
2. **2006-2016, es decir, los años después de la implementación del ALC:** Las exportaciones tendieron a crecer, e incluso desde 2010 superan el máximo obtenido en 1997, mientras que las importaciones, después de haber crecido entre 2006 y 2011, tendieron a decrecer. Debido a este incremento de las exportaciones junto con la disminución de las importaciones, el saldo comercial se mejoró progresivamente, hasta ser positivo en 2015 y 2016. Esto significa que solo en estos dos años Colombia ganó más oportunidades de venta en el mercado chileno (exportaciones) en comparación con las oportunidades de venta perdidas en el mercado nacional (importaciones).

En consecuencia, se puede afirmar que los productos CCM colombianos lograron mejorar su posicionamiento frente a sus homólogos chilenos en la zona bilateral de comercio conformada entre Colombia y Chile en el contexto de liberalización.

Gráfica 2a. Exportaciones, importaciones y saldo comercial de Cuero, Calzado y Marroquinería, Colombia frente a Chile, 1995-2016 (miles de dólares corrientes)



Fuente: elaboración propia con los datos proporcionados por UNCTADstat.

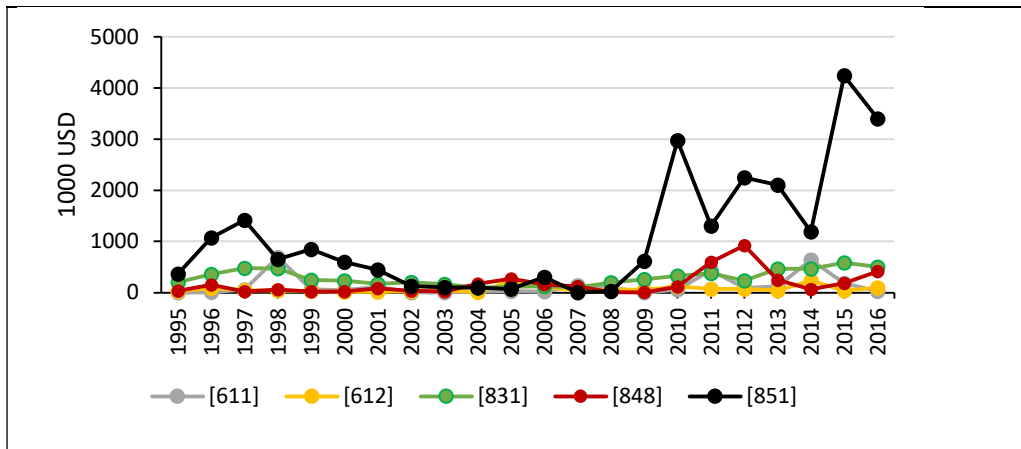
Más allá de esta visión general de los productos CCM en su conjunto, ¿cuáles serían los productos en particular que permitieron mejorar la tendencia del saldo comercial colombiano frente a Chile respecto a la rama en cuestión? Para dar respuesta a tal cuestionamiento, se estudian todas las categorías de productos que constituyen la rama CCM según la Clasificación Uniforme para el Comercio Internacional (CUCI) a tres dígitos en su tercera revisión. De esta manera, es posible identificar 8 categorías de productos:

- 2 categorías pertenecientes al capítulo de los materiales crudos no comestibles diferentes de los combustibles:
 - ✓ [211] “Cueros y pieles (excepto pieles finas), sin curtir”;
 - ✓ [212] “Pieles finas sin curtir, excepto las pieles y los cueros del grupo 211”.
- 3 categorías pertenecientes al capítulo de los artículos manufacturados clasificados principalmente según el material:
 - ✓ [611] “Cuero” (en planchas/hojas/tiras, depilado y curtido);
 - ✓ [612] “Manufacturas de cuero natural o sintético, n.e.p.; artículos de talabartería y guarnicionería”;
 - ✓ [613] “Pieles finas curtidas o adobadas, ensamblados o sin ensamblar (sin agregados de otros materiales)”.

- 3 categorías pertenecientes al capítulo de los artículos manufacturados designados como “diversos”:
 - ✓ [831] “Artículos de viajes, bolsos de mano y otros artículos análogos para contener objetos”
 - ✓ [848] “Prendas y accesorios de vestir que no sean de materias textiles”¹⁶
 - ✓ [851] “Calzado”

Entre estas 8 categorías, ¿existe una que impulse de manera decisiva el crecimiento en las exportaciones colombianas de productos CCM hacia Chile a partir de 2006? Como lo muestra la gráfica 3, la categoría [851], denominada calzado, sería la impulsora de tal tendencia. En efecto, se observa que las exportaciones totales tienden a crecer, precisamente cuando las exportaciones de calzado presentan la misma tendencia desde 2008, mientras que las exportaciones de las otras categorías tienden a mantenerse en el mismo nivel, sin tener la posibilidad de contribuir al aumento de las exportaciones totales de productos CCM. En este mismo contexto, cuando las exportaciones colombianas de productos CCM hacia Chile lograron un pico en 1997, las exportaciones de calzado también registraron un pico. Por consiguiente, es posible afirmar que la mejora tendencial en el saldo comercial de Colombia frente a Chile respecto a los productos CCM es impulsada por las exportaciones de calzado en particular. Esto se confirma por un coeficiente de correlación estadística igual a 0,973 entre las exportaciones de calzado y las exportaciones totales (cálculo de los autores).

Gráfica 2b. Exportaciones colombianas hacia Chile de la rama cuero, calzado y marroquinería, 1995-2016 (% de las exportaciones totales de la rama)



Fuente: elaboración propia con los datos proporcionados por UNCTADstat. Nota: las categorías [211], [212] y [613] no aparecen en la gráfica porque representan menos del 0,25% de las exportaciones totales, e incluso hasta el 0% dependiendo del año considerado.

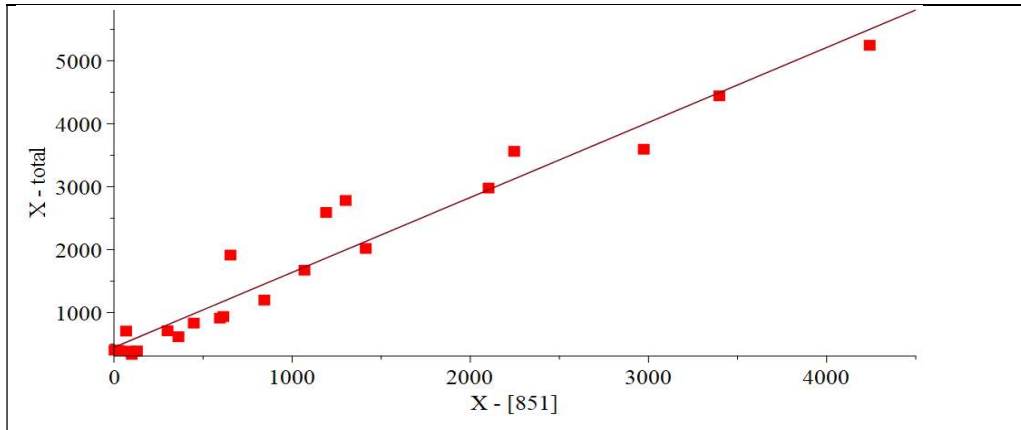
Igualmente, el rol del calzado como impulsor de las exportaciones de productos CCM puede ser constatado en la gráfica 3, en la que se presenta la dispersión de los puntos compuestos por el valor de las exportaciones de calzado (en abscisas) y el valor de las exportaciones totales (en ordenadas), para cada año desde 1995 hasta 2006.

Aplicando una regresión lineal $X_{CCM,t} = a + bX_{851,t} + \epsilon_t$, donde $X_{CCM,t}$ y $X_{851,t}$ representan respectivamente las exportaciones totales y las exportaciones de calzado en el año t (ϵ_t siendo el término de error en el año t), se obtienen las siguientes estimaciones de $a = 446,7404$ y $b = 1,1914$. Los dos coeficientes son significativos al 1%. En particular, se rechaza la hipótesis $b = 0$, lo que confirma que el valor tomado por $X_{851,t}$ constituye un factor explicativo del valor tomado por $X_{CCM,t}$. Ahora bien, el coeficiente R^2 es igual a 0,9444, lo que implica

¹⁶ Esta categoría no se conforma exclusivamente de productos CCM. Además de prendas y accesorios de vestir en cuero, incluso guantes, mitones, cinturones, correas, sombreros y otros artículos de peletería, la categoría [848] incorpora prendas y accesorios de vestir de plástico o caucho vulcanizado. Se decide considerar la categoría [848] para el desarrollo del presente trabajo, recordando al lector que alguna parte de los productos correspondientes no son exclusivamente de cuero.

que, por estar tan cerca de 1, las exportaciones de calzado constituyen por sí solas un factor explicativo de las exportaciones totales.

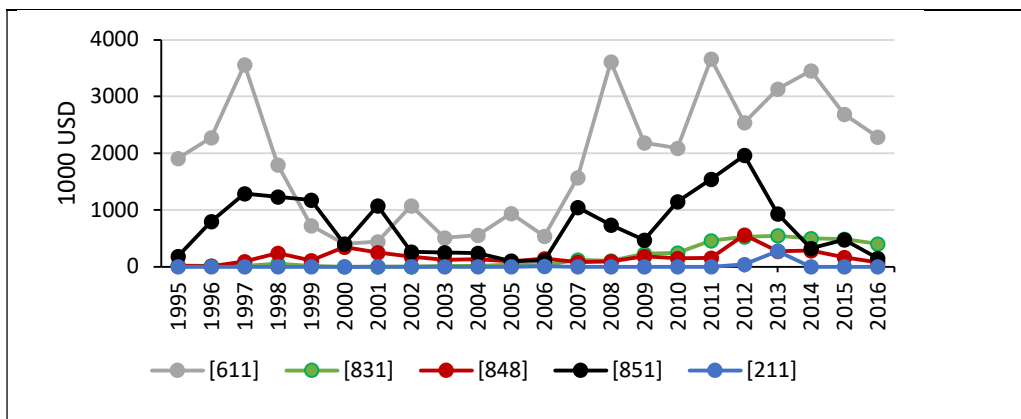
Gráfica 3. Correlación entre las exportaciones de calzado y las exportaciones totales de la rama CCM, Colombia frente a Chile en miles de dólares)



Fuente: elaboración propia con los datos proporcionados por UNCTADstat. Nota: cada punto representa, para un cierto año, el valor de las exportaciones de calzado (eje horizontal) y el valor de las exportaciones totales de la rama CCM (eje vertical), para Colombia hacia Chile. La línea recta, que da cuenta de la correlación entre ambas variables, es dada por $X_{CCM,t} = 446,7404 + 1,1914X_{851,t}$, donde $X_{CCM,t}$ representa las exportaciones totales en el año t y $X_{851,t}$ denota las exportaciones de calzado en el año t (coeficiente R^2 igual a 0,9444).

De la misma manera, ¿existiría una categoría de productos CCM que impulse de manera decisiva el decrecimiento en las importaciones colombianas de dichos productos desde Chile a partir de 2011? Como lo muestra la gráfica 4, tal categoría es nuevamente el calzado, con una reducción significativa en las importaciones a partir de 2012. Sin embargo, se puede constatar que simultáneamente las importaciones de la categoría [611] “cuero manufacturado” tendieron a comportarse de manera cíclica desde 2008, sin poder identificar una posible disminución. Además, el cuero manufacturado constituye el principal componente de las importaciones totales de productos CCM. Por ende, la reducción general en las importaciones esconde el alto nivel de importaciones de cuero respecto a los productos CCM, incluyendo el calzado.

Gráfica 4. Importaciones colombianas desde Chile de productos cuero, calzado y marroquinería, 1995-2016 (miles de dólares corrientes)



Fuente: elaboración propia con los datos proporcionados por UNCTADstat. Nota: las categorías no presentes en la gráfica representan menos del 0,25% de las importaciones totales, e incluso hasta el 0% dependiendo del año considerado.

2. VENTAJAS COMPARATIVAS: METODOLOGÍA DE MEDICIÓN Y ESTUDIO ECONOMÉTRICO DE SUS DETERMINANTES

En esta sección, se describen las metodologías aplicadas para cumplir con los objetivos planteados, a saber: Medir las ventajas comparativas y presentar un protocolo econométrico para justificar la elección del indicador (2.1) e identificar los determinantes de estas ventajas comparativas (2.2).

MEDICIÓN DE LAS VENTAJAS COMPARATIVAS Y ELECCIÓN DEL INDICADOR

Como se comentó en los apartados anteriores, la mejora tendencial en el saldo comercial de Colombia frente a Chile para los productos CCM fue impulsada principalmente por el calzado. Por esto, esta sección analiza la evolución en el tiempo de las *ventajas comparativas* de Colombia frente a Chile para los productos CCM, haciendo énfasis en el calzado. El objetivo es estudiar en qué medida dicha evolución contribuye a comprender la mejora arriba mencionada.

A través del concepto de ventajas comparativas, se hace referencia a la (in)capacidad para producir un bien o un servicio a menores costos relativos y con mayor diferenciación, lo que confiere competitividad y permite a los productores aumentar (o al menos mantener) sus ventas en el mercado nacional e internacional (Palley, 2008; Stellan & Danna-Buitrago, 2017a). Bajo esta óptica, la mejora tendencial en el saldo comercial de Colombia frente a Chile para los productos CCM debería asociarse a una mejora en las ventajas comparativas, especialmente en el calzado, lo que hubiese permitido estimular las exportaciones y reducir las importaciones totales de productos CCM.

Las ventajas comparativas se miden mediante la utilización de datos sobre los intercambios, bajo la hipótesis que estos últimos “revelan” el nivel de tales ventajas. En este trabajo, se utiliza un indicador de Ventajas Comparativas Reveladas (VCR) en términos de *Contribución al Saldo Comercial* (CSC), elaborado según los lineamientos de Lafay (1987, 1992) y De Saint-Vaulry (2008); véanse también Iapadre (2001), Marconi y Rolli (2008), Marconi (2012) e Ishchukova y Smutka (2013).

El indicador VCR-CSC se calcula de la siguiente manera:

$$CSC_{kit} = \frac{1}{GDP_{it}} \left[\frac{w_{rt}}{w_{kt}} (X_{kit} - M_{kit}) - w_{kt} \sum_{p \in K} \left(\frac{w_{rt}}{w_{pt}} (X_{pit} - M_{pit}) \right) \right] \quad (1)$$

$$\text{con } w_{ku} = \frac{\sum_{i \in J} (X_{kiu} + M_{kiu})}{\sum_{i \in J} \sum_{p \in K} (X_{piu} + M_{piu})}; u = t, r$$

El indicador VCR-CSC se calcula para un país i en una zona J de intercambios, un producto k en una nomenclatura K , y un periodo t en un conjunto T de varios periodos. El cálculo se hace a partir de dos tipos de variables:

- Las variables del tipo X_{kit} denotan las exportaciones del producto k realizadas por el país i en el periodo t ; y
- Las variables del tipo M_{kit} denotan las importaciones de k realizadas por i en t .

El indicador VCR-CSC consiste en calcular la diferencia entre:

1. $X_{kit} - M_{kit}$, a saber, el saldo comercial asociado a cualquier $\langle k, i, t \rangle$; y
2. Un saldo comercial teórico cuyo valor implica la ausencia de ventajas y de desventajas. Este saldo comercial teórico resulta de la ponderación de $\sum_{p \in K} (X_{pit} - M_{pit})$ por w_{kt} . Por una parte, $\sum_{p \in K} (X_{pit} - M_{pit})$ es el saldo comercial total asociado a $\langle i, t \rangle$. Por otra parte, w_{kt} es el cociente de $\sum_{i \in J} (X_{kit} + M_{kit})$ por $\sum_{i \in J} \sum_{p \in K} (X_{pit} + M_{pit})$, el cual corresponde al peso de k en el comercio total de la zona en t .

Una diferencia positiva significa que el saldo efectivo supera el saldo teórico, lo que revela la existencia de ventajas comparativas (indicador positivo); y viceversa. Al final, esta diferencia se divide por el Producto Interno Bruto (PIB) de i en t , denotado GDP_{it} .

De manera previa, X_{kit} y M_{kit} , al igual que cualquier otra variable X_{kip} y M_{kip} se ajusta siendo multiplicada por w_{kr}/w_{kt} para que k tenga el mismo peso en el comercio total de la zona independientemente de t . Este peso es el de un periodo de referencia denotado r . Más detalles sobre el marco teórico subyacente al indicador VCR-CSC, así como su mayor confiabilidad en comparación con los indicadores B y N , se encuentran en Danna-Buitrago (2017).

El indicador VCR-CSC no constituye la única manera de revelar las ventajas comparativas. Balassa (1965) propuso por primera vez un indicador VCR, el cual, incluso en su versión normalizada (Iapadre, 2001; Laursen, 2015), sigue siendo la referencia en la literatura. Véase, por ejemplo: Cristelli, Gabrielli, Tacchella, Caldarelli & Pietronero, (2013); Ofa & Karingi (2014); Yasmin & Altaf (2014); Moore, Pereira & Walrond (2015); Gibba (2017); Halilbašić & Brkić (2017); Shaul Hamid & Aslam (2017).

Balassa (1965) se basa en el peso de k en las exportaciones totales de i en t , es decir, el cociente de X_{kit} por $\sum_{p \in K} X_{pit}$, para revelar ventajas o desventajas. Este cociente no revela ni ventajas ni desventajas si es igual al mismo peso calculado para el conjunto de países pertenecientes a J , o sea, el cociente de $\sum_{i \in J} X_{kit}$ por $\sum_{i \in J} \sum_{p \in K} X_{pit}$. El indicador VCR de Balassa (1965), denotado por B , se calcula dividiendo el primer peso por el segundo:

$$B_{kit} = \frac{X_{kit}}{\sum_{p \in K} X_{pit}} / \frac{\sum_{i \in J} X_{kit}}{\sum_{i \in J} \sum_{p \in K} X_{pit}} \quad (2)$$

Por consiguiente, 1 es el valor neutro del indicador, lo que significa que no existen ventajas ni desventajas. Si $B_{kit} > 1$, el indicador revela ventajas comparativas. Por el contrario, si $B_{kit} \in [0; 1[$, el resultado señala desventajas comparativas. En cuanto a la versión normalizada del indicador, denotada N :

$$N_{kit} = \frac{B_{kit} - 1}{B_{kit} + 1} \quad (3)$$

Siendo cero el valor neutro del indicador, $N_{kit} \in]0; 1[$ revela ventajas comparativas y $N_{kit} \in [-1; 0[$ revela desventajas comparativas. Y N_{kit} corresponde a una aproximación de tipo logarítmico de B_{kit} .

Para justificar el uso del indicador VCR-CSC al medir las ventajas comparativas, específicamente en el caso de la zona bilateral de comercio Colombia-Chile, se implementa el siguiente protocolo:

1. Se calculan 6 veces una serie de indicadores VCR, a saber, con los indicadores B , N y CSC , tanto para Colombia como para Chile (tres tipos de indicadores VCR para 2 países diferentes). Para cada indicador y cada país, los cálculos se hacen en el marco de la zona $J = \{\text{CHL}; \text{COL}\}$ compuesta solo por los dos países en cuestión, para cada año desde 1995 hasta 2016 ($t \in \{1995; 1996; \dots; 2016\}$), y para las 255 categorías según la CUCI a tres cifras. En el caso del indicador CSC se toma como periodo de referencia $r = 1995$. Se toman los datos de las exportaciones y las importaciones de *UNCTADStat17* y los datos del PIB colombiano del Banco Mundial¹⁸.
2. Para cada indicador, se realiza la siguiente regresión:

$$VCR_{kit_1} = \alpha + \beta \cdot VCR_{kit_0} + \delta_i + \epsilon_{ik} \quad (4)$$

VCR_{kit_1} y VCR_{kit_0} representan los indicadores VCR calculados para $\langle k; i \rangle$ en dos periodos t_1 y t_0 respectivamente, con $t_1 > t_0$. Igualmente δ_i es una variable *dummy* que da cuenta de un efecto fijo por país y ϵ_{ik} es el término de error por país y por producto. La regresión (4) es una fórmula que resulta de una modificación de la regresión de Laursen (2015), la cual se realiza sin efecto fijo por país (δ_i) y se aplica para un solo país en lugar de varios países (en nuestro caso, Colombia y Chile).

¹⁷ unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx. Como se trata de una zona bilateral, las importaciones totales de Colombia dentro de la zona corresponden a las exportaciones totales de Chile dentro de la zona; y recíprocamente, las exportaciones totales de Colombia dentro de la zona corresponden a las importaciones totales de Chile dentro de la zona. En consecuencia, es suficiente tomar los datos de un solo país (entre los dos) respecto a las exportaciones e importaciones.

¹⁸ data.worldbank.org/indicator/NY.GDP.MKTP.CD.

3. Se deduce cuál indicador VCR es más consistente respecto a la *estacionariedad tendencial en el tiempo* de las ventajas/desventajas. Este concepto hace referencia al hecho de que las ventajas/desventajas comparativas tienden a cambiar solamente en horizontes de largo plazo (Lafay, 1987). Si un indicador VCR se caracteriza por alguna volatilidad en el tiempo, existe entonces un mayor riesgo de que tal volatilidad no traduzca exclusivamente un cambio en las ventajas/desventajas sino un cambio de tipo coyuntural. En el marco de la regresión (4), una *estacionariedad total* en el tiempo de las ventajas/desventajas entre t_0 y t_1 se traduce por las siguientes estimaciones:

- $\alpha = 0$;
- $\beta = 1$; y
- $\delta_i = 0$ para cualquier i fuera de la variable *dummy* para la cual se asigna por construcción el valor cero.

En efecto, en este caso un indicador VCR en t_1 es diferente del indicador en t_0 solo por el término de error. No obstante, la estacionariedad total solo constituye un referente teórico, pues las ventajas/desventajas pueden cambiar en el tiempo, bajo la condición de que cambien solo “un poco” para así ser consistente con la histéresis en los esquemas de especialización internacional. Por esta razón, nos focalizamos en la estacionariedad *tendencial* en el tiempo de las ventajas/desventajas entre t_0 y t_1 , lo que se traduce por las siguientes estimaciones:

- $|\alpha| \rightarrow 0$;
- $|\beta - 1| \rightarrow 0$; y
- $|\delta_i| \rightarrow 0$ para cualquier i fuera de la variable *dummy* para la cual se asigna por construcción el valor cero.

Se observa que las condiciones dadas por la estacionariedad tendencial son menos fuertes que las de la estacionariedad total. Todo es cuestión de ver cuál indicador se acerca más a las condiciones arriba mencionadas.

Se realiza la regresión con $t_0 = 1995$ y $t_1 = 2016$, en otras palabras, se comprueba la estacionariedad tendencial en el tiempo de las ventajas/desventajas sobre 22 años. Como lo muestra la tabla 1, los resultados de las regresiones indican que el indicador VCR-CSC es el más consistente en términos de la estacionariedad tendencial en el tiempo de las ventajas/desventajas en comparación con los dos otros indicadores (véanse también la gráfica 5). En efecto:

- Solo la regresión asociada al indicador CSC resulta en $|\alpha| \rightarrow 0$. De hecho, solo esta regresión conlleva a aceptar con certeza la hipótesis nula $\alpha = 0$ (con un p -valor igual a 1), lo que implica *a fortiori* $\alpha \rightarrow 0$, contrario a las dos otras regresiones.
- Además, la regresión asociada al indicador VCR-CSC resulta en el coeficiente β más cerca de 1: 1.125494 versus 0.3598445 (B) y 0.3492286 (N), es decir, una distancia de 0.125494 respecto a 1 versus 0.6401555 (B) y 0.6507714 (N).
- Respecto a los efectos fijos por país, se debe por definición atribuir cero a una de las variables *dummy*, tal como se mencionó anteriormente. Se selecciona δ_{CHL} como esta variable *dummy* “excluida”, la cual es por ende igual a cero (sin incidencia sobre los resultados). Ahora, la regresión realizada a partir del indicador VCR-CSC indica que δ_{COL} es la más cercana a cero. Además, solo con el indicador VCR-CSC se acepta con certeza la hipótesis nula $\delta_{COL} = 0$ (p -valor igual a 1), contrario a los dos otros indicadores (p -valores iguales a 0.762 y 0.699 respectivamente).

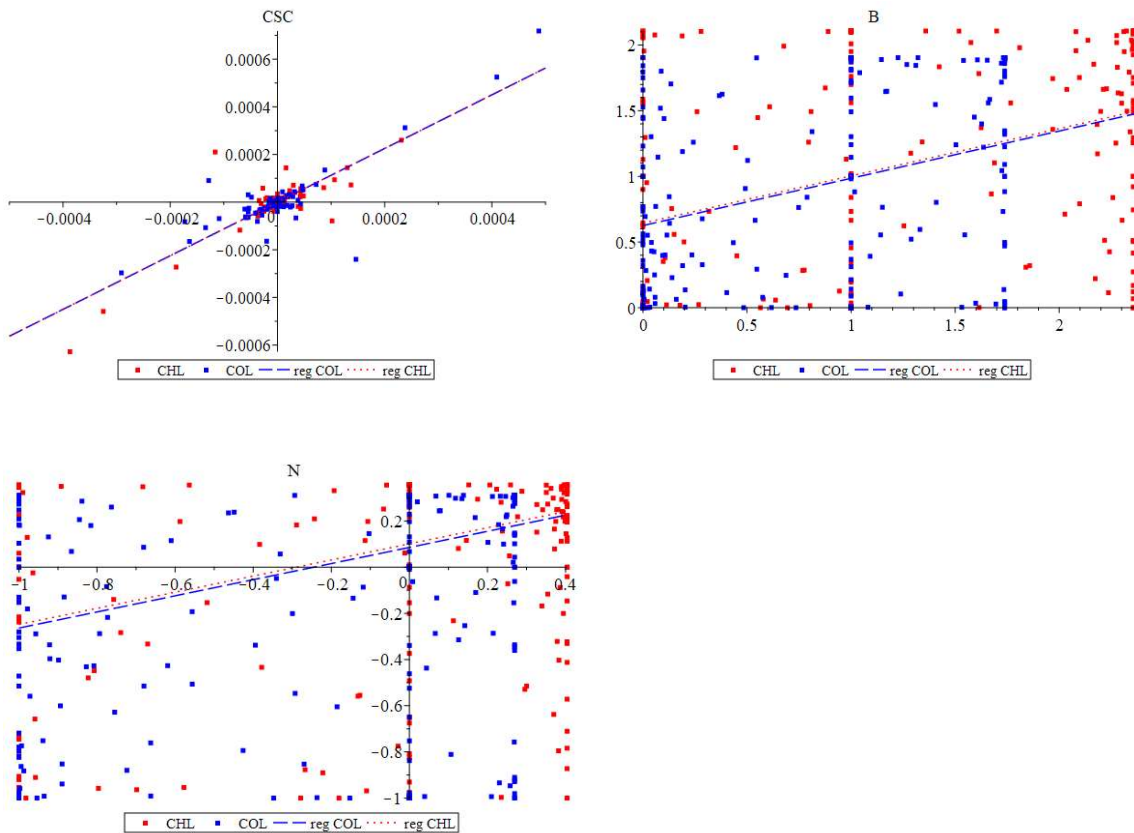
Tabla 1. Resultados de la regresión (4) para cada indicador VCR comprobado: Contribución al saldo comercial (CSC), indicador estándar (B) e indicador estándar normalizado (N)

	CSC	B	N
α	2.17e-21	0.6436871***	-0.1011033***
β	1.125494***	0.3598445***	0.3492286***
δ_{COL}	-5.27e-21	-0.183384	-0.155171

N	510	510	510
R^2	0.7141	0.1373	0.1239
F-statistic	633.22	40.34	35.84
Prob > F	0.0000	0.0000	0.0000
* significativo al 10%, ** significativo al 5%, *** significativo al 1%			
$ \alpha $	2.17e-21	0.6436871	-0.1011033
$ \beta - 1 $	0.125494	0.6401555	0.6507714
$ \delta_{COL} $	5.27e-21	.0183384	.0155171

Fuente: elaboración propia. Nota: por construcción, $\delta_{CHL} = 0$.

Gráfica 5. Representación de los resultados de la regresión (4) para cada indicador VCR comprobado: VCR-CSC, B y N



Fuente: elaboración propia con los datos proporcionados por UNCTADstat y World Bank Open Data. Nota: cada punto corresponde a $(VCR_{kit_0}; VCR_{kit_1})$ con $VCR = CSC, B$ o N ; $i = CHL, COL$; $t_0 = 1995$, $t_1 = 2016$; y k una de las 255 categorías de productos CUCI a tres dígitos.

Al final, se confirma que es posible utilizar el indicador VCR-CSC en el marco de este estudio para las 255 categorías de productos, los dos países y los 22 periodos. Luego y para el propósito de esta investigación, en la tercera sección se calculan y analizan los valores del indicador en el caso de las ocho categorías de productos que conforman la rama CCM desde el punto de vista de Colombia.

2.2 IDENTIFICACIÓN DE LOS DETERMINANTES DE LAS VENTAJAS COMPARATIVAS

Se plantea como hipótesis que existen ventajas comparativas de los productos colombianos CCM en el mercado chileno, y que estas son determinadas por el tamaño del mercado y las dotaciones factoriales, lo que representa una oportunidad para el crecimiento económico de Colombia. Para contrastar esta hipótesis, primero es necesario calcular las ventajas comparativas, mediante el método definido en el apartado antes presentada (2.1), y luego se identifican económicamente sus determinantes.

Así las cosas, para identificar los determinantes que influyen en la conformación de las ventajas comparativas de los productos colombianos CCM respecto a sus homólogos hechos en Chile, se estimó un modelo probit:

$$P_i = P(Y = 1 | X) = P(I_i^* \leq I_i) = P(Z_i \leq a + bX_i) = F(a + bX_i) \quad (5)$$

Donde $Z \sim \mathcal{N}(0, s^2)$ es una variable distribuida según una ley normal de media igual a cero, y F es la función de distribución normal acumulada, es (6)
 decir: $F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} \exp\left(-\frac{z^2}{2}\right) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a+bX_i} \exp\left(-\frac{z^2}{2}\right) dz =$

La variable dependiente, denotada Y , es un variable dicotómica que toma el valor de 1 si un indicador VCR-CSC calculado para los productos CCM colombianos en su conjunto, es positivo; y cero en otro caso (var Y). Las variables independientes son las siguientes:

- ITI: Índice de términos de intercambio de la mercancía para Colombia.
- RRNco: Total rentas de los recursos naturales como porcentaje del PIB para Colombia. Como proxy de la reprimarización de la economía colombiana.
- POP: La población de Chile como proxy del tamaño del mercado. El flujo de comercio bilateral entre dos países está positivamente relacionado con el tamaño de sus economías, puesto que se espera una mayor especialización de carácter intra-industrial, en la medida que el tamaño del mercado es mayor; puesto que el número de variedades de un grupo de bienes producidos en una economía en equilibrio es más numeroso, cuanto el tamaño del mercado interno sea más grande, si existen economías de escala y una distribución homogénea de las preferencias (Lancaster, 1980).
- EC: La diferencia del PIB per cápita de los dos países. Se toma esta variable como proxy de la diferencia de las dotaciones factoriales. El signo esperado difiere según el tipo de IIT.
- RRNcl: Total rentas de los recursos naturales como porcentaje del PIB para Chile. Como proxy de la reprimarización de la economía chilena.

3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS SOBRE LAS VENTAJAS COMPARATIVAS Y SUS DETERMINANTES

Después de confirmar en la sección 2, apartado 2.1, que es posible utilizar el indicador VCR-CSC en el marco de este estudio para las 255 categorías de productos, los dos países y los 22 periodos, se procede a calcular las ventajas comparativas y a analizar los valores del indicador en el caso de las ocho categorías de productos que conforman la rama CCM desde el punto de vista de Colombia (3.1). Luego, se presentan y discuten los resultados obtenidos de la estimación del modelo probit para determinar los factores que influyen en la conformación de tales ventajas (3.2).

VENTAJAS COMPARATIVAS DE LOS PRODUCTOS COLOMBIANOS CCM FRENTE A CHILE

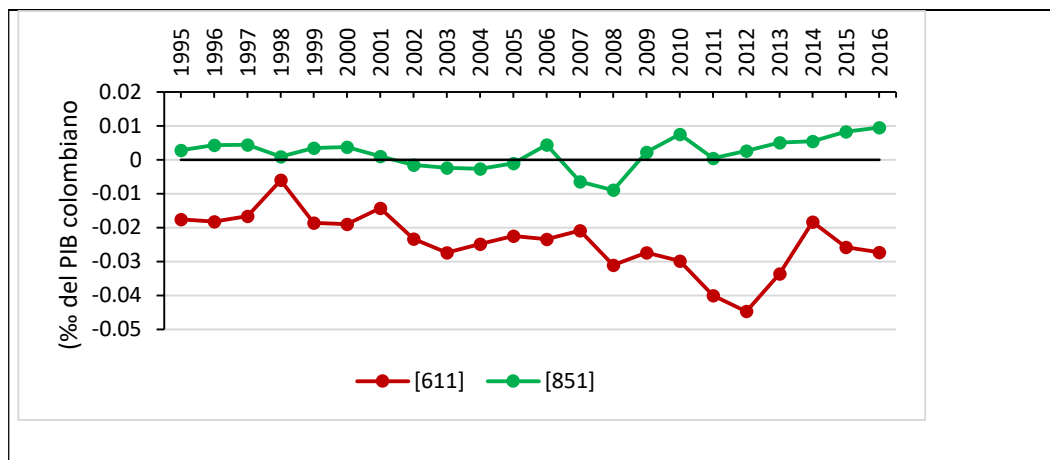
Dos resultados pueden observarse a partir de la gráfica 6, en relación con los dos tipos de productos que tuvieron un papel mayor en el posicionamiento de Colombia frente a Chile en el comercio de productos CCM, a saber, el calzado y el cuero manufacturado:

- El calzado colombiano se caracteriza por una mejora casi continua de su indicador VCR-CSC a partir de 2009, además de ser positivo, lo que se traduce en una consolidación progresiva de sus ventajas comparativas. Lo anterior, sucedió precisamente cuando el calzado empezó a impulsar de manera decisiva las exportaciones colombianas de productos CCM en su conjunto hacia Chile, permitiendo mejorar el saldo comercial de Colombia frente a tal país respecto a dichos productos. Del mismo modo, se constata que esta

mejora en el indicador VCR-CSC es consistente con el hecho de que las importaciones colombianas de calzado desde Chile disminuyeron a partir de 2012. Por ende, es posible indicar que el mejor posicionamiento de Colombia en el comercio de los productos CCM con Chile a través del calzado, resulta de una consolidación de las ventajas comparativas para este tipo de producto. Los productores de calzado lograron disminuir sus costos relativos y/o mejorar los factores de competitividad no vinculados a los precios, tales como la marca, la calidad o la innovación, lo que hizo posible estimular las exportaciones y simultáneamente limitar las importaciones¹⁹.

- Igualmente, se había mencionado que las importaciones colombianas de cuero manufacturado procedentes de Chile siguen siendo importantes e incluso impulsan las importaciones de productos CCM en su conjunto. Esto es consistente con el hecho que el indicador VCR-CSC del cuero colombiano, además de haber sido siempre negativo desde 1995, se caracteriza por una degradación casi continua, pese a la mejora registrada entre 2013 y 2015. Lo anterior, permite señalar que el cuero manufacturado colombiano no posee ventajas comparativas sino desventajas frente al cuero chileno, lo que perjudica el posicionamiento de Colombia en el comercio de productos CCM con Chile. De igual forma, estas desventajas se están consolidando, tal como lo indica la tendencia del indicador VCR-CSC a decrecer y mantenerse simultáneamente negativo. Por este motivo, Colombia se aleja cada vez más de especializarse en la producción de cuero manufacturado frente a Chile.

Gráfica 6. Indicador de Ventajas Comparativas Reveladas de Colombia frente a Chile, categorías CUCI [611] y [851], 1995-2016 (% del PIB colombiano).



Fuente: elaboración propia con los datos proporcionados por UNCTADstat y World Bank Open Data.

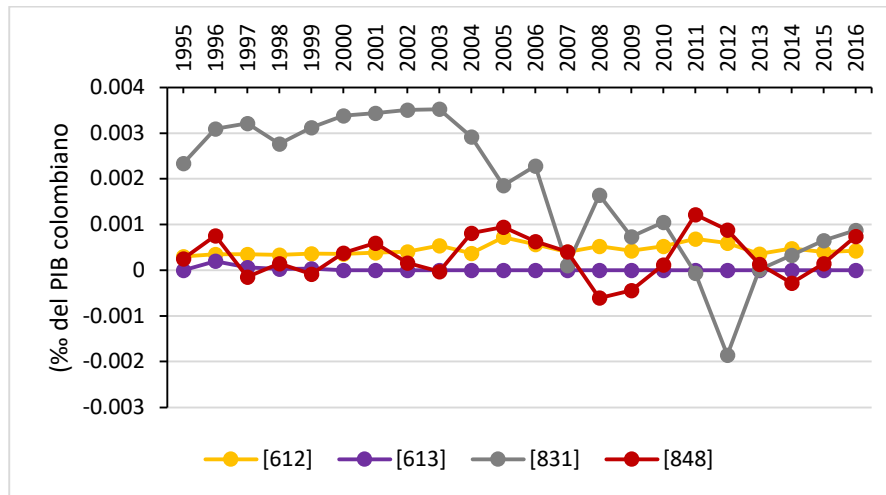
Respecto a los demás productos CCM, la gráfica 7 muestra que estos se caracterizan por disponer de ventajas/desventajas débiles e incluso iguales a cero. Esto es consistente con el hecho de que tales productos no han sido impulsores del comercio de productos CCM entre Colombia y Chile, desde el punto de vista de las exportaciones (calzado) ni de las importaciones (cuero manufacturado y calzado).

Sin embargo, se pueden mencionar dos fenómenos:

1. Los indicadores VCR-CSC en la gráfica 7 convergen en el tiempo hacia el intervalo $[0; 0,001]$, Teniendo en cuenta lo antes señalado, los productos correspondientes tienden a poseer ventajas comparativas, así sean débiles, o en el peor de casos no tienden a poseer desventajas comparativas.
2. La categoría [831], es decir, los artículos de viajes, los bolsos de mano y los otros artículos análogos para contener objetos, se caracteriza por la convergencia en cuestión. No obstante, tal convergencia implica que el indicador decrece, indicando que la categoría [831] ha perdido sus ventajas comparativas, así fueran débiles.

¹⁹ Para una exposición detallada de otros factores no vinculados a los precios, véase: Danna-Buitrago, Stellan, Mercado & Páez (2017) y Stellan & Danna-Buitrago (2017b).

Gráfica 7. Indicador de Ventajas Comparativas Reveladas de Colombia frente a Chile, categorías CUCI [612], [613], [831] y [848], 1995-2016 (% del PIB colombiano).



Fuente: elaboración propia con los datos proporcionados por UNCTADstat y World Bank Open Data. No se han reportado las categorías [211] y [212] debido a que estas se caracterizan por un indicador igual a cero en cada año, salvo algunas excepciones en donde el resultado del indicador VCR-CSC es casi igual a cero.

DETERMINANTES DE LAS VENTAJAS COMPARATIVAS DE LOS PRODUCTOS COLOMBIANOS CCM FRENTE A CHILE

A continuación, se presentan los resultados obtenidos de la estimación del modelo probit para determinar los factores que determinan la conformación de las ventajas comparativas de los productos CCM fabricados en Colombia frente a Chile. Para tales efectos, la tabla 2 reporta la estimación del modelo.

Tabla 2. Estimación del modelo probit

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ITI	0.1156119	0.0897569	1.29	0.198	-0.0603084	0.2915321
RRNco	-0.802506	0.6565436	-1.22	0.222	-2.089.308	0.4842958
POP	0.0041815	0.0021129	1.98	0.048	0.0000403	0.0083226
EC	1.54E-10	7.48E-11	2.06	0.039	7.66E-12	3.01E-10
RRNcl	-5.572.627	4.281.224	-1.3	0.193	-1.396.367	2.818.418
_cons	-1.253.403	5.783.553	-2.17	0.03	-2.386.959	-1.198.475
				Number of obs = 22		
Log likelihood = -7.341956				LR chi2(4) = 11.10		
				Prob > chi2 = 0.0255		
				Pseudo R2 = 0.4305		

Fuente: Elaboración propia.

Se eligió el modelo que minimiza la pérdida de información con el criterio de Akaike, tal como lo indica la tabla 3 que se presenta a continuación:

Tabla 3. Criterio de información de Akaike (AIC) y criterio bayesiano de información (BIC)

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	22	-1.289.096	-7.341.956	5	2.468.391	3.013.912

Fuente: Elaboración propia.

Luego, se realizó la prueba de bondad y ajuste de Pearson y se concluyó que el modelo presentaba un buen ajuste, tal como puede observarse en la tabla 4:

Tabla 4. Prueba “Goodness-of-fit” del modelo

Número de observaciones	22	Pearson chi2(16)	15.39
Número de <i>covariate patterns</i>	22	Prob > chi2	0.4961

Fuente: Elaboración propia.

Para detectar si el modelo está correctamente especificado, no se debería encontrar ningún predictor adicional que sea estadísticamente significativo excepto por casualidad. Se utilizó el valor predicho lineal (hat) y el valor predicho lineal cuadrado (hatsq) como los predictores para reconstruir el modelo. Por tanto, la variable hat debe ser un predictor estadísticamente significativo, pues es el valor predicho del modelo. Además, si el modelo está correctamente especificado, la variable hatsq no debería tener mucho poder predictivo, excepto por casualidad. Como hatsq no resulta significativo, entonces no existe evidencia de que se omitió la(s) variable(s) relevante(s) o la función de enlace no se ha especificado correctamente. Es lo que muestra la tabla 5.

Tabla 5. Modelo probit y valor predicho

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_hat	1.041.107	0.4921583	2.12	0.034	0.0764946 200.572
_hatsq	0.0672425	0.2360206	0.28	0.776	-0.3953493 0.5298344
cons	-0.0414999	0.4187274	-0.1	0.921	-0.8621904 0.7791907

Log likelihood = -7.3277753

Number of obs = 22
LR chi2(4) = 11.13
Prob > chi2 = 0.0038
Pseudo R2 = 0.4316

Fuente: Elaboración propia

La evidencia empírica arrojó que un incremento en la población chilena aumenta la probabilidad de que Colombia posea ventajas comparativas para los productos CCM en 0.0041815 como proxy del tamaño del mercado. En este sentido, se puede concluir que efectivamente el flujo de comercio bilateral de productos CCM entre los dos países está positivamente relacionado con el tamaño de sus economías.

La diferencia del PIB per cápita de los dos países resultó significativa a un nivel del 5% de significancia. El modelo estimado arroja que un aumento en la diferencia del PIB per cápita incrementa la probabilidad de que Colombia posea ventaja comparativa para los productos CCM en 1.54e-10. Por lo tanto, en cuanto la diferencia de las dotaciones factoriales, se corrobora lo que plantea la teoría económica.

El total de rentas naturales como porcentaje del PIB para los dos países y el índice de términos de intercambio de la mercancía para Colombia no resultaron significativos. Esto muestra que la conformación de ventajas comparativas en los productos CCM no depende de la reprimarización de las economías.

Gracias a estos resultados se confirma la hipótesis inicial, presentada en el apartado 2.2, en la que sea afirma que existen ventajas comparativas de los productos de cuero, calzado y marroquinería de Colombia en el mercado chileno, determinadas por el tamaño del mercado y las dotaciones factoriales, lo que representa una oportunidad para el crecimiento económico del país.

4. INFLUENCIA DEL PTP EN LAS VENTAJAS COMPARATIVAS DE PRODUCTOS COLOMBIANOS CCM FRENTE A CHILE

¿Qué otro(s) factor(es) podría(n) explicar los cambios en las ventajas comparativas resaltados en la sección 3, y por consiguiente la evolución en el posicionamiento de Colombia frente a Chile en el comercio de productos CCM (sección 2)? Es posible que el PTP, implementado en 2013, represente un factor determinante.

Para dar respuesta a estos cuestionamientos, esta sección inicia señalando los objetivos del PPT para el caso de los productos CCM:

- Para 2018, “desarrollar y consolidar un modelo de gestión integral que mejore la productividad y competitividad de las empresas del sector, transforme la formación, articule con la oferta institucional del Estado y genere círculos virtuosos entre productores y comercializadores (...)”.
- Para 2023, “fortalecer el desarrollo de productos de calzado y marroquinería de gama media y alta en cuero de calidad, así como en productos de materiales sintéticos, textiles y ecológicos, a partir de la incorporación del diseño, el trabajo conjunto entre eslabones de la cadena y la difusión de información y conocimiento asociado a las dinámicas de la moda”.
- Para 2028, “consolidar una cadena productiva que cuente con empresas “capaces”, y con marcas posicionadas en el mercado interno y en el mercado mundial, en un marco de alta apropiación tecnológica y desarrollo sostenible”²⁰.

Para alcanzar estos objetivos, se encuentran varias “líneas estratégicas” relacionadas a la gestión empresarial, el marco normativo, el capital humano y la infraestructura. Estas líneas apuntan al reconocimiento internacional de los productos CCM “hechos en Colombia”, es decir, consolidar una “marca país” que apoye el desempeño de los productos CCM, con énfasis en los productos CCM de gamas media y alta.

La marca país hace parte de los factores de competitividad no vinculados a los precios²¹. Los esfuerzos realizados desde el 2013 para desarrollar la marca país coinciden con la consolidación de las ventajas comparativas del calzado colombiano frente a Chile. Por tanto, es posible afirmar que el calzado colombiano pudo mejorar su competitividad frente a su homólogo chileno mediante la marca país, lo que precisamente se tradujo en la consolidación de sus ventajas comparativas y el estímulo de las exportaciones colombianas de calzado hacia Chile. En efecto, el PTP identificó el calzado de cuero para dama como una oportunidad de venta en Chile, frente a la competencia ejercida por productores de China, Vietnam y Brasil. Teniendo en cuenta lo anterior, el PTP apoya las iniciativas para que los productores colombianos tengan más en cuenta el diseño, el cual según los mismos términos del PTP, ha sido subestimado como factor de generación de valor agregado y posicionamiento de marcas y productos. Por esta razón, el PTP toma acciones específicas en la forma de:

- Apoyo mediante concursos nacionales y regionales de diseñadores.
- Apoyo a las iniciativas orientadas a la construcción de espacios que fortalezcan la actividad de diseño en el sector CCM.
- Realización de un análisis de la anatomía y la biomecánica del pie en Colombia, con la finalidad de obtener información que sirva para incrementar la calidad del calzado colombiano, en términos de confort y salud.

Igualmente, el PTP identificó una falta de estrategias de comercialización y mercadeo de productos CCM en general y de calzado en particular²². Por ello, toma en la actualidad acciones para que los productores colombianos manejen herramientas físicas y tecnológicas que fortalezcan la gestión comercial. Esto debería permitirles desarrollar competencias en *e-commerce*, sin dejar de lado la promoción de alianzas entre empresas internacionales y empresas colombianas, a fin de favorecer la implementación de los productos colombianos CCM en los mercados extranjeros, incluyendo Chile. Bajo esta misma óptica, es también posible mencionar las

²⁰ Documento titulado “Plan de negocios del Sector de Cuero, Calzado y Marroquinería”, disponible en www.ptp.com.co/contenido/categoria.aspx?catID=782.

²¹ Para más información sobre la marca país, véase: <https://www.colombia.co/>

²² Véase plan de negocios del PTP para los productos CCM, en www.ptp.com.co/ptp-sectores/historico/cuero-calzado-marroquineria.

diversas ayudas financieras que el PTP brinda a los productores colombianos de calzado (y otros productos CCM):

- Apoyo financiero mediante el “Capital semilla” para estimular emprendimientos dinámicos e innovadores²³.
- Líneas de créditos especiales otorgados por el Banco de Comercio Exterior de Colombia (Bancoldex) y diversas entidades públicas locales para la reconversión tecnológica y la construcción de nuevas fábricas²⁴.
- Cupo rotatorio de crédito para microempresas.

Finalmente, la multiplicación de medidas para desarrollar la marca país e impulsar las ventas en el extranjero, pudieron también haber tenido efectos positivos sobre las exportaciones colombianas de calzado hacia Chile.

Respecto al cuero, el PTP lo integra en sus objetivos, e incluso apunta a luchar contra una oferta limitada de insumos, tanto en cantidad como en calidad y variedad. En efecto, la baja disponibilidad de insumos, junto con las deficiencias en términos de calidad, afecta lógicamente la fabricación de productos con mayor valor agregado, sobre todo si se tratan de productos que pueden ser de media o alta gama, tales como los bolsos de mano, los artículos de viajes, las prendas y accesorios de vestir, y el mismo calzado²⁵. Por esta razón, la oferta colombiana de cuero no es suficiente y debe importarse, incluso desde Chile, para cubrir las necesidades en inputs de los productores colombianos. Al respecto, el PTP intenta superar esta situación, mediante las siguientes acciones:

- Fomento a la adopción de buenas prácticas de manejo, transporte y sacrificio de ganado con el objetivo de mejorar la calidad y el aprovechamiento de la materia prima
- Promoción de espacios para la comercialización de insumos en distintas regiones del país, con participación de productores nacionales.
- Programa de certificación de proveedores y desarrollo de un sistema de información que califique los proveedores en términos de calidad, oportunidad, disponibilidad y capacidad de reposición.
- Revisión periódica de la oferta de insumos, clave para el sector, con el ánimo de plantear acciones dentro del marco jurídico que permitan garantizar el abastecimiento de la cadena.
- Desarrollo de un esquema de compras de insumos sostenibles al por mayor con el propósito de bajar sus costos mediante una negociación de mayor escala con los proveedores.

Sin embargo y pese a las acciones antes señaladas, es posible indicar que la evolución de las importaciones colombianas de cuero (bruto, semi-terminado o terminado) y la consolidación correspondiente de las desventajas comparativas de este insumo hasta el momento, contradicen la eficiencia de este tipo de medidas. Quizás se deba esperar más tiempo para que se empiecen a ver los frutos de tales medidas en términos de consolidación de las ventajas comparativas del insumo en mención. Mientras tanto, el cuero colombiano seguirá perjudicando el posicionamiento del país frente a Chile en el comercio de productos CCM.

CONCLUSIONES

el mercado chileno representa cada vez más oportunidades para los productos colombianos de Cuero, Calzado y Marroquinería (CCM). Sin embargo, estas oportunidades conciernen principalmente el calzado, y no los productos CCM en general. En este sentido, el mercado chileno no genera oportunidades significativas de venta para las pieles finas, los artículos de viaje, los bolsos de mano ni las prendas y accesorios de vestir. Y el cuero manufacturado es el producto que impulsa de manera decisiva las importaciones colombianas desde Chile. Si

²³ Para conocer detalles sobre el capital semilla y cómo identificar qué fondo se ajusta a las necesidades de cada empresa, véase: www.bancoldex.com/como-identificar-que-fondo-se-ajusta-mi-empresa-302

²⁴ Sobre Bancóldex y sus líneas de crédito, véase www.bancoldex.com/soluciones-financieras/lineas-de-credito

²⁵ Para más información sobre el retraso en la adopción de nuevas tecnologías, los costos ambientales generados por las curtiembres, al igual que las deficiencias en el tratamiento del cuero, su transporte y almacenamiento, véase García y Maldonado (2013). Sobre la creciente exportación de cueros crudos y curtido con sales de cromo que ocasiona una baja oferta en el mercado colombiano, y por ende incita las importaciones de tal materia prima, ver SDDE (2014).

bien en 2015 y 2016 Colombia registró un saldo comercial excedentario frente a Chile para la rama CCM, dicho saldo habría sido más alto si el cuero manufacturado colombiano hubiera tenido una mejor posición frente a su homólogo chileno.

Lo anterior se confirma por un análisis de las ventajas comparativas de los productos CCM de Colombia frente a Chile. En efecto, se calcula un indicador de Ventajas Comparativas Reveladas (VCR) por los intercambios que toma la forma de la Contribución al Saldo Comercial (CSC). Se trata de una versión alternativa al indicador estándar *à la* Balassa (1965) y a la versión normalizada del indicador estándar (Laursen, 2015), debido a una mejor consistencia con la estacionaridad tendencial en el tiempo de las ventajas/desventajas. El indicador VCR-CSC revela que el calzado ha consolidado sus ventajas comparativas, sobre todo en los últimos años, mientras que el cuero manufacturado ha consolidado sus desventajas comparativas desde el primer año de cálculo del indicador (1995). Los demás productos analizados en el marco de este capítulo muestran generalmente ventajas comparativas, pero estas no son significativas.

Al estimar un modelo probit se pudo comprobar que un incremento en la población de chilena aumenta la probabilidad de que Colombia posea ventajas comparativas en los productos CCM en 0.0041815 como proxy del tamaño del mercado. Con base en esto, se puede concluir que el flujo de comercio bilateral entre los dos países está positivamente relacionado con el tamaño de sus economías. Además, se pudo evidenciar que un aumento en la diferencia del PIB per cápita incrementa igualmente la probabilidad de que Colombia posea ventajas comparativas en los productos CCM en 1.54×10^{-10} .

Además de medir y analizar las ventajas comparativas de los productos CCM de Colombia frente a Chile, así como algunos de sus determinantes, el presente trabajo discute el papel del apoyo proporcionado a los productores de la rama CCM por el Programa de Transformación Productiva (PTP). Ciertamente, este último pudo haber estimulado la expansión del calzado colombiano en Chile, a través de diversas medidas, pero no ha hecho posible que el cuero manufacturado deje de ser el principal producto importado por Colombia desde Chile, al igual que el principal producto con desventajas comparativas que se han consolidado en 23 años de relaciones comerciales bilaterales entre los dos países en mención.

Del presente análisis es posible hallar algunas recomendaciones para que los productos CCM colombianos (y no únicamente el calzado) puedan lograr un mejor posicionamiento en el mercado chileno. Primero, el PTP podría establecer prioridad a los productos de categoría [831] “Artículos de viajes, bolsos de mano y otros artículos análogos para contener objetos” y [848] “Prendas y accesorios de vestir que no sean de materias textiles”, debido a que son categorías con mayor valor agregado que podrían contribuir más a la conformación de un saldo comercial excedentario de Colombia frente a Chile. Esto, si se les compara con otras categorías de productos CCM que tienen también un débil desempeño en Chile, a saber: [211] “Cueros y pieles (excepto pieles finas), sin curtir”; [212] “Pieles finas sin curtir, excepto las pieles y los cueros del grupo 211”; [612] “Manufacturas de cuero natural o sintético, n.e.p.; artículos de talabartería y guarnicionería”; y [613] “Pieles finas curtidas o adobadas, ensamblados o sin ensamblar (sin agregados de otros materiales)”. Una futura línea de investigación sería analizar en qué medida la “marca país” tiene la capacidad de reforzar las ventajas comparativas y así estimular las exportaciones hacia Chile de los artículos de viaje, los bolsos de manos y las prendas y accesorios de vestir en cuero.

Asimismo, el débil desempeño de las categorías 211, 212, 612 y 613, en relación a sus ventajas comparativas poco significativas, no debe considerarse un hecho alarmante. En efecto, estas categorías constituyen las *materias primas* de las categorías 831, 848 y 851, por lo que es mejor que Colombia conserve estas materias para producir bienes de mayor valor agregado, en lugar de venderlas en el exterior y dejar a otros países la oportunidad de producir y vender dichos bienes. Ahora bien, respecto a la categoría 611 “cuero manufacturado”, que constituye igualmente una materia prima, pero que muestra desventajas comparativas, sería interesante estudiar la capacidad del PTP para reforzar el potencial productivo de tal categoría. Esto, con el fin de disminuir la dependencia de las importaciones de cuero manufacturado provenientes de Chile y así reducir sus desventajas comparativas.

Con todo lo anterior, se pueden evidenciar las ventajas comparativas del calzado (categoría 851) colombiano con respecto al mercado chileno. Aunque actualmente, es la categoría más representativa no se debe perder de vista la penetración en el mercado chileno con las otras categorías del CCM, teniendo en cuenta la calidad de los productos con la que se cuenta. El valor agregado de estos productos se transforma en una oportunidad para el crecimiento económico del país, ya que al aumentar las exportaciones aumentan los ingresos, lo a su vez se debe traducir en la mejora en la calidad de vida de los colombianos.

Bibliografía

- [1] Arango, B., Betancourt, J., & Martínez, L.F. (2015). implementación de herramientas para el diagnóstico de innovación en una empresa del sector calzado en Colombia. *Revista De Administração E Inovação*, 12(3), 310-329.
- [2] Balassa, B. (1965). Trade liberalization and 'revealed' comparative advantage. *Manchester School*, 33(2), 99-123.
- [3] Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., & Pietronero, L. (2013). Measuring the intangibles: a metrics for the economic complexity of countries and products. *Plos ONE*, 8(8), 1-20.
- [4] Danna-Buitrago, J. P. (2017). La Alianza del Pacífico+4 y la especialización regional de Colombia: una aproximación desde las ventajas comparativas. *Cuadernos de Administración*, 30(55), 163-192.
- [5] Danna-Buitrago, J.P., Stellan, R., Mercado, Á, & Páez, P. (2017). *Análisis retrospectivo de la integración internacional del sector manufacturero colombiano en el marco del Tratado de Libre Comercio con Estados Unidos*. Bogotá: Fundación Universitaria Los Libertadores.
- [6] Duarte, L.K., González, C.H., & Montoya, D.A. (2015). Colombia de cara al nuevo regionalismo renovado: la Alianza del Pacífico. *Punto de Vista*, 5(9), 137-162,
- [7] García, R.G. & Maldonado, A. (2013). Competitividad del calzado de cuero colombiano: perspectiva de la ventaja comparativa revelada (1980-2008). *Dimensión empresarial*, 11(1), 77-91.
- [8] Halilbašić, M., & Brkić, S. (2017). Export specialization of South East European countries in their trade with the European Union. *Economic Review: Journal Of Economics & Business*, 15(1), 75-87.
- [9] Forero, D. (2014). El sector del calzado en el barrio El Restrepo, Bogotá. Un análisis de caso a la luz de los sistemas productivos locales. *Equidad y Desarrollo*, 21, 97-123.
- [10] Gibba, A. (2017). Analysis of export competitiveness: the case of global nuts. *Albanian Journal Of Agricultural Sciences*, 16(2), 39-47.
- [11] Ishchukova, N., & Smutka, L. (2013). Comparative advantage: products mapping of the Russian agricultural exports. *Agris On-Line Papers In Economics & Informatics*, 5(3), 13-24.
- [12] Lafay, G. (1987). Avantage comparatif et compétitivité. *Économie Prospective Internationale* (29), 39-52.
- [13] Lafay, G. (1992). The measurement of revealed comparative advantages. En M. G. Dagenais & P.-A. Muet (Eds), *International Trade Modelling* (pp. 209-234). London: Chapman & Hall.
- [14] Iapadre, P.L. (2001). Measuring international specialization. *International Advances In Economic Research*, 7(2), 173-183.
- [15] De Saint-Vaulry, A. (2008). Base de données CHELEM-commerce international du CEPII. *Document de travail CEPII*, 9.
- [16] Lassudrie-Duchêne, B. (1971). La demande de différence et l'échange international. *Économies et Sociétés*, 5(6), 15-24.
- [17] Laursen, K. (2015). Revealed comparative advantage and the alternatives as measures of international specialization. *Eurasian Business Review*, 5(1), 99-115.
- [18] Lancaster, K. (1980). Intra-industry trade under perfect monopolistic competition. *Journal of international Economics*, 10(2), 151-175.
- [19] Marconi, D., & Rolli, V. (2008). Comparative advantage patterns and domestic determinants in emerging countries. *UNU-WIDER Research Paper*, (81), 1-37.
- [20] Marconi, D. (2012). Environmental regulation and revealed comparative advantages in Europe: is China a pollution haven? *Review of International Economics*, 20(3), 616-635.
- [21] Moore, W., Pereira, S., & Walrond, S. (2015). Comparative advantage and green goods in the Caribbean. *Social & Economic Studies*, 64(2), 59-90.

- [22] Ofa, S.V., & Karingi, S. (2014). Trade in intermediate inputs and trade facilitation in Africa's regional integration. *African Development Review*, 26(2), 96-110.
- [23] Páez, P., Jiménez, W., & Danna-Buitrago, J.D. (2018). La Competitividad de los Artículos de Calzado, Cuero y Marroquinería en Colombia: Revisión de la Literatura. *Diálogos de saberes*, (48), 1-36
- [24] Palley, T.I. (2008). Institutionalism and new trade theory: rethinking comparative advantage and trade policy. *Journal of Economic Issues*, 42(1), 195-208.
- [25] Shaul Hamid, M. F., & Aslam, M. (2017). Intra-regional trade effects of ASEAN free trade area in the textile and clothing industry. *Journal of Economic Integration*, 32(3), 660-688.
- [26] Secretaría Distrital de Desarrollo Económico de Bogotá (SDDE). (2014). *El Restrepo ampliado: la cadena productiva del cuero, calzado y marroquinería*. Bogotá: Dirección de Estudios Socio Económicos y Regulatorios, Secretaría Distrital de Desarrollo Económico, Alcaldía Mayor de Bogotá D.C.
- [27] Stellian, R., & Danna-Buitrago, J.P. (2017a). Competitividad de los productos agropecuarios colombianos en el marco del tratado de libre comercio con los Estados Unidos: análisis de las ventajas comparativas. *Revista CEPAL*, 122, 139-163.
- [28] Stellian, R. & Danna-Buitrago, J.P. (2017b). Protocolo de medición de las ventajas comparativas de los artículos manufacturados colombianos frente a Estados Unidos. En J.P Danna-Buitrago. (Ed.), *Análisis retrospectivo de la integración internacional del sector manufacturero colombiano en el marco del Tratado de Libre Comercio con Estados Unidos* (pp. 35-42). Bogotá: Los Libertadores.
- [29] Torres, J. (2013). Colombia: inserción y desequilibrios comerciales en la cuenca del pacífico. *Problemas del Desarrollo: Revista Latinoamericana de Economía*, 44(174), 139-165.
- [30] Velásquez, M.S., & Corrales, J.C. (2013). Identificación de factores de éxito para el sector cuero, calzado y marroquinería en Colombia, usando metodología Delphi: análisis estructural y juego de actores. *Informador Técnico*, 77(2), 136-146
- [31] Yasmin, B., & Altaf, S. (2014). Revealed comparative advantage of carpets and textile floor covering industry in Pakistan, India and China. *Journal Of Economic Cooperation And Development*, 35(4), 113-134.
- [32] Villegas, D. & Zapata, H. (2007). Competitividad sectorial internacional: Caso: sector del cuero y del calzado. *Entramado*, 3(1), 24-49.

Capítulo 9

pp 117-127

GESTIÓN DEL CONOCIMIENTO COMO DETERMINANTE DE LA CAPACIDAD DE INNOVACIÓN Y VENTAJA COMPETITIVA DE LAS ORGANIZACIONES

José Félix García Rodríguez^{1*}, Georgina Guadalupe Shriner Sierra², David Martínez Luis³ e Ignacio Caamal Cauch⁴

¹Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias Económico Administrativas

²Universidad Juárez Autónoma de Tabasco. División Académica de Educación y Artes

³Universidad Autónoma del Carmen. Facultad de Ciencias Económicas Administrativas

⁴Universidad Autónoma de Chapingo. División de Ciencias Económico Administrativas

RESUMEN

El conocimiento representa el activo intangible más importante de las organizaciones para impulsar la innovación y del conocimiento y capacidad de innovación existe un coeficiente de correlación de 0.921; una relación estimada de (1.267) y estandarizada de (.96), y un valor p (0.000), lo que proporciona evidencias de la presencia de una relación ventaja competitiva. La gestión del conocimiento se traduce en mayor eficiencia en el uso de recursos, mejor desempeño productivo y óptimos procesos innovadores. **Objetivo.** El propósito de la investigación es analizar la relación observada entre la gestión del conocimiento y la innovación. La hipótesis que se establece es que la capacidad de gestión del conocimiento se relaciona positivamente con la capacidad de innovación. **Método.** Se desarrolló una investigación cuantitativa de tipo descriptivo, explicativo y confirmatorio. Establecido el problema de investigación y fundamentadas teóricamente las variables generación de conocimiento e innovación, se procedió al análisis de la relación existente entre ambos constructos. Se utilizaron herramientas de análisis estadístico descriptivo, factorial exploratorio y confirmatorio. **Resultados.** Se encontró que entre las variables gestión positiva entre ambos constructos. **Conclusiones.** La gestión del conocimiento influye positivamente en la capacidad de innovación, lo que confirma la hipótesis planteada.

Palabras clave: Conocimiento, Innovación, Gestión del conocimiento, Capacidad de innovación

ABSTRACT

Knowledge represents the most important intangible asset of organizations to promote innovation and knowledge and innovation capacity. There is a correlation coefficient of 0.921; an estimated relationship of (1,267) and standardized of (.96), and a p-value (0.000), which provides evidence of the presence of a competitive advantage relationship. Knowledge management translates into greater efficiency in the use of resources, better productive performance and optimal innovative processes. Objective. The purpose of the research is to analyze the observed relationship between knowledge management and innovation. The

* Autor de correspondencia, jfgr55@hotmail.com

hypothesis that is established is that knowledge management capacity is positively related to innovation capacity. Method. A descriptive, explanatory and confirmatory quantitative investigation was developed. Once the research problem was established and the knowledge generation and innovation variables were theoretically based, we proceeded to analyze the relationship between both constructs. Descriptive, exploratory factorial and confirmatory statistical analysis tools were used. Results. It was found that among the variables positive management between both constructs. Conclusions. Knowledge management positively influences the capacity for innovation, which confirms the hypothesis raised.

Key words: Knowledge, Innovation, Knowledge Management, Innovation Capacity

INTRODUCCIÓN

Actualmente el conocimiento se ha convertido en un factor clave para el desarrollo económico y bienestar. Para poder lograrlo, las economías mundiales deben modernizarse y prepararse para competir en una economía mundial donde la generación, uso y difusión del conocimiento derivado de la ciencia, la tecnología y la innovación son determinantes del éxito económico y social. Por ello, los gobiernos, las empresas, las organizaciones públicas y privadas e instituciones educativas procuran mantenerse al día en conocimientos estratégicos capaces de generarles ventajas competitivas. De esta manera, la gestión de activos intangibles, como lo es el conocimiento, asume una importancia estratégica (Salette, Tcholakian y Selig 2013).

De Long y Fahey (2000) clasifican el conocimiento en tres tipos: a). Conocimiento humano (lo que los individuos saben, su habilidad o pericia para hacer e incluye conocimiento tácito y explícito); b). Conocimiento social (conocimiento colectivo que es más que la suma del conocimiento individual de los miembros del equipo y es en gran parte tácito, resultado de la labor conjunta; c). Conocimiento estructurado (está incorporado en los sistemas, procesos y herramientas de una organización). Se trata de conocimiento explícito, ya que existe independientemente del conocimiento humano, lo que lo convierte en un recurso organizativo fundamental. Por ello, el conocimiento, sobre todo el estructurado, es considerado un recurso clave de las instituciones, por lo que su gestión constituye una tarea básica de las organizaciones Grant (1996).

Por gestión del conocimiento (GC) puede entenderse la habilidad individual o colectiva para generar, difundir, compartir y utilizar el conocimiento, tanto tácito como explícito. La GC se ha convertido en actividad básica dentro de la organización, ya que le aporta valor estratégico (Barragán, 2009). Así, los procesos de innovación y ventaja competitiva de la organización serán más eficientes (Segarra, 2006). En suma, la GC se establece como una actividad estratégica para el desarrollo de toda organización, sobre todo en ambientes altamente competitivos (Drucker, 1993; Piri, Jasemi y Abdi, 2013, Lee Leong, Hew y Ooik, 2013). Frente al capital físico y financiero, el conocimiento resulta ser uno de los activos intangibles más importantes de la organización (Reza y Pahlavani, 2013). En síntesis, la GC se constituye en tarea permanente de la organización moderna (Lee, *et al* 2013; Darroch, 2005). A través de la adquisición, creación, intercambio y aplicación de conocimientos, las organizaciones aprovechan mejor sus activos y son más adaptables a los cambios externos (Nguyen, 2010).

Para que la organización pueda generar procesos de innovación a partir de los conocimientos que posee, es necesario que implemente procesos de gestión (Swan y Newell, 2000). La innovación es más eficiente cuando los trabajadores son provistos de una formación adecuada, así como de oportunidades para generar nuevas ideas (Bidmeshgipour, Khairuzzaman y Omar, 2012). Diversos estudios han investigado la relación existente entre la innovación y el capital humano, entendido éste como el conjunto de conocimientos, destrezas y habilidades que poseen los trabajadores de la organización (Bornay, De la Rosa, López y Valle, 2012). Existen fuertes evidencias del efecto directo y positivo de la calidad del capital humano en la innovación (Cabello, López y Valle, 2011).

En síntesis, la capacidad de una organización para innovar está directamente relacionada con los activos intelectuales y conocimientos que posee (Darroch, 2005).

Gestión del conocimiento

Dado que el conocimiento constituye el activo más importante de la organización para la innovación y ventaja competitiva (Nonaka y Takeuchi, 1995), las organizaciones se esfuerzan en lograr una eficiente GC que impulse su capacidad de innovación (CIN) (Davenport y Prusak, 1998). La relación que existe entre la GC y la CIN al interior de las organizaciones se ha estudiado ampliamente. Así, López y Meroño (2011) encontraron que la estrategia de GC tiene un impacto positivo sobre la innovación y el rendimiento productivo. El-Kot y Gamal (2011) confirmaron que existe una relación positiva y significativa entre la GC y la innovación organizacional, así como la ventaja competitiva sostenible. Por su parte, Mehrabani y Shajari (2012) observaron que la creación, organización, difusión y aplicación del conocimiento como actividades sustantivas de la GC se asocia directamente con la CIN. De la misma manera, Palacios, Gil y Garrigos (2009) hallaron que la introducción de un programa de GC en la organización coadyuva al desarrollo de habilidades de adquisición, transferencia, difusión y aplicación del conocimiento acumulado.

La literatura sobre GC de las organizaciones la sustenta en dos dimensiones: capacidad de infraestructura y capacidad de proceso. La primera corresponde a las actividades generales de la organización y comprende cuatro aspectos: estructura organizacional, cultura organizacional, recursos humanos y tecnología de la información. La segunda corresponde a la coordinación estructurada creada con el fin de gestionar eficazmente el conocimiento y resulta esencial, ya que permite a la organización capturar, procesar y transferir el conocimiento, así como gestionar eficazmente el conocimiento interno y externo (Gold, et al., 2001). Está integrada por la adquisición, conversión, aplicación y protección del conocimiento (Gold, Malhotra y Segars, 2001, Lee y Choi, 2003, Nguyen, 2010).

Asimismo, se destaca la importancia de las capacidades estructurales de la organización (cultura, estructura funcional, recursos humanos y tecnológicos) como aspectos estratégicos que impulsan los procesos de adquisición, retención, transferencia y utilización del conocimiento (Chuang, 2004; Lee y Choi, 2003). En síntesis, la GC debe entenderse como un mecanismo institucional que hace posible que el conocimiento sea creado, explotado y compartido (Palacios, Gil y Garrigos, 2009), y este proceso de intercambio de conocimientos conlleva a la generación de nuevas ideas, procesos y productos, es decir, a la innovación (Camelo, García, Sousa y Valle, 2011).

Innovación

La innovación tiene implícita la adopción de una nueva idea o comportamiento de una organización (Damanpour et al. 2009). A nivel organizacional, la innovación se entiende como la adopción por vez primera de una tecnología, estrategia, o práctica de gestión, o una significativa reestructuración o mejora de un proceso (Haiyang Li y Atuahene-Gima, 2001). La innovación puede adoptar muchas formas. De acuerdo con los resultados que produce, puede ser de proceso, producto y servicio (Prajogo y Pervaiz, 2006). De acuerdo con el nivel de alteración o cambio que introduce puede ser incremental y radical (Darroch, 2005).

Las innovaciones tecnológicas corresponden a las modificaciones incorporadas en los productos y procesos actuales basadas en la aplicación de tecnologías (Lee, *et al.* 2013). La innovación no técnica u organizativa involucra a la estructura funcional y los procesos administrativos y de gestión (Abdullah y Hassan, 2013). La innovación técnica, que se divide en innovación de productos (nuevos productos o servicios introducidos para cumplir con un usuario externo o necesidad del mercado), e innovación de proceso (se refiere a los nuevos elementos introducidos en las operaciones de producción o servicios de una organización) que puede mejorar las operaciones, ahorrar costos, aumentar la eficiencia, la productividad y el rendimiento en un corto tiempo (Shu, *et al.* 2012).

Damanpour, Walker y Avellaneda (2009) distinguen dos tipos de innovación de producto (de bienes y de servicios), y dos tipos de innovación de proceso: innovaciones en los procesos operativos (como los servicios al cliente, logística y adquisiciones), y las innovaciones en los procesos de gestión (como la planificación estratégica, el proyecto de gestión y evaluación de los empleados). Asimismo, clasifican tres tipos de innovación que son aplicables a las organizaciones de servicios: de servicio, de procesos tecnológicos y de procesos administrativos. Respecto a las innovaciones de servicio, Damanpour, et al, (2009) mencionan que en la investigación de la innovación no se ha distinguido por lo general entre las innovaciones de producto y las de servicio, y que esto se debe a que ambas tienen un enfoque externo, son impulsadas principalmente por los mercados, y sus resultados son la introducción de cambios en la producción de la organización para sus consumidores o clientes. Al igual que las innovaciones de productos, los motores de la innovación de servicios son la demanda y el deseo de introducir nuevos servicios para los mercados existentes o nuevos nichos de mercado.

En lo que corresponde a innovaciones de proceso, contrario a las innovaciones de productos o servicios, tienen un enfoque interno, ya que su objetivo es aumentar la eficiencia y la eficacia de la organización para facilitar la producción y la entrega de bienes o servicios a los clientes. Los nuevos procesos pueden estar asociados con el núcleo tecnológico o sistema técnico de la organización (innovaciones de procesos tecnológicos), o con el núcleo administrativo de la organización (innovaciones de proceso administrativo). Las innovaciones de procesos tecnológicos constituyen nuevos elementos introducidos en el sistema productivo y de servicios de la organización. Este tipo de innovación buscan reducir los tiempos de entrega, aumentar la flexibilidad operativa y reducir los costos de producción. Por lo tanto, las innovaciones de procesos tecnológicos modifican los procesos y sistemas operativos de la organización.

Finalmente, las innovaciones en los procesos administrativos corresponden a nuevos enfoques y prácticas para motivar y recompensar a los miembros de la organización, diseñar la estrategia y la estructura de las tareas y unidades, y modificar los procesos de gestión de la organización. Las innovaciones tecnológicas están directamente relacionadas con el trabajo y actividad principal de la organización para producir cambios en sus sistemas operativos, y las innovaciones administrativas están indirectamente relacionadas con la actividad básica de trabajo de la organización y afectan principalmente a sus sistemas de gestión. Las Innovaciones administrativas de proceso se refieren a los cambios en la estructura y los procesos de la organización, sistemas administrativos, conocimientos utilizados en la realización de los trabajos de gestión, y capacidad de gestión que le permiten funcionar y tener éxito mediante el uso de sus recursos con eficacia. Este tipo de innovación se conoce también como innovación administrativa (Damanpour, *et al.* (2009).

En síntesis, está claro que las organizaciones capaces de gestionar el conocimiento (CGC) obtienen mejores resultados, y que éste es el motor principal de innovación y ventaja competitiva (Nonaka y Takeuchi 1995). Para que el conocimiento sea creado y explotado, tiene que ser compartido. Este proceso de intercambio de conocimientos lleva a la generación de nuevas ideas, procesos y productos, es decir, a la innovación (Camelo, García, Sousa y Valle, 2011). Por ello, la CGC debe asumirse como tarea permanente para impulsar la CIN institucional (Palacios, *et al.*, 2009). Las investigaciones de Nonaka y Takeuchi (1995) sobre la creación de conocimiento organizativo conciben el conocimiento como requisito principal para la innovación y ventaja competitiva. En esta línea de investigación, el propósito central de esta investigación es analizar la relación existente entre la CGC y la CIN.

En el análisis, la capacidad de innovación se asume como variable dependiente y está integrada por la innovación de productos y la innovación de procesos. La primera se refiere al desarrollo o mejora de nuevos productos y servicios introducidos a mercados ya existentes o nuevos (Wang y Ahmed, 2004; Damanpour, 2009). La segunda corresponde a nuevos métodos de producción y/o enfoques de gestión (Wang y Ahmed, 2004; Damanpour, 2009). Por su parte, la CGC se asume como variable independiente, y se sustenta en la capacidad de infraestructura y la capacidad de proceso con que cuentan las organizaciones.

MÉTODO

Investigación con enfoque cuantitativo de tipo descriptivo y confirmatorio, cuyo propósito es confirmar la existencia de una relación positiva entre la capacidad de gestión del conocimiento (CGC) y la capacidad de innovación (CIN) en las organizaciones, asumiéndose la CGC como variable independiente y la CIN como variable dependiente. Una vez establecido el problema de investigación y fundamentadas teóricamente las variables CGC y CIN, se procedió a estudiar la relación existente entre ambos constructos. Para ello se utilizaron herramientas de análisis estadístico descriptivo, factorial exploratorio y confirmatorio. En la modelación y prueba de hipótesis establecidas se aplicó un análisis de Ecuaciones Estructurales (SEM), el cual de acuerdo a Herrero (2010) es más adecuado para este tipo de análisis. El propósito principal es explicar las covarianzas o correlaciones observadas entre un conjunto de variables medidas a través de un conjunto de variables latentes o factores (Bollen, 1989).

Diseño. En el levantamiento de la información de campo se aplicó un cuestionario estructurado, diseñado a partir de los objetivos de la investigación. Con base en los hallazgos de la revisión bibliográfica se procedió a la operacionalización de las variables CGC y CIN. El instrumento consta de 88 ítems en escala de Likert. Para medir la CGC se utilizó la escala producida por Nguyen (2010), en tanto que en la medición de la CIN se utilizó la escala de Al-Husseini, S. y Elbeltagi, I. (2012). La confiabilidad de los ítems se evaluó mediante estimación de coeficiente alfa de Cronbach, según escala propuesta por Nunnally (1978): menos de 0,6 (baja); entre 0,61 y 0,70 (adecuada); entre 0,71 a 0,80 (buena); mayor de 0,80 (alta). En la validación de constructo se utilizó Análisis Factorial Exploratorio (EFA) y Análisis Factorial Confirmatorio (CFA), según criterios de Hair et al., (2006). Finalmente, en el procesamiento y análisis de los datos se utilizó el programa SPSS versión 21, así como análisis de ecuaciones estructurales en el estudio de relaciones causales entre los datos obtenidos, para lo cual se utilizó el paquete estadístico AMOSS versión 20.

RESULTADOS

El análisis desarrollado considera la CGC como variable independiente compuesta por los constructos CEGC y CPGC, habiéndose validado que ambos constructos son unidimensionales y están relacionados. Respecto a la CIN, ésta es considerada como variable dependiente, confirmándose también su unidimensionalidad. Los resultados del SEM efectuado arrojaron un coeficiente de correlación de 0.921 entre ambas variables. Asimismo, la relación estimada (1.267), la estandarizada (.96) y el valor p (0.000), proporcionan evidencia de la existencia de una relación positiva entre ellas. Por lo tanto, se puede afirmar que la CGC influye positivamente en la CIN de las organizaciones.

En lo que corresponde a la relación existente entre la CEGC y la CPGC, se encontró evidencia suficiente para confirmar que la CEGC determina positivamente la CPGC. Esto es respaldado por la evidencia empírica, que muestra que las actividades generales de la organización englobadas en la CEGC son mecanismos que influyen determinantemente en la creación de conocimiento, así como en su protección e intercambio, conclusión coincidente con Leey Choi, 2003, y Nguyen, 2010. Asimismo, la dimensión de la CEGC que más influye en la CPGC resulta ser la cultura organizacional, misma que tiene una relación directa y significativa sobre la aplicación del conocimiento e indirecta con la adquisición y protección del conocimiento. De la misma manera, la estructura organizacional tiene también una relación indirecta sobre la conversión y aplicación del conocimiento.

DISCUSIÓN

La revisión teórica efectuada, así como la evidencia empírica derivada de la investigación demuestran que la GC contribuye de manera determinante para que las organizaciones apliquen de forma eficiente y efectiva el conocimiento y logren consolidar de mejor manera sus procesos de innovación. También se confirmó que existe una relación positiva y significativa entre la CGC y la CIN organizacional, y entre la innovación organizacional y la ventaja competitiva sostenible. Así, los resultados concuerdan con las conclusiones de Nonaka y Takeuchi (1995), quienes conciben el conocimiento como un recurso estratégico de la organización, así como para la innovación y ventaja competitiva sostenible (Grant, 1996). Asimismo, la puesta en marcha de programas de

GC en la organización tiene un efecto positivo no solo sobre la innovación, sino que ayudan también a desarrollar habilidades para la adquisición de conocimientos, así como en su conversión, difusión y aplicación interna (Palacios, Gil y Garrigos, 2009). En este sentido, la GC debe entenderse como un mecanismo institucional capaz de estimular la coordinación de conocimiento tácito y explícito que se difunde a través de la organización y su entorno. Por lo tanto, la CGC está fuertemente vinculada con la CIN dentro de las organizaciones (Palacios, et al., 2009).

Los hallazgos de la investigación empírica sobre la influencia positiva y significativa de la CGC en la CIN en las organizaciones coinciden con estudios preliminares de Noruzy, Dalfard, Azhdari, Nazari-Shirkouhi y Rezazadeh, 2012, Nouri, Mousavi, y Soltan, 2016. Los resultados demuestran también que la innovación requiere de actividades de creación de conocimiento a nivel organizacional (Shu, Page, Gao, y Jiang, 2012); otros resultados muestran que la aplicación del conocimiento tiene un efecto mediador en la innovación (Li, Liu, Wang, Li, y Guo, 2009). La estrategia de GC tiene impacto positivo sobre la innovación y el rendimiento de la organización, a través de un aumento en la CIN (López y Meroño, 2011). La creación y adquisición de conocimiento es determinante en el rendimiento innovador (Zhang, Shu, Jiang, y Malter, 2010), y la combinación del conocimiento impacta directamente la innovación de productos y de procesos (Shu, et al. 2012).

En relación al análisis de la CEGC, se observó que los resultados coinciden con Gold, Malhotra y Segars (2001) y Lee y Choi (2003), quienes indican la existencia de una relación directa entre la cultura de conocimiento y la tecnología, e indirectamente con el recurso humano de la organización. Respecto a la estructura organizativa como fuente de estímulo a la GC dentro de las organizaciones, el análisis efectuado indica la presencia de una relación indirecta con la adquisición, conversión y aplicación del conocimiento, resultado compatible con los hallazgos de Nonaka y Takeuchi (1995), Wang y Ahmed (2003) y Nguyen (2010). En relación a la Tecnología, Gold, et al., (2001) encuentran igualmente que ésta representa un elemento fundamental de la dimensión estructural necesaria para la creación de nuevos conocimientos, ya que permite superar las barreras de comunicación presentes en la organización educativa. Asimismo, Gold et al., (2001) y Allameh, Zare y Davoodi (2011) encuentran que la tecnología de la información determina el acceso al conocimiento dentro de la organización, y por lo tanto, ésta debe invertir en la infraestructura tecnológica adecuada que respalde las actividades científicas desarrolladas en la misma.

Respecto al hallazgo de que al interior de la estructura de las organizaciones los recursos humanos facilitan el proceso de intercambio de conocimientos entre sus diversas áreas, éste es compatible con Bharadwaj, *et al.* (2015), quienes encuentran que una estructura flexible permite la formación de equipos de trabajo colaborativo. Asimismo, Lee y Choi (2003) sostienen que el capital humano es la clave en la creación del conocimiento organizacional. Por ello, para Nguyen (2010) es fundamental el desarrollo continuo de las habilidades y competencias de los trabajadores de la organización.

En lo que respecta a cultura organizacional, el análisis demostró que se relaciona directamente con la aplicación del conocimiento e indirectamente con su adquisición y protección, lo que coincide con Gold, et al., (2001), quienes encuentran que el mayor obstáculo para una eficaz GC es la ausencia de cultura organizacional. Asimismo, respecto a las tecnologías de la información (TI), el análisis estableció la existencia de una relación

directa entre ésta y la adquisición de conocimientos, hallazgo que está en línea con los resultados de Lee y Choi (2003), quienes sostienen que las TI impactan positivamente en el conocimiento, facilitando la adquisición, almacenamiento e intercambio de conocimientos a gran escala, coadyuvando así al proceso de creación de conocimiento. Estos resultados son compatibles también con Hsu (2014) y Bharadwaj, Chauhan y Raman, (2015) quienes afirman que las TI constituyen un factor clave en la gestión del conocimiento.

En lo que corresponde a la CPGC, entendida como la capacidad de las organizaciones para crear nuevos conocimientos, el análisis arroja la existencia de una relación directa y significativa, resultado compatible con el alcanzado por Nonaka y Takeuchi (1995). Por su parte, Lee y Suh (2003) encontraron que los procesos de gestión del conocimiento están orientados a lograr el que el conocimiento existente sea útil para la organización. Finalmente, respecto a la CIN en las organizaciones, los resultados del estudio indican una relación directa y significativa entre la CGC y la CIN de productos y procesos, hallazgo concordante con Shu, *et al.* (2012), Al-Husseini y Elbeltagi (2012), quienes encuentran que la capacidad de una organización para combinar y utilizar distintos tipos de conocimiento es fundamental para lograr procesos efectivos de innovación. Asimismo, se encontró evidencia empírica de una relación directa entre la estructura organizacional y la innovación de procesos, hallazgo compatible con los de Dilnutt (2000), quien encuentra que la estructura organizativa influye directamente en la gestión de conocimientos e innovación dentro de la organización. Al respecto, Nonaka y Takeuchi (1995) establecen que ciertos tipos de estructura facilitan el intercambio y generación de conocimiento, así como el impulso de la innovación.

CONCLUSIONES

En un mundo más abierto, interconectado y competitivo, el conocimiento se ha convertido en factor clave para el desarrollo económico y social de todos los países del mundo. Por ello, el conocimiento, sobre todo el estructurado, es considerado como un recurso clave de las organizaciones, y su gestión y transferencia constituye un principio básico de la administración de las mismas. Los resultados de la investigación demuestran que las organizaciones que gestionan los conocimientos utilizan de manera más eficiente los recursos con que cuentan, son más innovadoras y tienen un mejor desempeño. De la misma manera, la capacidad de innovación de las organizaciones está estrechamente relacionada con los activos intelectuales y conocimientos que posee. Asimismo, existe una relación positiva y significativa entre su capacidad de generación de conocimiento y su capacidad de innovación, y esta capacidad de innovación impulsa su ventaja competitiva sostenible. Todas estas evidencias empíricas están ampliamente respaldadas en la revisión teórica efectuada.

En síntesis, la puesta en marcha de programas de generación de conocimiento en la organización tiene un efecto positivo no solo sobre su capacidad de innovación, sino que contribuyen también a desarrollar habilidades para la adquisición de conocimientos, así como en su conversión, difusión y aplicación interna. En este sentido, la gestión del conocimiento debe entenderse como un mecanismo institucional capaz de estimular la generación del conocimiento tácito y explícito que se difunde a través de la organización y su entorno. Por lo tanto, la capacidad de generación de conocimiento está fuertemente vinculada con la capacidad de innovación dentro de las organizaciones.

REFERENCIAS

- [1] Abdullah, S. y Hassan, W. (2013). The Impact of Human Resource Management Practices, Organisational Culture, Organisational Innovation and Knowledge Management on Organisational Performance in Large Saudi Organisations: Structural Equation Modeling With Conceptual Framework. *International Journal of Business and Management*, 8(22), 1-19. Doi: 10.5539/ijbm.v8n22p1
- [2] Al-Husseini, S. y Elbeltagi, I. (2012). Knowledge Sharing and Innovation: An Empirical Study in Iraqi Private Higher Education Institutions. International Conference on Intellectual Capital. Knowledge Management & Organizational Learning. 129-138.
- [3] Allameh, S., Zare, S. y Davoodi, S. (2011). Examining the Impact of KM Enablers on Knowledge. *Management Processes, Procedia Computer Science*, 3, 1211-1223. Doi: 10.1016/j.procs.2010.12.196
- [4] Barragan, A. (2009). Aproximación a una taxonomía de modelos de gestión del conocimiento. *Intangible Capital*, 5(1), 65-101. Doi: 10.3926/ic.2009.v5n1.p65-101
- [5] Bharadwaj, S., Chauhan, S. y Raman, A. (2015). Impact of Knowledge Management Capabilities on Knowledge Management Effectiveness in Indian Organizations. *VIKALPA, The Journal for Decision Makers*, 40(4), 421-434. Doi: 10.1177/0256090915613572
- [6] Bidmeshgipour M., Khairuzzaman W. y Omar, R. (2012). Knowledge management and organizational innovativeness in Iranian banking industry. *Knowledge Management & E-Learning: An International Journal*, 4 (4), 481-499.
- [7] Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley y Sons.
- [8] Bornay, M., De la Rosa, D., López, A. y Valle, R. (2012). Employment Relationships and Firm Innovation: The Double Role of Human Capital. *British Journal of Management*, 23, 223-240. Doi: 10.1111/j.1467-8551.2010.00735.x
- [9] Cabello, C., López, A. y Valle, R. (2011). Leveraging the innovative performance of human capital through HRM and social capital in Spanish firms. *The International Journal of Human Resource Management*, 22(4), 807-828. Doi: 10.1080/09585192.2011.555125
- [10] Camelo, C., García, J., Sousa, E. y Valle, R. (2011). The influence of human resource management on knowledge sharing and innovation in Spain: the mediating role of affective commitment. *The International Journal of Human Resource Management*, 22(7), 1442-1463. Doi: 10.1080/09585192.2011.561960
- [11] Chuang, Shu-Hui. (2004). A resource-based perspective on knowledge management capability and competitive advantage: an empirical investigation. *Expert Systems with Applications*, 27(3), 459-465. Doi: 10.1016/j.eswa.2004.05.008
- [12] Damanpour, F., Walker, R. y Avellaneda, C. (2009). Combinative Effects of Innovation Types and Organizational Performance: A Longitudinal Study of Service Organizations. *Journal of Management Studies*, 46(4), 650-675. Doi: 10.1111/j.1467-6486.2008.00814.x
- [13] Darroch, J. (2005). Knowledge management, innovation and firm performance. *Journal of knowledge Management*, 9(3), 101-115. Doi: 10.1108/13673270510602809
- [14] Davenport, T. y Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press.
- [15] De Long, D. y Fahey, L. (2000). Diagnosing cultural barriers to knowledge management. *Academy of Management Executive*, 14(4), 113-127. Doi: 10.5465/AME.2000.3979820
- [16] Dilnutt, RP (2000). "Knowledge management as practiced in Australian organizations: a case study approach". DBA thesis. Australia: Southern Cross University.
- [17] Drucker, P. F. (1993). *Innovation and Entrepreneurship (Harper Business edition, first edition 1986)*. New York; HarperCollins Publishers.
- [18] El-Kot, G. y Gamal, D. (2011). How Does Knowledge Management Drive Competitiveness in Egyptian Software Companies?. *Review of Management*, 1(4), 56-76.
- [19] Gold, A., Malhotra, A. y Segars, A. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems*, 18(1), 185-214. Doi: 10.1080/07421222.2001.11045669
- [20] Grant, R. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, 17, Special Issue: Knowledge and the Firm, 109-122. Doi: 10.1002/smj.4250171110
- [21] Hair, J. F., Gomez, S, M., Cano, D., y Frances P. E. (2006). *Análisis multivariante*. Madrid: Prentice Hall.

- [22] Haiyang, A-G. (2001). Product innovation strategy and the performance of new technology ventures in China. *Academy of Management Journal*, 44(6), 1123-1134. Doi: 10.2307/3069392
- [23] Herrero, J. (2010). El Análisis Factorial Confirmatorio en el estudio de la Estructura y Estabilidad de los Instrumentos de Evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14. *Psychosocial Intervention*, 19(3), 289-300.
- [24] Hsu, S. (2014). Effects of Organization Culture, Organizational Learning and IT Strategy on Knowledge Management and Performance. *The Journal of International Management Studies*, 9(1), 50-58.
- [25] Lee, H. y Suh, Y. (2003), Knowledge conversion with information technology of Korean companies. *Business Process Management Journal*, 9(3), 317-336. Doi: 10.1108/14637150310477911
- [26] Lee, H. y Choi, B. (2003). Knowledge management enablers, processes, and organizational performance: an integrative view and empirical examination. *Journal of Management Information Systems*, 20(1), 179-228. Doi: 10.1080/07421222.2003.11045756
- [27] Lee, V., Leong, L., Hew, T. y Ooik, K. (2013). Knowledge management: a key determinant in advancing technological innovation?. *Journal of Knowledge Management*, 17(6), 848-872. Doi: 10.1108/JKM-08-2013-0315
- [28] Li, Y., Liu, X., Wang, L., Li, M. y Guo, H. (2009). How Entrepreneurial Orientation Moderates the Effects of Knowledge Management on Innovation. *Systems Research and Behavioral Science*, 26, 645-660. Doi: 10.1002/sres.980
- [29] López, C. y Meroño A. L. (2011). Strategic knowledge management innovation and performance. *International Journal of Information Management*, 31(6), 502-509. Doi: 10.1016/j.ijinfomgt.2011.02.003
- [30] Mehrabani, S. E. y Shajari, M. (2012). Knowledge Management and Innovation Capacity. *Journal of Management Research*, 4(2), 164-177. Doi: 10.5296/jmr.v4i2.1390
- [31] Nguyen, T. (2010). *Knowledge management capability and competitive advantage: an empirical study of Vietnamese enterprises* (PhD thesis). Southern Cross University, Lismore, NSW.
- [32] Nonaka, I. y Takeuchi, H. (1995) "The knowledge-creating company: How Japanese companies create the dynamics of innovation". Oxford, London: Oxford University Press.
- [33] Noruzi, A., Dalfard, V. M., Azhdari, B., Nazari-Shirkouhi, S. y Rezazadeh, A. (2012). Relations between transformational leadership organizational learning knowledge management organizational innovation, and organizational performance: an empirical investigation of manufacturing firms. *International Journal of Advanced Manufacturing Technology*, 64, 1073-1085. Doi: 10.1007/s00170-012-4038-y
- [34] Nouri, B. A., Mousavi, M. M. y Soltan, M. (2016). Effect of Transformational Leadership and Knowledge Management Processes on Organizational Innovation in Ardabil University of Medical Sciences. *International Journal of Management, Accounting and Economics*, 3(11), 672-698.
- [35] Nunnally, J. C. (1978). *Psychometric Theory*. 2ª ed., New York: McGraw Hill.
- [36] Palacios, D., Gil, I. y Garrigos, F. (2009). The impact of knowledge management on innovation and entrepreneurship in the biotechnology and telecommunications industries. *Small Business Economics*, 32, 291-301. Doi: 10.1007/s11187-008-9146-6
- [37] Piri, M., Jasemi, M. y Abdi, M. (2013). Intellectual capital and knowledge management in the Iranian space industries. *VINE: The Journal of Information and Knowledge Management Systems*, 43(3), 341-356. Doi: 10.1108/VINE-08-2012-0036
- [38] Prajogo, D. y Pervaiz, A. (2006). Relationships between innovation stimulus, innovation capacity, and innovation performance. *R&D Management*, 36(5), 449-515. Doi: 10.1111/j.1467-9310.2006.00450.x
- [39] Reza, M. y Pahlavani, N. (2013). The effect of social capital on knowledge creation in Petrochemical Industry. *Management Science Letters*, 3, 879-884. Doi: 10.5267/j.msl.2013.01.034
- [40] Salete, J., Tcholakian, A. B. y Selig, P. M. (2013). A influência do capital intelectual no potencial de inovação das universidades. *Iberoamerican Journal of Industrial Engineering*, 5(9), 01-22.
- [41] Segarra, M. (2006). *Estudio de la naturaleza estratégica del conocimiento y las capacidades de gestión: aplicación a empresas innovadoras de Base Tecnología* (Tesis Doctoral). Universitat Jaume I, Castellon, España.
- [42] Shu, Ch., Page, A., Gao, S. y Jiang, X. (2012). Managerial Ties and Firm Innovation: Is Knowledge Creation a Missing Link?. *Product Development & Management Association*, 29(1), 125-143. Doi: 10.1111/j.1540-5885.2011.00883.x

- [43] Swan, J., y Newell, S. (2000). Linking Knowledge Management and innovation. *European Conference on Information Systems*, 173, 591-598.
- [44] Wang, C. y Ahmed, P. (2003). Structure and structural dimensions for knowledge-based organization's. *Measuring Business Excellence*, 7(1), 51-62. Doi: 10.1108/13683040310466726
- [45] Wang, C. y Ahmed, P. (2004). The development and validation of the organisational innovativeness construct using confirmatory factor analysis. *European Journal of Innovation Management*, 7(4), 303-313. Doi: 10.1108/14601060410565056
- [46] Zhang, H., Shu, Ch., Jiang, X. y Malter, A. J. (2010). Managing Knowledge for: The Role of Cooperation, Competition, and Alliance Nationality. *Journal of International Marketing*, 18(4), 74-94. Doi: 10.1509/jimk.18.4.74

Capítulo 10

pp 129-138

FORMACIÓN DEL PRECIO DE EXPORTACIÓN DEL LIMÓN PERSA DEL MUNICIPIO DE MISANTLA, ESTADO DE VERACRUZ, MÉXICO

Ignacio Caamal Cauich^{1*}, Verna Grisel Pat Fernández¹, José Félix García Rodríguez², José Antonio Ávila Dorantes¹ y Alberto Pérez Fernández²

¹Universidad Autónoma Chapingo, Chapingo, Estado de México, México.

²Universidad Juárez Autónoma de Tabasco, Tabasco, México.

³Universidad Autónoma del Carmen, Cd. del Carmen, Campeche, México

*E-mail: icaamal82@yahoo.com.mx

RESUMEN

El mayor productor y exportador de limón persa en el mundo es México. Los principales estados productores de limón persa en México son Veracruz, Oaxaca, Tabasco, Jalisco y Yucatán. En el estado de Veracruz el limón persa se cultiva principalmente en la región de Martínez de la Torre, en los municipios de Martínez de La Torre, Atzalan, San Rafael, Tlapacoyan, Papantla, Nautla, Misantla, Tecolutla, Espinal, Gutiérrez Zamora, Coxquihui y Vega de Alatorre. Las exportaciones de limón persa se destinan básicamente a Estados Unidos de América, Países Bajos y Japón. El canal de distribución principal que sigue el limón persa es del productor al mercado de subasta, del mercado de subasta a la empacadora y de la empacadora a la frontera para ingresar a los Estados Unidos de América y a puertos del Golfo de México para su transporte a Europa, principalmente a los Países Bajos, Reino Unido y Francia. Las mayores proporciones del precio de frontera o de puerto del limón persa las obtienen el productor, el bróker, la empacadora y el acopiador del mercado de subasta. El precio de exportación representa casi el triple del precio del productor, del cual se descuentan los costos de producción para obtener las utilidades del productor, que son positivas, al igual que las de los agentes de distribución, los cuales obtienen proporciones mayores de utilidades.

Palabras clave: productor, acopiador, empacador, precio.

FORMATION OF THE EXPORT PRICE OF THE PERSIAN LIME OF THE MUNICIPALITY OF MISANTLA, STATE OF VERACRUZ, MEXICO

ABSTRACT

The largest Persian lime producer and exporter in the world is Mexico. The main Persian lime producing states in Mexico are Veracruz, Oaxaca, Tabasco, Jalisco and Yucatan. In the state of Veracruz, the Persian lime grown mainly in the region of Martínez de la Torre, in the municipalities of Martínez de La Torre, Atzalan, San Rafael, Tlapacoyan, Papantla, Nautla, Misantla, Tecolutla, Espinal, Gutiérrez Zamora, Coxquihui and Vega de Alatorre. Persian lime exports are mainly destined for the United States of America, the Netherlands and Japan. The main distribution channel that follows Persian lime is from the producer to the auction market, from the auction market to the packer and from the packer to the border to enter the United States of America and ports in the Gulf of Mexico for transport to Europe, mainly to the Netherlands, the United Kingdom and France. The highest proportions of the Persian lime border price or port price are obtained by the producer, broker, packer and auction market collector. The export price accounts for almost three times the producer's price, from which production costs are discounted to obtain the producer's profits, which are positive, as well as those of distribution agents, who obtain higher proportions of profits.

Keywords: producer, collector, packer, price.

1. INTRODUCCIÓN

En México se cultivan más de 611 mil hectáreas de cítricos, que representan el 36% de la superficie sembrada de frutales y generan el 36% de la producción nacional frutícola. Dentro de los cítricos, la naranja ocupa el primer lugar, tanto en superficie cosechada como en volumen de producción; el segundo lugar le corresponde al limón, del cual se producen las tres variedades más importantes, mexicano, persa e italiano (SIACON, 2020).

El limón persa se produce en la cuenca del Golfo de México, principalmente en los estados de Veracruz, Oaxaca, Tabasco y Yucatán. El Distrito de Desarrollo Rural de Martínez de la Torre, compuesto por los municipios de Martínez de La Torre, Atzalan, San Rafael, Tlapacoyan, Papantla, Nautla, Misantla, Tecolutla, Espinal, Gutiérrez Zamora, Coxquihui y Vega de Alatorre, es la región más importante en la producción de limón persa, aportando más del 74% de la producción del estado de Veracruz, la producción de la región cumple con la calidad para la exportación (SIAP, 2020).

A pesar de que fue desde el año de 1975 cuando se comenzó a introducir el limón persa al mercado mexicano, es en la década de los ochentas cuando el cultivo se consolida y se inicia la producción en mayor escala comercial, desarrollándose principalmente en la región de Martínez de la Torre, estado de Veracruz, de donde se fue difundiendo ampliamente en la vertiente del Golfo de México, en los estados de Veracruz, Tabasco y Yucatán (CONCITVER, 2012).

El limón persa genera un volumen de producción similar a la del limón mexicano (Figura 1). La importancia del limón persa radica, en que se destina principalmente a la exportación y por consiguiente genera divisas, mientras que el limón mexicano tiene como destino el mercado nacional y el limón verdadero o italiano, se siembra en pequeñas extensiones, genera un reducido volumen de producción y es de consumo doméstico (Figura 1).

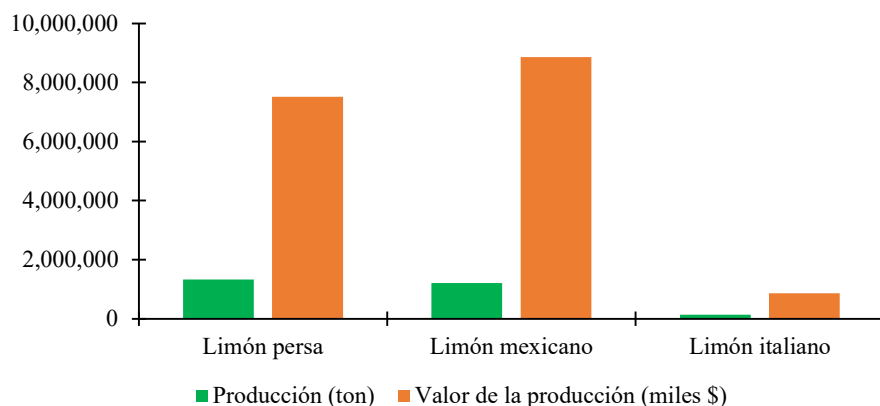


Figura 1. Volumen y valor de la producción de los diferentes tipos de limón en México.
Fuente: Elaborado con datos de SIACON, 2020.

México tiene excelentes condiciones geográficas y naturales en algunas regiones agrícolas, especialmente en la región de Martínez de la Torre, estado de Veracruz, la cual cuenta con las mejores condiciones geográficas y naturales para la producción de limón persa, las cuales se reflejan en la alta calidad de la fruta producida, en el bajo costo de producción y en la competitividad del producto. El limón persa se cosecha durante todo el año y genera ingresos permanentes para los productores de las diferentes regiones de México (Schwentesiuss y Gómez, 2005).

El canal de comercialización se refiere al conjunto de agentes por los que pasa el producto, desde que sale de la explotación agrícola hasta que llega al consumidor final; también se puede definir como la vía por la que pasa el producto, desde el productor hasta el consumidor final, pasando por varios agentes de comercialización. El estudio de los canales de comercialización, aparte de reflejar las rutas por donde circulan los productos, persigue la disminución de la incertidumbre y los riesgos que originan la deficiente coordinación de la comercialización. Los principales agentes de la cadena de comercialización son los acopiadores, mayoristas, que se clasifican en mayoristas de origen y mayoristas de destino, industria, minorista y detallista (García et al., 1990).

Los márgenes de comercialización se definen como el diferencial entre el precio que paga el consumidor final y el precio recibido por el productor. Se clasifican en: 1) margen total de comercialización, que se refiere a todo el proceso de comercialización, se forma por los márgenes parciales obtenidos por cada agente y se expresa en términos absolutos y relativos; 2) margen parcial de comercialización, que se refiere al margen que se va generando en las diferentes etapas del proceso de comercialización; 3) margen absoluto de comercialización, que representa la diferencia entre el precio de venta de un producto, efectuado por un determinado agente, y el pago realizado por el mismo en el momento de la compra; 4) margen relativo de comercialización, que se refiere al margen de comercialización en porcentaje, que resulta de dividir el margen total y el precio de venta final; y 5) margen relativo parcial, que se obtiene al dividir el margen parcial y el precio parcial de cada etapa de la comercialización (Tomek & Robinson, 2003).

El objetivo del trabajo fue analizar la cadena de comercialización internacional del limón persa que se produce en el municipio de Misantla, Veracruz, para identificar los canales, márgenes y problemas en el proceso de comercialización del limón persa y ofrecer alternativas de solución, para que los pequeños y medianos productores tengan una comercialización más eficiente y obtengan una mayor proporción del valor generado.

2. MATERIALES Y MÉTODOS

2.1. Obtención y sistematización de información de campo

La información se obtuvo a través de encuestas aplicadas a los agentes de comercialización directamente en campo. Se encuestaron a 73 acopiadores, los cuales se eligieron por muestreo dirigido, debido a que no existe una base de datos de dichos agentes. También se encuestaron a empacadores que se encuentran en el municipio de Misantla, con el mismo método de encuesta.

La información recabada se ordenó en bases de datos, de acuerdo con las características generales de los diferentes agentes que participan en el canal de comercialización, resaltando la importancia de los datos generales de producción, costos de comercialización y empaqueo, volúmenes comercializados, precios de compra y venta regional, nacional e internacional, origen y destino del producto, entre otros; con la finalidad de identificar los diferentes canales y calcular los márgenes y ganancias de comercialización de los agentes de comercialización.

Se identificaron las cuatro calidades de limón persa que se manejan en el municipio de Misantla (primera, empaque, segunda y tercera), la primera se destina al mercado de Japón, Europa y E.E.U.U., en temporada de alta producción; la calidad empaque se destina al mercado de Estados Unidos de América, la segunda para consumo nacional y la tercera para la industria juguera.

2.2. Márgenes de comercialización

Con base en García *et al.* (1990), el método usado para la identificación de los agentes y el cálculo de los márgenes de comercialización fue el método directo, por considerarse el más adecuado y consiste en los siguientes pasos:

- a) Seguir lotes estadísticamente representativos del producto agrícola (limón persa), desde que salen de la explotación hasta que llegan al consumidor.

- b) Tomar nota de los distintos costos, volúmenes y precios que se van produciendo a su paso por los distintos agentes.
- c) Limitar la investigación a lotes representativos del movimiento del producto agrícola.

Con base en la información obtenida se van identificando los canales (principales, secundarios y terciarios) y calculando los márgenes de comercialización. Los canales de comercialización se identificaron con base a los agentes de comercialización.

Para el cálculo de los márgenes de comercialización se obtuvieron los promedios ponderados de los precios de venta de limón persa de los distintos agentes de comercialización, con base en el precio promedio de cada agente, para lo cual se utilizó la siguiente fórmula:

$$\bar{P}_i = \frac{\sum_{i=1}^n P_i Q_i}{\sum_{i=1}^n Q_i} \quad (1)$$

Donde:

\bar{P}_i = Promedio ponderado del precio de limón persa del agente i

P_i = Precio de venta de limón persa del agente i

Q_i = Cantidad comercializada de limón persa del agente i

\sum_i^n = Sumatoria de las cantidades comercializadas de limón persa del agente i

Para obtener los promedios ponderados de cada agente, se utilizó la unidad de kilogramo para expresar las cantidades comercializadas.

Para analizar los márgenes absoluto y relativo de comercialización de limón persa de cada agente de los diferentes canales de comercialización identificados, se utilizaron las fórmulas siguientes:

Margen absoluto

$$M_i = \bar{P}_i - \bar{P}_j \quad (2)$$

Donde:

M_i = Margen absoluto de comercialización del agente i

\bar{P}_i = Promedio ponderado del precio de venta del agente i

\bar{P}_j = Promedio ponderado del precio de venta del agente j

Margen relativo

$$m_i = \frac{M_i}{P_c} * 100 \quad (3)$$

Donde:

m_i = Margen relativo de comercialización del agente i

M_i = Margen absoluto de comercialización del agente i

P_c = Precio al consumidor de limón persa

El margen de comercialización se compone del costo de la comercialización y de la ganancia que obtienen los diferentes agentes de comercialización.

$$Mi = Ci + Gi \quad (4)$$

Donde:

Mi = Margen de comercialización del agente i

Ci = Costo de comercialización del agente i

Gi = Ganancia de comercialización del agente i

2.3. Costos de comercialización y empaqueo

El cálculo de los costos de comercialización, se realizó con las siguientes fórmulas:

$$CTci = \sum CFi \quad (5)$$

$$\bar{C}ci = \frac{\sum_{i=1}^n CTiQi}{\sum_{i=1}^n Qi} \quad (6)$$

Donde:

CTci = Costo total de comercialización del agente i

CFi = Costo de operación por fase del agente i

$\bar{C}ci$ = Promedio ponderado de los costos de comercialización del agente i

Qi = Cantidad comercializada del agente i

\sum_i^n = Sumatoria de las cantidades comercializadas de limón persa del agente i

Para obtener los promedios ponderados de cada agente, se utilizó la unidad de kilogramo para expresar las cantidades comercializadas.

Para el cálculo de los costos de empaqueo se utilizaron las siguientes fórmulas:

$$CTei = \sum_i^n CFi \quad (7)$$

$$C\bar{T}e = \frac{\sum_{i=1}^n CTeiQi}{\sum_{i=1}^n Qi} \quad (8)$$

Donde:

CTei = Costo total de empaqueo de la empresa empacadora i

CFi = Costo de operación por fase de la empresa empacadora i

$C\bar{T}e$ = Promedio ponderado general del costo de empaqueo

Qi = Volumen empaqueo de la empresa empacadora i

\sum_i^n = Sumatoria de las cantidades comercializadas de limón persa del agente i

3. RESULTADOS

3.1. Caracterización de los productores de limón persa

En este apartado se caracterizó a los productores del municipio de Misantla por el tamaño de la superficie que poseen, por el tipo de propiedad de la parcela y por el rendimiento promedio obtenido.

3.1.1. Superficie sembrada

De acuerdo con la figura 2, el 72.6% de los productores de Misantla tienen superficies menores de 3 hectáreas; seguido por un 16.4% de los productores con parcelas que oscilan entre 3 y 6 hectáreas, y, por último, el 11.0% de los productores poseen parcelas con más de 6 hectáreas. El tamaño promedio de las parcelas de los productores es de 3.1 hectáreas.

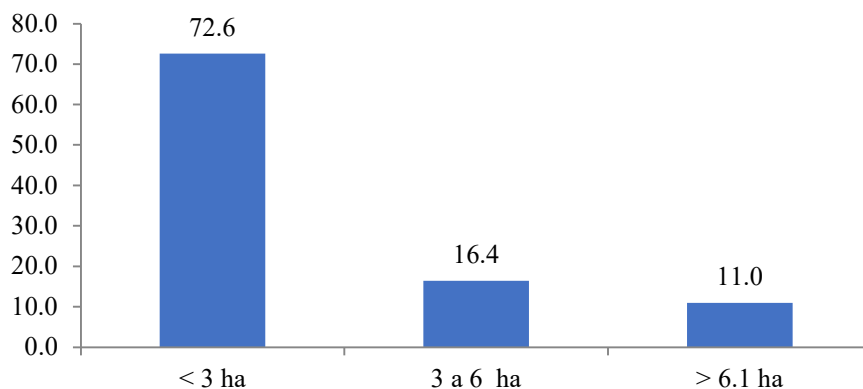


Figura 2. Distribución de productores por superficie en producción (%), 2012.

Fuente: Elaboración propia con información directa de campo, 2012.

3.1.2. Tipo de propiedad de los productores

Como se observa en la figura 3, el 87.5% de los productores de Misantla tienen parcelas con régimen ejidal y el 12.5% con régimen privado. Una de las ventajas de conservar el régimen ejidal es que les permite a los productores ser sujetos de los apoyos gubernamentales que se ofrecen en la producción de limón persa.

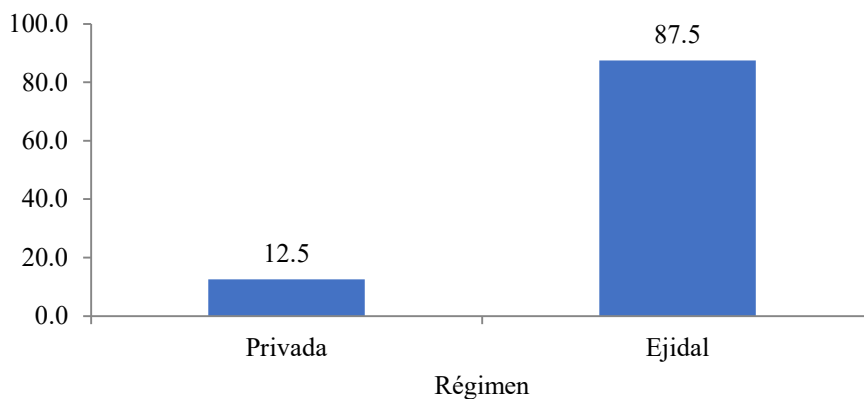


Figura 3. Productores de limón persa por tipo de propiedad (%), 2012.

Fuente: Elaboración propia con información directa de campo, 2012.

3.1.3. Rendimiento promedio de los productores

De acuerdo con la figura 4, el 21.9% de los productores de limón persa tienen un rendimiento de entre 1 y 6 toneladas por hectárea, el 45.2% presenta un rendimiento de entre 6.1 a 12 toneladas por hectárea y el 32.9% tienen un rendimiento de entre 12.1 a 26 toneladas por hectárea, siendo el promedio general en el municipio de Misantla de 12.6 toneladas por hectárea.

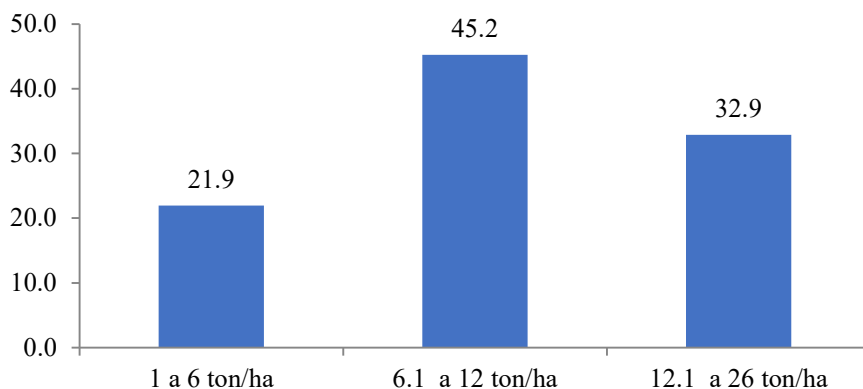


Figura 4. Distribución de productores por rendimiento (%), 2012.

Fuente: Elaboración propia con información directa de campo, 2012.

3.2. Calidad de limón persa

Por la calidad del limón persa obtenido a nivel de unidad de producción, la fruta se clasifica en cuatro categorías, de esta manera, se denomina como “primera” la fruta que se destina a los mercados de Europa, Japón y E.E.U.U., en periodo de alta producción; “empaque” la fruta que se destina al mercado de EE.UU.; “segunda” la fruta que se destina para el mercado nacional y “tercera” la que se destina para la industria juguera. Los aspectos que toman en cuenta los productores para realizar la clasificación y preseleccionado son: color, tamaño, estado de la fruta, entre otros. La producción de limón persa calidad primera y empaque que se exporta, se destina al mercado norteamericano, europeo y japonés; la calidad segunda se comercializa nacionalmente y la calidad tercera se envía a la industria de jugo y aceite.

3.3. Canales, márgenes y precios de exportación

3.3.1. Canal principal

En la tabla 1, se muestran los márgenes de comercialización y exportación del limón persa por calidad y destino de exportación, en la tabla se observa que los márgenes totales de comercialización aumentan, con base en la calidad de limón persa; teniendo mayor margen el producto destinado a Japón que el exportado a Europa y Estados Unidos, a su vez la fruta exportada a Europa registra un mayor margen con respecto a la exportada a Estados Unidos. El 59% de los productores de limón persa del municipio de Misantla participan en este canal de comercialización.

De acuerdo con la tabla 1, se puede observar que los productores tienen una participación en la formación del precio de frontera del limón persa de Misantla, con 33.9, 31.1 y 29.0% del precio del producto destinado a Estados Unidos, Europa y Japón, respectivamente, seguido de la participación del bróker y después de la empaedora, siendo estos de 23.2, 22.2 y 21.0% para el limón destinado a Estados Unidos, Europa y Japón, respectivamente, cuando la empaedora vende en zona de producción y cuando vende en frontera aumenta la participación en la formación del precio de exportación, ya que absorbe el margen correspondiente al bróker, agente exportador.

Tabla 1. Márgenes de comercialización de limón persa en el canal principal, datos anuales, 2012.

Concepto	Empaque EE.UU.			Primera Europa			Japón		
	P	MA	MR	P	MA	MR	P	MA	MR
	\$/kg	\$/kg	%	\$/kg	\$/kg	%	\$/kg	\$/kg	%
1. Costo de producción del productor	1.5		26.8	1.5		16.7	1.5		15.0
a) Margen del productor		0.4	7.1		1.3	14.4		1.4	14.0
2. Precio de venta del productor	1.9		33.9	2.8		31.1	2.9		29.0
b) Margen del acopiador en subasta		0.6	10.7		1.7	18.8		1.8	18.0
3. Precio de venta del acopiador en subasta	2.5			4.5			4.7		
c) Margen en empacadora		1.3	23.2		2.0	22.2		2.1	21.0
4. Precio de venta a nivel de empacadora	3.8			6.5			6.8		
d) Margen de la empacadora a nivel de frontera		1.8	32.1		2.5	27.8		3.2	32.0
5. Precio de venta de la empacadora al bróker	5.6			9.0			10.0		
e) Margen Total		4.1	100.0		7.5	100.0		8.5	100.0

Donde: P=Precio; MA=Margen absoluto; MR=Margen relativo.

Fuente: Elaboración propia con información directa de campo, 2012.

3.3.2. Canal secundario

El 30% de los productores de Misantla comercializan la producción a través del canal secundario. En la tabla 2, se observa que la participación del margen relativo de los productores en el precio de frontera del producto fue de 32.1, 30.0 y 28.0% para el producto destinado a Estados Unidos, Europa y Japón, respectivamente, seguido de la participación del bróker y después de la empacadora. Se observa una disminución de la participación del productor en la formación del precio de exportación con respecto al margen relativo que presenta el productor en el canal principal de 33.9, 31.1 y 29.0% para Estados Unidos, Europa y Japón, respectivamente. Por otro lado, las empacadoras participan con 23.2, 22.2 y 21.0% para Estados Unidos, Europa y Japón, respectivamente, cuando la empacadora vende en zona de producción y cuando vende en frontera aumenta la participación en la formación del precio de exportación, ya que absorbe el margen correspondiente al bróker, agente exportador.

Tabla 2. Márgenes de comercialización de limón persa en el canal secundario, datos anuales, 2012.

Concepto	Empaque EE.UU.			Primera Europa			Japón		
	P	MA	MR	P	MA	MR	P	MA	MR
	\$/kg	\$/kg	%	\$/kg	\$/kg	%	\$/kg	\$/kg	%
1. Costo de producción del productor	1.5		26.8	1.5		16.7	1.5		15.0
a) Margen del productor		0.3	5.4		1.2	13.3		1.3	13.0
2. Precio de venta del productor	1.8		32.1	2.7		30.0	2.8		28.0
b) Margen del Acopiador local		0.7	12.5		1.8	20.0		1.9	19.0
3. Precio de venta del acopiador local	2.5			4.5			4.7		
c) Margen en empacadora		1.3	23.2		2.0	22.2		2.1	21.0
4. Precio de venta de la empacadora	3.8			6.5			6.8		
d) Margen de la empacadora a nivel de frontera		1.8	32.1		2.5	27.8		3.2	32.0
5. Precio de venta en frontera	5.6			9.0			10.0		
e) Margen total		4.1	100.0		7.5	100.0		8.5	100.0

Fuente: Elaboración propia con información directa de campo, 2012.

3.3.3. Canal terciario

El 11% de los productores comercializan la producción a través del canal terciario, el cual es el más recomendable, debido a que el productor comercializa la producción directamente con las empacadoras, desplaza a los acopiadores y se queda con un mayor margen de comercialización y aumenta su participación en el precio de exportación, y por consecuencia aumenta las ganancias que obtiene considerablemente.

Por lo tanto, este canal de comercialización es el más recomendable, aunque no todos los productores tienen la capacidad para participar en este circuito de comercialización, ya que las empacadoras requieren de grandes volúmenes y calidad del producto y la mayoría de los productores poseen unidades de explotación muy pequeñas y generan bajos volúmenes de producción, sin embargo, esta exigencia puede ser superada con la organización de los productores para obtener y manejar un mayor volumen de producto.

En la tabla 3 se observa que la participación del margen relativo del productor en el precio de frontera del producto fue de 44.6, 48.9 y 46.0% para el producto destinado a Estados Unidos, Europa y Japón, respectivamente, seguido de la participación del bróker y después de la empacadora. Se observa un incremento de la participación del productor con respecto al margen relativo que presenta el productor en el canal principal de 33.9, 31.1 y 29.0% para Estados Unidos, Europa y Japón, respectivamente. Mientras que las empacadoras participan con 23.2, 35.5 y 22.0% para Estados Unidos, Europa y Japón, respectivamente, cuando la empacadora vende en zona de producción y cuando vende en frontera aumenta la participación en la formación del precio de exportación, ya que absorbe el margen correspondiente al bróker, agente exportador.

Tabla 3. Márgenes de comercialización de limón persa en el canal terciario, datos anuales, 2012.

Concepto	Empaque EE.UU.			Primera Europa			Japón		
	P	MA	MR	P	MA	MR	P	MA	MR
	\$/kg	\$/kg	%	\$/kg	\$/kg	%	\$/kg	\$/kg	%
1. Costo de producción del productor	1.5		26.8	1.5		16.7	1.5		15.0
a) Margen del productor		1.0	17.9		2.9	32.2		3.1	31.0
2. Precio de venta del productor	2.5		44.6	4.4		48.9	4.6		46.0
b) Margen en empacadora		1.3	23.2		3.2	35.5		2.2	22.0
3. Precio de venta de la empacadora	3.8			7.6			6.8		
c) Margen de la empacadora a nivel de frontera		1.8	32.1		1.4	15.6		3.2	32.0
4. Precio de venta en frontera	5.6			9.0			10.0		100.0
d) Margen total		4.1	100.0		7.5	100.0		8.5	

Fuente: Elaboración propia con información directa de campo, 2012.

La participación de los costos de producción y de comercialización en promedio en el margen absoluto del productor es de aproximadamente 64%, debido a los altos precios de los insumos, y el 36%, restante, corresponde a la ganancia. El beneficio que obtiene el productor por kilogramo, aumenta en relación a las calidades del limón que comercializa, obteniendo una mayor ganancia con el producto que destina a Europa, que el destinado a EE.UU. y una ganancia mayor por el que se destina a Japón.

4. CONCLUSIONES

En el municipio de Misantla, Veracruz la mayoría de los productores de limón persa son pequeños, tienen bajos rendimientos y poseen parcelas con régimen ejidal. De los tres canales de exportación que existen, el tercero es el más benéfico al productor, debido a que el productor vende la producción directamente con las empresas empacadoras y desplaza al acopiador, lo que genera mayor participación del productor en el precio en frontera del limón persa.

El canal de comercialización principal es el menos benéfico al productor, por la alta presencia de intermediarios. Los productores que comercializan el limón persa en el canal principal, obtienen un menor margen de comercialización y tienen menor participación en el precio en frontera del limón persa. La mayor participación en el precio en frontera del limón persa es la del productor, seguido por el bróker y la empacadora, la cual agrega valor al producto: selecciona, limpia, encera, empaca y almacena el producto en cámaras de refrigeración y en ocasiones lo transporta a la frontera, apropiándose del margen del bróker, agente exportador.

Los productores tienen menor ganancia que los acopiadores y las empresas empacadoras, debido a los altos costos de producción. Los acopiadores y las empresas empacadoras tienen una mayor ganancia, debido a que tienen un mayor diferencial entre precios y costos.

REFERENCIAS

- [1] CONSEJO ESTATAL CITRÍCOLA DE VERACRUZ A. C. (CONCITVER). (2012): Portal de internet. Disponible en: <<http://www.concitver.com>> Consultado el 6 de julio de 2012.
- [2] García M., R., García D., G. y Montero H., R. (1990): *Notas sobre mercados y comercialización de productos agrícolas*. México: Centro de Economía CEICADES CRECIDATH. Colegio de Postgraduados.
- [3] Schwentesius R., R y Gómez C., M. A. (2005): *Limón persa. Tendencias en el mercado mexicano*. México: CIESTAAM/UACH y Banco Mundial.
- [4] Servicio de Información Agroalimentaria y Pesquera (SIAP). (2020): *Anuario estadístico de la producción agrícola*. México: SAGARPA. [En línea] Consultado el 26 de septiembre de 2020. Disponible en: <<https://nube.siap.gob.mx/index.php/s/Qv6klxhXkDv2uma>>.
- [5] Sistema de Información Agroalimentaria de Consulta (SIACON). (2020): *Agrícola estatal*. México: SIAP-SAGARPA: [En línea] Consultado el 25 de septiembre de 2020. Disponible en: <<https://nube.siap.gob.mx/index.php/s/Qv6klxhXkDv2uma>>.
- [6] Tomek, W. G. and Robinson, K. L. (2003): *Agricultural product prices*. 4 Ed. NY: Cornell University Press. Ithaca and London.

Capítulo 11

pp 139-155

EVOLUCIÓN DE LAS TIC, DESARROLLO DE LA INDUSTRIA 4.0 Y SU APLICACIÓN EN EL SECTOR TURÍSTICO

M. Andreina Moros Ochoa¹, Paula Viviana Robayo Acuña² y Gilmer Yovani Castro Nieto³

¹CESA Colegio de Estudios Superiores de Administración.

²Universidad Nacional de Colombia. Docente investigadora, Fundación Universitaria Konrad Lorenz.

³Pontificia Universidad Javeriana (Bogotá, Colombia)

RESUMEN

Con el desarrollo de las Tecnologías de la Información y la Comunicación (en adelante TIC) se ha revolucionado a la sociedad y los sectores económicos, razón por la cual el objetivo de este trabajo es analizar la evolución de las TIC y el desarrollo de la Sociedad de la Información (en adelante SI) y posteriormente cómo ha sido la aplicación las TIC y del desarrollo de la Industria 4.0 en el sector turístico. Se aplicó el método deductivo de carácter cualitativo y descriptivo, en el cual se realizó revisión bibliométrica de la base de datos Scopus, y se analizó a través de la herramienta Biblioshyne del programa estadístico R y la función Bibliometrix, la producción científica que se ha publicado en los últimos años, lo que permitió evidenciar el desarrollo de la tecnología y su aplicación en el turismo, representa grandes retos para todos los actores que intervienen en este sector productivo.

Palabras clave: Tecnologías de la Información y la Comunicación- TIC; Sociedad de la Información; Internet; Industria 4.0; Sector turístico.

ABSTRACT

With the development of Information and Communication Technologies (hereinafter ICT), society and economic sectors have been revolutionized, which is why the objective of this work is to analyze the evolution of ICT and the development of the Information Society (hereinafter IS) and subsequently how ICT and the development of Industry 4.0 have been applied in the tourism sector. The qualitative and descriptive deductive method was applied, in which a bibliometric revision of the Scopus database was carried out, and the scientific production that has been published in the last few years was analysed through the Biblioshyne tool of the R statistical program and the Bibliometrix function, which allowed for evidence of the development of the technology and its application in tourism, represents great challenges for all the actors that intervene in this productive sector.

Keywords: ICT; Society of the Information; Internet; Industry 4.0.; Tourism sector.

INTRODUCCIÓN

Las TIC han revolucionado los sectores donde se desarrolla el ser humano. Por esta razón, el objetivo de este trabajo es analizar el avance de las TIC, el desarrollo de SI, la innovación y la industria 4.0 en el sector turístico. Con este trabajo se pretende contestar la siguiente pregunta de investigación: ¿Cuál es la incidencia de las TIC y el desarrollo de la industria 4.0 en el sector turístico?

Para contestar la pregunta de investigación, se aplica una metodología de revisión de literatura que permite resumir el avance de las TIC (desde la aparición de internet como factor de desarrollo de la SI) a nivel mundial y después, mediante una aproximación bibliográfica analizar el avance de las TIC, y la industria 4.0 en el sector turístico. Este trabajo permitió concluir que la tecnología avanza rápidamente y que es necesario introducirlas en todos los sectores productivos (especialmente en el sector turístico), lo que representa grandes retos para las empresas y los colaboradores que intervienen en el proceso, en busca de ofrecer calidad de servicio.

Así, este trabajo se encuentra dividido en las siguientes secciones: marco teórico que comprende los conceptos básicos: las TIC y el desarrollo de la SI, la innovación y la industria 4.0 en el sector turístico; metodología que hace referencia a las fases metodológicas propuestas; resultados y, finalmente, las conclusiones.

REFERENTE TEÓRICO

En esta sección se analizan los principales conceptos que componen el tema de investigación.

Las Tecnologías de Información y Comunicación (TIC) y el desarrollo de la Sociedad de la Información (SI)

Para contextualizar la evolución de la SI, es necesario hacer referencia al concepto de TIC, el cual, a pesar de ser muy amplio, se resumirán los aspectos más importantes desde sus orígenes.

El origen del concepto TIC está en la “tecnología”, así a inicios del siglo XXI uno de los estudiosos del tema, Castells (2001, pág 60) la define como “el uso del conocimiento científico para especificar modos de hacer cosas de una manera reproducible”. Para ese momento el autor incluye definiciones como microelectrónica, informática en general y optoelectrónica, que eran conceptos muy novedosos para la época.

Posteriormente, Cornella (2003) opina que las TI pasaron de ser sólo instrumentos de automatización de procesos, a ser instrumentos que les permiten a las empresas realizar más cosas de nuevas maneras.

El avance de las TIC se podría remontar a los primeros avances tecnológicos que datan de inicio del siglo XIX, momento en el que se dio origen a la revolución tecnológica, pero en este trabajo sólo se considerará lo referido a los avances tecnológicos desde la aparición de Internet. Sin embargo, es importante recordar brevemente la fecha en que salieron a la luz importantes inventos de aplicación masiva, que cambiaron el proceso de comunicación a distancia y que se pueden considerar la base de Internet, los cuales fueron: Telégrafo (en 1837 por Samuel Morse); Teléfono (en 1876 por Alexander Gram Bell); Radio (en 1898 por Guillermo Marconi); Televisión (en 1884 Paul Nipkow patenta su disco de exploración lumínica y en 1926 Jhon Logie Baird lo mejora y realiza su primera transmisión de imágenes en 1928).

Internet permitió mejorar la calidad de vida, acortando distancias y rompiendo fronteras, permitiendo el manejo de grandes cantidades de información en un lapso de tiempo que antes era casi inimaginable, lo cual agiliza la comunicación entre las personas, dando origen a las llamadas TIC. En la Tabla 1 se resumen la evolución de las TIC hasta el siglo XX en la que coinciden los autores de esa época Carballar (1994), Negroponte (1995) y Ramonet (2001).

Tabla 1. Evolución de las TIC hasta el siglo XX

AÑO	AUTOR	INVENTO
1946	Mauchly y Eckert	Electronic Numerical Integrator and Calculator (ENIAC)
1951	Mauchly y Eckert	Versión Comercial UNIVAC-1
1947	Bardeen, Brattain y Shockley	Transistor
1957	Jhon Kilby y Bob Noyce	Circuito integrado
1958	Sperry Rand	Ordenador mainframe
1969	Bolt Beranek y Newman (BBN)	Red de ordenadores ARPANET. Posteriormente Internet
1971	Ted Off Ray Tomlinson	Ordenador en un chip Origen de Internet y del correo electrónico (a través de la arroba)
1975	Ed Roberts	Calculadoras con microprocesador, base para el diseño de Apple
1981	IBM	Personal Computer PC
1984	Apple	Macintosh Apple
1989	Tim Berners-Lee	Creación de HTTP y HTML. Origen del World Wide Web
1994	Marc Andreessen	Creación del Netscape Navigator
1995	Microsoft	Internet Explorer
1990/2000		Digitalización: TV digital y CD de audio Redes Digitales de Servicios Integrados Banda Ancha (RDSI-BA) MODEM, CDROM

Fuente: Moros (2011).

Se observa que la tecnología en el campo electrónico avanzó aceleradamente en las últimas décadas del siglo XX.

Este proceso, narrado brevemente, logró que Internet se convirtiera en un sistema mediático, en la medida en que pasa a ser un sistema integrado que revolucionó igualmente la televisión, el cine, la edición, la música, los videojuegos, la información, el deporte, la mensajería electrónica, entre otras muchas cosas. Es evidente que, después de la aparición de internet, en las últimas dos décadas los avances en TIC han sido constantes y revolucionarias. En el Gráfico 1 se resume su clasificación.

Como se ha apreciado hasta ahora, la aparición de las TIC y su uso masivo han implicado diferentes etapas en la sociedad moderna. Algunos autores en la primera década de este siglo coincidieron en que gracias a las TIC la sociedad tradicional experimentó una transformación fundamental, en la cual la fabricación de productos tangibles poco a poco puede ser superada por los productos intangibles, como los servicios en los que el manejo y análisis de la información dieron origen a la SI (Castells, 2001; Cornella, 2002,2003; Pérez, 2004).

Tomando en consideración estas ideas, no se debe obviar el hecho de que el desarrollo y penetración de las TIC en la sociedad no se ha dado en las mismas proporciones en todas las zonas geográficas debido a diferentes factores socioeconómicos. Ante esta situación de desigualdad de condiciones entre los países y regiones, es importante tener en cuenta que un desarrollo de la SI mal enfocado o desequilibrado, puede producir grandes “gaps” entre los sectores de la población que no tienen fácil acceso a los avances tecnológicos, produciendo el efecto contrario al que se busca con el uso de las TIC, lo que ocasiona un crecimiento de las diferencias sociales y económicas, y dando origen las llamadas “Brecha Digital” y “Brecha cognitiva” (Grisales, 2011).

Gráfico 1. Clasificación de TIC



Elaboración propia a partir de Cano y Baena (2015)

Esta desigualdad, ha impulsado la enorme competencia a nivel mundial por un buen posicionamiento en el uso de Internet, lo que ha llevado a las compañías operadoras y prestadoras del servicio a rediseñar sus planes de acción, creando alianzas estratégicas, fusiones y multiplicación de redes de manera escalonada. En la Tabla 2 que se presenta a continuación, se puede observar los datos sobre el uso de internet a nivel mundial.

Tabla 2. Usuarios de internet en el mundo

LUGAR	POBLACION ESTIMADA 2019	USUARIOS (Ultima data) 2019	% DE USUARIOS EN EL MUNDO	% USUARIOS POR CONTINENTE
América		2019		
Total Norte América	366.496.802	327.568.628	4.8	89.4 (Norteamérica)
Estados Unidos	329.093.110	292.822.868		
Total Centro América	181.886.939	112.668.952	2.5	72.8 (Centroamérica)
México	132.328.035	88.000.000		
Total Sur América	428.240.515	306.349.946	7.4	48.7 (Suramérica)
Brasil	210.867.954	149.057.635		
Colombia	49,464,683	31,275,567		
Argentina	44.688.864	41.586.960		
Chile	18.197.209	14.108.392		
Total Caribe	44.191.318	20.986.379	0.5	4.6 (Suramérica)

Total América	1.015.592.658	783.909.293	13.3	
Europa				
Unión Europea	510.381.379	461.255.831	11,3	
España	46.441.049	42.961.230		9.3 (Unión Europea)
Resto de Europa	318.791.628	266.304.031		
Total Europa	829.173.007	727.559.862	16	
Asia				
China	1.415.045.928	772.000.000		37.1 (Asia)
Japón	126.874.745	118.626.672		5.2 (Asia)
Hong Kong	7.490.776	6.698.252		0.3 (Asia)
India	1.368.737.513	560.000.000		24.3 (Asia)
Resto de Asia	1.323.823.828	843.144.935		
Total Asia	4.241.972.790	2.300.469.859	55	
TOTAL MUNDO	7.716.223.209	4.536.248.808	100	

Fuente: Elaboración propia con datos de la página www.internetworldstats.com

Estos datos muestran diferencias en los números con respecto a los usuarios de internet. Norte América representa el 4,8% del mundo y el mayor porcentaje lo representa USA con un 89,4%; Sur América representa el 7,4% del mundo y el mayor porcentaje lo representa Brasil con un 48,7%, Colombia representa el 11,6%; Europa representa el 16% del mundo, dentro de Europa la UE representa el 11,3% y Asia representa el 55% del mundo y el mayor porcentaje corresponde a China con un 37,1% (de Asia). De esta manera se puede evidenciar que el número de usuarios en el mundo es muy diferente, se infiere que algunos de estos porcentajes se deben al número de habitantes por continente, por lo tanto, hay que seguir trabajando en el tema.

Es evidente que dentro de cada continente y cada país la penetración de internet no se ha dado de la misma forma. Aun así, sería impensable para los nativos digitales (definidos por García et al (2007) como los individuos que han crecido inmersos en la tecnología digital) concebir el mundo sin internet, para los inmigrantes digitales (definidos por el mismo autor, como aquellos que están en proceso de adaptación y cambios de hábitos empujados por esta generación), todos estos avances han supuesto una nueva forma de hacer las cosas y un aprendizaje continuo.

La mayoría de las TIC son utilizadas por los nativos digitales de manera autónoma y autodidacta, es decir, están en su quehacer diario y las ejecutan con total normalidad. Esta situación ha obligado al resto de las personas a convertirse en inmigrantes digitales adoptando la utilización de estas herramientas mediante un proceso de adaptación y aprendizaje, en una evolución constante de la llamada SI.

En esta competencia y búsqueda de posicionamiento en el mercado de internet han surgido las redes sociales (social media), como lo es el caso de Facebook que se convirtió en la mayor comunidad virtual existente en el mundo y además de forma gratuita. A través de esta red millones de usuarios pueden estar comunicados sin fronteras e intercambiar gran cantidad de información en tiempo real.

De esta manera se puede observar que las TIC han transformado las comunicaciones del ser humano, al igual que el procesamiento de la información y acelerando el proceso de la globalización de la economía, debido a la constante modificación de los productos y servicios ofrecidos en busca de una mejor calidad para ser competitivos en el mercado. Todo ello ha originado conceptos como: Social Media o redes sociales (Instagram, Twitter, LinkedIn, Snapchat, Qzone, Skype, Badoo, BaiduTieba, WeChat, Pinterest, Flickr, Spotify, Vine, entre otras); Mobile Computing (computación móvil); Transformación Digital; Internet of Things (IoT); Inteligencia artificial (IA); Realidad aumentada; Realidad virtual; e-Commerce, Business Intelligence (BI); Big Data y Analytics, entre otros, lo cual origina innovación y da paso a la llamada Industria 4.0.

Así, una vez resumido el avance de las TIC y el desarrollo de la SI, para continuar con el aporte se analizará desde la teoría su aplicación para el desarrollo del turismo, que es la industria de servicios más grande del mundo, considerándose uno de los sectores de mayor importancia en muchos países y uno de los principales motores de la economía mundial (Devi, 2015).

De otra parte, el sector turístico es sin duda uno de los más importantes para la economía de los países, según la Organización Mundial de Turismo, representa el 10% al PIB. Además, las llegadas internacionales a América aumentaron en 7 millones en 2016, hasta alcanzar casi los 200 millones y la demanda de viajes permaneció fuerte en la región. Por subregiones, América del Sur lideró con un 7% más, aportando Colombia un 11%, en 2018 llegaron 4.281.851 visitantes no residentes, lo que representa un incremento de 7,7% respecto al año 2017 (OMT, 2018). En el mismo año, el turismo generó 1.974.000 empleos, lo que significa 71.000 nuevos empleos respecto a lo registrado durante el año 2017, representando un incremento del 3,74%, lo que representa que un 4% de la empleabilidad del país está en el sector turístico (CITUR, 2019).

A continuación, se indica la metodología que se consideró más apropiada para cumplir con el objetivo de este trabajo.

METODOLOGÍA

El presente estudio forma parte de una investigación deductiva de carácter cualitativo y descriptivo, en el cual se lleva a cabo el análisis de un conjunto de 845 referencias obtenidas de la base de datos Scopus® el pasado 6 de diciembre de 2019 con la ecuación de búsqueda (*TITLE-ABS-KEY ("information and communication technologies" OR "ICT") AND TITLE-ABS-KEY (tourism)*).

Para el análisis se incluyeron las referencias publicadas e indexadas, teniendo en cuenta para cada documento el tipo de fuente, el título, la revista, la fecha de publicación, los nombres y afiliaciones de los autores y por supuesto, el resumen. En cuanto al periodo analizado, este se asumió desde la primera publicación registrada en la base de datos utilizada, con respecto al tema y las palabras claves usadas en la búsqueda, es decir, desde el año 1985 hasta diciembre de 2019, inclusive 7 artículos aceptados y publicados en revistas para el año 2020. A continuación, en la Tabla 3 se muestra la información básica del conjunto de datos:

Tabla 3. Información básica del conjunto de datos

<i>Ítem</i>	<i>Valor</i>
<i>Documentos</i>	<i>845</i>
<i>Fuentes (Revistas, libros, etc.)</i>	<i>503</i>
<i>Palabras clave</i>	<i>2281</i>
<i>Periodo</i>	<i>1985-2020</i>
<i>Promedio de citas por documento</i>	<i>8502</i>
<i>Autores</i>	<i>1884</i>

Fuente: elaboración propia (2019).

La información de los estudios recopilados se analizó utilizando la herramienta Biblioshiny¹, que se refiere a una interfaz para el paquete R bibliometrix. El archivo del conjunto de datos fue cargado en la Biblioshiny, el cual extrae toda la información bibliométrica del conjunto de datos y proporciona información relacionada con el análisis descriptivo, la estructura conceptual, la estructura intelectual y la estructura social (Aria & Cuccurullo, 2017).

Posteriormente, se realizó un análisis de cada uno de los resúmenes de las referencias encontradas con el fin de identificar cuál ha sido la incidencia de las TIC como factor fundamental de la SI y la industria 4.0 específicamente en el sector turístico, obteniendo los resultados que se describen en el siguiente apartado.

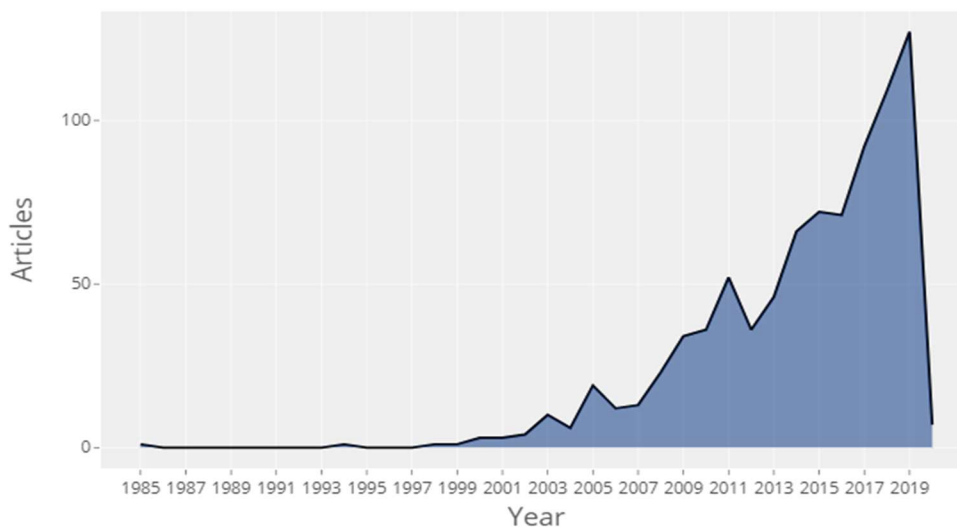
RESULTADOS

Las TIC en el sector turístico

Inicialmente se analiza el avance de las TIC en el sector turístico, el cual tampoco se ha dado de la misma forma y al mismo tiempo. Parra y Santana (2015) comentan que el tema a nivel académico se ha generado interés de investigación a nivel nacional e internacional sobre la relación Turismo-TIC, las cuales han sido agrupadas así: las TIC y su relación con la demanda turística, las TIC y su relación con la innovación turística y las TIC y su relación con los nuevos modelos de negocios turísticos.

Lo anterior se ve reflejado en la bibliometría realizada de la cual se puede observar que la temática abordada (TIC y turismo), ha tenido un importante crecimiento de interés en la literatura con una tasa anual de crecimiento del 8,45%. Tal como se puede visualizar en el Gráfico 2, pasó de un artículo en 1985 a 19 en el 2005 y 72 en 2015. De otra parte, en los últimos 10 años, es decir, a partir del 2010 hasta la fecha, se han publicado un total de 714 referencias, lo que implica que se ha publicado el 84% de la totalidad de artículos desde 1985 de acuerdo con la base de datos de Scopus® y Bibliometrix.

Gráfico 2. Documentos publicados por año sobre TIC y turismo

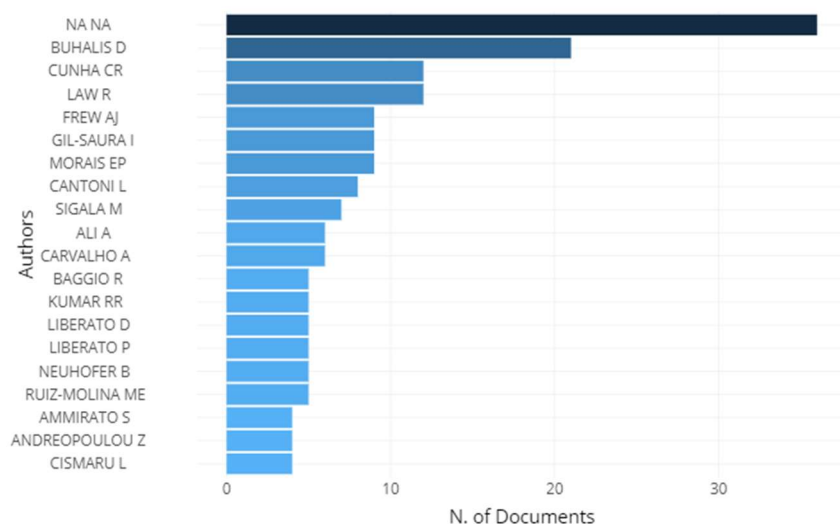


Fuente: elaboración propia, a partir de datos obtenidos en Scopus® y R Bibliometrix

¹ <http://www.bibliometrix.org/biblioshiny.html>

En el Gráfico 3 se pueden visualizar los autores con más documentos escritos sobre TIC y su relación con el sector turístico, entre ellos destaca: Dimitrios Buhalis quien es Catedrático de Turismo y Subdirector del Centro Internacional de Investigación en Turismo y Hostelería (ICTHR) en la Escuela de Gestión de Servicios de la Universidad de Bournemouth (Reino Unido) y que desde 1993 ha venido realizando diversos estudios sobre turismo y la tecnología, con un gran impacto para el sector. La mayoría de los otros autores, iniciaron sus trabajos a partir del 2010 a 2012, como: Cunha, CR., Morais, Ep., Ali, A. y Carvalho, A., y entre los más recientes, el autor Liberato.

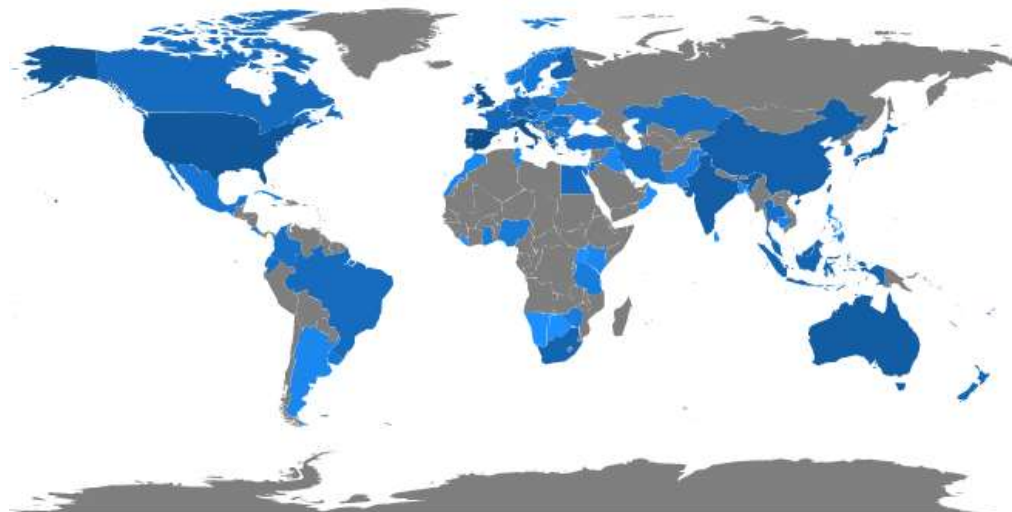
Gráfico 3. Número de documentos por autor sobre TIC y turismo



Fuente: elaboración propia, a partir de datos obtenidos en Scopus® y R Bibliometrix

En la ilustración 1, se puede observar que, entre los países con mayor número de artículos sobre el tema, se encuentran España con 171 documentos, Italia con 116, Reino Unido con 116, Portugal con 84 y Estados Unidos con 73, aportando el 66,3% del total de referencias publicadas a la fecha según Scopus® (845).

Ilustración 1. Publicaciones por países



Fuente: elaboración propia, a partir de datos obtenidos en Scopus® y R Bibliometrix

Tabla 4. Estudios relevantes relacionados con TIC en el sector turístico

Investigadores	Propósitos / Resultados del estudio
García, M. J. (2011)	Opina que las TIC han promovido rápidos cambios en el turismo en aspectos como: la estrategia de branding (o de marca), la influencia de las TIC's en la oferta turística (agencias de viaje, tour operadores, etc.) que ven internet como un nuevo canal de comercialización-venta y desde el punto de vista del consumidor para obtener variedad de información sobre: destinos, precios, ofertas, características del servicio, etc. influyendo, a su vez, en su decisión de compra, sus motivaciones, su comportamiento, sus gustos, etc.
Albacete y Herrera (2012)	Sintetizan las ventajas de las TIC para la optimización de la eficiencia y reducción de los costes operativos; nuevas relaciones con los clientes que mejoran el servicio, la calidad y la satisfacción; perfeccionamiento de las relaciones con los proveedores; acceso a mercados más globales, y planificación de los servicios turísticos más eficiente y ajustada a las necesidades de la demanda.
Manero, C. B., García-González, M., García-Uceda, M. E., & Grijalba, J. M. M. (2012)	Desde la revisión de la bibliografía opinan que las TIC han generado cambios en el sistema de distribución en el sector turístico e identifica tres criterios de dichos cambios: la estructura del sistema, la posición de poder de los operadores y los procesos productivos y productos-servicios turísticos.
Bulchand y Melián, (2014)	Clasificaron un catálogo de TIC hoteleras en cuatro grupos, según su tipo de contribución a la mejora del rendimiento hotelero: incremento de la productividad de los trabajadores, mejora de los servicios hoteleros, mejora de la comercialización y marketing e incremento de los ingresos.
Guevara, P. (2015)	Comenta que en el sector turístico que se ha producido una implantación de TIC en cuanto a hardware e Internet, pero no hay gran avance respecto a la implantación de software relacionado con los procesos de gestión y comercialización, por lo cual hay que seguir trabajando en ello.
Caro, Luque y Zayas (2015)	Comentan que estamos ante un turista con capacidades, habilidades y costumbres relacionadas con la web 2.0 que ha pasado de ser un mero lector a ser lector/escritor con gran influencia en las marcas turísticas.
Cerezo, A. y Guevara, A. (2015)	Resaltan un fuerte nexo entre TIC y turismo que se puede percibir mediante indicadores como el volumen de comercio electrónico en turismo, el uso de las tecnologías móviles por los viajeros durante todas las fases del proceso de compra y consumo y el grado de implantación de las TIC en las empresas turísticas (tecnologías móviles, ciudades y destinos inteligentes, medios sociales, etc.)
Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015)	Hacen referencia a que las agencias de viajes en línea (OTA), siguen dominando la planificación de viajes. Sin embargo, destacan cuatro tendencias: el uso de Internet para la planificación de viajes está más que saturado; existen contrastes y similitudes entre generaciones; las redes sociales, los dispositivos móviles y los canales de emergencia admiten nuevos comportamientos. Así, estas tendencias certifican la necesidad de adoptar la tecnología en todos los procesos del sector turístico.

<p>Baidal, J. A. I., Monzonís, F. J. S., & Sánchez, D. G. (2016)</p>	<p>Aconsejan repensar las fórmulas tradicionales de gestión turística local y evolucionar hacia estrategias más innovadoras como el enfoque en destinos inteligentes que se refieren a destinos que se adaptan al cambio y que aprenden continuamente a mejorar su sostenibilidad y su competitividad, con capacidad para influir o, incluso, para crear su propio futuro. Los destinos inteligentes deben contemplar entre otras cosas: sistemas de sensorización de espacios turísticos muy aparentes pero de impacto limitado, páginas web de última generación, asistente virtual en web, códigos QR, sistemas de accesibilidad, sistemas de geolocalización, apps, pantallas táctiles, wi-fi de acceso libre en espacios públicos, sistemas de reservas en línea, sistemas de inteligencia empresarial en destino, video guías, encuestas en línea a empresas del destino, mostradores táctiles, realidad aumentada.</p>
<p>Aparicio (2017)</p>	<p>Asegura que el turista o usuario hace uso de la web desde el momento que inicia la búsqueda del destino, luego pasa por la elección y preparación del viaje, así como la terminación del mismo. También hace referencia a que una persona que viaja entre uno o más países o diferentes localidades en su país, principalmente cuando es un viajero independiente demandan información acerca de los destinos, productos y servicios turísticos ofrecidos en el lugar que visitará y menciona como relevante las diferentes formas de promocionar un servicio intangible de acuerdo al nuevo paradigma de consumo y contratación del servicio, por esta razón, las TICS aportan de manera primordial a los sistemas de información ya que vienen transformando la manera para que los turistas planeen y realicen sus viajes y concluye que los canales de distribución turísticos generan cambios en los productos desarrollando promociones y ofertas ya que para el viajero: Lo puede hacer en cualquier lugar y momento, ahorra tiempo, lo puede hacer en cualquier dispositivo y recibe millones de ofertas de diferentes tipos de productos y hace referencia a las siguientes opciones:</p> <ul style="list-style-type: none"> • Portal turístico: Ofrece diferentes tipos de servicios como reservas o compra de productos. • Wireless Application System: Son dispositivos inalámbricos para el despliegue de contenidos básicamente son orientados a darle independencia al consumidor. • Computadoras de bolsillo y Telefonía celular: Información instantánea para el desarrollo continuo en las redes de comunicación. • Web TV: Este tipo de mecanismo es utilizado para acceder al internet de manera inalámbrica y es una aplicación que se presenta en los hoteles como parte de uno de los servicios que se ofrecen en las habitaciones permitiendo al viajero tener acceso sin necesidad de que se tenga que conectar de una computadora o tenerse que desplazar a un centro de negocios. • Televisión Interactiva: Este sistema igual al de Web TV es un aparato conectado al televisor ya sea por celular u otro medio con la diferencia que este sistema está diseñado para presentar programas sobre destinos turísticos en los programas de televisión de una forma convencional donde el usuario es libre de navegar por los destinos o actividades que más le interesen.

Rahmani, K., Gnoth, J., & Mather, D. (2017)	Realizaron una investigación en Nueva Zelanda para demostrar el comportamiento de los turistas en la web 2.0 a través de la cual, documentan cada vez más sus historias lo que les ayuda a mejorar sus experiencias. Para este estudio aplicaron la herramienta de Lingüística de Corpus (CL) para extraer significados ocultos en conjuntos de datos tanto en estudios exploratorios y demostrar cómo se producen los datos en línea y qué los caracteriza. El rastreo de estas historias les permite a los gerentes entender la mente y los sentimientos de los turistas sobre las experiencias que ofrecen, lo que les permite diseñar estrategias sobre cómo posicionar y administrar un destino.
Yen, C. L. A., & Tang, C. H. H. (2019)	Analizan los efectos del rendimiento de los atributos del hotel a través del "eWOM" (Word of Mouth), es decir "el boca a boca" electrónico, que influye en las actitudes de los consumidores, percepciones, intenciones y la elección de productos en el contexto de planificación turística. Se refiere a la importancia que tienen los comentarios sobre los hoteles en las redes.
Buhalis, D., Harwood, T., Bogicevic, V., (...), Beldona, S., Hofacker, C. (2019)	Proporciona ejemplos de las industrias del turismo y la hostelería como un contexto de gestión de servicios dependiente de la información, expresando que las perturbaciones tecnológicas como la Internet de las cosas (IoT) y los dispositivos autónomos, las capacidades analíticas mejoradas (inteligencia artificial) y los medios de comunicación enriquecidos (realidad virtual y aumentada) están creando entornos inteligentes que están transformando las estructuras, los procesos y las prácticas de la industria.
Fan, D.X.F., Hsu, C.H.C., Lin, B. (2020)	En la era de la conectividad, el desarrollo de la tecnología de la información y las comunicaciones ha cambiado enormemente la forma en que las personas viajan, se comportan y aprecian la experiencia.

Fuente: elaboración propia (2019), a partir de los diferentes autores descritos en la tabla.

En estos trabajos surgen los conceptos que dan lugar al desarrollo de la Industria 4.0, la cual se refiere a la digitalización de los procesos de fabricación que afectan a todos los ámbitos económicos: comercio, industria, turismo y administración pública.

La industria 4.0 incluye avances en la tecnología como el internet de las cosas (IoT), 5G, cloud computing, robótica, realidad virtual que están transformando productos, procesos y modelos de negocios en todos los sectores, creando finalmente nuevos patrones industriales a medida que cambian las cadenas de valor globales (Marini, 2016; Sung, 2018). Por su parte, Zhong, Xu, Klotz & Newman (2017) y Alcácer & Cruz-Machado (2019) coinciden que los recursos típicos se convierten en objetos inteligentes para que puedan sentir, actuar y comportarse en un entorno inteligente mediante un proceso interoperable, integrado, adaptado, optimizado, orientado a servicios que se correlaciona con algoritmos, Big Data (BD), internet de las cosas (IoT), los sistemas cibernéticos (CPS), Cloud Computing (CC) y la tecnología de la información y las comunicaciones (TIC) que se utilizan para permitir la fabricación inteligente, la interoperabilidad autónoma, la agilidad, la flexibilidad, la toma de decisiones, la eficiencia o la reducción de costos.

Nick, Pongracz & Radacs (2018), comentan que la última generación de sistemas de innovación abierta resulta en cambios revolucionarios en casi todas las industrias, así como en otras áreas de la vida. Desde la industria automotriz hasta el sector energético y el turismo lo cuales considera recién llegados, utilizan herramientas de la innovación disruptiva que, con velocidades anteriores sin precedentes, transforman su poder y generan las

llamadas “Smart Cities” o ciudades inteligentes y su enfoque se basa en poner las últimas tecnologías para servir a la sostenibilidad social, económica y ecológica de los habitantes y empresas de la ciudad y afirman que la industria 4.0 de manera directa o indirecta, tendría una influencia fundamental en las ciudades inteligentes y su entorno.

Peraković, Periša, & Zorić (2019) opinan que la tecnología de la información y la comunicación sigue teniendo un impacto positivo en muchas etapas del entorno de fabricación. La inteligencia está a punto de ser compartida desde el principio hasta el final de la cadena de suministro. El Internet de las cosas (IoT) está agregando inteligencia a los puntos finales, los grandes datos se están convirtiendo en la nueva forma de administrar un negocio y la Cloud Computing (CC) se está convirtiendo en un nuevo centro de datos.

De esta manera, se observa en todo el análisis realizado, la coincidencia en los conceptos que dan lugar a la Industria 4.0, por lo que se considera importante hacer referencia a los más importantes:

- Big Data (BD) es una tendencia de los últimos tiempos y se refiere grandes cantidades de datos y el análisis que se realiza de esa información. Sin embargo, aunque su nombre así lo indica, en algunas ocasiones, no necesariamente deben ser “grandes” cantidades de datos. Al respecto, Berman (2013) opina que los principios del Big Data son preparar, compartir y analizar información, lo que contribuye a evitar errores en proyectos que requieren grandes cantidades de datos. Por su parte, Aguilar (2016) define Big Data como la confluencia de una multitud de tendencias tecnológicas que venían madurando desde la primera década del siglo y que han irrumpido en organizaciones y empresas, en particular, y en la sociedad, en general. Además, la mayoría de los autores coinciden en las principales características de Big Data entre las que se encuentran: Volumen de los datos; Velocidad con la que se generan los datos; Variedad de los datos; Valor y Veracidad de los datos.
- Internet de las cosas (IoT), al respecto Kshetri (2016) afirma que se refiere a la red de objetos físicos "o cosas" (p.ej., máquinas, dispositivos y aplicaciones, animales o gente) integrado con la electrónica, software y sensores, que poseen identificadores únicos, con la capacidad de transferir datos a través del web, con apenas intervención del ser humano.
- Machine Learning utiliza modelos de analítica y algoritmos que aprenden de los datos de forma iterativa, permitiendo a los sistemas encontrar información oculta sin que se hayan programado explícitamente los lugares donde buscarla”. Al respecto, Sun et al., (2019) formulan un modelo de pronóstico de la demanda turística a destinos populares en China mediante el uso del aprendizaje automático y los índices de búsqueda en Internet y comparó su desempeño de previsión con los resultados de búsqueda generados por Google y Baidu, respectivamente.
- eWOM se refiere al boca-oído electrónico (Word of Mouth) que genera un comportamiento donde los consumidores leen las opiniones de otros usuarios para informarse sobre las características de un producto o servicio recibiendo un mayor control en las valoraciones que se obtienen online. Algunas empresas ofrecen en sus propias páginas web un espacio virtual en el que los usuarios pueden publicar sus opiniones e intercambiar experiencias acerca de los productos y servicios ofertados con lo cual, se incorporan la confianza en el vendedor online o la utilidad de la página web como variables importantes para explicar los comportamientos de compra de los consumidores en el entorno virtual y un antecedente de la intención de recompra. Es decir, se convierte en una ventaja de internet como herramienta que pone al alcance de los consumidores una serie de plataformas virtuales donde compartir opiniones y valoraciones con otros usuarios y acceder a un gran volumen de datos (Vallejo, J. M., Redondo, Y. P., & Acerete, A, 2015).

Al realizar una búsqueda de los trabajos realizados en los últimos años en las principales bases de datos (Scopus, ScienceDirect), específicamente sobre industria 4.0 en el sector turístico se pueden resaltar los siguientes estudios:

Leea, Chen, Leea, Xua, Lia & Zhao (2015), hacen referencia a que la industria 4.0 pone más énfasis en la creación de redes e interconexiones entre los activos existentes mediante el Internet de las cosas (IoT). En su estudio intentaron representar conceptos de inno-servicio con los elementos comunes en diferentes categorías basadas en la tecnología IoT y los conceptos de sistemas cibernéticos (CPS) e identificaron seis categorías para las funciones principales del distrito de una mega ciudad: (1) turismo MICE (reuniones, incentivos, conferencias, exposiciones) y alojamiento, (2) alimentos, bebidas y supermercados, (3) sectores culturales y creativos y librerías, (4) muebles y electrodomésticos, (5) ropa, accesorios y cosméticos, (6) cine, recreación y deportes. Además, basándose en los 54 casos de servicios inteligentes en el entorno global de las seis clasificaciones, las soluciones más valiosas de la base de casos se adaptan y reutilizan para satisfacer las necesidades y la situación actual. Por lo tanto, seis temas de compras inteligentes están diseñados como servicios basados en CPS entre las seis clasificaciones anteriores, que incluyen: (1) Servicios de integración de la calle creativa y cultural, (2) Servicios de comida y restaurantes inteligentes, (3) Maquillaje futuro y rápidos servicios de moda, (4) Servicios de marketing experiencial de vida inteligente, (5) Servicios innovadores basados en entretenimiento, (6) Servicios integrados de turismo.

Syahril, Bruri & Minghat (2018) hicieron un estudio sobre el impacto en la demanda de puestos de trabajo que tiene la revolución industrial 4.0 y concluyeron que la ocupación se transformará mediante la automatización y digitalización, pero no se eliminará. Las áreas que tienen riesgo de ser automatizadas en la Educación y Formación Técnica Vocacional (TVET) en Indonesia son la administración, el turismo y la salud. Lo que hace pensar que no todo son ventajas también se genera un impacto social que es importante tener en cuenta para futuras investigaciones.

De esta manera, la investigación permite dar respuesta a la pregunta que se planteó inicialmente: ¿Cuál es la incidencia de las TIC y el desarrollo de la industria 4.0, en el sector turístico? con la afirmación que la incidencia de las TIC ha sido determinante en los últimos años, lo que ha dado paso al desarrollo de la industria 4.0 en el sector turístico y supone un reto para todos los actores de la actividad, quienes deben adoptar estas tecnologías para seguir en un mercado cada vez más competitivo.

CONCLUSIONES

Después de este análisis se puede afirmar que las TIC han influido en todos los sectores donde se desenvuelve el ser humano, lo que ha permitido el desarrollo de la llamada “Sociedad de la Información”, la cual tiene cada vez más tecnología en sus actividades diarias, por lo tanto, está más informada y es más exigente con lo que espera obtener. Sin embargo, se debe tener en cuenta que la evolución de la Sociedad de la Información no ha sido uniforme en todo el mundo, lo que ha dado origen a las llamadas “Brecha digital” y “Brecha cognitiva”.

Las TIC juega un papel vital en las industrias turísticas (hoteles, líneas aéreas, transporte terrestre, restaurantes, servicios de guías, servicios de traducción, etc.) no sólo para ser competitivas sino para sobrevivir en un mercado cada vez más agresivo. Existen innovaciones tanto en los procesos como en los servicios en sí, por ello, es necesario seguir culturizando a los clientes sobre la utilización de todos los elementos digitales, entre los que se pueden mencionar: páginas web con asistente virtual, códigos QR, sistemas de accesibilidad, sistemas de geolocalización, apps, pantallas táctiles, Wi-fi de acceso libre en espacios públicos, sistemas de reservas en línea, sistemas de inteligencia empresarial en destino, video guías, encuestas en línea a empresas del destino, mostradores táctiles, realidad aumentada, internet de las cosas, inteligencia artificial.

Álvarez (2016) refuerza estos hallazgos señalando que las inversiones en TIC mejoran los resultados de innovación, independientemente de las medidas de innovación empleadas y confirman que éstas son más importantes en la industria de servicios. Además, señala que existen factores determinantes para el aprovechamiento efectivo de las TIC, y de ahí el de las innovaciones: el tamaño de las empresas, el capital humano, la experiencia con procesos online y externalidades potenciales.

Además, mediante la tecnología se recopilan grandes cantidades de información, lo que da origen a un Big Data. Esta información tiene gran valor porque permite conocer el comportamiento y gusto de los consumidores. Por otro lado, se origina la huella digital que dejan los clientes en sus búsquedas y contrataciones en la web, la técnica utilizada para la extracción de estos datos y analizar el eWOM se denomina Webscraping. Los datos extraídos de las webs tienen un gran potencial porque permiten realizar investigaciones cualitativas y cuantitativas acerca de las tendencias del mercado. Sin embargo, es necesario analizar toda esa gran cantidad de información que se genera a través de la tecnología, por lo cual otro reto es aprender a hacerlo de forma eficaz y eficiente mediante “Analytics”, con herramientas novedosas que permiten tomar decisiones acertadas basadas en información en tiempo real.

Por todo lo anterior, se considera que la Industria 4.0 específicamente en el sector turístico, permite la innovación en: los procesos, la oferta de servicios y la comunicación con los clientes. Este trabajo permitió identificar todos los retos a los que se enfrenta el sector turístico en busca de la calidad. No sólo debe adoptar innovación en sus actividades, sino estar al día en los cambios tecnológicos que se desarrollan a gran velocidad. Por lo tanto, es importante continuar investigando en un sector dinámico y con grandes posibilidades como lo es el turístico.

REFERENCIAS

- [1] Aguilar, L. J. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.
- [2] Albacete, C. A. y Herrera, M. (2012). *Las tecnologías de la información y la comunicación en el sector turístico: Una aproximación al uso de las redes sociales*. Madrid: Centro de Estudios Ramón Areces.
- [3] Alcácer, V., & Cruz-Machado, V. (2019). Scanning the industry 4.0: A literature review on technologies for manufacturing systems. *Engineering Science and Technology, an International Journal*.
- [4] Álvarez, R. (2016). *The Impact of R&D and ICT Investment on Innovation and Productivity in Chilean Firms*. Inter-American Development Bank.
- [5] Aparicio, C. G. M. (2017). Una nueva forma de hacer negocios en el turismo: los buscadores en la web. *International journal of scientific management and tourism*, 3(2), 101-120.
- [6] Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- [7] Baidal, J. A. I., Monzonís, F. J. S., & Sánchez, D. G. (2016). Gestión turística y tecnologías de la información y la comunicación (TIC): El nuevo enfoque de los destinos inteligentes. *Documents d'anàlisi geogràfica*, 62(2), 327-346.
- [8] Berman, Jules (2013). *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Elsevier Books Reference.
- [9] Buhalis, D., Harwood, T., Bogicevic, V., (...), Beldona, S., Hofacker, C. (2019). Technological disruptions in services: lessons from tourism and hospitality. *Journal of Service Management* 30(4), pp. 484-506. <https://doi.org/10.1108/JOSM-12-2018-0398>
- [10] Bulchand, J., y Melián, S. (2014). Catálogo de TIC en hoteles y su contribución al rendimiento organizativo. En *TURITEC 2014*
- [11] Cano, José A y Baena, José J (2015) Tendencias en el uso de las tecnologías de información y comunicación para la negociación internacional. *Estudios Gerenciales* Vol 3, Pág 335–346.
- [12] Carballar, José (1994). *Internet. El mundo en sus manos*. Editorial Ra-ma, Madrid.
- [13] Castells, Manuel (2001). *La era de la información: Economía, sociedad y cultura*. Vol.3, 2da edición, Alianza Editorial, Madrid.

- [14] Caro, J. L., Luque, A., & Zayas, B. (2015). Nuevas tecnologías para la interpretación y promoción de los recursos turísticos culturales. *PASOS Revista de turismo y Patrimonio Cultural*, 13(4), 931 – 945. Disponible en: <https://www.redalyc.org/articulo.oa?id=881/88140288014>
- [15] Cerezo, A. y Guevara, A. (2015). El papel estratégico de las tecnologías de la información y las comunicaciones en el turismo. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 2(2), 52 – 69. Disponible en: <http://uajournals.com/ojs/index.php/ijisebc/article/view/109/97>
- [16] Centro de información turística de Colombia (CITUR). (2019). Estadísticas Nacionales. Ministerio de Industria, Comercio y Turismo. Disponible en: <http://www.citur.gov.co/estadisticas> (08/09/2019)
- [17] Cornella, Alfons (2002). *Infonomía.com. La gestión inteligente de la información en las organizaciones*. Ediciones Deusto, Bilbao.
- [18] Cornella, Alfons (2003). *Hacia la empresa en red. Gestión 2000*, Barcelona.
- [19] Devi, S. V. (2015). “A study on tourist satisfaction and their preference to act as responsible tourist for sustainable tourism”. *International Journal on Global Business Management & Research*, 3(2), 65-71.
- [20] Fan, D.X.F.; Hsu, C.H.C.; Lin, B. (2020). Tourists’ experiential value co-creation through online social contacts: Customer-dominant logic perspective. *Journal of Business Research* 108, pp. 163-173
- [21] García, M. J. (2011). Las TIC en el sector turístico: branding, oferta y demanda. *Estudios Turísticos*, (187), 151-166.
- [22] García, F., Portillo, J., Romo, J., & Benito, M. (2007, September). Nativos digitales y modelos de aprendizaje. In SPDECE.
- [23] Grisales, Nancy (2011). La brecha cognitiva: una realidad educativa que va más allá de la brecha digital entre las instituciones urbanas y rurales de Manizales. *Revista Latinoamericana de Estudios Educativos (Colombia)*, vol. 7, núm. 2, julio-diciembre, 2011, pp.37-56. Disponible en: <http://www.redalyc.org/pdf/1341/134125454004.pdf>
- [24] Guevara Plaza, A. (2015). El turismo ante el avance de las TIC. In *20 retos para el turismo en España* (pp. 75–90).
- [25] Internet World Stast (2018) Disponible en: <https://www.internetworldstats.com/stats.htm> [Consulta día: 07-11-2019].
- [26] Kshetri, Nir (2016). The economics of the Internet of Things in the Global South. *Third World Quarterly*, DOI: 10.1080/01436597.2016.1191942.
- [27] Leea, Chen, Leea, Xua, Lia & Zhao (2015). Accelerating Retail-Innovation Design for Smart Services via Foresight Approach and Case-Based Design. *Transdisciplinary Engineering: A Paradigm Shift*, 813.
- [28] Manero, C. B., García-González, M., García-Uceda, M. E., & Grijalba, J. M. M. (2012). Modelización de los cambios en el sistema de distribución del sector turístico debidos a la incorporación de las tecnologías. *Cuadernos de Economía y Dirección de la Empresa*, 15(3), 117-129.
- [29] Marini, D. (2016). «Industria 4.0»: A first criticañ reflection. *L'industria*, (3), 383-386.
- [30] Moros, M (2011). Impacto de las TIC en la calidad de servicio del sector financiero. Caso de estudio: Banco Sofitasa-Venezuela. Tesis Doctoral. Universidad de Deusto, España.
- [31] Negroponte, Nicholas (1995). *El mundo digital. El futuro que ha llegado*. Ediciones B, Barcelona.
- [32] Nick, G., Pongracz, F., & Radacs, E. (2018). Interpretation of disruptive Innovation in the era of Smart Cities of the Fourth Industrial Revolution. *Deturope-The Central European Journal of Regional Development and Tourism*, 10(1), 53-70.
- [33] Organización Mundial de Turismo (2018). ¿Por qué el turismo? Disponible en: <http://www2.unwto.org/es>

- [34] Parra López, E., y Santana Talavera, A. (2015). Tecnologías de la información y la comunicación en turismo. 20 años de la actividad turística en España (pp. 279–293).
- [35] Peraković, D., Periša, M., & Zorić, P. (2019, June). Challenges and Issues of ICT in Industry 4.0. In *Design, Simulation, Manufacturing: The Innovation Exchange* (pp. 259-269). Springer, Cham.
- [36] Pérez, Carlota (2004). *Revoluciones tecnológicas y capital financiero. La dinámica de las grandes burbujas financieras y las épocas de bonanza*. Siglo XXI editores, S.A, México.
- [37] Rahmani, K., Gnoth, J., & Mather, D. (2017). Tourists' Participation on Web 2.0: A Corpus Linguistic Analysis of Experiences. *Journal of Travel Research*, 0047287517732425.
- [38] Ramonet, I. (2001). Impacto de la globalización en los países en desarrollo. *Revista Memoria*, 143.
- [39] Sun, S., Wei, Y., Tsui, K.L. & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, 70, 1-10
- [40] Sung, T. K. (2018). Industry 4.0: A Korea perspective. *Technological forecasting and social change*, 132, 40-45.
- [41] Syahril Chairani, V., Bruri, M., & Minghat, D. (2018). Literature Review: Some of TVET Area Will Be Eliminated Due to Industrial Revolution 4.0, is That True?. *International Journal of Engineering & Technology*, 7(4.33), 161-165. doi:<http://dx.doi.org/10.14419/ijet.v7i4.33.23523>
- [42] Vallejo, J. M., Redondo, Y. P., & Acerete, A. U. (2015). Las características del boca-oido electrónico y su influencia en la intención de recompra online. *Revista Europea de Dirección y Economía de la Empresa*, 24(2), 61-75.
- [43] Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, 22, 244-249.
- [44] Yen, C. L. A., & Tang, C. H. H. (2019). The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviors. *International Journal of Hospitality Management*, 76, 9-18.
- [45] Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, 3(5), 616-630.