
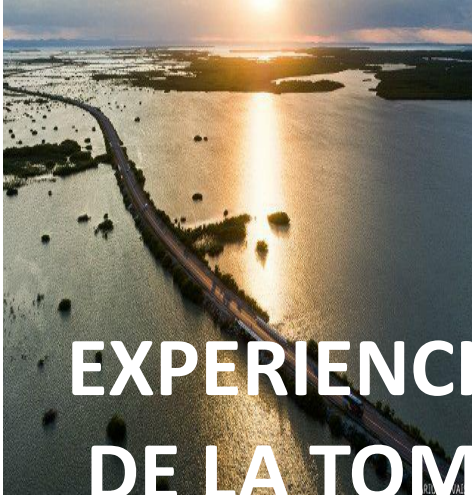


---



# EXPERIENCIAS EN LA MODELACIÓN DE LA TOMA DE DECISIONES EN LA SALUD HUMANA, MEDIO AMBIENTE Y DESARROLLO HUMANO

Editores  
Carlos N. Bouza Herrera  
Claudio Castro López  
José F. García Rodríguez  
María del M. Rueda García



TOMA DE DECISIONES EN LA  
SALUD HUMANA, MEDIO  
AMBIENTE  
Y  
DESARROLLO HUMANO

Tomo 2

ISBN 978-84-608-4246-0

Editores

Carlos N. Bouza Herrera, Universidad de La Habana, Cuba.

Claudio, R. Castro López, Universidad Veracruzana, México.

José F. García Rodríguez, Universidad Juárez Autónoma de Tabasco, México.

Maria del M. Rueda García, Universidad de Granada, España.



2016

---

© Universidad de Granada, España  
Universidad de La Habana, Cuba.  
Universidad Juárez Autónoma de Tabasco, México.  
Universidad Veracruzana, México.

Editores

Carlos N. Bouza Herrera.

Claudio, R. Castro López.

José F. García Rodríguez.

María del M. Rueda García.

ISBN 978-84-608-4246-0

Este libro ha sido subvencionado por el grupo de investigación FQM365 de la Junta de Andalucía, el Proyecto MTM2012-35650 del Ministerio de Educación de España, el Proyecto CITMA Modelos Matemáticos para el Estudio de Medio Ambiente, Salud y Desarrollo Humano, el Grupo GRASHUMEDIA de MATCOM-UH y RIDECA.

---

## Prólogo

La elaboración y diseño de modelos cuantitativos para el medio ambiente necesariamente toman en consideración los aspectos de la salud y del desarrollo social. Este desarrollo incluye la elaboración de modelos teóricos cuya aplicación en estos campos es importante. La aceleración de los procesos de cambio y creciente interrelación, a fin de responder a las necesidades del desarrollo de la economía y la sociedad a corto, mediano y largo plazo, suscitan un incentivo para el desarrollo de nuevos conceptos y problemáticas que conllevan la necesidad de cuantificar. Con ello se estimula constantemente el desarrollo de modelos matemáticos, aplicados o importante en el desarrollo de la investigación enmarcada en la RIDECA (Red Iberoamericana de Estudios Cuantitativos Aplicados) en el 2015.

En la actualidad hay una gran vinculación entre los temas de investigación que hemos estado desarrollando en el marco de esta red. La correspondencia de las investigaciones, con la proyección del desarrollo económico y social en los que se involucran los investigadores incorporados a RIDECA, tiene como pilares fundamentales, la integración, la racionalidad y el fortalecimiento de las actividades conjuntas. Estas se reflejan en este libro que es el segundo de una serie de 3. Esperamos que estos trabajos den al lector una valoración de esta cooperación, que va más allá de la organización de eventos científicos, al trascenderles con estas obras de autor. El análisis de estas contribuciones permitirá valorar lo que hemos y tendremos que avanzar en el aprendizaje, de cómo desarrollar nuestras potencialidades y colaborar en equipos interdisciplinarios. Nótese que muchos de estos trabajos incorporan a médicos, biólogos economistas, químicos además de los esperados especialistas de matemática e informática. .

Las Universidades constituyen actores fundamentales y motores principales de la labor investigativa aquí presentada.

Por otra parte, el objetivo de la cooperación ha sido el fomentar la generación, asimilación y aplicación de conocimientos y tecnologías; así como estimular y propiciar el aprendizaje y la innovación, vinculados con la problemática medio ambiental.

Los trabajos aquí presentados se enmarcan en diferentes campos de la ciencia, la tecnología y la innovación, con especial énfasis en la modelación matemática del manejo de los recursos humanos y naturales, el desarrollo de softwares y aplicaciones en investigaciones sociales, económicas y humanísticas. Todas ellas se enrumban, de una u otra forma, en la línea de la preservación del medio ambiente.

En el área pura del medio ambiente se tratan problemas de contaminación, biodiversidad y otros problemas relacionados. En los problemas médicos se trata de problemáticas asociadas a enfermedades de amplio impacto social como el Sida y los trastornos cardiovasculares, entre otros. Por otra parte la temática del desarrollo humano se enfoca hacia varios problemas de índole económica de México y Panamá, entre ellos la pobreza, el manejo de recursos y aportes de índole teórico que son de inmediata aplicación en diversos sectores. En este aspecto se puede señalar la existencia de contribuciones que tratan de los problemas de pérdida y falseo de la información.

Esperamos que esta nueva obra complemente la de la anterior editada con este fin bajo los auspicios de RIDECA

Los editores

# ÍNDICE DE CAPÍTULOS

<b>MEDICINA</b>			
<b>Capítulo</b>	<b>Título</b>	<b>Autores</b>	<b>Páginas</b>
1	SISTEMA EXPERTO PARA LA DETECCIÓN DE DENGUE CLÁSICO Y HEMORRÁGICO	López González, E. C., N. A. Pérez, L. R. Marcial Castillo, M. Rivera Martínez & L. Sandoval Solís	8
2	APLICACIÓN DE LA REGRESIÓN MULTINIVEL EN LA DETERMINACIÓN DE FACTORES ASOCIADOS A LA ADIPOSIDAD CENTRAL EN PACIENTES CON VIH/SIDA	León Sánchez M.A., & E. M. Linares Guerra	16
3	MODELO PARA LA EVALUACIÓN DEL RIESGO CARDIOVASCULAR PARA CUBA	Lahera Pérez, G.	27
4	ANÁLISIS ESPACIAL DE MODELOS PARA LA ESTIMACIÓN DE RIESGO RELATIVO BAJO EL PARADIGMA BAYESIANO	Raíces Cruz, I., V. Sistachs Vega, A. Tuero Iglesias*	41
<b>MEDIO AMBIENTE</b>			
5	MÉTODO PARA DETERMINAR LAS FASES MINERALES DEL CLÍNKER Y SU INFLUENCIA EN REDUCIR LOS DAÑOS AL MEDIO AMBIENTE	Alvarez Bravo C.A., M. E. Cortés Cortés & M. Moreira	51
6	MODELOS DE REGRESIÓN LINEAL PARA DIÁMETRO Y ÁREA FOLIAR EN <i>FRAXINUS UHDEI</i> (WENZ.) EN RESPUESTA A CONCENTRACIONES DE PARTÍCULAS ATMOSFÉRICAS EN BOSQUES URBANOS.	Sánchez Landero, L.A., S. Hernández González, M. Y. Zavaleta Sánchez & M. Á. López López	60
7	THE MULTINOMIAL SUPER POPULATION MODEL FOR BIODIVERSITY INDEXES	Bouza Herrera C., D. Covarrubias Melgar & M.Li	71

Capítulo	Título	Autores	Páginas
	DESARROLLO HUMANO		
8	ALGORITMO Y PSEUDOCÓDIGO PARA CALCULAR EL VECTOR ELÉCTRICO INTEGRAL DE LA ACTIVIDAD VENTRICULAR CARDIACA EN EQUINO	Camejo Zamora, M:C:, A. Pompa Núñez & Y. Vázquez Alfonso	79
9	PLATAFORMAS DE APRENDIZAJE: UN ENTORNO PARA LA MEJORA DE LA DOCENCIA EN EL GRADO DE ENFERMERÍA	González-Jiménez E. & M. Á. Montero-Alonso	85
10	INCLUSIÓN FINANCIERA A TRAVES DEL CRÉDITO DE DESARROLLO HUMANO COMO HERRAMIENTA PARA EL FOMENTO DE CAPACIDADES Y EMPRENDIMIENTO	Salas E. E., V. del Cisne Espinoza, A. Samaniego & V. A. Martínez-Fernández	92
11	A RANKED SET SAMPLING ALTERNATIVE FOR SINGLE IMPUTATION IN A HOT DECK ENVIRONMENT	Allende S., C. Bouza & I. Jeelani	106
12	REVISIÓN SISTEMÁTICA DE LOS MÉTODOS DE IMPUTACIÓN DE DATOS FALTANTES	Viada C, C. Bouza, J. Ballesteros, M. Fors, M. Robaina & R. Uranga	115
13	APLICACIÓN DEL MODELO DE HORVITZ DE RESPUESTA ALEATORIZADA AL CONSUMO DE DROGAS	Cobo, B.	132
14	PANAMÁ: MODELACIÓN DE LA DEMANDA AGREGADA COMO ARISTA DEL DESARROLLO HUMANO	Bracho, R.	139
15	CARACTERIZACIÓN DE LOS NIVELES DE DESARROLLO ECONÓMICO DE LAS REGIONES DEL CENTRO SUR DE MÉXICO	Caamal Cauich I., J. F.García Rodríguez, V. Gricel-Pat Fernández, D. Martínez Luis & F. Jerónimo Ascencio	151
16	EL DISEÑO DEL RECORRIDO PARA LA RECOGIDA DE RESIDUOS SOLIDOS. CASO DE ESTUDIO: MUNICIPIO PLAZA DE LA REVOLUCION	<i>Campbell Amos J. &amp; S. Allende Alonso</i>	168
17	APROXIMACIÓN A LA SITUACIÓN DE POBREZA Y DESIGUALDAD EN MÉXICO. FACTORES CAUSALES Y ALTERNATIVAS DE SOLUCIÓN	García Rodríguez, J. F., I. Caamal Cauich, R. G. B. Primo & O.Priego Hernández*	180
18	FACTORES QUE IMPACTAN EL CRECIMIENTO ECONÓMICO: UN ESTUDIO COMPARATIVO ENTRE PAÍSES INDUSTRIALIZADOS Y ECONOMÍAS EMERGENTES USANDO MODELACIÓN MULTINIVEL	Ortiz Betancourt, I., M. M. Ojeda Ramírez & P. Tapia Blásquez	202

# ÍNDICE DE AUTORES

<b>Autores</b>	<b>Institución</b>
Allende/Alonso, Sira M.	Universidad de La Habana, Cuba
Alvarez Bravo, Carlos A.o	Universidad de Cienfuegos, Cuba
B. Primo, Rilton Goncalo	Centro de Estudios por la Amistad de Latinoamérica, Asia y África (CEALA), Brasil.
Ballesteros, Javier	University of the Basque Country, UPV/EHU Dept of Neuroscience, España
Bouza-Herrera, Carlos N.	Universidad de La Habana, Cuba
Bracho, René	Universidad de Panamá, Panamá, Ciudad de Panamá, Pánama
Caamal Cauich, Ignacio	Universidad Autónoma Chapingo. México
Camejo Zamora, Maikel	Universidad Agraria de la Habana, Cuba
Campbell Amos, Joanna	Universidad de La Habana, Cuba
Carmen Viada	Department of Clinical Information Management, Center of Molecular Immunology, Cuba
Cobo, B.	Departamento de Estadística e Investigación Operativa, Universidad de Granada, España
Cortés Cortés, Manuel E	Universidad de Cienfuegos, Cuba
Covarrubias Melgar, Dante	Universidad Autónoma de Guerrero, México
Espinoza, Viviana del Cisne	Dpto. Ciencias Empresariales., Universidad Técnica Particular de Loja. Ecuador
Fors, Martha	Department of Clinical Information Management, Center of Molecular Immunology, Cuba
García Rodríguez, José F.	Universidad Juárez Autónoma de Tabasco, México
González-Jiménez, Emilio	Departamento de Enfermería. Facultad de Enfermería (Campus de Melilla). Universidad de Granada.
Hernández González, Sergio	Facultad de Estadística e Informática, Universidad Veracruzana, México
Jeelani, Iqbal	Sher-e-Kashmir University of Agricultural Sciences and Technology, Agricultural Statistics, India
Jerónimo Ascencio, Felipe	Universidad Autónoma Chapingo. México
León Sánchez, María A.	Universidad de Pinar del Río, Cuba
Li, Marja	Software development division, Institute of Computing Training, China
Linares Guerra, Elisa M.	Universidad de Pinar del Río, Cuba

<b>Autores</b>	<b>Institución</b>
López López, Miguel Á.	Colegio de Postgraduados, Campus Montecillo, México
Martínez Luis, David	Universidad Autónoma del Carmen, Campeche. México
Martínez-Fernández, Valentín. A.	Dpto. Ciencias Empresariales, Universidad de A Coruña, España.
Montero-Alonso, Miguel Á.	Departamento de Estadística e I.O. Facultad de Ciencias Sociales (Campus de Melilla). Universidad de Granada, España
Moreira, Mario	Cementos Cienfuegos S.A, Cuba
Pat Fernández, Verna	Universidad Autónoma Chapingo. México
Pompa Núñez, Alberto	Universidad Agraria de la Habana, Cuba
Priego Hernández, Oscar	Universidad Juárez Autónoma de Tabasco, México
Robaina, Mayteé	Centro Nacional Coordinador de Ensayos Clínicos, Cuba
Salas, Eulalia. E.	Dpto. Ciencias Empresariales., Universidad Técnica Particular de Loja. Ecuador
Samaniego, Aurora	Dpto. Ciencias Empresariales., Universidad Técnica Particular de Loja. Ecuador
Sánchez Landero, Luz A.	Facultad de Estadística e Informática, Universidad Veracruzana & Colegio de Postgraduados, Campus Montecillo, México
Uranga, Rolando	Centro Nacional Coordinador de Ensayos Clínicos, Cuba
Vázquez Alfonso, Yasser	Universidad Agraria de la Habana, Cuba
Zavaleta Sánchez, María Y.	Facultad de Estadística e Informática, Universidad Veracruzana, México
López González, Elena Cristian	Universidad Autónoma de Puebla, México
Pérez, Nayeli Angel	Puebla, Puebla, México
Marcial Castillo, Luis René	Universidad Autónoma de Puebla, México
Rivera Martínez, Marcela	Puebla, Puebla, México
Sandoval Solís, Lourdes	Universidad Autónoma de Puebla, México
Lahera Pérez, Germán	Departamento de Matemática Aplicada de la Facultad de Matemática y Computación, Universidad de la Habana, cUBA
Raíces Cruz , Ivette	Centro de Ingeniería Genética y Biotecnología, Cuba
Tuero Iglesias, Angela	Centro de Ingeniería Genética y Biotecnología, Cuba
Sistachs Vega, Vivian	Universidad de La Habana, Cuba



## **SISTEMA EXPERTO PARA LA DETECCIÓN DE DENGUE CLÁSICO Y HEMORRÁGICO**

Elena Cristian López González, Nayeli Angel Pérez, Luis René Marcial Castillo,  
Marcela Rivera Martínez & Lourdes Sandoval Solís  
Benemérita Universidad Autónoma de Puebla  
Puebla, Puebla, México

### **ABSTRACT:**

Expert systems are the most common type of artificial intelligence systems for clinical routine, the goal of this work is to provide a tool to the general population, which is able to provide a more accurate diagnosis for infectious disease dengue, because it can be confused with a number of viral diseases. The codification of knowledge is done by rules. The expert system is the result of the experience of human experts and clinical manuals of Secretaría de Salud and private laboratories, which help in strengthening the clinical decision making. The developed system is implemented in the programming language SWI-Prolog and the results are validated with the monitoring of human experts.

**KEYWORDS:** Dengue fever, dengue hemorrhagic fever, expert system, diagnostic.

### **RESUMEN:**

Los sistemas expertos son el tipo más común de los sistemas de inteligencia artificial para la rutina clínica, el objetivo de éste trabajo es proveer de una herramienta a la población en general, la cual sea capaz de brindar un diagnóstico más acertado para la enfermedad infecciosa del dengue, debido a que puede confundirse con un sin número de enfermedades virales. La codificación del conocimiento se realiza mediante reglas. El sistema experto es el resultado de la experiencia de expertos humanos y manuales clínicos de la secretaría de salud pública y laboratorios privados, que ayudan en el fortalecimiento de la toma de decisiones clínicas. El sistema desarrollado está implementado en el lenguaje de programación Swi-Prolog y los resultados se validan con la supervisión de los expertos humanos.

## **1. INTRODUCCIÓN**

Más del 70% de la carga de morbilidad por la enfermedad del dengue se concentra en Asia Sudoriental y en el Pacífico Occidental. En los últimos años, la incidencia y la gravedad de la enfermedad han aumentado rápidamente en Latinoamérica y el Caribe. En las regiones de África y el Mediterráneo Oriental también se han registrado más brotes de dengue en los últimos 10 años. En 2010 se notificó la transmisión indígena del dengue en dos países de Europa. Al aumento mundial del dengue han contribuido la urbanización, los movimientos rápidos de personas y bienes, las condiciones climáticas favorables y la falta de personal capacitado. En el caso de México se dio el rebrote de esta enfermedad en 1967 (Badii, Landeros, Cerna, Abreu, 2007).

En México la prevención y control del dengue ha tomado relevancia, considerándosele uno de los programas prioritarios en salud pública. Las estrategias básicas de prevención están enfocadas al control del vector y éstas se fundamentan en la bionomía del mosquito (Thrión, 2010).

### **Sistema experto**

Un sistema experto es capaz de procesar y memorizar información, aprender y razonar en situaciones deterministas e inciertas, imitar el razonamiento de un experto humano para tomar decisiones apropiadas en una tarea particular y explicar por qué se han tomado tales decisiones (Castillo, Gutiérrez and Hadia, 1997). Este tipo de sistema va de la mano con un experto humano, no se pretende sustituir a éste sino que los sistemas sirvan de apoyo. Un problema debe ser resuelto mediante un sistema experto cuando (Prasad and Sagar, 2011):

- a) El problema puede resolverse solamente por un conocimiento experto en vez de usar algoritmos particulares.
- b) Se tiene acceso a un experto que puede dar forma a los conocimientos necesarios para resolver el problema, por lo que la intervención de este experto dará al sistema la experiencia que necesita.

Los sistemas basados en el conocimiento o sistemas expertos son el tipo más común de los sistemas de inteligencia artificial usados en la rutina clínica, contienen los conocimientos médicos, por

---

lo general, sobre una tarea muy específica y son capaces de razonar con los datos de paciente individuales para llegar a conclusiones razonadas.

Aunque hay diversas variaciones, el conocimiento dentro de un sistema experto típicamente es representado en forma de un conjunto de reglas. Existen diferentes tareas clínicas en las que un sistema experto se puede aplicar como son:

1. Asistencia en el diagnóstico. Cuando el caso del paciente es complejo, raro o la persona que hace el diagnóstico no tiene experiencia, entonces, un sistema experto puede obtener un diagnóstico oportuno tomando en cuenta la base de datos del paciente.
2. Planificación y terapia crítica. El sistema experto es capaz de buscar para encontrar inconsistencias, errores y omisiones en un plan de tratamiento existente o se puede utilizar para formular un tratamiento basado en la condición específica del paciente.

### **Enfermedades vectoriales**

Las enfermedades vectoriales son padecimientos cuya transmisión se realiza mediante agentes vectoriales o vectores. Los vectores son insectos con el potencial de transmitir enfermedades a los humanos; para convertirse en portador un vector recibe el organismo patógeno de un portador infectado, es decir, es decir humano o animal, la transmisión en la mayoría de los casos se realiza mediante picaduras ya que el contagio se realiza mediante alimentación sanguínea. Estas enfermedades son más frecuentes en zonas tropicales y subtropicales y en lugares con problemas de acceso al agua potable y al saneamiento (Frenk, 2001).

Existen diferentes enfermedades transmitidas por vectores, entre las cuales se pueden encontrar:

- 1) Dengue. El dengue es una infección vírica transmitida por mosquitos. Causa síntomas gripales y en ocasiones evoluciona hasta convertirse en un cuadro potencialmente mortal llamado dengue grave.
- 2) Esquistosomiasis. Enfermedad parasitaria crónica causada por gusanos. Corren el riesgo de contraer la infección las personas que realizan actividades agrícolas, domésticas o recreativas que las exponen a aguas infestadas.
- 3) Filariasis linfática. Conocida generalmente como elefantiasis, la filariasis linfática puede alterar el sistema linfático y producir una hipertrofia anormal de algunas partes del cuerpo, dolor y discapacidad grave.
- 4) Fiebre hemorrágica de Crimea-Congo. Enfermedad muy extendida causada por un virus transmitido por garrapatas. Este virus causa graves brotes de fiebre hemorrágica viral, con una tasa de letalidad del 10%-40%.
- 5) Fiebre amarilla. Enfermedad vírica aguda, hemorrágica, transmitida por mosquitos infectados. El término "amarilla" alude a la ictericia que presentan algunos pacientes. El virus es endémico en las zonas tropicales de África y América Latina.
- 6) Enfermedad de Chagas (tripanosomiasis americana). Enfermedad potencialmente mortal causada por el parásito protozoo *Trypanosoma cruzi*. Se encuentra sobre todo en América Latina.
- 7) Paludismo. El paludismo es causado por un parásito denominado *Plasmodium* que se transmite a través de la picadura de mosquitos infectados. En el organismo humano, los parásitos se multiplican en el hígado y después infectan los glóbulos rojos.
- 8) Tripanosomiasis africana humana. La tripanosomiasis africana humana, también llamada enfermedad del sueño, es una parasitosis transmitida por la mosca tsetse.

La enfermedad con mayor crecimiento en el mundo es el dengue, cuya incidencia se ha multiplicado por 30 en los últimos 50 años. Dicha enfermedad se trasmite mediante la picadura de las hembras infectadas de mosquitos del género *Aedes*.

Hay cuatro serotipos de virus del dengue (DEN 1, DEN 2, DEN 3 y DEN 4) (Contreras, Villanueva, Salinas, Enríquez, Pérez, 2013). Los síntomas aparecen 3–14 días (promedio de 4–7 días) después de la picadura infectiva (Pizarro, 2009). El dengue es una enfermedad similar a la gripe que afecta a lactantes, niños pequeños y adultos.

---

No hay ningún tratamiento específico contra el dengue. El dengue grave es una complicación potencialmente mortal, pero su diagnóstico clínico precoz y una atención clínica cuidadosa por personal médico y de enfermería experimentada suele salvar la vida de los pacientes.

Este trabajo se enfoca en detectar la enfermedad del dengue clásico y dengue hemorrágico así como otras enfermedades virales: influenza, sarampión, varicela, salmonelosis y brucelosis, esperando en el futuro aplicarlo a pacientes reales así como su implantación en algún sistema de salud.

En la sección 2, se presentan los antecedentes del trabajo, en la sección 3, se muestra la representación del conocimiento, la sección 4 proporciona la implementación y pruebas realizadas, en la sección 5 se presentan las conclusiones y finalmente se listan las referencias.

## 2. ANTECEDENTES

Durante el 2013 el Centro Nacional de Vigilancia Epidemiológica y Control de enfermedades (CENAVECE), reportó un incremento en los casos de dengue en un 60.5 por ciento en comparación del año anterior. Entre los estados con mayor incidencia se encuentran Tabasco, Guerrero, Veracruz y Quintana Roo (Frenk J., 2015). Levantando la alerta en el 60% del territorio mexicano, ya que es el porcentaje del territorio que presenta las condiciones adecuadas para el crecimiento del mosquito transmisor del dengue. El dengue afecta actualmente a 112 países en los que se presentan más de 100 millones de casos de dengue clásico y medio millón de casos de fiebre hemorrágica, con varios miles de defunciones anuales (Frenk, 2015).

Las altas tasas de morbilidad y mortalidad, son capaces de bloquear las actividades económicas de ciudades y países así como colapsar los servicios de salud de cualquier región cuando se presentan epidemias de fiebre hemorrágica por esta enfermedad. Es por ello, que se decide realizar un sistema experto que ayude a la detección oportuna de los tipos de dengue mencionados anteriormente y que proporcione recomendaciones, de una forma fácil y sin usar grandes cantidades de recursos económicos.

## 3. REPRESENTACIÓN DEL CONOCIMIENTO

Un sistema experto puede considerarse como una máquina pensante la cual podrá resolver problemas de un área específica tal como un especialista humano lo haría, además de ser capaz de tomar decisiones más adecuadas y precisas las que serán compartidas con el humanos y/o otros sistemas expertos. Se debe tomar en cuenta que para una máquina es imposible tratar la información del mismo modo que lo hace un humano, es por eso, que la información transferida por el médico deberá ser previamente tratada y representada a una estructura entendible por el sistema experto. Para poder modelar el conocimiento a un nivel óptimo se utilizará el concepto de Sistema Basado en Reglas, que funciona con implicaciones de la forma SI-ENTONCES, las cuales formarán un sistema de inferencia en donde el "SI" representara condiciones o antecedentes y el "ENTONCES" indicará las consecuencias o restricciones del problema. Aplicando este concepto al sistema experto se obtiene un sistema basado en reglas mismo que consta de un conjunto de implicaciones ligadas entre sí por respuestas afirmativas y/o negativas. Tal que el "SI" es un conjunto  $S$  de síntomas del dengue clásico y/o hemorrágico y "Entonces" es el conjunto  $D$  de diagnósticos que son producidos por un uno más elementos del conjunto  $S$ . Una vez que se tienen todas las implicaciones, es necesario relacionarlas unas con otras mediante conectores booleanos para poder obtener un árbol de inferencias que será interpretado y utilizado por el sistema experto. El árbol de inferencias muestra una mayor flexibilidad de representación mediante módulos de código, los cuales forman parte de un gran motor de inferencia y diagnóstico de padecimientos, como el dengue.

A continuación se presenta la codificación en forma algorítmica del conocimiento aportado por los expertos humanos: Dra. Estela Rosete Quintana y Dr. José Evodio López Castillo que laboran en consultorios particulares en la ciudad de Puebla, Puebla y Perote, Veracruz respectivamente.

- 1) Si el paciente presenta escalofríos, ir a 2. En otro caso, el paciente no presenta síntomas suficientes para un diagnóstico, salir.

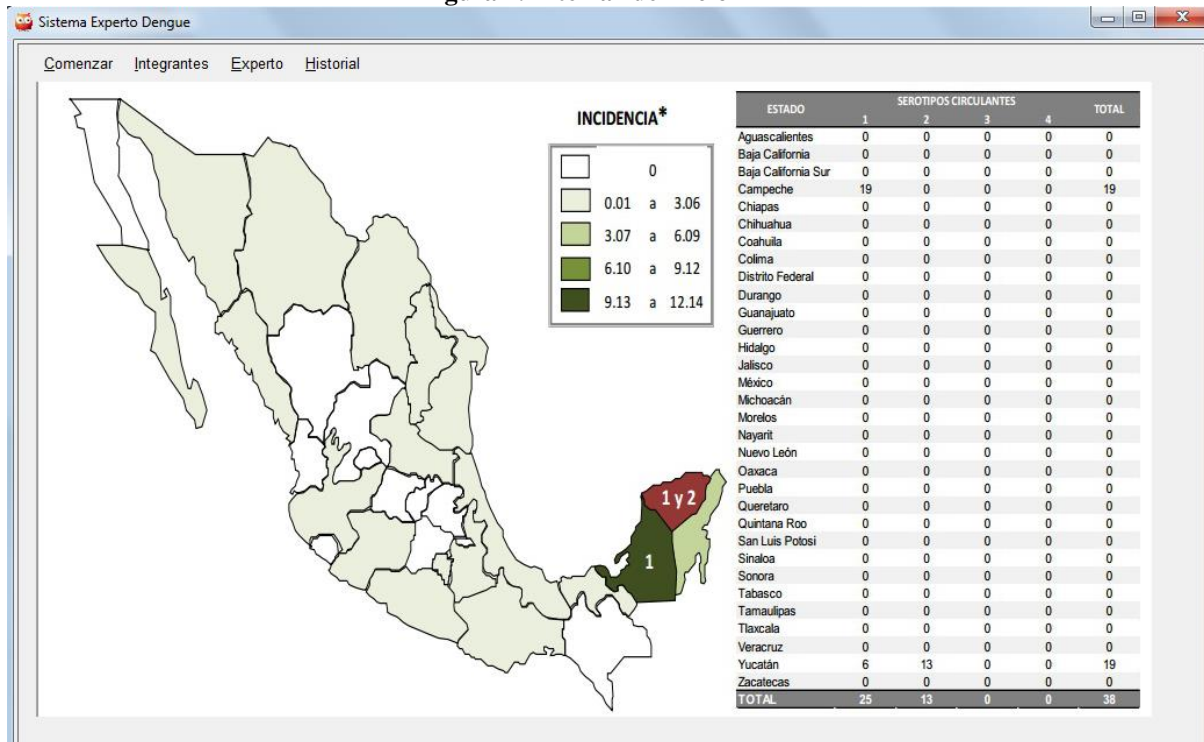
- 
- 2) Si el periodo de incubación es de un lapso de 3 a 4 días en el cual la persona tiembla de frío en periodos de 15 minutos a 1 hora, ir al paso 3. En otro caso ir al paso 4.
  - 3) Si el paciente presenta fiebre repentina mayor a los 39.5°C. El enfermo se pone muy débil, colorado, a veces se “pierde” y presenta síntomas como: dolor agudo de cabeza, dolor ocular, dolor de espalda, ir al paso 5. En otro caso ir al paso 6.
  - 4) Si el paciente presenta diarrea, entonces el paciente tiene Salmonelosis, Salir. En otro caso el paciente tiene Brucelosis, Salir.
  - 5) Si el paciente presenta erupción que inicia en tronco y se extiende a todo el cuerpo, ir al paso 7. En otro caso, ir al paso 8.
  - 6) Si el paciente tiene astenia (cansancio crónico), escalofríos, sudoración, dolor muscular, dolor en las articulaciones y dolor abdominal con constipación, el paciente tiene Influenza, Salir.
  - 7) Si el paciente tiene comezón y no presenta descamación como las enfermedades propias de la infancia, ir al paso 9. En otro caso, el paciente tiene Varicela, Salir.
  - 8) Si el paciente presenta rubor en frente y mejilla y palidez peribucal, la erupción es más intensa en pliegues cutáneos, axilas e ingles y sitios de presión, el paciente tiene Sarampión, Salir.
  - 9) Si el paciente presenta aumento en la temperatura de su piel, ir al paso 10. En otro caso, ir al paso 11.
  - 10) El paciente presenta síntomas de dengue clásico, Grupo A. Recomendaciones: Observación en unidad de salud y/o manejo ambulatorio. Tratamiento inmediato: Hidratación oral, analgésicos y antipiréticos, informar al paciente y familiares sobre los síntomas del Grupo C de los cuales deberá permanecer alerta. Realizar prueba del torniquete y seguir con análisis de los síntomas, ir al paso 12.
  - 11) Si el paciente presenta lesiones purpúreas en la piel que desaparecen en 3 días. Ir al paso 13.
  - 12) Si en el paciente observa lo siguiente:
    - a. Si el paciente presenta sangrados tras realizada la prueba del torniquete, ir al paso 14. En otro caso, ir al paso 15.
    - b. Si el paciente presenta al menos uno de estos síntomas: Dolor Abdominal intenso y sostenido, vómitos persistentes y frecuentes, descenso brusco de la temperatura, irritabilidad, somnolencia o ambos en forma alternante, derrames serosos, frialdad de la piel, ir al paso 18. En otro caso, ir al paso 15.
  - 13) El paciente tiene Rubeola, Salir.
  - 14) Realizar estudios: Hematocrítico, hemoglobina, plaquetas, si los resultados muestran alteración en base a los niveles comunes, ir al paso 16, En otro caso, ir al paso 17.
  - 15) Manejo Ambulatorio. Instruir a pacientes acerca de la vigilancia de sangrados, petequias (pequeñas manchas color púrpura en la piel, que no desaparecen con la presión) o signos de alarma para consultar inmediatamente al médico.
  - 16) El paciente presenta síntomas de Fiebre Hemorrágica de Dengue, Grupo B. Recomendaciones: Hospitalización con vigilancia estrecha. Tratamiento inmediato: tratamiento local compresivo si hay sangrado en mucosas, vigilar signos de alarma del Grupo C y signos de choque pertenecientes al Grupo D, vigilar ascitis o hidrotórax por ultrasonido, hematocritos y recuento plaquetarios seriados: buscar hemoconcentración. Seguir con el diagnóstico, ir al paso 12b.
  - 17) El paciente tiene dengue clásico con manifestaciones hemorrágicas. Seguir con el diagnóstico, ir al paso 12b.
  - 18) El paciente presenta signos de Dengue hemorrágico, Grupo C. Recomendaciones: Hospitalización para tratamiento rehidratante y vigilancia estrecha. Tratamiento inmediato: Infusiones intravenosas cristaloides, monitorización clínica y de laboratorio, iniciar oxigenoterapia, vigilar signos iniciales de choque, estrechamiento de la presión del pulso. Seguir con el diagnóstico, ir al paso 19.
  - 19) Si el paciente presenta signos de choque: Palidez, frialdad, sudoración, taquicardia, oliguria (disminución en la producción de orina), ir al paso 20, En otro caso continuar con las observaciones mientras dure la etapa de peligro, ir al paso 18.
  - 20) El paciente presenta signos de Dengue con choque, Grupo D. Recomendación: Hospitalización inmediata en terapia intensiva. Tratamiento inmediato: Aporte de soluciones cristaloides, oxigenoterapia, hacer prevención del edema pulmonar, vigilar hematemesis y hemorragia pulmonar.

#### 4. IMPLEMENTACIÓN Y PRUEBAS

La implementación computacional se desarrolló bajo el lenguaje de programación Swi-Prolog y la librería gráfica XPCE.

Swi-Prolog es el entorno de desarrollo del lenguaje lógico Prolog, este proyecto inicio en el año de 1987, su principal desarrollador fue Jan Wielemaker, dicho proyecto es popular en el desarrollo e implementación de soluciones que involucran inteligencia artificial. Swi-Prolog se encuentra vigente en su versión 7.1. Su valor está en que posee un rico conjunto de características, bibliotecas (incluyendo su propia biblioteca para GUI, XPCE), herramientas (incluyendo un IDE) y una documentación extensiva. SWI-Prolog funciona en las plataformas Unix, Windows y Macintosh.

**Figura 1. Interfaz de inicio**



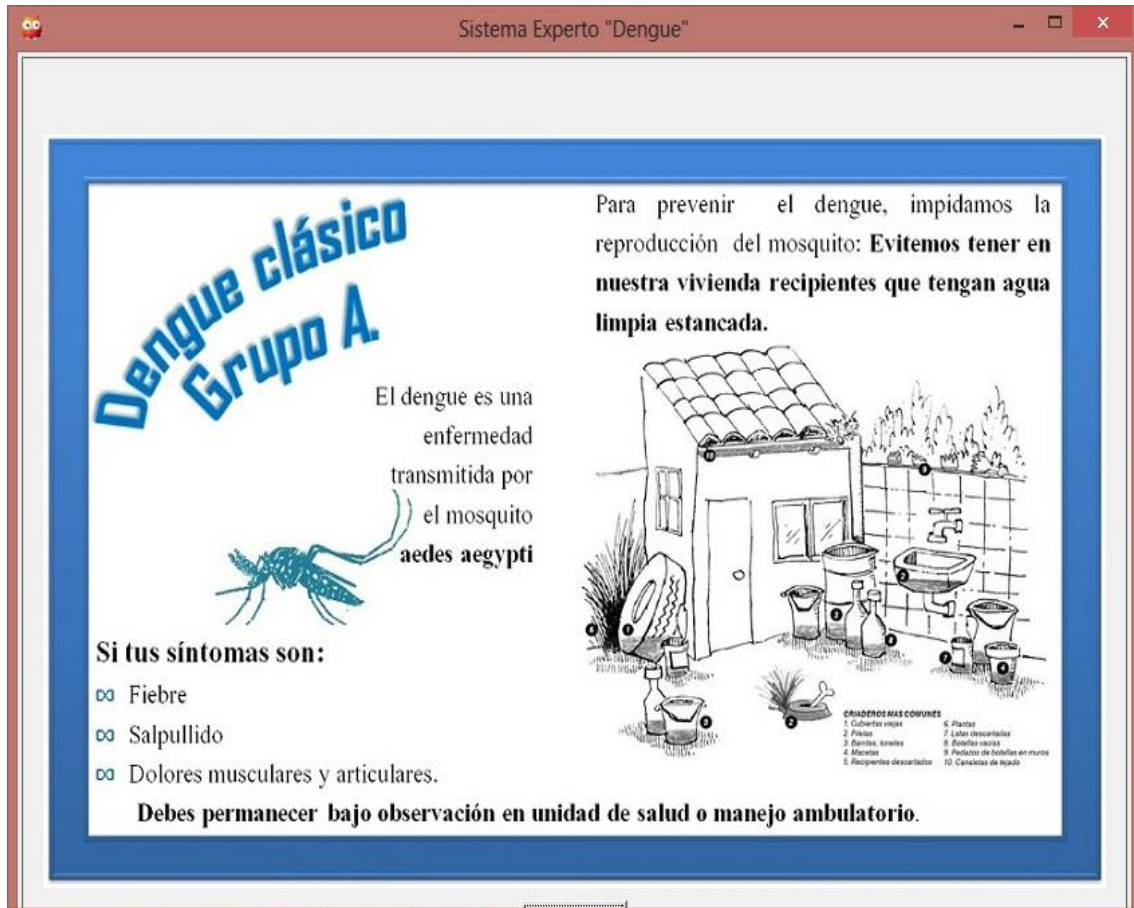
El sistema experto final cuenta con 32 ventanas y es capaz de diagnosticar 6 padecimientos virales además de los diferentes tipos de dengue, es decir un total de 11 diagnósticos diferentes. Los colores usados son el color blanco a 30% con azul a 70%, esto de acuerdo a lo sugerido en el libro Psicología del Color de Eva Heller, en el cual se hace referencia a esta combinación como los colores de la salud.

Queda por mencionar que el sistema es capaz de almacenar la información del paciente así como las reglas usadas en cada diagnóstico.

La interfaz de inicio se muestra en la figura 1, donde *Comenzar* lanza a pantalla la pregunta del paso uno en el algoritmo de diagnóstico, *Integrantes* muestra a los colaboradores del proyecto e información de contacto, *Experto* proporciona información sobre el experto humano que brindó su experiencia y conocimiento sobre el tema, *Historial* muestra una serie de ventanas en las cuales se pide proporcionar la información del paciente para poder llevar un control de nombre, edad, sexo, estos datos serán almacenados en un archivo de tal modo que pueda ser utilizada para estudios posteriores.

, el cual era el resultado esperado. Observe la figura 2 en la cual se muestra la ventana del diagnóstico.

**Figura 2. Diagnóstico de Dengue clásico.**



Para mostrar la eficacia del sistema se realizaron una serie de pruebas las cuales fueron supervisadas por la experta humana: Dra. Estela Rosete Quintana. A continuación, se muestran los resultados de una de estas pruebas: Para esta corrida el experto propone un “paciente piloto” que durante 3 días tuvo escalofríos y temblores, posteriormente presentó una fiebre repentina de 40 °C con dolores oculares, erupciones en el tronco y cuerpo, comecón sin señales de descamación, un aumento drástico y repentino en la temperatura de la piel, el cual por la experiencia del médico será indudablemente diagnosticado con Dengue clásico. Una vez planteados los síntomas y el resultado esperado se procedió a iniciar con el diagnóstico mediante el sistema experto, el paso 1 pregunta por un paciente con escalofríos a lo cual se responde “sí” ya que es uno de los primeros síntomas que se plantearon en el “paciente piloto”, así se llega al paso 2 del algoritmo en el cual se pregunta por un paciente con fiebre repentina mayor a los 39.5 °C lo cual concuerda con los signos del “paciente piloto” ya que se mencionaba fiebre de 40 °C, en el paso 5 se pregunta si el paciente presenta erupciones iniciadas en el tronco las cuales se extienden por el resto del cuerpo, a este paso se respondió afirmativamente ya que el “paciente piloto” presentaba erupciones en el tronco y cuerpo, además comecón sin descamación lo cual respondía “sí” a la pregunta del paso 7, en el paso 9 se pregunta si el paciente presenta aumento en la temperatura de su piel, lo cual responde afirmativo a los signos planteados del “paciente piloto”, después de esta última pregunta se obtuvo un paciente diagnosticado con Dengue catalogado en el Grupo A

## 5. CONCLUSIONES

El uso de la lógica de primer orden, es útil en el desarrollo de Sistemas Expertos que sirven para mejorar la calidad de vida de la población, en este caso ayudar a diagnosticar de manera oportuna el Dengue clásico, hemorrágico, sus diferentes grupo y otras enfermedades virales.

La experiencia que aporta el experto humano es de mucha utilidad para poder diferenciar enfermedades que tienen síntomas parecidos.

---

## REFERENCIAS

1. Badii M.H., Landeros J. M., Cerna E. y Abreu J. L. (2007), "Ecology and history of dengue in Americas", *International Journal of Good Conscience*. 2(2) : 248-273.
2. Castillo E., Gutiérrez J. M., Hadi A. S., (1997) *Expert Systems and Probabilistic Network Models*, Springer, New York.
3. Contreras I., Villanueva J., Salinas M., Enríquez M., Pérez R. (2013), "Patogénesis del dengue hemorrágico síndrome de choque del dengue", *Electronic Journal of Biomedicine*. 1:37-49.
4. Frenk J. (2001), *Enfermedades transmitidas por vector*, Secretaría de Salud de México.
5. Frenk J. (2015), *Dengue clásico y fiebre hemorrágica*, URL: <http://www.cenaprece.salud.gob.mx/descargas/pdf/dengue.pdf>, fecha de consulta: Junio de 2015, fecha de actualización: Marzo de 2015.
6. Pizarro D. (2009), "Dengue, dengue hemorrágico", *Acta pediátrica costarricense*. 21(1): 8-17.
7. Prasad K., Sagar Y., (2011), *An Approach to Develop Expert Systems in Medical Diagnosis Using Machine Learning Algorithms (ASTHMA) and a Performance Study*, *International Journal on Soft Computing (IJSC)*, PA, 26-33.
8. Thrión J. (2010), *El mosquito Aedes aegypti y el dengue en México*, *Bayer Environmental Science*.

## **APLICACIÓN DE LA REGRESIÓN MULTINIVEL EN LA DETERMINACIÓN DE FACTORES ASOCIADOS A LA ADIPOSIDAD CENTRAL EN PACIENTES CON VIH/SIDA**

María Amparo León Sánchez\* & Elisa Maritza Linares Guerra\*\*

\*Universidad de Pinar del Río, Hermanos Saíz Montes de Oca  
Departamento de Matemática, Pinar del Río.

\*\*Universidad de Ciencias Médicas de Pinar del Río. Facultad de Ciencias Médicas., Pinar del Río, Cuba

### **RESUMEN**

La presente investigación se desarrolló con el objetivo de utilizar la regresión multinivel en la identificación de los factores asociados a los cambios ocurridos en los indicadores antropométricos de adiposidad central en pacientes con VIH / SIDA, a lo largo del tiempo. El estudio incluyó 159 pacientes a los que se midió el perímetro de la cintura (PC) y la cadera para crear el Índice Cintura Cadera (ICC), en tres momentos diferentes durante cuatro años. Para cada etapa del modelo se analizó la importancia de la reducción de la desviación, la significación de los coeficientes y el efecto de la nueva variable en los componentes aleatorios. Los valores iniciales de las variables dependientes PC<sub>0</sub> e ICC<sub>0</sub> tuvieron una contribución relevante en los modelos, lo que demuestra que su evolución está muy relacionada con el estado inicial del paciente. El tiempo de infección con el VIH resultó un factor predictor de incremento del PC y del ICC, independientemente del sexo y del estado de la TAV; mientras que a mayor tiempo de exposición a los antirretrovirales, mayor es el riesgo de lipoacumulación central. La ICC refleja el cambio de los individuos mejor que el PC.

### **ABSTRACT**

This research was carried out with the aim to use multilevel regression in the identification of factors associated to changes occurred in anthropometric indicators of central adiposity in patients with HIV/AIDS, through the time. The study included 159 patients for whom the perimeter of the waist (PC) and the hip were measured to create the Waist Hip Index (ICC), in three different moments, during four years. For every stage of the model were analyzed the significance of the reduction in the deviance and coefficients, as well as the effect of the new variable in random components. The initial values for the dependent variables PC<sub>0</sub> and ICC<sub>0</sub> had a relevant contribution in the models; it demonstrates their high relationship with patients' initial state. The infection time was the predictor of the increased for the IC and the IPC, with independence of the sex and state of the TAV. While higher is the exposure time to antiretroviral, higher is the central lipoaccumulation risk. The ICC reflected individuals' change better than the PC.

**KEY WORDS:** repeated measures, random coefficients, HIV/AIDS, central adiposity.

## **1. INTRODUCCIÓN**

Los métodos estadísticos que estudian variables continuas se centran básicamente en el registro de una única variable por paciente y los más empleados son el análisis de la varianza y las técnicas de regresión, sin embargo lo más frecuente es encontrar estudios longitudinales en que se registran para cada paciente varias observaciones de la misma variable por lo que es razonable pensar que la variabilidad entre las medidas de cada sujeto sea menor que entre los diferentes sujetos, y no pueden por tanto considerarse como observaciones independientes, supuesto básico para estimar un modelo de regresión clásico, por otro lado mediante el análisis de la varianza es posible contemplar diseños en los que existen medidas repetidas sobre el mismo elemento de observación, pero tienen el inconveniente de que el número de observaciones por elemento debe ser idéntico (balanceado), requisito que salvo en los estudios experimentales es muy difícil de cumplir, y aún en éstos es posible que se den pérdidas que rompen el equilibrio del diseño. Los modelos multinivel,<sup>[1]</sup> jerárquicos<sup>[2]</sup> o de coeficientes aleatorios<sup>[3]</sup> garantizan resultados confiables aun cuando se violan las suposiciones inherentes a la inferencia clásica. Hoy día son muchos los trabajos en que se utiliza la regresión multinivel en estudios de salud, algunos son estudios teóricos que constituyen contribuciones importantes para la preparación de los especialistas en estas técnicas<sup>[4, 5, 6, 7]</sup>, otros son aplicaciones de gran utilidad sobre todo para estudiar los cambios que ocurren en el tiempo en pacientes afectados por algunas patologías, ejemplos de estos estudios: una aplicación sobre hipertensión arterial<sup>[8]</sup>, para evaluar dos fármacos antihipertensivos en pacientes nuevos, divididos en dos grupos a los que se les asigna el tratamiento de forma aleatoria<sup>[9]</sup>, una aplicación de la regresión logística multinivel en pacientes diabéticos de atención primaria<sup>[10]</sup>, para analizar la influencia de dos dispositivos de inducción anestésica sobre la tensión arterial<sup>[11]</sup>, en el estudio sobre el bajo peso al nacer<sup>[12]</sup>, en la asociación entre el bienestar material de la provincia de residencia y la inactividad física durante el tiempo libre en la población española mayor de 15 años<sup>[13]</sup>, o en el análisis de los cambios en la función cognitiva de una cohorte de personas mayores, el estudio



---

«Envejecer en Leganés», <sup>[14]</sup>. Si bien es amplia la aplicación de la regresión multinivel en salud, no se reportan estudios de su utilización en el campo del VIH/sida, en la valoración de los cambios que con el tiempo desencadena la terapia antirretroviral (TAV) en la composición corporal de estos individuos.

La exposición a los antirretrovirales ha provocado cambios en la distribución de la grasa corporal de los pacientes con VIH/sida, lo que trae como consecuencia pérdida subcutánea y acumulación central del tejido adiposo, como parte del síndrome de lipodistrofia asociado a los antirretrovirales <sup>[15]</sup>. Sin embargo, la adiposidad central no es solo un efecto adverso de los antirretrovirales, depende además de otros factores como el sexo <sup>[16]</sup>, la edad del paciente <sup>[17]</sup>, la acción directa del virus y el tiempo de exposición a los fármacos antirretrovirales <sup>[18; 19]</sup>. Por otra parte la localización de la grasa corporal en la región abdominal, se ha convertido en un importante indicador de morbilidad y muerte, al estar asociado a Diabetes mellitus tipo II, dislipidemias <sup>[20]</sup>, y coronariopatías <sup>[21]</sup>, entre otras situaciones morbosas <sup>[22]</sup> de ahí la necesidad de detectar precozmente el exceso de grasa en la región abdominal del cuerpo en las personas con VIH/sida expuestas al TAV, lo que puede lograrse por técnicas radiológicas o mediciones antropométricas como las circunferencias de la cintura y la cadera. El monitoreo sistemático de estas dimensiones corporales, y la identificación de factores asociados, permitirá la detección precoz de obesidad central o troncal, y el riesgo consecuente de morbilidad que ello implica, de manera fácil, económica y no invasiva. Por lo anteriormente expuesto, el presente trabajo tiene como objetivo utilizar la regresión multinivel, un modelo de medidas repetidas, como herramienta matemática para identificar factores asociados a los cambios que en función del tiempo, ocurren en indicadores antropométricos de adiposidad central, en pacientes con VIH/sida.

### 3. MATERIALES Y MÉTODOS

Se trabajó con datos de 159 pacientes VIH/sida, con una edad media de  $32,7 \pm 8,6$  años; 115 del sexo masculino y 44 del femenino, 72 expuestos a los antirretrovirales (antes de ser incorporados al estudio, o durante el tiempo que duró el mismo) y 87 sin TAV. Los pacientes que recibieron TAV tuvieron un tiempo medio de exposición de  $3,47 \pm 2,48$  años. A todos se les realizó la determinación de variables antropométricas en tres momentos comprendidos dentro de un período de 4 años.

Criterios de inclusión: Individuos adultos de ambos sexos con confirmación de diagnóstico serológico positivo al VIH, con atención ambulatoria, que acudieron a las citas realizadas por el “Departamento Provincial de ITS/VIH/sida” de la provincia de Pinar de Río, para la realización de las evaluaciones inmunológicas y nutricionales y además aceptaron participar en el estudio.

Criterios de exclusión y/o salida: Niños, mujeres embarazadas, adultos con algún tipo de enfermedad metabólica o uso de esteroides anabólicos, así como los fallecidos en el período de tiempo que duró el seguimiento.

Determinaciones Antropométricas: Se cumplieron los requisitos y recomendaciones generales de la técnica antropométrica descritos por Díaz Sánchez [23] relacionados con las condiciones basales de la técnica antropométrica, y el procedimiento técnico para las mediciones de las circunferencias de la cintura y cadera. A partir de las mediciones anteriores se determinó el índice cintura cadera [24].

Variables utilizadas en el estudio, abreviatura para su operacionalización y unidad de medida:

- 1 Perímetro de la cintura [PC] (cm)
- 2 Índice Cintura Cadera [ICC] (%)
- 3 Perímetro de la cintura en el momento inicial del estudio [PC\_0] (cm)
- 4 Índice Cintura Cadera en el momento inicial del estudio [ICC\_0] (%)
- 5 Edad del paciente en el momento que le fue diagnosticada la enfermedad [ed\_diag] (años)
- 6 Tiempo de infección antes de comenzar el estudio [t\_inf] (años)

TAV)	7	Estado de la terapia antirretroviral [TAV] (0 para Sin TAV, 1 Para Con TAV)
	8	Tiempo de exposición a la TAV en el momento de la medición [t ] (años)
	9	Sexo (0 para Masculino, 1 para Femenino)

En este trabajo se utilizó la modelación multinivel para describir los cambios en las variables antropométricas, mediante dos ecuaciones: la primera modela la trayectoria individual o evolución de cada individuo y la segunda describe la variación de las trayectorias entre individuos e identifica los factores que explican las diferencias entre los mismos.

En el análisis de datos longitudinales se utilizó un modelo jerárquico con dos niveles:

Nivel 1: las medidas repetidas, en este caso son las determinaciones de las variables antropométricas.

Nivel 2: el individuo.

El nivel 1 describe la evolución de cada individuo mediante una función matemática cuyos parámetros son a su vez las variables de resultado en el modelo de nivel 2. En este caso las evaluaciones de las variables antropométricas en los diferentes momentos actúan como medidas repetidas y las variables restantes son los posibles predictores.

Se trabajó en dos escenarios, primeramente se ajustaron los modelos con todo el conjunto de datos (N=159) y en un segundo escenario se trabajó sólo con los pacientes expuestos a TAV (N=72).

Para hacer una formulación general se definió como variable resultado  $Y_{ij}$  comenzando por el modelo más simple sin predictores y se añadieron las variables restantes hasta conseguir el modelo de mayor poder explicativo.

$$Y_{ij} \sim N(XB, \Omega)$$

$$Y_{ij} = \beta_{0j} + \sum_{i=1}^p \beta_{ij} X_{ij} + \sum_{j=1}^k \beta_j Z_j + \epsilon_{ij} + u_j$$

$Y_{ij}$  variable resultado de la medición en la ocasión  $i$  del individuo  $j$

$X_{ij}$ : predictores del nivel 1

$Z_j$ : predictores de nivel 2

$[u_j] \sim N(0, \Omega_u) = [\sigma_u^2] = \sigma_{individuos}^2$ : variabilidad entre individuos

$[\epsilon_{ij}] \sim N(0, \Omega_\epsilon) = [\sigma_\epsilon^2] = \sigma_{medidas}^2$ : variabilidad entre ocasiones del individuo  $j$

En la modelación se estudiaron también las interacciones entre niveles por lo que se relajó el nivel de significación al 10%. Los datos fueron procesados con el programa MLwiN 2.02.

### 3. RESULTADOS

Los indicadores de adiposidad central tales como la circunferencia de la cintura y el ICC pueden alterarse por efecto de la infección por VIH, la TAV y otros factores inherentes al individuo en cuestión. A la vista de los resultados descriptivos en la tabla 1 donde se valora el factor estado de la TAV, no se aprecian diferencias sustanciales entre los valores medios de los grupos con y sin TAV para ninguno de dichos indicadores antropométricos, no obstante, para el PC hay una diferencia entre los valores medios totales de 1,4 cm, superior en los pacientes expuestos a TAV, mientras que el valor medio del Índice Cintura Cadera en los pacientes sometidos a TAV supera en 1,3 % a la media de los no expuestos.

El rápido aumento de la circunferencia de la cintura fue uno de los primeros efectos secundarios e inusuales provocados por la TAV<sup>[25]</sup> posteriormente varias investigaciones relacionadas con los cambios en la composición corporal de las personas con VIH/sida, han demostrado que tanto el PC como el ICC aumentan después de la administración de los fármacos antirretrovirales (ARV)<sup>[26; 27; 18]</sup>.

En el presente estudio, aunque de forma ligera, se observa un aumento de ambos indicadores de adiposidad central en el grupo expuesto a los ARV cuando analizamos los resultados de los tres momentos en los que se realiza la medición, tal y como muestra la tabla 1.

**TABLA 1. Medidas descriptivas para las variables antropométricas de adiposidad abdominal en pacientes con VIH/sida.**

Variables Antropométricas		TAV							
		0 (N= 87)				1 (N=72)			
		Oc_1	Oc_2	Oc_3	Total	Oc_1	Oc_2	Oc_3	Total
PC (cm)	Media	82,9 0	84,2 0	84,9 0	84,0 0	83,8 0	85,6 0	86,8 0	85,4 0
	Desviación estándar	10,7 0	10,7 0	10,5 0	10,6 0	10,1 0	10,5 0	11,7 0	10,8 0
ICC (%)	Media	88,1 0	88,2 0	89,7 0	88,7 0	89,2 0	89,7 0	91,0 0	90,0 0
	Desviación estándar	6,14	5,27	5,85	5,77	6,87	6,30	7,49	6,90

0: Sin TAV 1: Con TAV

Las tablas de la 2 a la 5 presentan los resúmenes de los modelos jerárquicos estimados para ambas variables, Muestran además los parámetros estimados con sus respectivos errores estándar y la estimación de la varianza entre individuos ( $\sigma_{\text{individuos}}^2$ ) y la varianza entre medidas repetidas de cada individuo ( $\sigma_{\text{medidas}}^2$ )

Para la modelación se comenzó siempre por el modelo no condicional o vacío, sin variables predictoras, las variables restantes se incorporaron una a la vez al igual que las interacciones entre niveles. En cada paso se analizó la significación de la disminución de la desviación, y de los coeficientes además del efecto de la nueva variable en los componentes de la varianza.

El ICC se trabajó en por ciento sólo para facilitar la comparación de valores de las varianzas debido a la exactitud en las cifras decimales que da la salida del programa.

### 3.1. Escenario 1 para el perímetro de la cintura, N=159

El modelo no condicional o vacío predice el resultado dentro de la unidad de nivel 1 (medidas repetidas) con un solo parámetro de nivel 2, la intersección en el origen (constante) que representa el resultado promedio para el individuo j. A partir del modelo no condicional o vacío se calcula también el coeficiente de Correlación Intraclass (CCI) o cociente entre la variabilidad entre individuos y la variabilidad total y expresa el porcentaje de la variabilidad total que es atribuible a las diferencias entre individuos. En estos casos los coeficientes toman valores elevados próximo a 1 (100%) lo que indica que las trayectorias de diferentes personas son muy variables, esto sugiere que las diferencias podrían deberse a las características del individuo, lo que resta para llegar a 1 es lo que el individuo puede evolucionar con el transcurso del tiempo. El valor de desviación del modelo nulo constituye una referencia para comparar con otros modelos en que se incluyen variables predictoras.

Las tablas 2 y 3 contienen los modelos para el total de pacientes de la muestra, se analizaron como posibles predictores: la exposición a la TAV, el sexo, el tiempo de infección antes de comenzar el estudio, la edad del paciente en el diagnóstico confirmatorio de VIH+ y el valor de la variable resultado en la primera evaluación: perímetro de la cintura en el momento cero (PC\_0) e índice cintura cadera en el momento cero (ICC\_0).

En el caso del perímetro de la cintura (tabla 2), el 89,56% de la variabilidad fue atribuible a las diferencias entre individuos (Coeficiente de Correlación Intraclass = 89,68%), resultaron significativas a nivel individual, contribuyendo positivamente a la variable

dependiente: el sexo femenino (  $P = 0,07$  ), el perímetro de la cintura en el primer momento del estudio ( $p < 0,01$ ) y el tiempo de infección ( $p < 0,01$ ), Los resultados demuestran que las mujeres incrementan en 2,72 cm su PC con relación a los hombres; el perímetro de la cintura de cada integrante de la muestra en un momento determinado será el 94% del valor del PC en el tiempo cero y en la medida que el tiempo de infección aumenta, también lo hace el perímetro de la cintura a razón de 0,63 cm por año. Con la inclusión del tiempo de infección, la varianza entre individuos se reduce en un 7,63% y la disminución en la desviación resultó significativa al 1%.

Las diferencias en la composición corporal entre hombres y mujeres a consecuencia del dimorfismo sexual son bien conocidas [28], la grasa abdominal tiene una tendencia de acumularse en las mujeres en las regiones periféricas como las caderas, mientras que en los hombres hay mayor acumulación en el área abdominal [29], no obstante, se ha comprobado que tanto la terapia antirretroviral, como la propia infección por VIH, desencadenan alteraciones metabólicas que resultan en una acumulación de la grasa en la región abdominal en hombres y mujeres por igual, [27]. En un estudio realizado en un contexto similar al de la presente investigación, se encontró una asociación significativa entre el sexo, el tratamiento antirretroviral y el PC, experimentando las mujeres con TAV un aumento significativo de dicha dimensión corporal con relación a los hombres, [27], aspecto este coincidente con los resultados de esta investigación. Es importante destacar que en el presente estudio ni la edad del individuo al diagnóstico confirmatorio de seropositividad al VIH, ni el estado de la TAV, representaron factores predictores de los cambios en el PC, ya que no se alcanzó significación estadística al introducir ambas variables del nivel 2, sin embargo, el tiempo de infección con VIH tuvo un aporte significativo a los cambios del PC ( $P < 0,01$ ). Estos resultados coinciden con lo reportado por Brown y colaboradores en el año 2007 [31], quienes comprobaron que el aumento de la circunferencia de la cintura es más atribuible a la acción del VIH que a la propia TAV. La inclusión en el modelo 4 de PC\_0 y el sexo disminuyó la varianza entre individuos en un 90,3%. La disminución en la desviación resultó significativa al 1% con los modelos 1, 2 y 3, de manera que para predecir los cambios en el perímetro de la cintura de personas con VIH/sida, independientemente de otros factores que pudieran estar influyendo en los mismos, se propone utilizar el modelo 4, donde además del sexo, se tiene en cuenta la medida del PC en el momento inicial (tiempo cero), es decir este modelo parte de los cambios inherentes de esta dimensión antropométrica en respuesta al dimorfismo sexual y de la medida del PC en una primera evaluación, la cual está garantizando una disminución significativa de la variabilidad entre individuos, y tiene implícito en sí el resto de los factores que afectan esta dimensión corporal.

**Tabla 2. Coeficientes y errores estándar de los modelos multinivel para el Perímetro de la cintura (N =159)**

Coeficientes (error estándar)					
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Constante	85,27 (0,84)***	84,52(0,98)***	6,91(2,61)***	82,57(1,11)***	6,43(2,53)***
Sexo		2,72(1,86)*			2,30(0,70)***
PC_0			0,94(0,03)***		0,94(0,03)***
t_inf				0,63(0,18)***	
Efectos aleatorios					
( $\sigma^2_{individuos}$ )	106,35(12,51)	104,87(12,35)	11,36(1,93)	98,23 (11,61)	10,30(1,82)
( $\sigma^2_{medida}$ )	10,23(1,15)	10,23(1,15)	10,23(1,15)	10,23(1,15)	10,23(1,15)
Desviación	2131,85	2129,73	1827,81	2119,83	1817,29
Diferencia en desviación					
Modelo nulo		2,12 NS	304,04***	12,02***	314,56***
Modelo 1					312,44***
Modelo 2					10,52***
Coeficiente de Correlación Intraclase: 89,68%					

NS: No Significativo      \* $p < 0,1$       \*\*  $p < 0,05$       \*\*\* $p < 0,01$

### 3.2. Escenario 2 para el perímetro de la cintura, N=72

Los resultados para el perímetro de la cintura en la muestra de pacientes expuestos a la TAV, se representan en la tabla 3.

Los resultados muestran una pequeña disminución de 0,38%, del Coeficiente de Correlación Intraclase con respecto al modelo que incluye la totalidad de los pacientes, la diferencia entre individuos se reduce aunque discretamente cuando se estudian solo pacientes expuestos a TAV. Del modelo 1 se concluye que el tiempo de exposición a TAV, significativa con  $P < 0,01$ , provoca un aumento promedio en el perímetro de la cintura de 0,957 cm por año, esta variable  $t$  provoca una reducción en la variabilidad entre individuos de 3,1% y 2,98% en la variabilidad dentro de las medidas repetidas de cada individuo, la disminución en desviación respecto al modelo nulo resultó significativa con  $P = 0,03$ . Estos resultados reafirman lo encontrado por otros investigadores con relación al papel del tiempo de exposición a los ARV como factor asociado a la lipodistrofia en pacientes con VIH/sida [18].

En el modelo 2 la inclusión del tiempo de infección antes de comenzar el estudio, significativa con  $P < 0,01$ , provoca un aumento promedio en el perímetro de la cintura de 0,786 cm por año, esta variable,  $t_{inf}$ , provoca una reducción en la variabilidad entre individuos de 10,21% en relación con el modelo nulo. La diferencia en desviación resultó significativa con  $P < 0,01$ . En este caso quedó comprobada la acción directa del VIH sobre la composición corporal del individuo infectado, específicamente sobre el PC, el cual tiende significativamente al aumento, independientemente de la exposición a los antirretrovirales, tal y como han encontrado otros investigadores [30].

El modelo 3 incluye como predictor a la determinación del PC en el momento cero,  $PC_0$ , esto provoca una reducción en la variabilidad entre individuos relevante, de 87,44%, lo que nos induce a afirmar que la variación en el PC de los individuos depende casi exclusivamente de sus valores iniciales, incluso los interceptos no son significativos en los modelos 3 y 4.

**Tabla 3. Coeficientes y errores estándar de los modelos multinivel para el Perímetro de la cintura de pacientes con VIH/sida con TAV (N =72).**

Coeficientes (error estándar)	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Constante	86,19(1,27)***	83,10(1,97)***	81,20(2,16)***	4,70 (4,46) NS	3,40(4,21) NS
Sexo					3,25(1,06)***
$t$		0,957(0,045)***			
$t_{inf}$			0,786(0,284)***		0,975(0,05)***
$PC_0$				0,973(0,053)***	
Efectos aleatorios					
$(\sigma^2_{individuos})$	109,08(19,30)	105,66(18,67)	97,94(17,44)	13,70(3,54)	11,35(3,17)
$(\sigma^2_{medidas})$	13,06(2,18)	12,67(2,12)	10,23(1,15)	13,06(2,18)	13,06(2,17)
Desviación	985,60	981,08	978,30	860,10	851,22
Diferencia en desviación					
Modelo nulo		4,52**	7,3***	125,5***	134,38***
Moldeo 1					8,88***
Coeficiente de Correlación Intraclase:	89,30%				

Los resultados encontrados en el grupo expuesto a los ARV permiten sugerir que a la hora de predecir los cambios en el PC de individuos VIH/sida con TAV, se pueden usar indistintamente como posibles factores predictores, la medida del valor del PC en un momento inicial, tal y como muestra el modelo 3 de la tabla 3, o independientemente de ello, utilizar los predictores sexo y tiempo de infección, presentes en el modelo 4. Ambos modelos son similares

al lograr disminuir significativamente la variación entre los individuos, y disminuir también de forma similar la desviación con relación al resto de los modelos. El modelo 4 tiene como principal ventaja que permite predecir la evolución en el tiempo.

### 3.3. Escenario 1 para el índice cintura/cadera, N=159

Al igual que el PC, el ICC puede ser utilizado para estimar la grasa localizada en la región central del cuerpo<sup>[31]</sup>. Con este indicador (tabla 4) resultaron significativas al introducirse individualmente las siguientes variables: sexo, tiempo de infección antes de comenzar el estudio y la edad del paciente en el momento del diagnóstico confirmatorio de VIH+, todas con  $P < 0,01$ , la exposición a la TAV resultó con una probabilidad de 0,08, La contribución de la TAV a la variable resultado es positiva, modelo 2, lo que significa que las trayectorias del ICC en el caso de los pacientes sometidos a tratamientos antirretrovirales están por encima de los que no lo reciben y esta diferencia es significativa con un incremento promedio de 1,39 %. Si tenemos en cuenta que para el total de la muestra, la TAV no resultó un factor determinante en los cambios del PC, y que el ICC no es más que la relación entre la circunferencia de la cintura y la de la cadera, la explicación a este resultado pudiera estar en un aumento más lento de la circunferencia de la cadera en el grupo expuesto a los antirretrovirales tal y como reportó Brown y colaboradores en el año 2006<sup>[32]</sup>.h Estos investigadores demostraron durante un período de observación de 4 años, que la circunferencia de la cadera aumentó más lentamente en el grupo tratado con TAV que en el grupo VIH-negativos, produciéndose un aumento más rápido del ICC en el grupo con TAV, de manera que el aumento de la tasa de cambio del ICC en personas infectadas por el VIH que reciben TAV en comparación con los VIH-negativos es atribuible a los aumentos más lentos en la circunferencia de la cadera en lugar de un aumento de la tasa de cambio en la circunferencia de la cintura.

Si bien la introducción de estas variables disminuyeron discretamente la varianza entre individuos (5% el sexo, 1,7% el estado de la TAV, 5,3% el tiempo con TAV, 6,9% el tiempo de infección antes de comenzar el estudio y 4,5% la edad del paciente en el momento que le fue diagnosticada la infección por VIH), modelos 1 a 5, la disminución más marcada se logra con la inclusión de la determinación del ICC en el momento cero ICC\_0, modelo 5, que provocó una disminución del 77,35% lo que indica que el estado inicial del paciente es más relevante en el estudio que el resto de las variables. La disminución en desviación es también altamente significativa con respecto al modelo nulo.

**Tabla 4a. Coeficientes y errores estándar de los modelos multinivel para el Índice Cintura Cadera (N = 159)**

Coeficientes (error estándar)				
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3
Constante	89,55(0,46)***	90,28(0,53)***	88,93(0,61)***	88,24(0,61)***
Sexo		-2,64(1,00)***		
TAV			1,39(0,91)*	
t_inf				0,31(0,10)***
ed_diag				
Efectos aleatorios				
( $\sigma^2_{individuos}$ )	27,51(3,79)	26,12(3,64)	27,04(3,74)	25,61 (3,58)
( $\sigma^2_{modelos}$ )	11,60(1,30)	11,60(1,30)	11,60(1,30)	11,60(1,30)
Diferencia en desviación				
Desviación	1959,78	1952,99	1957,49	1950,39
Modelo nulo		6,79***	2,29 NS	9,39***

NS: No Significativo \*  $p < 0,10$  \*\* $p < 0,05$  \*\*\* $p < 0,01$

La introducción de una segunda variable aumenta el poder explicativo de los modelos del 1 al 4. En el modelo 6 donde se introducen sexo y exposición a la TAV, se logra una disminución del 7,5% en la varianza entre individuos en relación con el modelo nulo, 2,6% respecto al modelo 1 que solo contiene el sexo y 5,9% en relación con el modelo 2 que solo contiene la exposición a la TAV. En el modelo 7 que contiene las variables sexo, edad del paciente en el momento de la confirmación

diagnóstica de VIH+ y tiempo de infección todos los coeficientes resultaron significativos, las diferencias en desviación con respecto a los modelos nulo, 1 y 4 resultaron también significativas con  $p < 0,01$  y se redujo la varianza entre individuos en 22,46%. El modelo 5 tiene la ventaja de predecir con una única variable pero no se incluye ninguna variable temporal, por lo que se sugiere utilizar el modelo 7 para predecir los cambios.

**Tabla 4b. Coeficientes y errores estándar de los modelos multinivel para el Índice Cintura Cadera (N = 159)**

Coeficientes (error estándar)				
	Modelo 4	Modelo 5	Modelo 6	Modelo 7
Constante	85,93(1,52)***	26,32(3,78)***	89,58(0,54)***	82,78(1,69)***
Sexo			-2,85(1,00)***	-2,69(0,93)***
TAV			1,68(0,90)**	
t_inf				0,46(0,098)***
ed_diag	0,13(0,051)***			0,20(0,05)***
ICC_0		0,71(0,043)***		
Efectos aleatorios				
( $\sigma^2_{individuos}$ )	26,26(3,65)	6,23(1,50)	25,43(3,56)	21,33(3,11)
( $\sigma^2_{medidas}$ )	11,60(1,30)	11,60(1,30)	11,60(1,30)	11,60(1,30)
Desviación	1953,68	1797,77	1949,52	1927,10
Diferencia en desviación				
Modelo nulo	6,1**	162,01***	10,26***	32,68***
Modelo1			3,47*	25,89***
Modelo2			7,97***	
Modelo 4				23,29***
Coeficiente de Correlación Intraclase: 89,68%				

NS: No Significativo \*  $p < 0,10$  \*\* $p < 0,05$  \*\*\* $p < 0,01$

### 3.4. Escenario 2 para el índice cintura/cadera, N=72

**Tabla 5a. Coeficientes y errores estándar de los modelos multinivel con un solo predictor para el Índice Cintura Cadera de pacientes VIH/sida con TAV (N = 72)**

Coeficientes (error estándar)						
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Constante	90,31(0,75)***	91,38(0,89)***	87,74(1,17)***	87,97(1,30)***	86,29(2,67)***	25,96(6,13)***
Sexo		-3,21(1,54)***				
t			0,80(0,29)***			
t_inf				0,368(0,17)***		
ed_diag					0,148(0,095)**	
ICC_0						0,721(0,068)***
Efectos aleatorios						
( $\sigma^2_{individuos}$ )	32,30(6,78)	30,01(6,41)	30,01(6,34)	29,85(6,38)	30,98(6,57)	8,07(2,92)
( $\sigma^2_{medidas}$ )	15,33(2,56)	15,33(2,56)	14,75(2,47)	15,33(2,56)	15,33(2,56)	15,33(2,56)
Desviación	920,64	916,39	913,04	916,09	918,23	853,52
Diferencia en desviación						
Modelo nulo		4,25**	7,60***	4,55**	2,41 NS	67,12***
Coeficiente de Correlación Intraclase: 67,81%						

NS: No Significativo \*  $p < 0,10$  \*\* $p < 0,05$  \*\*\* $p < 0,01$

Los resultados para el índice cintura/cadera en la muestra de pacientes expuestos a la TAV, se representan en las tablas 5a y 5b. En los modelos con un solo predictor (tabla 5a): sexo, tiempo de exposición a la TAV, tiempo de infección, y edad al diagnóstico de seropositividad al VIH, modelos del 1 al 4, todas las variables tuvieron un aporte significativo al menos al 5%,

Los aportes a la disminución de la desviación en relación con el modelo nulo fue significativo en todos los modelos excepto el que tiene como único predictor edad del paciente en el momento de la confirmación diagnóstica de VIH+. Las contribuciones de todos los coeficientes al ICC en todos los modelos fueron positivas con excepción de la edad. En el modelo 5 que tiene como único predictor la determinación en el momento inicial del estudio, ICC\_0 corrobora la fortaleza del estado inicial del paciente en la modelación, esta variable provoca una disminución en la varianza entre individuos del 75% respecto al modelo nulo y la disminución en desviación es del 7,2%. Otro modelo de especial interés por su calidad y relación con otras variables del estudio resultó el modelo 9 (tabla 5b), obsérvese que la diferencia en desviación con todos los modelos restantes resultó significativa, y los coeficientes de todas las variables también son estadísticamente significativos. Según este modelo y haciendo un análisis de los coeficientes de los predictores los hombres ganan como promedio 2,93% más de ICC que las mujeres. Independientemente del sexo, por cada año de exposición a la terapia el ICC aumenta en 0,46%, por cada año de infección con el VIH en 0,39% y por cada año más que tenga la persona en el momento del diagnóstico VIH+, el incremento en ICC es como media 0,22 %. El CCI de este modelo con los pacientes expuestos a terapia disminuyó a 67,81 de un 70,34 que tenía el modelo con la información completa, de manera que el considerar en el estudio solo al grupo con TAV, disminuyó significativamente la posibilidad de cambios en el ICC a consecuencia de la variabilidad entre los sujetos, para depender más de los factores evaluados.

**Tabla 5b .Coeficientes y errores estándar de los modelos multinivel con más de un predictor para el Índice Cintura Cadera de pacientes con VIH/sida con TAV (N = 72).**

Coeficientes (error estándar)				
	Modelo 6	Modelo 7	Modelo 8	Modelo 9
Constante	88,83(1,24)***	88,94(1,32)***	84,38(2,89)***	81,27(3,29)***
Sexo	-	-	-	-
t	3,21(1,48)**	3,49(1,98)***	2,83(1,47)**	2,93(1,43)**
t_inf	0,79(0,28)***			0,46(0,33)
t_inf		0,399(0,164)***	0,842(0,275)***	
ed_diag			0,154(0,090)**	0,39(0,21)**
ICC_0				0,22(0,096)***
Efectos aleatorios				
( $\sigma^2_{\text{entre individuos}}$ )	27,72(5,97)	27,16(5,94)	26,43(5,76)	24,34(5,43)
( $\sigma^2_{\text{entre años}}$ )	14,75(2,47)	15,33(2,56)	14,72(2,46)	14,95(2,5)
Desviación	908,49	910,72	905,65	902,46
Diferencia en desviación				
Modelo nulo	12,15**	9,92***	14,99***	18,18***
Modelo 1			10,74***	13,93***
Modelo 2	7,9***	5,67**		
Modelo 3	4,55***			10,58**
Modelo 4	7,6***		12,58***	3,63***
Modelo 5		5,37**		15,77***
Modelo 6			5,07**	6,03**
Modelo 7				3,19*

NS: No Significativo \* p < 0,10 \*\*p < 0,05 \*\*\*p < 0,0

#### 4. CONCLUSIONES.

Se comprobó la influencia de un conjunto de factores que se asocian a los cambios que con el tiempo ocurren en indicadores antropométricos de adiposidad central en personas con VIH/sida. Independientemente de la TAV, el PC aumentó significativamente en las mujeres mientras que el ICC



---

aumentó en el sexo masculino, el tiempo de infección con el VIH resultó un factor predictor de incremento del PC y el ICC independientemente del sexo y del estado de la TAV, mientras que tiempo de exposición a los antirretrovirales mayor riesgo de lipoacumulación central.

Para predecir los cambios en el PC y el ICC, el mejor predictor, independientemente que el individuo esté o no expuesto a los antirretrovirales, lo constituye el propio indicador de adiposidad central en un primer momento, antes de iniciar un estudio de seguimiento o evolución, de manera que para evaluar los cambios que ocurren en indicadores de adiposidad central en personas con VIH/sida, deberán preferirse los estudios longitudinales o de seguimiento, siendo el mejor predictor el valor del indicador en la primera evaluación aun cuando sean más útiles aquellos modelos con variables asociadas al tiempo que facilitan evaluación sistemática de los cambios.

## REFERENCIAS

1. GOLDSTEIN, H. (2010). **Multilevel statistical models (4th ed.)**. New York: Wiley Series in Probability and Statistics. 381 p.
2. RAUDENBUSH, S., y BRYK, A. (2002). **Hierarchical linear models: Applications and data analysis methods** (2nd ed.). Sage: Thousand Oaks.
3. LANGFORD, N. (1993). **Random coefficient models**. New York: Oxford University Press.
4. SAEZ, M (2001) El problema de las medidas repetidas. Análisis longitudinal en epidemiología. **Gaceta Sanitaria** ; 15 (4): 347-352
5. ARNAU, J. Y BONO R. (2008) Estudios longitudinales. Modelos de diseño y Análisis. **Estudios de psicología**; 2 -1: 32-41
6. CATALÁN, M.J., GALINDO-VILLARDÓN M.P. (2003) Utilización de los modelos multinivel en investigación sanitaria. **Gaceta Sanitaria**; 17(Supl 3):35-52.
7. STOEL, R., VAN DEN WITTENBOER, G., & HOOK, J. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. **Metodología de las Ciencias del Comportamiento**, 5, 21-42.
8. SOCIEDAD ESPAÑOLA DE BIOQUÍMICA CLÍNICA Y PATOLOGÍA MOLECULAR SEQC. (2007) Notas Estudios longitudinales. Disponible en <http://www.seh-lilha.org/mixedmodels.html>. Consultada 22-07-2015
9. MOLINERO, L.E. (2001) Análisis de estudios longitudinales, datos agrupados y medidas repetidas. **Asociación Sociedad Española de hipertensión**
10. APARICIO A. Y MORERA M. (2007) La conveniencia del análisis multinivel para la investigación en salud. Una aplicación para Costa Rica; **Población y salud en Mesoamérica** 4(2) Informe técnico 1
11. MONTERO, M., CASTELLS, E., LANTIGUA I. (2007) Modelos multinivel: una aplicación a datos longitudinales en una investigación médica. **Revista Investigación Operacional** Vol., 28 No2., 170-178
12. HACHUEL, L., BOGGIO, G., BORRA, V. (2006) Uso de modelos logit mixtos para el estudio del bajo peso al nacer en Rosario. **Undécimas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística, Universidad Nacional de Rosario.**
13. CRUZ, P., REGIDOR, E., GUTIÉRREZ-FISAC, J.L., MARTÍNEZ, D., CALLE, M.E., DOMÍNGUEZ, V. (2005) Bienestar material de la provincia de residencia. **Gaceta Sanitaria** 19(6):424-432.
14. ZUNZUNEGUI, M., GARCÍA DE YÉBENES, M., FORSTER, M., AGUILAR CONESA, M., RODRÍGUEZ, A. y OTERO, A. (2004). Aplicaciones de los modelos multinivel al análisis de medidas repetidas en estudios longitudinales. **Revista Española de Salud Pública**, 78(2):177-188.
15. [FREITAS P](#), [CARVALHO D](#), [SANTOS AC](#), [MATOS MJ](#), [MADUREIRA AJ](#), [MARQUES R](#), [MARTINEZ E](#), [SARMENTO A](#), [MEDINA JL](#). (2012). Prevalence of obesity and its relationship to clinical lipodystrophy in HIV-infected adults on anti-retroviral therapy. **Journal of Endocrinological Investigation** 35(11):964-70.
16. SORLI REDÓ ML, KNOBEL FREUD H, MONTERO M, JERICÓ ALBA C, GUELARGRIMBERG A, BOTET MONTOYA JP. (2007). Impacto del sexo en el síndrome de lipodistrofia en pacientes con infección por el VIH y su asociación con factores de riesgo cardiovascular. **Anales de Medicina Interna (Madrid)**. 24 (4).
17. WU CH, HESHKA S, WANG J, PIERSON RN, HEYMSFIELD SB, LAFERRÈRE B, ET AL. (2007). Truncal fat in relation to total body fat: influences of age, sex, ethnicity and fatness. **International Journal of Obesity (London)**. 31(9):1384-1391.

- 
18. CASTRO SANSORES C, SANTOS RIVERO A, SALAZAR RENDÓN JC, DÍAZ RODRÍGUEZ I, GÓNGORA BIACHI R, GONZÁLEZ MARTÍNEZ P. (2008) .Síndrome de lipodistrofia en pacientes con infección por VIH que reciben tratamiento antirretroviral. **Medicina Interna de México**;24(1):8-15.
19. MAMAN D, GLYNN JR, CRAMPIN AC, KRANZER K, SAUL J, JAHN A, MWINUKA V, NGWIRA MH, MVULA H, MUNTHALI F, MCGRATH N (2012). Very early anthropometric changes after antiretroviral therapy predict subsequent survival, in Karonga, Malawi. **Open AIDS J**. 6: 36-44.
20. SAMARAS K. (2012). The burden of diabetes and hyperlipidemia in treated HIV infection and approaches for cardiometabolic care. **Current HIV/AIDS Report**. 9(3):206-17.
21. PICONI S, PARISOTTO S, RIZZARDINI G, PASSERINI S, MERAVIGLIA P, SCHIAVINI M, NIERO F, BIASIN M, BONFANTI P, RICCI ED, TRABATTONI D, CLERICI M (2012). Atherosclerosis is Associated with Multiple Pathogenic Mechanisms in HIV-Infected Antiretroviral-Naïve or -Treated Individuals. **AIDS**. Disponible en: <http://preview.ncbi.nlm.nih.gov/pubmed/23079800>
22. KRISHNAN S, SCHOUTEN JT, ATKINSON B, BROWN T, WOHL D, MCCOMSEY GA, GLESBY MJ, SHIKUMA C, HAUBRICH R, TEBAS P, CAMPBELL TB, JACOBSON DL (2012). Metabolic Syndrome Before and After Initiation of Antiretroviral Therapy in Treatment-Naive HIV-Infected Individuals. **Journal of Acquired Immune Deficiency Syndromes**. 61(3):381-389.
23. DÍAZ SÁNCHEZ ME (2008). Manual de técnicas antropométricas para estudios nutricionales. Instituto de Nutrición e Higiene de los Alimentos. La Habana. Cuba.
24. BERDASCO GÓMEZ A. (2002). Evaluación del estado nutricional del adulto mediante la antropometría. **Revista Cubana de Alimentación y Nutrición** 16(2):146-52.
25. BAKER R. (1998). ¿ Es TARSA responsable de los efectos secundarios anormales? **Boletín de tratamientos experimentales contra el SIDA**: 23-31.
26. ENA J, BENITO C, YACER P, PASQUAU F, AMADOR C. (2004). Distribución anormal de la **grasa corporal** y tipo de tratamiento antirretroviral como predictores de riesgo de enfermedad cardiovascular en pacientes infectados por el virus de la inmunodeficiencia humana. **Medicina Clínica (Barcelona)**; 122:721-26.
27. LINARES GUERRA EM, ACOSTA NUÑE N, HERNÁNDEZ RODRÍGUEZ Y, SANABRIA NEGRÍN J, JEREZ HERNÁNDEZ E, PLÁCRU A. (2008). Adiposidad abdominal y riesgo de morbilidad en personas de la provincia de Pinar del Río que viven con VIH/sida. **Revista Cubana de Alimentación y Nutrición [en línea]**. 2008 [acceso 14 Sep 2012]; 18(1):43-52. Disponible en: [http://www.revicubalimantanut.sld.cu/Vol\\_18\\_1/Art\\_18\\_3\\_43\\_52.pdf](http://www.revicubalimantanut.sld.cu/Vol_18_1/Art_18_3_43_52.pdf)
28. WELLS JC. (2007). Sexual dimorphism of body composition. **Best Practice & Research Clinical Endocrinology Metabolism**: 21(3):415-30.
29. JACOBSON DL, KNOX T, SPIEGELMAN D, SKINNER S, GORBACH S, WANKE C. (2005) Prevalence of, evolution of, and risk factors for fat atrophy and fat deposition in a cohort of HIV-infected men and women. **Clinical Infectious Diseases**: 40(12):1837-45
30. BROWN TT, CHU H, WANG Z, PALELLA FJ, KINGSLEY L, WITT MD, DOBS AS. (2007). Longitudinal increases in waist circumference are associated with HIV-serostatus, independent of antiretroviral therapy. **AIDS**. 21(13):1731-8.
31. VAN GREEVENBROEK MM, SCHALKWIJK CG, STEHOUWER CD. (2013). Obesity-associated low-grade inflammation in type 2 diabetes mellitus: causes and consequences. **Netherlands Journal of Medicine**. 71(4):174-87.
32. BROWN TT, WANG Z, CHU H, PALELLA FJ, KINGSLEY L, WITT MD, DOBS AS. (2006). Longitudinal anthropometric changes in HIV-infected and HIV-uninfected men. **Journal of Acquired Immune Deficiency Syndromes** 43(3):356-62.

## MODELO PARA LA EVALUACIÓN DEL RIESGO CARDIOVASCULAR PARA CUBA

Germán Lahera Pérez

Departamento de Matemática Aplicada de la Facultad de Matemática y Computación  
Universidad de la Habana

### ABSTRACT

This paper presents a proposal of a model that allows to predict the risk of suffering cardiovascular disease on Cuban patients. On the model construction were analyzed the data collected on the health care facility "Heroes de Corynthia" using Ordinal Logistic Regression and thus obtaining the model that best fit the data. The goal of the investigation is to obtain a model that emits results as close to the Cuban reality as possible and to be able to calculate of a more exact manner the real number of people with high risk of suffering cardiovascular disease on the country. The idea for this paper occur from the results obtained by forward studies that an elevate percent of the Cuban people has low risk to suffer a cardiovascular disease. At last it is shown the model implementation in the software MATLAB trough an interface allowing to calculate the cardiovascular risk of a patient.

**KEYWORDS:** Cardiovascular risk, ordinal Logistic regression.

### RESUMEN

En este artículo se presenta un modelo que permite predecir el riesgo de padecer una enfermedad cardiovascular en pacientes cubanos. Para la construcción del mismo, se analizaron los datos recolectados de la población de "Héroes de Corynthia" utilizando Regresión Logística Ordinal y se obtuvo el modelo que mejor se ajustaba a los datos. El objetivo de la investigación realizada es que el modelo obtenido arroje resultados más cercanos a la realidad cubana y ayude a calcular de forma más exacta el porcentaje real de personas con alto riesgo en el país. Todo esto como consecuencia de los resultados poco fiables obtenidos por estudios anteriores, que planteaban que un elevado por ciento de los cubanos tiene bajo riesgo de padecer una enfermedad cardiovascular. Por último se muestra el modelo implementado en MATLAB mediante una interfaz que permite calcular el riesgo cardiovascular de un paciente.

## 1. INTRODUCCIÓN

Las enfermedades cardiovasculares (ECV) son una de las causas principales de morbilidad y mortalidad a escala mundial [16]. Se calcula que en el 2030 morirán cerca de 23,3 millones de personas por ECV. Cuba, no queda exenta de esta realidad con un comportamiento similar a los países desarrollados, lo cual se debe, en gran medida, al predominio de una pirámide poblacional envejecida.

Numerosos estudios epidemiológicos realizados en los últimos 20 años han coincidido, y puesto en evidencia, la asociación de factores tales como: sedentarismo, estrés, hipertensión arterial, diabetes mellitus, obesidad, hipercolesterolemia y tabaquismo, entre otros, con la aterosclerosis, motivo por el cual han recibido el nombre de factores de riesgo coronario (FRC) [16]. Cuando estos factores de riesgo se asocian, tienen un efecto sinérgico, de modo que un individuo con varios factores de ellos considerados de grado leve, puede estar en riesgo de sufrir un episodio cardiovascular de origen isquémico superior al que tendría con un único factor grave.

El riesgo coronario o cardiovascular es la probabilidad de presentar una enfermedad coronaria o cardiovascular en un período de tiempo determinado, generalmente de 5 o 10 años. Existen dos formas de calcular el riesgo cardiovascular: por métodos cualitativos y métodos cuantitativos; los cualitativos se basan en la suma de factores de riesgo y clasifican al individuo en riesgo leve, moderado y alto; los cuantitativos dan un número, que es la probabilidad de presentar un evento cardiovascular en un determinado tiempo [16]. La forma de cálculo es a través de programas informáticos, basados en ecuaciones de predicción de riesgo, o en las llamadas tablas de riesgo cardiovascular.

Predecir la probabilidad que tiene una persona de desarrollar un accidente cardiovascular, es un verdadero reto en el ámbito médico científico que, hasta ahora, no ha sido resuelto de forma adecuada. Se han obtenido varios modelos logísticos predictivos basados en los resultados de estudios epidemiológicos de cohorte prospectivos a partir de los cuales se han diseñado modelos de predicción clínica y se han construido tablas de riesgo, pero no existe un acuerdo, de cuál es la tabla de riesgo más

---

adecuada para utilizar [16]. Las tablas y fórmulas que se utilizan para calcular el riesgo cardiovascular (RC) de un individuo incluyen como variables la mayor parte de los FRC clásicos.

En los últimos años, múltiples organismos y sociedades científicas apoyan la idea de que la prevención cardiovascular debe afrontarse desde la valoración del RC que tiene un individuo de experimentar un evento de esta naturaleza, o sea, se ha recomendado la estimación del riesgo cardiovascular para clasificar a las personas en los distintos grupos de riesgo. Se plantea que, la evaluación del riesgo cardiovascular constituye una herramienta útil para identificar a los individuos de alto riesgo cardiovascular y así poder implementar medidas intensivas sobre ellos; sin embargo, el problema es que las herramientas propuestas y evaluadas hasta el momento actual son imperfectas, no se ajustan a las características de todas las poblaciones, o simplemente su aplicación es costosa por necesitar de datos de análisis de laboratorio, lo que las convierte en inapropiadas para ciertos sectores sin recursos del mundo [1] [16].

En el año 2007, la OMS propone el uso de sus tablas con colesterol y sin colesterol en países donde no fuese posible hacer la medición del mismo, pero no utilizar las mediciones de glucosa y colesterol son dos barreras importantes en el manejo del riesgo cardiovascular en los países de muy bajos recursos y su no medición no asegura necesariamente el comienzo y sostenibilidad del tratamiento de los pacientes de alto riesgo [1].

Entonces se hace imprescindible la investigación científica para la creación de nuevas técnicas, métodos y herramientas cada vez más sencillas de aplicar, efectivas y económicas, tomando en cuenta la actual crisis económica mundial, para así prevenir precozmente un accidente cardiovascular y salvar más vidas humanas.

En Cuba se han desarrollado varios estudios para determinar el RC, entre ellos, uno realizado por el Instituto de Cardiología y Cirugía Cardiovascular (ICCCV) en la población de trabajadores de la fábrica “Miguel Oramas” (La Polar) durante el año 2002 y otro en los trabajadores del “Hotel Cohíba” en el año 2006. En estos dos estudios se utilizaron las tablas de RC de Framingham [9]. En el 2008 se llevó a cabo un estudio piloto en el área de salud “Héroes de Corynthia” auspiciado por la OMS, el cual tenía como objetivo validar las tablas de RC propuestas por esta organización y que permitían, por primera vez, poder compararnos con Estados Unidos y Canadá (países con semejantes niveles de salud) utilizando un mismo instrumento [1].

Según los resultados de la aplicación de las tablas de RC de la OMS más del 90% de la población fue clasificada de riesgo bajo, pero teniendo en cuenta que la mortalidad por ECV en esta área de salud es alta, hace suponer que estas tablas subestiman el riesgo, lo que pudiera deberse a que no contemplan algunos factores de mayor prevalencia en la población cubana [1].

Los pocos estudios sobre RC realizados en Cuba (los cuales intentan validar las tablas de riesgo de Framingham clásicas, las de la OMS y las de Gaziano en la población cubana) no están basados en estudios de cohorte prospectivos, que evalúen si estas tablas se ajustan bien a la población cubana. Además, muestran una gran variabilidad en la detección de los individuos en alto y mediano riesgo, siendo las de la OMS las de menor porcentaje en alto riesgo (alrededor de un 3%) y al utilizar las de Gaziano, reportan entre un 13% y un 29% de la población entre 40-74 años de edad en ese nivel. Por lo que estos resultados no nos permiten llegar a esclarecer cual tabla deberíamos utilizar [1].

Poder contar en Cuba con un modelo para la evaluación del riesgo cardiovascular que no subestime ni sobreestime el RC de padecer una ECV, es uno de los objetivos de trabajo del MINSAP debido a su importancia para poder establecer las estrategias de intervención, con prioridad en los individuos de alto y moderado riesgo. Además a partir de este modelo, se podrán construir tablas de riesgo autóctonas más precisas, algo muy necesario para el sistema de salud cubano [1].

El modelo resultante tendría un impacto social directo, debido a que en Cuba, las ECV representaron el 36 % de todas las causas de muertes en el 2012 y 37% en el año 2013 [1]. Contar con modelos de riesgo cardiovascular que permitan estratificar la población cubana en diferentes niveles de

---

riesgo, contribuirá a una mejor organización del trabajo de los médicos de la Atención Primaria para establecer intervenciones más eficientes y efectivas.

En el ámbito económico, contribuirá a la reducción de los costos en salud por estas enfermedades, al contar con un instrumento que permitirá identificar tempranamente a los que tienen un alto y mediano riesgo y poder priorizar los recursos.

El principal aporte científico radica en que se contará con un modelo para la evaluación de riesgo cardiovascular para Cuba, lo que resulta necesario, ya que no se cuenta con modelos autóctonos y los utilizados hasta ahora no predicen de manera correcta el riesgo real en la población cubana.

Debido a las características de los modelos de regresión logística, que permiten clasificar individuos dentro de las categorías (presente/ausente) de la variable dependiente, según la probabilidad que tenga de pertenecer a cada una de ellas dada la presencia de determinadas covariables y su objetivo primordial, que es el de modelar cómo influye en la probabilidad de aparición de un suceso, la presencia o no de diversos factores y el valor o nivel de los mismos, además de que también puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías [14]; no cabe duda que la regresión logística es la herramienta estadística más útil para la obtención de un modelo con las características presentadas anteriormente.

Para la construcción del modelo se cuenta con una muestra brindada por el MINSAP, de alrededor de 300 personas entre 40 y 70 años del área de salud “Héroes de Corynthia”, de las cuales tenemos los valores de la edad, el sexo, si es fumador o no, si es diabético o no, los valores de la PAS (Presión Arterial Sistólica), del IMC (Índice de Masa Corporal), de los triglicéridos y su colesterol. Además, se tiene en la muestra una clasificación del riesgo de cada individuo basándose en los resultados de la prueba del GYM<sup>1</sup>. El GYM será utilizado como variable dependiente en la Regresión Logística Ordinal.

El objetivo general es construir un modelo que permita estimar el RC en Cuba. Los objetivos específicos son:

1. Realizar un estudio del modelo de Regresión Logística Ordinal.
2. Identificar el modelo de mejor ajuste para la evaluación del riesgo cardiovascular en la población cubana.
3. Construir una interfaz de usuario usando el software MATLAB que tenga el modelo implementado y permita calcular el RC teniendo información del paciente sobre algunos de los llamados factores de riesgo.

## **2. REGRESIÓN LOGÍSTICA ORDINAL**

Los Métodos de Regresión son esenciales para cualquier análisis de datos que se realice donde se describa la relación entre una variable respuesta y una o varias variables explicatorias independientes. Además, en diferentes ramas de la ciencia se presentan problemas donde se hace necesario trabajar con variables respuestas que por su naturaleza no son continuas. La Regresión Logística es un método efectivo para trabajar con este tipo de problemas, el cual expresa la probabilidad de que ocurra el hecho en cuestión, en función de ciertas variables independientes que se presumen relevantes o influyentes. Estas variables explicatorias independientes reciben el nombre de covariables y pueden ser dicotómicas, nominales, ordinales o continuas. Esta flexibilidad en el tipo de variable le da un mayor atractivo al método [3].

Cuando los eventos cuya probabilidad se desea explicar corresponden a variables dependientes categóricas ordinales, es decir, aquellas cuyos valores no sólo diferencian a los individuos sino que también permiten establecer un orden entre estos, es recomendable el empleo de la regresión logística ordinal, pues con la aplicación de una técnica de predicción que solamente considere la variable dependiente como nominal (por ejemplo, la regresión logística multinomial, el análisis discriminante o

---

<sup>1</sup> El GYM es un examen médico que se realiza en la carótida común, cuyos resultados sirven como predictores de eventos coronarios, es no invasivo, no radioactivo y fácil de realizar.

las redes neuronales), no se captaría cabalmente la influencia de las variables explicativas sobre la variable dependiente, al no considerar la información acerca de las diferencias de orden entre las categorías de esta última [3].

### 3.1. Modelo de Regresión Logística

Si queremos modelar una variable respuesta categórica  $Y$ , de categorías  $y_1, \dots, y_g$ , con un conjunto de variables explicativas (factores o covariables)  $X = (X_1, \dots, X_m)$ , mediante un modelo lineal general<sup>2</sup>, podemos plantearnos las opciones siguientes [2]:

$g$	¿ $Y$ ordinal?	Regresión	Modelamos
2	No importa	Logística	$f(P(Y = y_2 \mathbf{X})) = \alpha + \beta' \mathbf{X}$
$\geq 3$	No	Multinomial	$f(P(Y = y_j \mathbf{X}_i)) = \alpha_j + \beta'_j \mathbf{X}_i, \quad j = 2, \dots, g; \quad i = 1, \dots, n$
$\geq 3$	Sí	Ordinal	$f(\gamma_j(\mathbf{X})) = f(P(Y \leq y_j \mathbf{X})) = \alpha_j + \beta' \mathbf{X}, \quad j = 1, \dots, g - 1$

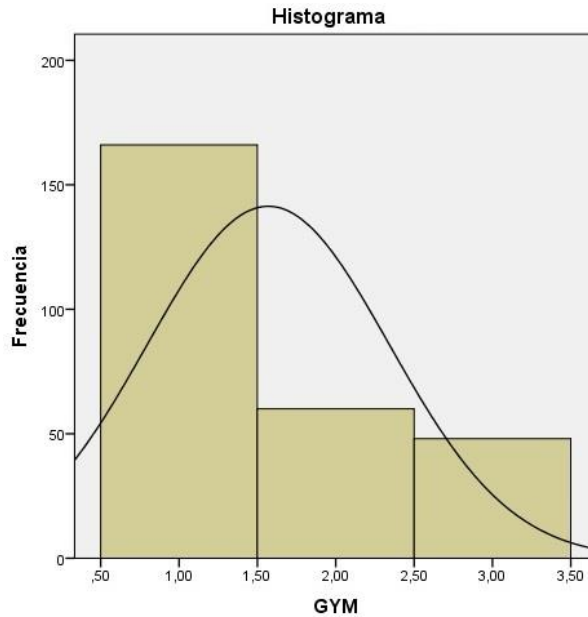
donde  $f(\cdot)$  es la llamada función de enlace,  $\alpha_j + \beta' \mathbf{X}$  es el predictor lineal y  $\alpha_j$  y  $\beta = \beta_1, \dots, \beta_m$  parámetros a estimar. El modelo de regresión ordinal es el adecuado para modelar una variable respuesta ordinal  $Y$ , con categorías ordenadas  $y_1, \dots, y_g$ ,  $g \geq 3$ .

La RLO (Regresión Logística Ordinal) utiliza una función de enlace para relacionar de forma lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la categoría  $i$  de la variable ordinal, y la probabilidad que la variable tome un valor mayor que la categoría  $i$  [3]. Las funciones de unión más empleadas son la Logit y la Cloglog o Log-log complementario (el software SPSS Statistics de la IBM tiene como opciones de funciones de enlace, además, las funciones Cauchit, Log-log negativo y Probit [7]). No existe un criterio que defina con claridad que caso es el más adecuado para cada una de estas funciones, y cuando existen dudas sobre cuál emplear, generalmente se utilizan ambas y se comparan los resultados para escoger los más satisfactorios [5]. No obstante McCullagh [12] plantea que la unión con Logit es la más adecuada para analizar datos ordinales cuya distribución de frecuencia es uniforme a lo largo de todas las categorías, mientras que la unión Cloglog es preferible para analizar datos categóricos cuyas categorías de mayor valor son las más probables.

En el caso de estudio que se muestra en esta investigación los valores de la variable ordinal representan la probabilidad de riesgo de padecer una ECV (baja, media, alta), siendo la categoría baja la más probable según el histograma de frecuencia (ver figura al final del epígrafe), por ende para que sea plausible considerar la unión Logit como la más satisfactoria, se realizó un muestreo aleatorio simple para obtener 60 pacientes del grupo clasificado con riesgo bajo según el GYM y de esta forma, obtener una muestra con distribución de frecuencia uniforme a lo largo de todas las categorías.

Existen diferentes modelos de regresión logística que tienen en cuenta un orden para la variable respuesta [3]. Tres modelos usados comúnmente son: el de categoría-adyacente (adjacent-category), el de razón-continuación (continuation-ratio) y el de razón de probabilidad proporcional (proportional odds). La diferencia principal entre cada uno de estos modelos radica en la razón de probabilidad que utilizan, donde comparan cada respuesta con diferentes variantes del resto de las respuestas. Por ser el modelo más utilizado en la práctica y el que más se ajusta a las características del problema en cuestión (ya que compara las respuestas mayores que  $k$  con las menores o iguales), el modelo de razón de probabilidad proporcional fue el seleccionado para esta investigación.

<sup>2</sup> El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados. En estadística, el modelo lineal generalizado (MLG) es una flexible generalización de la regresión de mínimos cuadrados ordinaria. Relaciona la distribución aleatoria de la variable dependiente en el experimento (la "función de distribución") con la parte sistemática (no aleatoria) (o "predictor lineal") a través de una función llamada la "función de enlace".



### 3.2. Modelo de Razón de Probabilidad Proporcional

La expresión de la función *Logit* para la RLO es la siguiente:

$$\ln(O_i) = \alpha_i + \beta'X$$

En esta ecuación,  $O_i$  es la “razón de probabilidad” (odds) asociada a la categoría  $i$  de la variable dependiente, debido a que se seleccionó el modelo de razón de probabilidad proporcional,  $O_i$  tiene la expresión descrita anteriormente:

$$O_i = \frac{P(\text{riesgo} \leq \text{categ. } i | X)}{P(\text{riesgo} > \text{categ. } i | X)}$$

que es equivalente a:

$$O_i = \frac{P(\text{riesgo} \leq \text{categ. } i | X)}{1 - P(\text{riesgo} \leq \text{categ. } i | X)}$$

Con el término “valor” en las expresiones (2) y (3) se hace referencia a cualquier valor de la variable dependiente. Como se observa, las probabilidades de estas expresiones son condicionales, es decir, dados los valores de las covariables. En la ecuación (1),  $\alpha_i$  es el intercepto asociado a la ecuación que modela la razón de probabilidad de la categoría  $i$ , y  $\beta'$  es el vector de coeficientes de la ecuación de regresión traspuesto [10].

Despejando de la función logarítmica (1) tenemos que:

$$\frac{P(\text{riesgo} \leq \text{categ. } i | X)}{P(\text{riesgo} > \text{categ. } i | X)} = e^{\alpha_i + \beta'X}$$

de donde deriva que:

$$P(\text{riesgo} \leq \text{categ. } i | X) = \frac{e^{\alpha_i + \beta'X}}{1 + e^{\alpha_i + \beta'X}}$$

La última expresión es de suma utilidad, pues posibilita estimar a través de la ecuación obtenida y dado un conjunto de valores de las covariables, la posibilidad de que la variable dependiente sea menor o igual que un valor. Si existen  $p$  covariables, existen  $p$  coeficientes, por lo que  $\beta'X$  es la combinación lineal  $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ . Si la variable dependiente tiene  $k$  categorías, existen  $k - 1$

---

ecuaciones, ya que a la categoría mayor no se le asocia  $O_i$  al ser la probabilidad acumulada hasta ésta igual a uno.

El modelo de regresión ordinal descrito anteriormente es denominado “modelo logit acumulado” [3] ya que es construido basándose en las probabilidades acumuladas de la variable respuesta dados los valores de las variables explicativas. También es denominado “modelo de razón de probabilidad proporcional” (“proportional odds model” en inglés) ya que los coeficientes de regresión son independientes de las categorías de la variable dependiente, siendo los mismos en las  $k - 1$  ecuaciones que se forman para las  $k$  categorías. Esto implica asumir que la relación entre las covariables y la variable dependiente ordinal, es independiente de las categorías de esta última, y por tanto que los cambios en las covariables provocan el mismo cambio en la razón de probabilidad acumulada de todas las categorías.

Por tanto, existen  $k - 1$  ecuaciones con los mismos coeficientes acompañando a las covariables, que sólo se diferencian en el valor del intercepto  $\alpha_i$ .

### 3.3. Estimación de los Parámetros del Modelo

Para estimar los coeficientes de la ecuación de regresión (o sea los  $\alpha_j$  y los  $\beta_i$ ) se utilizará la estimación de máxima verosimilitud<sup>3</sup> (ver [4]). Este método consiste en encontrar los valores de los parámetros que maximizan la función de verosimilitudes, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente  $Y$  proporcionados por los datos de la muestra. Si se desea profundizar en la forma que se emplea este método de estimación se puede consultar [15].

### 3.4. Pruebas de Hipótesis para la Significación de los Parámetros

Las pruebas de hipótesis sobre los parámetros del modelo, son de gran importancia para probar la significación de cada una de las covariables. La significación de los coeficientes de las covariables en cualquier modelo, busca responder la pregunta de que si el modelo que incluye ciertas covariables dice más acerca de la variable respuesta que el modelo que no las contiene.

### 3.5. Significación Global del Modelo

Un primer aspecto importante es comprobar la significación global del modelo, lo que significa evaluar si en conjunto las covariables tienen un efecto estadísticamente significativo sobre la variable dependiente.

Se quiere probar la siguiente hipótesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \text{ vs. } H_A: \text{ existe } i \text{ tal que } \beta_i \neq 0, i = 1, \dots, p$$

Con este fin se emplea el estadístico  $G$ , calculado según la fórmula:

$$G = -2 \ln \left[ \frac{\text{verosimilitud de la muestra sin covariables}}{\text{verosimilitud con la ecuación que incluye algunas covariables}} \right]$$

Este estadístico sigue distribución chi-cuadrado con tantos grados de libertad como covariables existan. Cuando las predicciones de la variable dependiente que se hacen con el modelo que incluye todas las covariables superan las predicciones que se realizan sin considerar éstas, el valor de  $G$  tiende a ser grande y conlleva a rechazar la hipótesis nula  $H_0$ , concluyendo que al menos una de las covariables tiene efecto sobre la variable dependiente y por lo tanto, que la probabilidad de ocurrencia de los valores que representan esta variable, varía para alguna de las combinaciones de valores de las covariables [5][14][15].

---

<sup>3</sup> Se considera que el método de máxima verosimilitud, tal y como hoy lo conocemos e interpretamos, fue propuesto por Fisher (1890-1962), aunque ya de una forma mucho más artificiosa fue inicialmente atisbado por Bernoulli (1700-1782), cuyo planteamiento fue revisado y modificado por su coetáneo y amigo, el gran matemático Euler (1707-1783).



---

### 3.6. Significación Individual de las Covariables

La significación individual de cada covariable generalmente se analiza a través de la prueba de hipótesis de Wald, la cual se basa en el valor de significación del coeficiente estimado para cada variable. Esta prueba se utiliza cuando se ha rechazado la hipótesis nula de la prueba de significación global del modelo.

Las hipótesis son:

$$H_0: \beta_i = 0 \text{ vs. } H_A: \beta_i \neq 0$$

El estadígrafo de prueba utilizado es el siguiente:

$$Z_{Wald} = \frac{\beta_j}{ES(\beta_j)}$$

donde  $\beta_j$  es el coeficiente de regresión estimado para la variable independiente  $j$ .

Bajo la hipótesis de que el coeficiente  $\beta_j = 0$  para la variable  $j$ , la razón entre la estimación de este coeficiente ( $\beta_j$ ) y el error estándar<sup>4</sup> de esta estimación ( $ES(\beta_j)$ ), sigue una distribución normal estándar. Mayores valores de este estadígrafo indican que el coeficiente  $\beta_j$  es distinto de cero, o sea se rechaza  $H_0$ , y por ende la covariable tiene efecto sobre la probabilidad de ocurrencia de los valores de la variable dependiente. En modelos con errores estándar grandes, el estadístico de Wald puede proporcionar falsas ausencias de significación [5] [14] [15].

### 3.7. Odds Ratio

Los odds asociados a cierto suceso se definen como la razón entre la probabilidad de que dicho suceso ocurra y la probabilidad de que no ocurra, es decir, es el número que expresa cuanto más probable es que se produzca el suceso frente a que no se produzca [15].

Si llamamos  $E$  a dicho suceso,  $p(E)$  a la probabilidad de que ocurra y  $O(E)$  al correspondiente odds entonces:

$$O(E) = \frac{p(E)}{1 - p(E)}$$

Por otra parte, conocido el valor del odds de un suceso se puede calcular su probabilidad puesto que:

$$p(E) = \frac{O(E)}{1 + O(E)}$$

Entre la probabilidad de un suceso y el correspondiente odds hay una relación directa evidente, si la probabilidad aumenta, el odds también aumenta y viceversa.

A la razón entre dos odds se le llama odds ratio y no es más que la razón del odds correspondiente a un suceso bajo cierta condición, entre la que le corresponde bajo otra.

Como el efecto que tiene una determinada covariable es el mismo para todas las categorías de la variable dependiente, para cada covariable se determina un solo odds ratio. El odds ratio es utilizado para interpretar el efecto de las covariables sobre la variable objeto de estudio. Si éste es igual a 1 indica que la covariable no tiene efecto. Si es menor que 1, lo cual sucede cuando el coeficiente de la covariable es negativo, indica que, si las otras variables independientes permanecen constantes, los cambios en la covariable analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable objeto de estudio. Valores de odds ratio mayores que 1, muestran que las variaciones en la covariable disminuyen la probabilidad de obtener categorías de mayor valor de la variable dependiente [5] [15].

### 3.8. Variables dummy

---

<sup>4</sup> El error estándar es la desviación estándar de la distribución muestral de un estadístico.

Si la covariable es nominal u ordinal y tiene más de dos categorías, para su inclusión en el modelo debe realizarse una transformación de la misma en varias covariables dicotómicas ficticias o de diseño (las llamadas variables *dummy*), de forma tal que una de las categorías se tomaría como categoría de referencia. Con esta transformación, cada categoría entraría en el modelo de forma individual [3] [14] [15].

Las variables *dummy* las denotaremos por  $D_1, \dots, D_{c-1}$  y se definen de la siguiente manera: Si la variable tiene  $c$  categorías o clases, entonces se pueden construir  $c - 1$  variables *dummy* que son dicotómicas.

Si el sujeto pertenece a la primera categoría, entonces las  $c - 1$  variables valen 0, es decir  $D_1, \dots, D_{c-1} = 0$ , si se halla en la segunda categoría entonces  $D_1 = 1$  y los restantes valen 0, si se encuentra en la tercera categoría  $D_2 = 1$  y las restantes valen 0, y así sucesivamente hasta llegar a la última categoría, en la cual  $D_{c-1}$  es la única que vale 1.

Como notación tenemos que si la  $i$ -ésima covariable  $X_i$  tiene  $C_i$  niveles ( $C_i \geq 2$ ), entonces las  $C_i - 1$  variables *dummy* se denotarán por  $D_{ij}$  y los coeficientes por  $\beta_{ij}$  donde  $j = 1, \dots, C_i - 1$ . Teniendo en cuenta esto la función *logit* tendrá la siguiente expresión:

$$g(X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \sum_{j=1}^{C_i-1} \beta_{ij} D_{ij} + \dots + \beta_p X_p$$

### 3.9. Selección de las Variables

En esta sección se responde la pregunta: ¿Del conjunto de variables que pueda tener un estudio, qué variables deben introducirse en el modelo?

El modelo debe ser el más reducido que explique los datos (principio de parsimonia<sup>5</sup>), y que además sea clínicamente congruente e interpretable [14]. Hay que tener en cuenta que un mayor número de variables en el modelo implicará mayores errores estándar. Deben incluirse todas aquellas variables que se consideren clínicamente importantes para el modelo, con independencia de si en un análisis previo se demostró o no su significación estadística. Por otro lado, no debería dejarse de incluir toda variable que en un análisis previo demostrara una relación “suficiente”, o sea, se desea evitar que se deje 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 las covariables.

Una cuestión importante a tener en cuenta es el correcto manejo de las variables transformadas en varias variables *dummy*. Siempre que se decida incluir (o excluir) una de estas variables, todas sus correspondientes variables *dummy* deben ser incluidas (o excluidas) en bloque.

Una vez se dispone de un modelo inicial, debe procederse a su reducción hasta obtener el modelo más reducido que siga explicando los datos.

Otro elemento que siempre debe considerarse, es la suficiencia del tamaño muestral para el número de covariables que se desea incluir en el modelo. Modelos excesivamente grandes para muestras con tamaños muestrales relativamente pequeños implicarán errores estándar grandes o coeficientes estimados falsamente muy elevados (sobreajuste) (ver [14]).

### 3.10. Ajuste del Modelo

<sup>5</sup> La navaja de Ockham, principio de economía o principio de parsimonia, es un principio metodológico y filosófico atribuido a Guillermo de Ockham (1280-1349), según el cual, “en igualdad de condiciones, la explicación más sencilla suele ser la correcta”. Esto implica que, cuando dos teorías en igualdad de condiciones tienen las mismas consecuencias, la teoría más simple tiene más probabilidades de ser correcta que la compleja.

Las pruebas de bondad de ajuste deben realizarse una vez que el modelo de regresión logística contenga las variables que son significativas para él, con el propósito de comprobar si el modelo se ajusta adecuadamente a los datos usados y examinar la efectividad del modelo en la descripción y estimación (predicción) de la variable respuesta.

Usualmente para evaluar el ajuste se construye una tabla de contingencia cuyas filas representan los valores de la variable objeto de estudio y cuyas columnas representan las posibles combinaciones de valores de las covariables. El empleo de una prueba de bondad de ajuste para comparar en cada celda de la tabla los valores observados y los valores predichos según el modelo es útil para valorar la calidad de ajuste. Si la frecuencia predicha para las combinaciones según el modelo, difiere significativamente de la frecuencia con la cual ocurren realmente los valores en estas combinaciones, existe evidencia de falta de ajuste. Un modelo se dice que presenta un buen ajuste a los datos si los valores predichos por él reflejan de forma adecuada a los valores observados [5].

Una forma de medir la adecuación de un modelo es proporcionando medidas globales de bondad de ajuste a través de pruebas o test estadísticos contruidos con tal fin. Lo anterior es válido para cualquier modelo de regresión, pero en el caso particular de la regresión logística no existe uniformidad en cuanto al test a utilizar.

### 3.11. Prueba basada en el estadístico Chi-Cuadrado de Pearson

Uno de los test estadísticos más utilizados es la prueba Chi-Cuadrado [8]. Este método se caracteriza por recurrir a una prueba de hipótesis formal para comprobar lo adecuado del modelo a los datos.

Las hipótesis a probar son:

$H_0$ : los datos se ajustan al modelo supuesto vs.  $H_A$ : los datos no se ajustan al modelo supuesto.

El estadístico chi-cuadrado de Pearson tiene la siguiente forma:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

En la expresión anterior se considera que la variable dependiente tiene  $k$  categorías y que se forman  $m$  combinaciones de valores con las covariables, de manera que:

- $O_{ij}$  es la frecuencia observada de la  $i$ -ésima categoría de la variable dependiente en la  $j$ -ésima combinación de valores de las covariables.
- $E_{ij}$  es la frecuencia esperada de la  $i$ -ésima categoría de la variable dependiente en la  $j$ -ésima combinación de valores de las covariables.

Mientras mayor sea el valor del estadístico  $\chi^2$  mayor sospecha de falta de ajuste.

### 3.11 Prueba basada en el estadístico devianza D

Otro test muy utilizado manteniendo las mismas hipótesis a probar que en el caso anterior, es el basado en la devianza  $D$  [8]. El estadístico  $D$  tiene la siguiente expresión:

$$D = -2 \sum_{i=1}^n \left( y_i \log \left( \frac{\hat{p}}{y_i} \right) + (1 - y_i) \log \left( \frac{1 - \hat{p}}{1 - y_i} \right) \right)$$

Para poder aplicar tanto el test basado en la devianza como el basado en el estadístico chi-cuadrado, tiene que verificarse que el número de observaciones para cada combinación de las covariables sea grande, es por ello, que no es recomendable aplicar estos métodos en el caso de covariables continuas.

Respecto a las ventajas de estos test, destaca su implementación en todos los programas estadísticos dada la simplicidad en su cálculo, mostrándose casi siempre tanto el valor del estadístico como el  $p$  valor asociado.

### 3.12. Medidas de Bondad de Ajuste Pseudo $R^2$

Distintas medidas de bondad de ajuste similares al  $R^2$  usado en modelos de regresión lineal fueron propuestas para medir la bondad de ajuste de un modelo de regresión logística.

Cox y Snell (1989) propusieron una medida tratando de generalizar el  $R^2$  que hoy lleva su nombre [8]:

$$R_g^2 = 1 - \left( \frac{\hat{L}_c}{\hat{L}_0} \right)^{\frac{2}{n}}$$

siendo  $\hat{L}_c$  la log-verosimilitud del modelo evaluado en los coeficientes estimados de máxima verosimilitud ( $\beta_i$ ),  $\hat{L}_0$  es la log-verosimilitud del modelo que solo incluye la constante (el intercepto) y  $n$  el número de casos.

Otra medida muy utilizada fue propuesta por Nagelkerke (1991) ajustando el valor máximo de  $R_g^2$  a 1 [8]:

$$\bar{R}_g^2 = \frac{R_g^2}{\max(R_g^2)}$$

donde  $\max(R_g^2) = 1 - \hat{L}_0^{\frac{2}{n}}$ .

Mayores valores implican un mayor ajuste y asociación entre la variable dependiente y las covariables [6]. Estas medidas toman valores entre 0 y 1 teóricamente, aunque en la práctica es muy difícil que tomen valores cercanos a 1.

## 4. MODELO DE RLO PARA LA EVALUACIÓN DEL RC para Cuba

Se quiere construir un modelo de Regresión Logística Ordinal que permita pronosticar el riesgo de padecer una enfermedad cardiovascular. La variable dependiente usada será la variable ordinal GYM<sup>6</sup>, cuyos valores posibles son: “bajo”, “medio” y “alto”. Como covariables se utilizarán diferentes parámetros que miden el estado de salud de un individuo (colesterol), hábitos (fuma o no) y características que diferencian a un organismo de otro (edad, sexo). Para ajustar el modelo se utilizará la base de datos de la población de la localidad “Héroes de Corynthia”.

### 4.1. Obtención del Modelo

Tenemos 8 covariables que son: edad, fumador (si o no), PAS, diabético (si o no), IMC (índice de masa corporal), sexo (1 masculino, 0 femenino), valor triglicéridos bien (si o no), valor del colesterol bien (si o no). Como se puede observar, 5 son binarias, la edad y el IMC se transformaron de continuas a ordinales (1, 2, 3) y la covariable PAS es continua. Las dos covariables ordinales se transforman en variables dummies.

Para aplicar la RLO al conjunto de datos se usó el software IBM SPSS Statistics (versión 20). Según las especificidades del software para realizar una RLO, como variable dependiente tomaremos el GYM, como covariables la continua (PAS) y como factores el resto, o sea, las variables ordinales (edad e IMC) y las binarias. Como función de enlace se usará la *logit* y el SPSS Statistics usa la estimación de máxima verosimilitud para estimar los parámetros por defecto [7] [10].

<sup>6</sup> En la base de datos brindada por el MINSAP esta variable es ordinal y sus categorías son 1 (riesgo bajo, GYM < 0.86 mm), 2 (riesgo moderado, GYM entre 0.89 y 1.49 mm) y 3 (riesgo alto GYM > 1.49 mm).

Debido a los resultados en la Prueba de Wald luego de realizar la RLO con todas las covariables, se decidió eliminar del modelo la covariable “Triglicéridos” por su resultado extremadamente pequeño en dicha prueba (0.005) y su valor muy elevado de significación (0.941), que nos lleva a la conclusión de no rechazar la hipótesis nula.

Al aplicar la RLO sin considerar la covariable “Triglicéridos” los resultados fueron los siguientes:

**Información sobre el ajuste de los modelos**

Modelo	-2 log de la verosimilitud	Chi-cuadrado	gl	Sig.
Sólo intersección	367,374			
Final	289,514	77,860	9	,000

Función de vínculo: Logit.

Como se puede observar, al ser el valor de significación del estadígrafo *G* sumamente pequeño, se concluye que al menos una de las covariables tiene efecto sobre la variable dependiente, y por lo tanto que la probabilidad de ocurrencia de los valores que representan esta variable, varía para alguna de las combinaciones de valores de las covariables.

Los resultados de las pruebas de bondad de ajuste de los estadígrafos chi-cuadrado y *D* indicaron que no hay buen ajuste (valores de significación 0.000017 y 0.000029 respectivamente), pero existen varios factores que atentan contra estos estadísticos haciendo que no sean realmente significativos y por lo tanto no deben ser tomados como resultados fiables. Primero, en el modelo tenemos una variable continua (PAS), cuya presencia afecta el cálculo de estos estadígrafos, debido a que la frecuencia de cada categoría de esta variable es muy baja. Esto último se ve representado también en varias covariables que no son continuas debido al tamaño relativamente pequeño de la muestra, que hace que muchas de las categorías de estas covariables no estén ampliamente representadas. Además, hay un número significativo de casillas en el modelo con frecuencia cero, algo que afecta también a estos estadígrafos. Debido a lo anterior para analizar el ajuste del modelo, se construyó la tabla de contingencias entre el GYM (valores observados) y la categoría de respuesta pronosticada (valores predichos) utilizando la misma muestra y se obtuvo el porcentaje de las bien clasificadas y mal clasificadas. Además se calcularon también los estadísticos pseudo *R*<sup>2</sup>.

Categorías	Mal Clasificadas
1	3
2	7
3	4
Total	14
%	8.3

Table 1: Tabla Mal Clasificadas

De aquí se obtiene que el modelo presenta buen ajuste (91.7% bien clasificadas), o sea, que las probabilidades de ocurrencia de los valores de la variable dependiente que se estiman según el modelo para las diferentes combinaciones de las covariables, no divergen significativamente de la frecuencia con la cual ocurren en la muestra los valores de la variable dependiente para estas combinaciones.

**Pseudo R-cuadrado**

Cox y Snell	,371
Nagelkerke	,418

Las medidas de Cox-Snell y Nagelkerke sobre la bondad de ajuste indican un valor de explicación de aproximadamente el 40%, lo cual, basándonos en la bibliografía consultada, significa un ajuste aceptable.

Debido a estos últimos resultados, llegamos a la conclusión de que el modelo obtenido presenta buen ajuste y es el modelo de mejor ajuste entre los modelos que consideran las variables clínicamente importantes para el RC.

Analizando los valores de la Prueba de Wald y su valor de significación (ver figura 7) se obtiene que las covariables más significativas son la edad, el IMC y si es diabético o no. La covariable “Fumar” aunque no tuvo un resultado muy bueno en dicha prueba se mantiene en el modelo debido a que es clínicamente importante para el mismo. Además la estimación de la covariable “Fumar” es negativa y las covariables cuyas estimaciones son negativas indican que el odds ratio asociado es menor que uno, lo cual indica que, si las otras variables independientes permanecen constantes, los cambios en la covariable analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable dependiente. Esto permite plantear, que casi todas las covariables tienen odds ratio asociado menor que uno.

Estimaciones de los parámetros

		Estimación	Error típ.	Wald	gl	Sig.	Intervalo de confianza 95%	
							Límite inferior	Límite superior
Umbral	[GYM = 1,00]	-.208	1,488	,020	1	,889	-3,125	2,708
	[GYM = 2,00]	1,978	1,499	1,740	1	,187	-.961	4,916
Ubicación	PAS	,020	,009	5,287	1	,021	,003	,038
	[Edad=1,00]	-2,055	,475	18,749	1	,000	-2,985	-1,125
	[Edad=2,00]	-1,048	,393	7,134	1	,008	-1,818	-,279
	[Edad=3,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[IMC=1,00]	,954	,470	4,122	1	,042	,033	1,876
	[IMC=2,00]	1,513	,434	12,131	1	,000	,662	2,364
	[IMC=3,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Sexo=,00]	-.838	,348	5,821	1	,016	-1,520	-,157
	[Sexo=1,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Colesterol=,00]	-.695	,357	3,780	1	,052	-1,395	,006
	[Colesterol=1,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Diabético=,00]	-1,537	,478	10,336	1	,001	-2,474	-,600
	[Diabético=1,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Fumar=,00]	-.303	,341	,789	1	,374	-,972	,366
[Fumar=1,00]	0 <sup>a</sup>	.	.	0	.	.	.	

Función de vínculo: Logit.

a. Este parámetro se establece en cero porque es redundante.

Como tenemos tres categorías (bajo, medio, alto), el modelo tiene dos ecuaciones. El modelo obtenido es el siguiente sistema:

$$P(\text{riesgo} \leq \text{categ. 1} | X) = \frac{e^{-0.208 + B}}{1 + e^{-0.208 + B} + e^{-1.978 + B}}$$

$$P(\text{riesgo} \leq \text{categ. 2} | X) = \frac{e^{-1.978 + B}}{1 + e^{-0.208 + B} + e^{-1.978 + B}}$$

donde

$$B = 0.020X_1 - 2.005D_{21} - 1.048D_{22} + 0.954D_{31} + 1.513D_{32} - 0.838X_4 - 0.695X_5 - 1.537X_6 - 0.303X_7$$

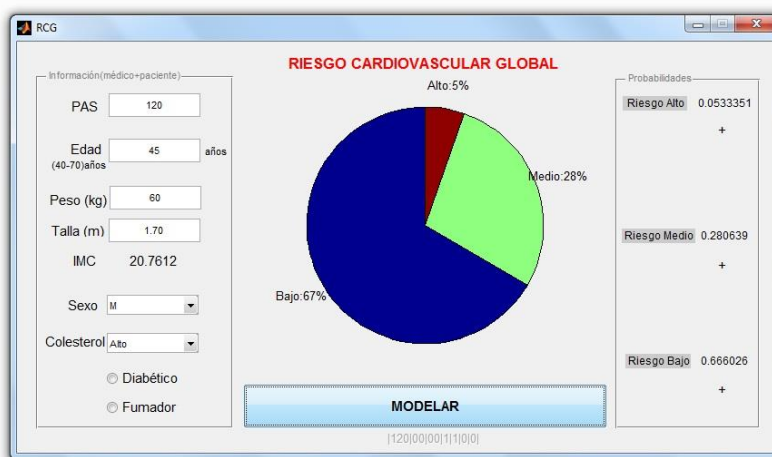
La primera ecuación calcula la probabilidad para la categoría riesgo bajo y la segunda calcula la probabilidad para la categoría riesgo medio y bajo (recordemos que el modelo es el logit acumulado), por lo que, para obtener la probabilidad de la categoría riesgo medio, debemos a la probabilidad obtenida por la segunda ecuación, restarle la probabilidad de riesgo bajo, o sea, la primera ecuación.

$$P(\text{riesgo medio} | X) = \frac{e^{-1.978 + B}}{1 + e^{-0.208 + B} + e^{-1.978 + B}} - \frac{e^{-0.208 + B}}{1 + e^{-0.208 + B} + e^{-1.978 + B}}$$

La probabilidad de la categoría riesgo alto, se obtiene de restarle al valor 1 las probabilidades de las categorías riesgo bajo y medio.

## 5. IMPLEMENTACIÓN EN MATLAB

Después de obtenida la expresión del modelo de mejor ajuste que permite predecir el riesgo de padecer una enfermedad cardiovascular, se implementó el mismo en MATLAB y se construyó una interfaz usando el mismo software [11], que permite ingresar los parámetros de las variables del modelo escogido y obtener la probabilidad de cada categoría de la variable RC (bajo, medio y alto) y un gráfico de pastel que muestra estas probabilidades de una forma más asequible.



Cuando se carga por primera vez, esta muestra ciertos parámetros por defecto. Como se puede observar en el panel de la izquierda, el médico selecciona o introduce los parámetros del paciente, si entra un valor no numérico o exagerado para el rango posible de medición de estos factores, la interfaz le notificará.

Introduciendo la talla y el peso, la interfaz brinda como salida el cálculo del IMC, lo cual puede servir para un diagnóstico más completo (este resultado es el utilizado por el modelo). Como la interfaz indica, el peso debe ser en kilogramos y la talla en metros, MATLAB para separar los decimales utiliza el punto y no la coma.

En el centro de la interfaz después de entrados los parámetros deseados y presionado el botón **Modelar**, se plotea un gráfico de pastel con las probabilidades de cada una de las categorías de riesgo cardiovascular del paciente analizado. También se muestra en porcentos estas probabilidades y se aclara a que categoría pertenece. El color azul corresponde a riesgo bajo, el verde a riesgo medio y el rojo a riesgo alto.

En el panel derecho se brindan estas probabilidades en forma numérica con seis cifras decimales. Debajo de las mismas aparece un signo más o menos, el cual significa que esa categoría del riesgo aumentó o disminuyó al modificar los parámetros del paciente. Al presionar el botón **Modelar** por primera vez, siempre estos indicadores mostrarán un signo + en todas las categorías.

## 6. CONCLUSIONES

Usando Regresión Logística Ordinal se ha obtenido un modelo que, basado en datos de 7 covariables, permite clasificar de forma cualitativa y cuantitativa el riesgo que tiene una persona que vive en Cuba de padecer una enfermedad cardiovascular.

El modelo obtenido presenta buen ajuste y las covariables incluidas son significativas para la variable dependiente.

Se ha construido en MATLAB una interfaz que permite de forma sencilla para el personal de la salud, calcular el riesgo cardiovascular de un paciente.

## REFERENCIAS

- 
1. Armas, N. (2015). Programa ramal: determinantes de salud, riesgos y prevención de enfermedades en grupos vulnerables. Proyecto: propuesta de modelo de evaluación de riesgo cardiovascular para la población cubana. MINSAP, 1-11.
  2. Barrera, J. (2009). Introducción a la regresión ordinal. Universidad Autónoma de Barcelona.
  3. Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. (2013). Applied logistic regression. Third Edition, John Wiley and Sons, Inc, New Jersey.
  4. Held, L., Sabanés, D. (2014). Applied statistical inference, likelihood and bayes. Springer, Berlin.
  5. Heredia, J., Rodríguez, A.G., Vilalta, J. (2012). Empleo de la regresión logística ordinal para la predicción del rendimiento académico. Revista Inv. Oper, 33, 252-267.
  6. Hu, B., Shao, J., Palta, M. (2006). Pseudo- $R^2$  in logistic regression model. University of Wisconsin, Madison.
  7. IBM. (2011). Manual de usuario del IBM SPSS. Statistics, 69-89.
  8. Iglesias, T. (2013). Métodos de bondad de ajuste en regresión logística. Tesis de Máster en Estadística Aplicada, Universidad de Granada.
  9. Kannel, W.B., McGee, D., Gordon, T. (1975). A general cardiovascular risk profile: the framingham study. Bethesda, Maryland.
  10. Martínez, D. (2004). El análisis de la regresión a través de SPSS. Universidad de Granada.
  11. MathWorks. (2012). Ayuda del software MATLAB.
  12. McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society, 42, 109-142.
  13. Minitab. (2005). Ayuda del software.
  14. Paz, J.M. (2009). Regresión logística.
  15. Sistachs, V. (2005). Un estudio del modelo de regresión logística binario bajo el paradigma bayesiano. Tesis de Doctorado, Universidad de la Habana.
  16. Suárez, C. (2006). Protocolos riesgo cardiovascular. Sociedad Española de Medicina Interna, Elsevier, 2da Edición.



# ANÁLISIS ESPACIAL DE MODELOS PARA LA ESTIMACIÓN DE RIESGO RELATIVO BAJO EL PARADIGMA BAYESIANO

Ivette Raíces Cruz\*, Vivian Sistachs Vega\*\*, Angela Tuero Iglesias\*

\*Centro de Ingeniería Genética y Biotecnología

\*\* Universidad de La Habana

## ABSTRACT

Hierarchical Bayesian models were used to estimate the Relative Risk (RR), specifically the models: Log-Normal, CAR and Convolution. This allowed identifying the geographic distribution of risk patterns at 168 municipalities from Cuba in 2011 for mortality associated first with respiratory cancer and second with pneumonia and influenza. We used DIC as a goodness of fit criteria. We obtain as a result that the Convolution model maps shows clearly geographic distribution of risk patterns to die for both diseases. This model was selected as the best in both cases. The software used was WinBUGS 1.4.

**KEYWORDS:** Bayesian Models, Relative Risk, Log-Normal Model, CAR Model, Convolution Model, Respiratory Cancer, Pneumonia and Influenza

## RESUMEN

Se emplearon modelos jerárquicos Bayesianos para la estimación de Riesgo Relativo (RR), específicamente los modelos: Log-Normal, CAR y Convolución. Esto permitió identificar los patrones de riesgo en la distribución geográfica en los 168 municipios de Cuba en el 2011 para la mortalidad por cáncer de las vías respiratorias y para la mortalidad por neumonía e influenza. Se utilizó como criterio de selección de modelo el DIC. Como resultados se obtuvo que los mapas del modelo Convolución evidencian claros patrones de distribución geográfica del riesgo de morir por ambas enfermedades. Este modelo fue seleccionado como el mejor en ambos casos. El software utilizado fue WinBUGS 1.4.

## 1. INTRODUCCIÓN

En Cuba, el cáncer de las vías respiratorias (tráquea, bronquios y pulmón) y la neumonía e influenza fueron la primera y cuarta causa de mortalidad en el 2012 respectivamente, ver [1]. Por esta razón, es de interés para los decisores de salud encontrar los patrones de Riesgo Relativo (RR) de estas enfermedades.

En la actualidad, los métodos jerárquicos Bayesianos están siendo utilizados en las ciencias aplicadas. Su empleo se ha visto favorecido por el desarrollo de herramientas de simulación y computación; ver [7, 12].

En este artículo se emplean modelos jerárquicos Bayesianos para el estudio de la mortalidad de las enfermedades mencionadas anteriormente. Específicamente los modelos utilizados son: Log-Normal, CAR (autorregresivo condicional) y Convolución. El modelo Log-Normal no considera la correlación espacial entre las áreas, por eso aquí se emplea para estudiar la distribución de la mortalidad por cáncer de las vías respiratorias, ya que esta es una enfermedad no transmisible. En cambio, el modelo CAR, sí toma en cuenta la correlación espacial entre las áreas, por esa razón este se usa para estudiar la neumonía e influenza, enfermedad transmisible. También se utiliza el modelo Convolución, pues este es una combinación de los dos anteriores. Además se emplea como criterio de selección de modelo el DIC (Criterio de Información Deviance) y el software WinBUGS 1.4. Asimismo, se usa la prueba de convergencia de Gelman y Rubin.

## 2. MODELOS PARA LA ESTIMACIÓN DE RR

En los últimos años se ha podido avanzar en la reformulación y tratamiento del cálculo y mapeo del RR para la morbi-mortalidad de enfermedades a través del enfoque Bayesiano. Este enfoque cuantifica la incertidumbre mediante la probabilidad, donde los parámetros de los modelos son tratados como variables aleatorias. La base de este paradigma es encontrar la distribución *a posteriori* de los parámetros; ver [3].

Para la estimación de RR bajo el paradigma Bayesiano se utilizaron modelos jerárquicos Bayesianos ya que incorporan estructura compleja al modelo, debido a que su modelación toma en cuenta diversos parámetros, donde se suponen diferentes distribuciones para ellos en dependencia del

problema a tratar. Además este paradigma permite combinar información y fuentes de variabilidad, para unificar los modelos; ver [8].

Un modelo jerárquico de Bayes,  $(f(x|\theta), \pi(\theta))$ , es un modelo estadístico Bayesiano donde la distribución *a priori*  $\pi(\theta)$  se descompone en distribuciones condicionales

$$\pi_1(\theta_1|\theta_1), \pi_2(\theta_2|\theta_1), \dots, \pi_n(\theta_n|\theta_{n-1}),$$

y una distribución marginal  $\pi_{n+1}(\theta_n)$  tal que:

$$\pi(\theta) = \int_{\theta_1 \times \dots \times \theta_n} \pi_1(\theta_1|\theta_1) \pi_2(\theta_2|\theta_1) \dots \pi_n(\theta_n|\theta_{n-1}) \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n$$

los parámetros  $\theta_i$  son llamados hiperparámetros de nivel  $i$  ( $1 \leq i \leq n$ ); ver [6].

Una de las ventajas de utilizar los modelos jerárquicos Bayesianos es que usándolos la distribución *a priori* se puede descomponer en dos partes principales: una que se refiere a información estructural o suposiciones acerca del modelo y otra que se refiere a la información subjetiva o no informativa de los parámetro en niveles más altos de jerarquía en el modelo. Otra ventaja es que esa estructura jerárquica lleva a un análisis más robusto, reduciendo la subjetividad. Finalmente, la estructura jerárquica simplifica la interpretación y el cómputo del modelo mientras que la distribución *a posteriori* correspondiente es simplificada, produciendo distribuciones condicionales más simples; ver [2, 4].

## 2.1. Modelo Log-Normal

El modelo Log-Normal asume que los casos observados  $Y_i = \overline{1, N}$  tienen una distribución Poisson con media  $E_i \xi_i$ , donde  $E_i$  son los casos esperados, este será un valor conocido y  $\xi_i$  es el RR desconocido para el área  $i$ ; ver [6].

$$[Y_i/E_i \xi_i] \sim \text{Poisson}(E_i \xi_i).$$

Comenzaremos denotando por  $\eta_i$  al logaritmo natural del RR desconocido para el área  $i$ :

$$\eta_i = \log \xi_i,$$

donde  $\eta_i$  puede ser modelado como la suma de una media global, denotada por  $\mu$ , que expresa el nivel global del log-riesgo relativo por todo el mapa y un efecto específico de área, denotado por  $v_i$ , que representa la diferencia entre el log-riesgo relativo para el área  $i$  y la media global. Por lo tanto para la  $i$ -ésima área genérica podemos escribir; ver [6],

$$\log \xi_i = \eta_i = \mu + v_i,$$

luego

$$[Y_i/E_i, \mu, v_i] \sim \text{Poisson}(E_i e^{\mu+v_i}).$$

En este modelo hay intercambiabilidad de áreas, es decir no hay estructura de correlación, donde  $v_i$  es el efecto de heterogeneidad no correlacionada. En el modelo se supone que este parámetro seguirá una distribución no informativa: específicamente distribución Normal con media 0 y varianza  $\frac{1}{h}$ , es decir:

$$v_i \sim N\left(0, \frac{1}{h}\right),$$

ya que esta es una de las empleadas cuando los parámetros pueden tomar valores en un amplio rango, esta distribución también ha sido utilizada en otros trabajos; ver [6, 11]. De esta forma todas las áreas tienen el mismo peso en el modelo.

También se supone que la distribución *a priori* del hiperparámetro  $\frac{1}{h}$  sigue una distribución Gamma con

parámetros  $\alpha$  y  $\beta$  conocidos, es decir:  $\frac{1}{h} \sim \text{Gamma}(\alpha, \beta)$ .

## 2.2. Modelo Autorregresivo Condicional (CAR)

La intercambiabilidad estricta debe ser descartada cuando hay creencias *a priori* de correlaciones espaciales entre los RR de las diferentes regiones.

De manera similar que en el modelo anterior los  $Y_i$  son mutuamente independientes y se distribuyen de acuerdo a una distribución de Poisson con media  $E_i \xi_i$ :

$$[Y_i/E_i \xi_i] \sim \text{Poisson}(E_i \xi_i),$$

donde  $\eta_i$  es el logaritmo natural del RR desconocido para el área  $i$  como se realizó en el modelo anterior:

$$\eta_i = \log \xi_i,$$

Además se asume que  $\eta_i$  puede ser modelado como la suma de una media global, denotada por  $\mu$ , que expresa el nivel global del log-riesgo relativo por todo el mapa y un efecto específico de área,

denotado por  $\phi_i$ , que representa la diferencia entre el log-riesgo relativo para el área  $i$  y la media global. Aquí a diferencia del modelo anterior los  $\phi_i$  representan el efecto de heterogeneidad correlacionada, es decir, reflejan correlación espacial:

$$\log \xi_i = \eta_i = \mu + \phi_i$$

El modelo resultante es:

$$[Y_i/E_i, \mu, \phi_i] \sim \text{Poisson}(E_i e^{\mu + \phi_i}).$$

En este modelo hay estructura de correlación, donde  $\phi_i$  es el efecto de correlación espacial, es decir, los RR de dos áreas vecinas pueden ser más parecidos entre sí, que las que no lo son, posibilidad debida a la dependencia de factores de riesgo no medidos que varían suavemente con la variación geográfica.

Un modelo que toma en cuenta esta correlación es el modelo *a priori* autorregresivo condicional (CAR del inglés); ver [6]:

$$[\phi_i | \phi_j, i \neq j, \lambda]_{CAR} \sim N\left(\bar{\phi}_i, \frac{1}{\lambda n_i}\right),$$

$$\bar{\phi}_i = \frac{1}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} \phi_j,$$

$$w_{ij} = \begin{cases} 1 & \text{si } i \text{ y } j \text{ son contiguas} \\ 0 & \text{si no} \end{cases}$$

En este se supone que la distribución condicional del efecto individual genérico  $\phi_i$  dado todos los restantes  $\phi_j$  con  $j \neq i$ , es Normal con media  $\bar{\phi}_i$  y varianza  $\frac{1}{\lambda n_i}$ , donde el parámetro  $\lambda$  representa la variabilidad geográfica y  $n_i$  es la cantidad de áreas vecinas del área  $i$ . El modelo *a priori* CAR tiende a desplazar cada estimación de los efectos de áreas individuales hacia un efecto medio local  $\bar{\phi}_i$ .

### 2.3. Modelo Convolución

A menudo es importante emplear ambos efectos, el de heterogeneidad correlacionada y el de heterogeneidad no correlacionada dentro de la especificación de la función predictora  $\eta_i$ . La racionalidad de esto se haya en la suposición de que los efectos no observados dentro de un área de estudio pueden tomar una variedad de formas. Es siempre prudente incluir un efecto de heterogeneidad no correlacionada a tener en cuenta para la variación extra no correlacionada. Sin embargo, sin conocimiento *a priori* de la confusión no observada, no hay razón para excluir cualquier efecto del análisis y es simple incluir ambos efectos dentro de una formulación de modelo aditivo tal como:

$$\eta_i = \mu + v_i + \phi_i;$$

donde  $v_i$  representa el efecto de no correlación y  $\phi_i$  la correlación espacial.

En general estos dos efectos no son identificados, pero como estamos normalmente interesados en el efecto total de confusión no observado, entonces la suma de los efectos es la componente importante y la misma está bien identificada; ver [5].

Este modelo es una combinación de los modelos explicados anteriormente, resultando el modelo Convolución:

$$[Y_i/E_i \xi_i] \sim \text{Poisson}(E_i \xi_i),$$

$$\ln \xi_i = \eta_i = \mu + v_i + \phi_i,$$

$$\xi_i = \exp(\mu + v_i + \phi_i),$$

$$\ln E_i \xi_i = \ln E_i + \mu + v_i + \phi_i$$

Donde  $\mu$  es la media global,  $v_i$  es el efecto de no correlación y  $\phi_i$  es el efecto de correlación espacial; ver [6, 13].

## 3. ANÁLISIS DE LOS RESULTADOS

Para la estimación del RR se accedió a los datos de la mortalidad por cáncer de las vías respiratorias y los de la mortalidad por neumonía e influenza considerando los 168 municipios de Cuba en el año 2011. Se utilizó el software WinBUGS 1.4 para simular los valores de los parámetros e hiperparámetros utilizando el método Monte Carlos vía Cadenas de Markov con el algoritmo muestreador Gibbs; ver [9, 10].

### 3.1. Resultados de los modelos para la mortalidad por cáncer de las Vías Respiratorias

A continuación se presentan los resultados de los modelos Log-Normal y Convolución para la mortalidad por cáncer de las vías respiratorias. El primer modelo se considera debido a que no toma en cuenta la correlación espacial entre las áreas y a que la enfermedad a estudiar es no trasmisible. El segundo modelo se emplea para tomar en cuenta ambos efectos, el de correlación y el de no correlación espacial entre las áreas.

### Resultados modelo Log-Normal

La siguiente tabla refleja el intervalo de credibilidad, que es el intervalo que contiene al parámetro con una probabilidad del 95%, este viene dado por el valor que toma en el cuartil 2.5% como límite inferior y el valor en el cuartil 97.5% como límite superior. Además se muestra la desviación estándar, el error de las cadenas de Markov, que mide la variabilidad de los parámetros, así como la media y la mediana de los parámetros del modelo Log-Normal para el cáncer de las vías respiratorias:

**Tabla 1 Parámetros estimados del modelo Log-Normal**

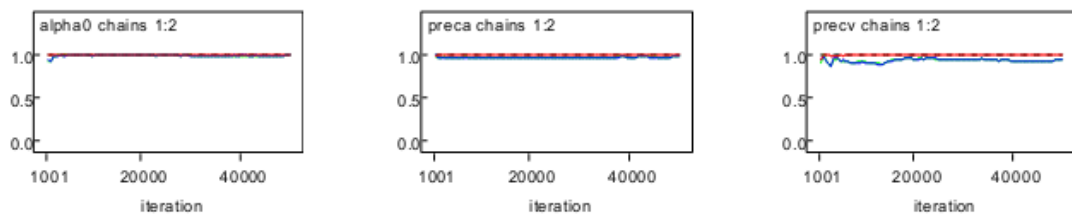
do	No	Me	DS	Er	2.5	Me	97.
	dia			ror CM	%	diana	5%
ha0	alp	-	0.0	7.8	-	-	0.0
		0.006268	1547	86E-5	0.038	0.005913	2337
ca	pre	167	173	5.9	40.	112	634
		2.0	5.0	08	05	4.0	6.0
	pre	88.	64.	1.5	37.	74.	218
cv		66	46	75	32	59	.3

Los parámetros  $\alpha_0$  y  $preca$  se refieren a  $\mu$  y el parámetro  $precv$  a  $v_i$ .

Se puede observar que los errores de las cadenas de Markov de los parámetros son pequeños respecto a la desviación estándar, por tanto la estimación de la media posterior fue realizada con alta precisión.

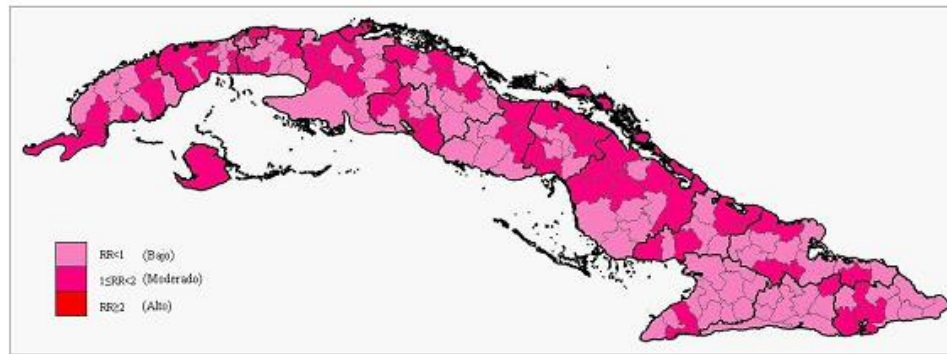
En la Figura 1 se muestra la convergencia de los parámetros estimados en el modelo Log-Normal.

**Figura 1 Convergencia de los parámetros del modelo Log-Normal**



Se puede observar que los 3 parámetros convergen. Se aprecia que la línea roja se hace constante en 1 y las líneas azul y verde se estabilizan.

**Figura 2 Mapa del RR a través del modelo Log-Normal**



En la Figura 2 aparecen representadas en 3 colores diferentes el RR, donde se puede apreciar que hay municipios de bajo y moderado riesgo en todas las regiones del país. Se evidencian de manera global patrones de riesgo de manera más aislada.

### 3.1.2. Resultados modelo Convolución

La Tabla 2 evidencia los parámetros estimados del modelo Convolución para el cáncer de las vías respiratorias.

**Tabla 2 Parámetros estimados del modelo Convolución**

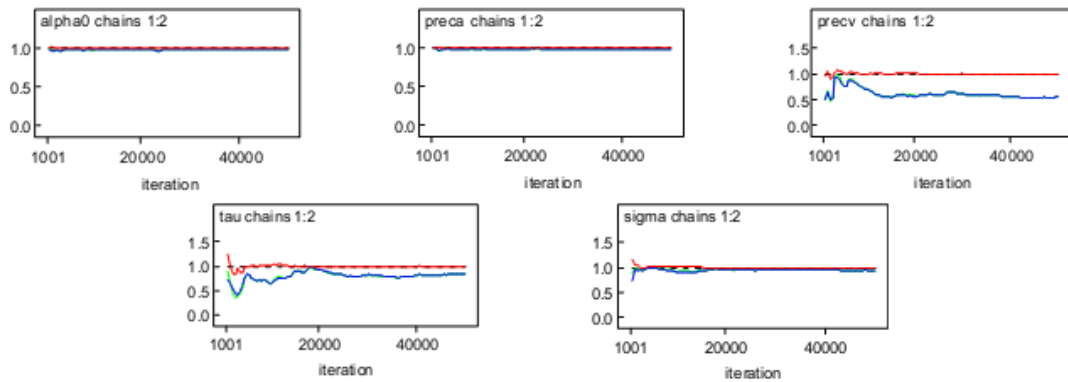
do	No	Me	DS	Er	2.5	Me	97.
dia				ror CM	%	diana	5%
alp	-	0.0	8.3	-	-	0.0	0.0
ha0	0.00953	1436	49E-5	-	0.03889	0.00918	1783
pre	164	170	6.0	39.	110	626	626
ca	5.0	2.0	23	12	4.0	7.0	7.0
pre	733	114	38.	62.	294	410	410
cv	.6	2.0	27	36	.8	4.0	4.0
tau	422	648	22.	47.	223	209	209
sig	.5	.9	49	06	.7	1.0	1.0
ma	7146	3264	0.0	0.0	0.0	0.0	0.1
			01124	2188	6686	458	458

Los parámetros  $\alpha_0$  y  $\text{preca}$  se refieren a  $\mu$  y el parámetro  $\text{precv}$  a  $v_i$  y los parámetros tau y sigma se refieren a  $\phi_i$ .

Se puede apreciar que los errores de las cadenas de Markov de algunos de los parámetros del modelo no son tan pequeños, no obstante los valores son pequeños con relación a la desviación estándar, por lo que la estimación de la media posterior fue estimada con alta precisión.

Al igual que en el modelo anterior, la Figura 3 muestra la convergencia de los parámetros estimados en el modelo Convolución.

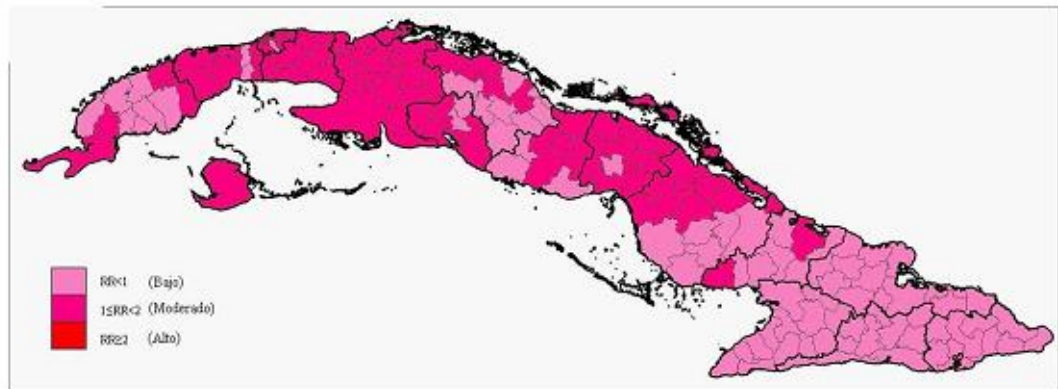
**Figura 3 Convergencia de los parámetros del modelo Convolución**



Aquí se observa que todos los parámetros convergen, aunque algunos de ellos lo hicieron tardíamente, el resultado no se afecta debido a que se realizaron una gran cantidad de iteraciones (49 000).

De igual manera los RR obtenidos para cada municipio se agruparon en 3 intervalos cuya amplitud se calculó teniendo en cuenta las diferentes escalas del riesgo.

**Figura 4 Mapa del RR a través del modelo Convolución**



En la Figura 4 se aprecia con mayor claridad patrones de distribución geográfica. Aquí, la estimación del RR en un área está fuertemente influenciada por la estimación de áreas adyacentes geográficamente e influenciada indirectamente por las otras áreas del mapa, haciendo que las estimaciones individuales estén más desplazadas hacia un valor medio local y no global. Se observa que en las zonas occidental y central predomina el riesgo moderado de mortalidad por cáncer de las vías respiratorias, en cambio las provincias orientales presentan bajo riesgo.

### 3.1.3. Comparación de modelos

La Tabla 3 muestra los resultados de los dos modelos analizados anteriormente a través del criterio DIC. Se obtuvo como resultado que el modelo de menor DIC fue el de Convolución:

**Tabla 3 Comparación de modelos**

Modelo	Mo	DIC
Log-Normal		051.880
Convolución		046.410

### 3.2. Resultados de los modelos para la mortalidad por Neumonía e Influenza

A continuación se presentan los resultados de los modelos CAR y Convolución para la mortalidad por neumonía e influenza. El primer modelo se emplea porque considera la correlación

espacial entre las áreas y a que la enfermedad a estudiar es transmisible. El segundo modelo, se utilizó para tomar en cuenta los dos aspectos la correlación espacial y la no correlación espacial.

### Resultado modelo CAR

En la Tabla 4 se reflejan los parámetros estimados del modelo CAR para la neumonía e influenza.

**Tabla 4 Parámetros estimados del modelo CAR**

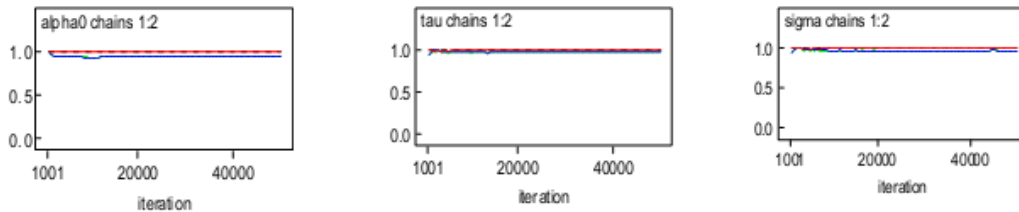
do	No	Me	DS	Er	2.5	Me	97.
dia				ror CM	%	diana	5%
ha0	alp	-	0.0	7.2	-	-	-
	0.07664	1936	13E-5	0.1148	0.07649	0.03905	
tau	1.4	0.2	0.0	1.0	1.4	1.9	
	99	276	01222	97	83	91	
sig	0.8	0.0	3.3	0.7	0.8	0.9	
	239	6252	71E-4	087	212	546	

El parámetro  $\alpha_0$  se refiere a  $\mu$  y los parámetros tau y sigma se refieren a  $\phi_i$ .

Se aprecia que los errores de las cadenas de Markov de los parámetros son pequeños respecto a la desviación estándar, obteniéndose estimaciones de la media posterior con alta precisión.

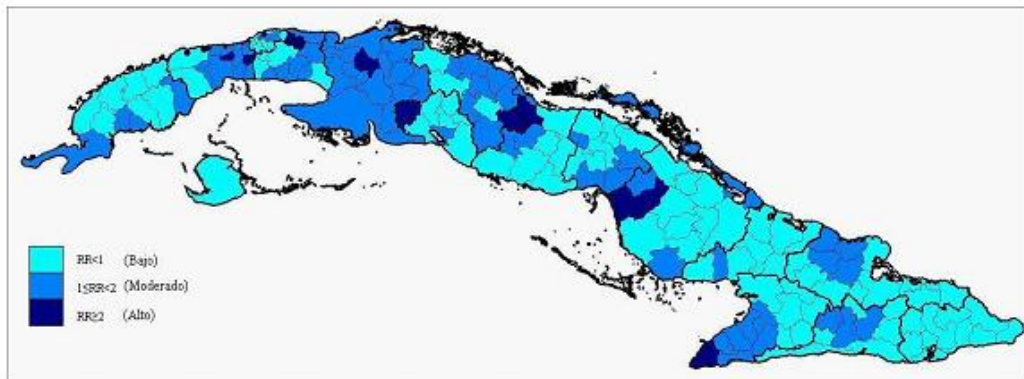
La Figura 5 muestra la convergencia de los parámetros estimados en el modelo CAR, donde se observa que convergen todos los parámetros.

**Figura 5 Convergencia de los parámetros del modelo CAR**



De manera similar los RR obtenidos para cada municipio se agruparon en 3 intervalos cuya amplitud se calculó teniendo en cuenta las diferentes escalas del riesgo. Se cambió el color para enfatizar que son los resultados de otra enfermedad.

**Figura 6 Mapa del RR a través del modelo CAR**



En la Figura 6 se observa que existen municipios de bajo, moderado y alto riesgo de mortalidad observándose claros patrones de distribución geográfica. La estimación del RR en un área está fuertemente influenciada por la estimación de áreas adyacentes geográficamente y solo influenciada indirectamente por las otras áreas del mapa, hecho que hace que las estimaciones individuales estén más desplazadas hacia un valor medio local que hacia un valor medio global. Los municipios de alto riesgo de mortalidad son: Guanajay, San Antonio de los Baños, Jaruco, Jovellanos, Remedios, Placetas, Aguada

de Pasajeros, Florida y Niquero pertenecientes a las provincias de Artemisa, Mayabeque, Matanzas, Villa Clara, Cienfuegos, Camagüey y Granma.

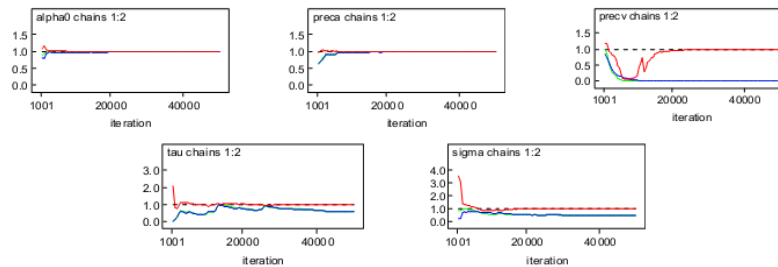
### 3.2.1. Resultados modelo Convolución

La Tabla 5 muestra los parámetros estimados del modelo Convolución para la neumonía e influenza.

**Tabla 5 Parámetros estimados del modelo Convolución**

Nodo	Media	DS	Error CM	2.5%	Mediana	97.5%
alpha0	-0.04769	0.03355	3.921E-4	-0.1191	-0.04531	0.009654
preca	840.9	1182.0	7.814	11.46	416.2	4206.0
precv	19.56	210.6	7.144	4.5	6.716	16.42
tau	68.71	299.3	12.25	2.696	15.42	468.4
sigma	0.269	0.1441	0.005683	0.04625	0.2547	0.6091

**Figura 7 Convergencia de los parámetros del modelo Convolución**



Los parámetros alpha0 y preca se refieren a  $\mu$ , el parámetro precv a  $v_i$  y los parámetros tau y sigma se refieren a  $\phi_i$ .

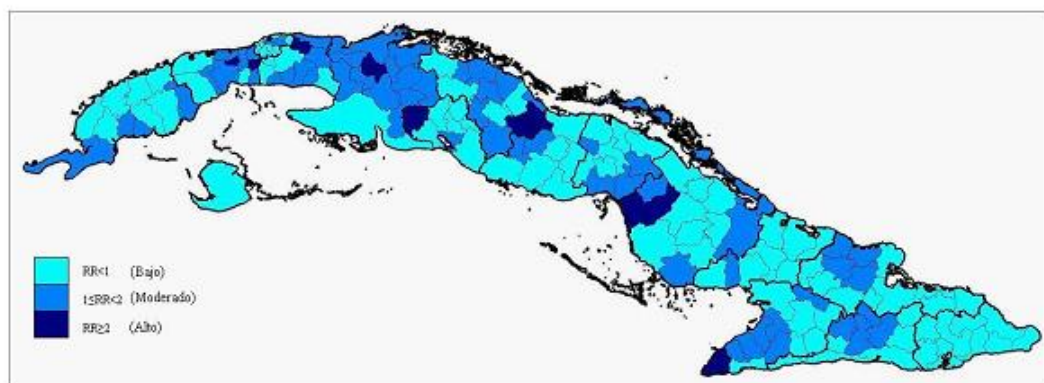
Se puede apreciar que los errores de las cadenas de Markov son pequeños en relación a la desviación estándar, por tanto las estimaciones obtenidas de la media posterior tienen una alta precisión.

En la Figura 7 se evidencia la historia, la densidad y la convergencia de los parámetros estimados en el modelo Convolución.

Aquí se evidencia la convergencia de todos los parámetros. También se puede observar que al igual que en el modelo Convolución anterior, algunos parámetros lo hicieron más tarde que otros, pero el resultado no se afectó por la gran cantidad de iteraciones (49000) realizadas.

De manera similar los RR obtenidos para cada municipio se agruparon en 3 intervalos cuya amplitud se calculó teniendo en cuenta las diversas escalas del riesgo.

**Figura 8 Mapa del RR a través del modelo Convolución**



En la Figura 8 al igual que en el mapa del modelo Convolución anterior se puede observar que existen claros patrones de mortalidad por neumonía e influenza en los diferentes municipios del país. Se puede apreciar la existencia de municipios de bajo, moderado y alto riesgo. Los municipios de alto riesgo



de mortalidad son: Guanajay, San Antonio de los Baños, Jaruco, Jovellanos, Remedios, Placetas, Aguada de Pasajeros, Florida y Niquero pertenecientes a las provincias de Artemisa, Mayabeque, Matanzas, Villa Clara, Cienfuegos, Camagüey y Granma.

### 3.2.2. Comparación de modelos

La Tabla 6 muestra los resultados del DIC para los dos modelos analizados anteriormente. Tomando en cuenta los resultados obtenidos, el menor valor de DIC lo tuvo el modelo Convolución.

**Tabla 6 Comparación de modelos Neumonía e Influenza**

Modelo	DIC
CAR	1114.72 0
Convolución	1114.58 0

Podemos apreciar que el modelo Convolución es el más apropiado a utilizar en ambas enfermedades analizadas. Este modelo permite incorporar al modelo efectos de correlación espacial y de no correlación que están incidiendo en la mortalidad de las enfermedades a estudiar.

## 4. CONCLUSIONES

- El uso del modelo Convolución permitió incluir en un mismo modelo la no correlación y la dependencia espacial en la estimación de los riesgos relativos. Este modelo resultó ser seleccionado como el mejor en ambas enfermedades analizadas.
- Los resultados obtenidos a través del modelo Convolución evidencian que en las provincias occidentales y centrales predomina el riesgo moderado de mortalidad por cáncer de las vías respiratorias.
- Los resultados obtenidos a través del modelo Convolución reflejan que los municipios de alto riesgo de mortalidad por neumonía e influenza son: Guanajay, San Antonio de los Baños, Jaruco, Jovellanos, Remedios, Placetas, Aguada de Pasajeros, Florida y Niquero pertenecientes a las provincias de Artemisa, Mayabeque, Matanzas, Villa Clara, Cienfuegos, Camagüey y Granma.

## REFERENCIAS

- [1] Anuario Estadístico de Salud 2011. (2012): Dirección Nacional de Registros Médicos y Estadísticas de Salud. La Habana, 31-82.
- [2] Carlin B. P. and Louis T. A. (2009) Bayesian methods for data analysis, 3<sup>rd</sup> ed., 59-79. Chapman & Hall Book.
- [3] Congdon P. (2006): Bayesian Statistical Modelling, 2<sup>nd</sup> ed., 1-5. J. Wiley & Sons, Ltd.
- [4] Congdon P. (2014) Applied Bayesian Modelling 2<sup>nd</sup> ed., 34-89. Wiley.
- [5] Eberley, L. and B. P. Carlin (2000): Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine* 19, 2279-2294.
- [6] Lawson, A. B. (2013): Bayesian Disease Mapping, Hierarchical Modeling in spatial Epidemiology 2<sup>nd</sup> ed., 73-94. Chapman & Hall/CRC.
- [7] Pfeiffer, D. U., T. P. Robinson, M. Stevenson, K. B. Stevens, D. J. Rogers y A. C. A. Clements (2008): Spatial Analysis in Epidemiology, 70-73. New York: Springer.
- [8] Robert C. P. (2007): The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, 2<sup>nd</sup> ed., 458-462. Springer Texts in Statistics.
- [9] Spiegelhalter, D., Thomas, A. y Best N. WinBUGS 1.4.3 Software. URL <http://www.mrc-bsu.cam.ac.uk/bugs>
- [10] Spiegelhalter, D., Thomas, A., Best, N. y Lunn, D. (2007): WinBUGS manual. Cambridge, UK: MRC Biostatistic Unit. versión 1.4.3.
- [11] Tango T. (2010): Statistical methods for disease clustering, 44-46. Springer

- 
- [12] Venkatesan, P., Srinivasan, R. y Dharuman C. (2012): Bayesian Conditional auto regressive model for mapping tuberculosis prevalence in India. *International Journal of Pharmaceutical Studies and Research*, Vol. 3, No. 1, Enero-Marzo.
- [13] Wang G., Minnis R., Belant J. y Wax C. (2010): Dry weather induces outbreaks of human West Nile virus infections. *BMC Infectious Disease*, 10:38.

## MÉTODO PARA DETERMINAR LAS FASES MINERALES DEL CLÍNKER Y SU INFLUENCIA EN REDUCIR LOS DAÑOS AL MEDIO AMBIENTE

Carlos Alberto Alvarez Bravo\*, Manuel E. Cortés Cortés<sup>7</sup>\* & . Mario Moreira\*\*

\*Dpto. de Matemática, Facultad de Ingeniería, Universidad de Cienfuegos, Cienfuegos, Cuba,

Dpto. de Matemática, Facultad de Ingeniería, Universidad de Cienfuegos, Cienfuegos, Cuba

\*\*Ingeniero de Proceso y Mejora Continua, Cementos Cienfuegos S.A, Cienfuegos, Cuba,

### ABSTRACT

The calculation of the percentages of the mineral phases of the clinker in the production of cement is an interesting and great-benefit theme in the present-day moments, the fundamental objective of the following investigation is rephrasing Bogue's Equations to calculate the percentage of the mineral phases of the clinker in the production of cement of the Mixed Enterprise Cementos Cienfuegos S.A (CCSA) located in the province of Cienfuegos, and with it influencing the reduction of the damages to the Environment derived of the manufacturing process of the aforementioned product. Calculation comes true from the knowledge of the composition of the main oxides (CaO, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, Fe<sub>2</sub>O<sub>3</sub> y SO<sub>3</sub>) in the clinker, but the interest of the specialists is realizing the same before the obtaining of the clinker and from the percentages of the oxides in the raw material to take the decisions in the process of production of the same, guaranteeing the quality of the product and a decrease in the losses for low-quality productions and in the damages to the Environment, from there the need to value the models of the relation between the percentages of each Oxide in the raw material and in the clinker. On the basis of these relations Bogue's equations to obtain the percentages of the mineral phases of the clinker in terms of the percentages of the main oxides in the raw material get modified, before the production of the same. The models valued, of great benefit in the control of the quality of the clinker corroborate themselves as to their statistical validity, stability and precision, influencing the reduction of the damages to the Environment derived of the manufacturing process of the afore.

**KEY WORDS:** Clinker, Calculation of clinker mineral phases, Bogue's equations, Linear Regression.

### RESUMEN

El cálculo de los porcentajes de las fases minerales del clínker en la producción de cemento es un tema interesante y de gran utilidad en los momentos actuales, el objetivo fundamental de la siguiente investigación es reformular las Ecuaciones de Bogue para calcular el porcentaje de las fases minerales del clínker en la producción de cemento de la Empresa Mixta "Cementos Cienfuegos S.A" (CCSA) ubicada en la provincia de Cienfuegos, y con ello influir en la reducción de los daños al Medio Ambiente derivados del proceso de fabricación de dicho producto. El cálculo se realiza a partir del conocimiento de la composición de los principales óxidos (CaO, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, Fe<sub>2</sub>O<sub>3</sub> y SO<sub>3</sub>) en el clínker, pero el interés de los especialistas es realizar el mismo antes de la obtención del clínker y a partir de los porcentajes de los óxidos en la materia prima para tomar las decisiones en el proceso de producción del mismo, garantizando la calidad del producto y una disminución en las pérdidas por producciones de baja calidad y en los daños al Medio Ambiente, de ahí la necesidad de estimar los modelos de la relación entre los porcentajes de cada óxido en la materia prima y en el clínker. En base a estas relaciones se modifican las ecuaciones de Bogue para obtener los porcentajes de las fases minerales del clínker en función de los porcentajes de los principales óxidos en la materia prima, antes de la producción del mismo. Los modelos estimados, de gran utilidad en el control de la calidad del clínker se corroboran en cuanto a su validez estadística, estabilidad y precisión, influyendo en la reducción de los daños al Medio Ambiente derivados del proceso de fabricación de dicho producto.

## 1. INTRODUCCIÓN

Según [1] "el cemento Portland es un producto finamente molido, altamente reactivo y compuesto de clínker, yeso y algunos materiales de adición" Para la producción cemento se necesita que ocurran una serie de procesos físico químicos que dan origen a un producto intermedio extraordinariamente valioso llamado clínker que representa las combinaciones estequiométricas de los óxidos fundamentales de las materias primas, a este proceso se le denomina proceso de clinkerización, proceso que solo tiene lugar en determinadas condiciones que están creadas previamente en el horno a temperaturas cercanas a los 1450 C°.

Generalmente, las materias primas (piedra calcárea y materiales arcillosos) para la producción de cemento proceden de recursos no renovables y su extracción tiene un notable impacto ambiental, como suele suceder con todas las extracciones de minerales. En lo referente al proceso industrial, la obtención del clínker es la parte más importante en la fabricación de cemento e implica un elevado consumo de energía y, posteriormente, emisiones importantes de gases y polvo al molerlo [2].

---

<sup>7</sup> Corresponding author: [mcortes@ucf.edu.cu](mailto:mcortes@ucf.edu.cu)

La eficiencia del proceso de clinkerización, (calidad del clínker), a la salida del horno, depende de la temperatura del horno, el tiempo de residencia de la harina (materias primas homogeneizadas) en el horno y los parámetros de calidad de la harina, o sea, la granulometría y la correcta composición de los óxidos: CaO; Fe<sub>2</sub>O<sub>3</sub>; Si<sub>2</sub>O, Al<sub>2</sub>O<sub>3</sub>y SO<sub>3</sub> [1].

Las desviaciones en los parámetros de calidad de la harina que entra al horno (considerando la granulometría constante) implican variaciones en el proceso de cocción y por tanto variaciones de la eficiencia del proceso de clinkerización en cuanto a lo que se refiere a las principales fases “minerales” constituyentes esenciales del clínker. Las principales fases en el clínker son: alita (C3S), belita (C2S), celita (C3A) y ferrita (C4AF), además, pueden estar presentes cristales de cal libre, periclasa y sulfatos alcalinos, entre otros [1]. Las proporciones, la cristalinidad y la textura de estas fases minerales en el clínker controlan propiedades tan importantes en el cemento como: fraguado, calor de hidratación, reactividad y desarrollo de resistencias [3, 4]. De ahí la importancia de cuantificarlas con precisión para así evitar que el producto tenga que ser producido reiteradamente y con ello impedir la extracción de recursos no renovables, un elevado consumo de energía y emisiones de gases contaminantes nuevamente, todo lo cual es perjudicial al Medio Ambiente.

R. H. Bogue desarrolló un proceso de cálculo según el cual, a partir del análisis químico, se puede calcular el contenido en minerales del clínker (en porcentaje), sobre todo, de alita (C3S), belita (C2S), celita (C3A) y ferrita (C4AF). A las ecuaciones encontradas por Bogue se les conoce en la actualidad como las “Ecuaciones de Bogue” [5]. Estas ecuaciones fueron planteadas así:

$$C3S = 4.071C_aO - (7.6S_iO_2 + 6.718Al_2O_3 + 1.43Fe_2O_3 + 2.852SO_3) \quad (1)$$

$$C2S = 2.867S_iO_2 - 0.7544C_3S \quad (2)$$

$$C3A = 2.65Al_2O_3 - 1.692Fe_2O_3 \quad (3)$$

$$C4AF = 3.043Fe_2O_3 \quad (4)$$

El cálculo del contenido en minerales del clínker mediante las Ecuaciones de Bogue se realiza a partir del conocimiento de la composición de los principales óxidos (CaO, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, Fe<sub>2</sub>O<sub>3</sub> y SO<sub>3</sub>) en el clínker, por tanto si se tiene en cuenta que el horno instalado es de 220 ton/h de clínker (con una significativa inercia) es de interés para los especialistas de Cementos Cienfuegos realizar el cálculo antes de la obtención del clínker y de ahí la necesidad de predecir los porcentajes de los óxidos en el clínker, lo cual se realiza a partir de encontrar como depende el porcentaje de cada uno de los óxidos en el clínker del porcentaje de estos óxidos en la harina, para de esta forma poder hacer correcciones en los gráficos de control del proceso y así evitar elevadas pérdidas económicas por productos fuera de especificaciones, consumo de energía innecesario y emisiones atmosféricas (Información aportada por los especialistas de la empresa).

## 2

### 2. ANÁLISIS DE ESTUDIOS REALIZADOS

Para la cuantificación de los porcentajes de las fases minerales del clínker se han desarrollado métodos tales como [1] la Difracción de Rayos-X [6], químico-Cálculo Potencial de Bogue [7, 8] y microscopía óptica [3, 9-15]. Este último, se realiza mediante conteo manual de puntos [16] o análisis digital de Imágenes [17].

En Cuba (según los especialistas de la empresa CCSA) y el mundo el método clásico de cuantificar los porcentajes de los minerales del clínker es usando las ecuaciones propuestas por Bogue hace cerca de un siglo [1], conocidas como Cálculo Potencial de Bogue [7]. El cálculo se realiza a partir del conocimiento de los porcentajes de los principales óxidos que están presentes en el clínker. En su formulación estas ecuaciones asumen materias primas con pureza y reacción entre ellas del 100%, lo cual no es cierto para la mayoría de las cementeras, donde se tienen diferentes combinaciones de materias primas y procesos de clinkerización no totalmente controlados. Además el error se incrementa por la formación de compuestos menores y por la presencia de soluciones sólidas entre los componentes principales y menores [18].

[19], [4] y [18] muestran como algunos investigadores han concluido que los cálculos con las ecuaciones de Bogue generalmente subestiman el contenido de alita y sobreestiman el de belita y celita hasta en un 10% agregando, que el conteo de puntos mediante microscopía óptica puede producir un resultado más preciso para estas fases.

Autores como [20] identifican como una dificultad en el método óptico la cuantificación de los aluminatos (celitas) y ferroaluminatos (ferritas) presentes en la fase intersticial debido principalmente al tamaño tan pequeño de los cristales, los cuales pueden llegar incluso a ser amorfos. Ellos proponen para subsanar esta dificultad el empleo de Difracción de Rayos X Cuantitativa (QXRD) la cual puede ser difícil porque este es un material multifases y varios picos se superponen, sin embargo, ellos plantean que con los desarrollos del método de Rietveld se pueden minimizar o eliminar estos errores.

Aceptando que esta dificultad existe, es importante resaltar que una ventaja comparativa que tiene la microscopía óptica sobre estas otras técnicas es que además de cuantificar las fases permite ver las texturas y las alteraciones en los cristales como son: el tamaño de los cristales, distribución de las fases dentro de la muestra, clúster, retrogradaciones, zonaciones, maclas, etc. [4, 10, 14], y ahora con el desarrollo de sistemas de análisis digital de imágenes (ADI) esto se puede hacer en muy corto tiempo [17].

No obstante a lo planteado por [19], [4] y [18], la práctica ha demostrado que en la fábrica de cemento de Cienfuegos Cementos Cienfuegos S.A los cálculos con las ecuaciones de Bogue para cuantificar los porcentajes de los minerales del clínker brindan resultados suficientemente buenos (Información aportada por los especialistas de la empresa).

En el mundo, se han realizado variaciones a las ecuaciones de Bogue ya que cada cementera tiene materias primas y procesos de producción con características propias. Es por esto que autores como [21], [22] y [8] han propuesto modificaciones a dichas ecuaciones a partir de las composiciones químicas y mineralógicas para la dosificación y corrección de la pasta, pero ninguna de éstas ha recibido una aceptación general [1] mostrando que el problema es propio de cada planta.

Todas estas modificaciones han estado en función de la modificación de los coeficientes de las ecuaciones a partir del conocimiento de los porcentajes de los principales óxidos en el clínker, pero ninguna de ellas considera estas modificaciones a partir del conocimiento de los porcentajes de los principales óxidos en la harina.

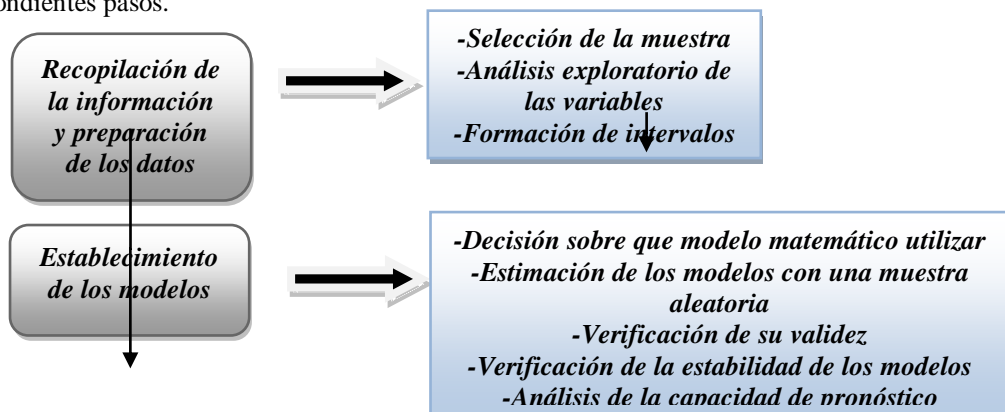
En Colombia ya se conoce la existencia de una modificación de estas ecuaciones en una industria cementera con gran aceptación. Pero al igual que las anteriores esta modificación se realizó a partir del conocimiento de los porcentajes de los principales óxidos en el clínker.

### 3. REFORMULACIÓN DE LAS ECUACIONES DE BOGUE

El método empleado se organiza metodológicamente en dos etapas básicas:

1. Recopilación de la información y preparación de los datos.
2. Establecimiento de los modelos matemáticos para el cálculo de los porcentajes de las fases minerales del clínker.

Cada una de las etapas consta de un conjunto de pasos. La siguiente figura muestra las etapas con sus correspondientes pasos.



**Figura 1. Etapas y pasos del método para el cálculo de los porcentajes de las fases minerales del clínker. Fuente: Elaboración propia.**

Sea

$$CaO = f(CaO_H) \quad (5)$$

la relación funcional entre el óxido de calcio en el clínker y el óxido de calcio en la harina, donde:

$CaO$  es el óxido de calcio en el clínker y  $CaO_H$  es el óxido de calcio en la harina,

$$SiO_2 = f(SiO_{2H}) \quad (6)$$

la relación funcional entre el dióxido de sílice en el clínker y el dióxido de sílice en la harina, donde:

$SiO_2$  es el dióxido de sílice en el clínker y  $SiO_{2H}$  es el dióxido de sílice en la harina,

$$Al_2O_3 = f(Al_2O_{3H}) \quad (7)$$

la relación funcional entre el óxido de aluminio en el clínker y el óxido de aluminio en la harina, donde:

$Al_2O_3$  es el óxido de aluminio en el clínker y  $Al_2O_{3H}$  es el óxido de aluminio en la harina,

$$Fe_2O_3 = f(Fe_2O_{3H}) \quad (8)$$

la relación funcional entre el óxido de hierro en el clínker y el óxido de hierro en la harina, donde:

$Fe_2O_3$  es el óxido de hierro en el clínker y  $Fe_2O_{3H}$  es el óxido de hierro en la harina y

$$SO_3 = f(SO_{3H}) \quad (9)$$

la relación funcional entre el trióxido de azufre en el clínker y el trióxido de azufre en la harina, donde:

$SO_3$  es el trióxido de azufre en el clínker y  $SO_{3H}$  es el trióxido de azufre en la harina, si sustituimos

(5), (6), (7), (8) y (9) en (1), (2), (3), (4) y (5), obtenemos:

$$C3S = 4.071 f(C_aOH) - \left( 7.6 f(S_iO_{2H}) + 6.718 f(Al_2O_{3H}) + \right. \\ \left. + 1.43 f(Fe_2O_{3H}) + 2.852 f(SO_{3H}) \right) \quad (10)$$

$$C2S = 2.867 f(S_iO_{2H}) - 0.7544 C_3S \quad (11)$$

$$C3A = 2.65 f(Al_2O_{3H}) - 1.692 f(Fe_2O_{3H}) \quad (12)$$

$$C4AF = 3.043 f(Fe_2O_{3H}) \quad (13)$$

y con estas ecuaciones (10), (11), (12) y (13) puede predecirse el porcentaje de cada una de las fases minerales del clínker a partir del conocimiento de los porcentajes de los principales óxidos en la harina, pero para la obtención de las ecuaciones anteriores se debe recopilar información y preparar los datos relacionados con los porcentajes de los óxidos tanto en la harina como en el clínker y establecer los modelos (10), (11), (12) y (13) después de hallar las relaciones funcionales (5), (6), (7), (8) y (9).

#### 4. APLICACIÓN DEL MÉTODO

Los resultados de la aplicación del método se presentan a continuación sobre la base de los resultados obtenidos a partir de la ejecución de cada uno de los pasos de cada etapa.

##### 4.1. Resultados de la primera etapa

En la investigación se trabaja con toda la información generada en el período de enero de 2013 a abril de 2013 el cual se considera representativo del proceso productivo y de la calidad de los materiales que se utilizan. Los datos se corresponden con los resultados de los análisis químicos que se realizan en la fábrica y están contenidos en una base de datos en formato .xls que contiene un total de 1117 observaciones de cada una de las variables mencionadas con anterioridad.

Como resultado del análisis exploratorio de las variables los estadísticos descriptivos (media, desviación típica, mínimo y máximo) y los intervalos de confianza para la media de las variables que se utilizan en el estudio se contrastan con los requerimientos técnicos de estas variables en la producción del clínker observándose que los intervalos de confianza para la media de todas las variables son un subconjunto del intervalo que garantiza los requisitos de calidad del cliente y además se aprecia que el

---

valor típico de cada una de las variables de acuerdo con los requisitos del cliente no está incluido en el intervalo de confianza de su variable correspondiente.

A partir de los gráficos de tallos y hojas y los diagramas de cajas de las variables que se utilizan en el estudio, se excluyen de los análisis para obtener los modelos dados por las ecuaciones (5), (6), (7), (8) y (9) los valores atípicos encontrados para cada una de las variables, eliminándose 24 observaciones de la variable  $\text{SiO}_{2\text{H}}$ , 43 de la variable  $\text{Al}_2\text{O}_{3\text{H}}$ , 10 de la variable  $\text{CaO}_{\text{H}}$ , 8 de la variable  $\text{SO}_{3\text{H}}$ , 41 de la variable  $\text{SiO}_2$ , 30 de la variable  $\text{Al}_2\text{O}_3$ , 9 de la variable  $\text{Fe}_2\text{O}_3$ , 38 de la variable  $\text{CaO}$  y 37 de la variable  $\text{SO}_3$ . La observación de los gráficos de tallos y hojas permite además sospechar una posible normalidad de las variables en estudio.

En el estudio de las correlaciones bivariadas de Pearson y los diagramas de dispersión entre las variables correspondientes de las variables que se utilizan en el estudio se puede apreciar que los coeficientes de correlación de Pearson entre las variables dependientes e independientes son estadísticamente significativos pero muy débiles, mientras que los gráficos de dispersión entre estas variables no muestran otras relaciones funcionales además de estas relaciones lineales, comportamiento que no favorece la obtención de modelos que permitan estimar con exactitud los porcentajes de los óxidos en el clínker a partir de los porcentajes de los óxidos en la harina, por tanto y en base al comportamiento de los óxidos en el clínker que muestran que cuando en la harina se producen incrementos pequeños, menores o iguales a 0.1% para la variable  $\text{CaO}_{\text{H}}$ , 0.02% para las variables  $\text{SO}_{3\text{H}}$ ,  $\text{Al}_2\text{O}_{3\text{H}}$  y  $\text{Fe}_2\text{O}_{3\text{H}}$  y en el caso del  $\text{SiO}_{2\text{H}}$  inferiores o iguales a 0.05%, estos incrementos no inciden significativamente en los porcentajes de estos óxidos en el clínker, se procede a formar intervalos de amplitud menor o igual a 0.1 para la variable  $\text{CaO}_{\text{H}}$ , menor o igual a 0.02 para las variables  $\text{SO}_{3\text{H}}$ ,  $\text{Al}_2\text{O}_{3\text{H}}$  y  $\text{Fe}_2\text{O}_{3\text{H}}$  y menor o igual a 0.05 para la variable  $\text{SiO}_{2\text{H}}$  y a tomar como valores para las variables independientes los puntos medios de cada uno de estos intervalos y como valores correspondientes de las variables dependientes los puntos medios de los intervalos correspondientes y así se formaron 22 intervalos para la variable  $\text{CaO}_{\text{H}}$ , 16 intervalos para las variables  $\text{SiO}_{2\text{H}}$  y  $\text{SO}_{3\text{H}}$ , 25 intervalos para la variable  $\text{Al}_2\text{O}_{3\text{H}}$  y 23 intervalos para la variable  $\text{Fe}_2\text{O}_{3\text{H}}$  y se crearon las variables:

*MediaCaOCAT* : que representa la media del porcentaje de óxido de calcio en el clínker para cada intervalo de la variable  $\text{CaO}_{\text{H}}$ ,

*MediaCaOHCAT* : que representa la media del porcentaje de óxido de calcio en la harina para cada intervalo de la variable  $\text{CaO}_{\text{H}}$ ,

*MediaSiO<sub>2</sub>CAT* : que representa la media del porcentaje de dióxido de sílice en el clínker para cada intervalo de la variable  $\text{SiO}_{2\text{H}}$ ,

*MediaSiO<sub>2</sub>HCAT* que representa la media del porcentaje de dióxido de sílice en la harina  $\text{SiO}_{2\text{H}}$  para cada intervalo de la variable  $\text{SiO}_{2\text{H}}$ ,

*MediaAl<sub>2</sub>O<sub>3</sub>CAT* que representa la media del porcentaje de óxido de aluminio en el clínker para cada intervalo de la variable  $\text{Al}_2\text{O}_{3\text{H}}$ ,

*MediaAl<sub>2</sub>O<sub>3</sub>HCAT* que representa la media del porcentaje de óxido de aluminio en la harina para cada intervalo de la variable  $\text{Al}_2\text{O}_{3\text{H}}$ ,

*MediaFe<sub>2</sub>O<sub>3</sub>CAT* que representa la media del porcentaje de óxido de hierro en el clínker para cada intervalo de la variable  $\text{Fe}_2\text{O}_{3\text{H}}$ ,

*MediaFe<sub>2</sub>O<sub>3</sub>HCAT* que representa la media del porcentaje de óxido de hierro en la harina para cada intervalo de la variable  $\text{Fe}_2\text{O}_{3\text{H}}$ ,

*MediaSO<sub>3</sub>CAT* que representa la media del trióxido de azufre en el clínker para cada intervalo de la variable  $\text{SO}_{3\text{H}}$

*MediaSO<sub>3</sub>HCAT* que representa la media del porcentaje de trióxido de azufre en la harina para cada intervalo de la variable  $\text{SO}_{3\text{H}}$ .

## 4.2. Resultados de la segunda etapa

Dado que en el problema tratado no se puede obtener un ajuste exacto a todos los puntos ya que existen valores de las variables independientes para los cuales se tienen más de un valor de las variables dependientes se decide realizar un análisis de regresión, más específicamente un análisis de regresión lineal simple para estimar los modelos:

$$CaO_i = \beta_{11} + \beta_{21}CaOH_i + \varepsilon_i \quad (14)$$

$$SiO_{2i} = \beta_{11} + \beta_{21}SiO_2H_i + \varepsilon_i \quad (15)$$

$$SO_{3i} = \beta_{11} + \beta_{21}SO_3H_i + \varepsilon_i \quad (16)$$

$$Al_2O_{3i} = \beta_{11} + \beta_{21}Al_2O_3H_i + \varepsilon_i \quad (17)$$

$$Fe_2O_{3i} = \beta_{11} + \beta_{21}Fe_2O_3H_i + \varepsilon_i \quad (18)$$

La estimación de los modelos (14), (15), (16), (17) y (18) con una muestra aleatoria del 50% de los datos, y con las variables *MediaCaOCAT*, *MediaCaOHCAT*, *MediaSiO<sub>2</sub>CAT*, *MediaSiO<sub>2</sub>HCAT*, *MediaAl<sub>2</sub>O<sub>3</sub>CAT*, *MediaAl<sub>2</sub>O<sub>3</sub>HCAT*, *MediaFe<sub>2</sub>O<sub>3</sub>CAT*, *MediaFe<sub>2</sub>O<sub>3</sub>HCAT*, *MediaSO<sub>3</sub>CAT*, *MediaSO<sub>3</sub>HCAT* obtenidas por los intervalos formados para cada variable, arrojó como resultado la obtención de los modelos dados por las ecuaciones siguientes:

$$CaO = 60.812 + 0.117CaOH \quad (19)$$

$$SiO_2 = 18.478 + 0.194SiO_2H \quad (20)$$

$$SO_3 = 1.460 + 0.510SO_3H \quad (21)$$

$$Al_2O_3 = 2.236 + 0.884Al_2O_3H \quad (22)$$

$$Fe_2O_3 = 1.221 + 1.044Fe_2O_3H \quad (23)$$

A los modelos (19), (20), (21), (22) y (23) se le comprueba la validez con la prueba *T* de Student, la cual arroja resultados satisfactorios, ya que se puede constatar que la significación del estadígrafo es menor en todos los casos que el nivel de significación  $\alpha = 0.05$  y por tanto se concluye que todos los modelos son válidos.

Con el 50% de los datos no seleccionados para la obtención de los modelos (19), (20), (21), (22) y (23) se obtienen los modelos:

$$CaO = 61.171 + 0.108CaOH \quad (24)$$

$$SiO_2 = 19.247 + 0.138SiO_2H \quad (25)$$

$$SO_3 = 1.430 + 0.569SO_3H \quad (26)$$

$$Al_2O_3 = 1.920 + 0.988Al_2O_3H \quad (27)$$

$$Fe_2O_3 = 1.593 + 0.874Fe_2O_3H \quad (28)$$

y se compara el parámetro correspondiente a la variable independiente en dichos modelos contra el valor del parámetro obtenido en los primeros modelos estimados (19), (20), (21), (22) y (23) utilizando la prueba *T* de Student la cual arroja como resultado que los modelos no cambian significativamente con el cambio de la muestra, es decir son estables en la población en estudio.

Para verificar la capacidad de pronóstico de los modelos (19), (20), (21), (22) y (23) se aplica la prueba *T* de Student en 2 poblaciones mediante un diseño de muestras pareadas, donde una población está constituida por los valores de la variable dependiente en los datos no utilizados en la estimación de los modelos y la segunda población está dada por los pronósticos de estos datos realizados a partir de los modelos estimados, la hipótesis nula en esta prueba es que las medias en ambas poblaciones son iguales lo que se reduce a verificar la hipótesis nula de que la diferencia de las medias en ambas poblaciones es cero o lo que es lo mismo, que la media de los errores es cero. Esta prueba se aplica para cada uno de los modelos y arroja como resultado que los modelos tienen una buena capacidad de pronóstico, por tanto se



puede concluir que todos los modelos estimados pronostican adecuadamente, además en todos los casos se puede apreciar que los intervalos de confianza para la media de los pronósticos de cada uno de los modelos son un subconjunto de los intervalos que garantizan los requisitos del cliente.

El análisis de los supuestos de cada uno de los modelos (19), (20), (21), (22) y (23) también arroja resultados satisfactorios.

Una vez realizado el análisis de los supuestos de los modelos dados por las ecuaciones (19), (20), (21), (22) y (23) se obtienen los modelos:

$$CaO = 61.642 + 0.097CaOH \quad (29)$$

$$SiO_2 = 18.791 + 0.171SiO_2H \quad (30)$$

$$SO_3 = 1.444 + 0.541SO_3H \quad (31)$$

$$Al_2O_3 = 2.162 + 0.908Al_2O_3H \quad (32)$$

$$Fe_2O_3 = 1.303 + 1.007Fe_2O_3H \quad (33)$$

con el 100% de los datos. Estos modelos (29), (30), (31), (32) y (33) representan las relaciones funcionales (5), (6), (7), (8) y (9). A estos modelos también se les realiza un análisis de los supuestos, obteniéndose en todos los casos resultados satisfactorios.

A partir de la sustitución de los modelos (29), (30), (31), (32) y (33) en las ecuaciones (1), (2), (3) y (4) se obtienen los modelos:

$$C3S = 87.627088 + 0.394887 C_aOH - \left( \begin{array}{l} 1.2996SiO_2H + 6.099944 Al_2O_3H + \\ + 1.44001Fe_2O_3H + 1.542932 SO_3H \end{array} \right) \quad (34)$$

$$C2S = 53.873797 + 0.490257 S_iO_2H - 0.7544C_3S \quad (35)$$

$$C3A = 3.524624 + 2.4062Al_2O_3H - 1.703844 Fe_2O_3H \quad (36)$$

$$C4AF = 3.965029 + 3.064301 Fe_2O_3H \quad (37)$$

los cuales representan las ecuaciones (10), (11), (12) y (13) y permiten calcular los porcentajes de las fases minerales del clínker a partir del conocimiento de los porcentajes de los óxidos en la harina.

La validación de los modelos (34), (35), (36) y (37) se realiza mediante la verificación de la capacidad de pronóstico con una muestra adicional de 400 observaciones no empleadas en la estimación aplicando la prueba *T* de Student a los errores dados por la diferencia de los pronósticos de estos modelos con los valores reales de la variable dependiente. Esta prueba arrojó resultados satisfactorios.

El método estudiado permite reducir daños al Medio Ambiente puesto que en el proceso productivo se eliminan producciones por productos fuera de especificaciones dadas por los clientes.

En resumen, al aplicar el método matemático para el cálculo de las fases minerales del clínker se logra la obtención de 5 modelos que permiten predecir los porcentajes de los principales óxidos en el clínker a partir del conocimiento de los porcentajes de estos óxidos en la harina, estos modelos son:

$$CaO = 61.642 + 0.097CaOH \quad (29)$$

$$SiO_2 = 18.791 + 0.171SiO_2H \quad (30)$$

$$SO_3 = 1.444 + 0.541SO_3H \quad (31)$$

$$Al_2O_3 = 2.162 + 0.908Al_2O_3H \quad (32)$$

$$Fe_2O_3 = 1.303 + 1.007Fe_2O_3H \quad (33)$$

y mediante la sustitución de estos modelos en las Ecuaciones de Bogue se obtiene un conjunto de ecuaciones que permiten predecir los porcentajes de las fases minerales del clínker también a partir del conocimiento de los porcentajes de los óxidos en la harina. Estas ecuaciones son:

$$C3S = 87.627088 + 0.394887 C_a OH - \left( \begin{array}{l} 1.2996 SiO_2 H + 6.099944 Al_2 O_3 H + \\ + 1.44001 Fe_2 O_3 H + 1.542932 SO_3 H \end{array} \right) \quad (34)$$

$$C2S = 53.873797 + 0.490257 S_i O_2 H - 0.7544 C_3 S \quad (35)$$

$$C3A = 3.524624 + 2.4062 Al_2 O_3 H - 1.703844 Fe_2 O_3 H \quad (36)$$

$$C4AF = 3.965029 + 3.064301 Fe_2 O_3 H \quad (37)$$

## 5. CONCLUSIONES

Los métodos que se usan en la actualidad para la cuantificación de los porcentajes de las fases minerales del clínker no responden a las necesidades actuales de los especialistas de la Empresa Mixta Cementos Cienfuegos S.A. Lo más utilizado para la cuantificación de los mencionados porcentajes son las Ecuaciones de Bogue, las cuales usan para el cálculo de los porcentajes de las fases minerales los porcentajes de los principales óxidos en el clínker y no en la harina, por tanto en esta investigación se establece un método matemático para el cálculo de los porcentajes de las fases minerales del clínker sustentado en 5 modelos que permiten realizar dichos cálculos a partir del conocimiento de los porcentajes de los principales óxidos en la harina. El método se valida comparando los resultados de la aplicación del mismo con los resultados obtenidos por la empresa por análisis químico e influye en la reducción de los daños al Medio Ambiente derivados del proceso de fabricación del cemento debido a que se elimina la necesidad de realizar nuevas producciones por productos fuera de especificaciones.

## 6. RECOMENDACIONES

Que se continúe utilizando el método propuesto en la Empresa Mixta Cementos Cienfuegos S.A. Aplicar el método en otras industrias cementeras del país para así evitar pérdidas económicas por producciones de baja calidad y daños innecesarios al medio ambiente. Que se divulgue la experiencia de esta investigación otras empresas cementeras de otros países.

## REFERENCIAS

1. J. I. Tobón, "Replanteamiento de las Ecuaciones de Bogue en el cálculo mineralógico del clínker para una cementera colombiana," 2006.
2. A. Agostini, "Características Medioambientales y Riesgos en la salud por los Materiales de Construcción en las Edificaciones," 2007.
3. Holderbank, "Química y mineralogía de las materias primas del cemento: Influencia de las propiedades de las materias primas en el proceso de fabricación del cemento," in *Curso del Cemento* Bogotá, 1975, p. 46.
4. F. Glasser, "The burning of Portland cement," 1998.
5. W. H. Duda, *Manual tecnológico del cemento*, 1997.
6. A. C1365-98, "Standard test method for determination of the proportion of phases in Portland clinker using X-ray diffraction analysis." vol. 4, 1998.
7. A. C150-94, "Standard Specification for Portland Cement. ANEXO A1. Calculation of potential cement phase composition." vol. 4, 1994.
8. M. Clark, "Bogue vs Chromy," *International cement review*, 2002.
9. F. Calderón, "Resumen de microscopía de clínker Medellín," p. 16, 1977.
10. E. Fundal, "Microscopy of cement raw mix and clinker," F.L.Smidth, Copenhagen 25, 1979.
11. J. R. Camara, "Análise microscópica de clínker," Sao Paulo 16, 1988.
12. L. F. Rodriguez, "Análisis microscópico del clínker de cemento portland," Medellín 53, 1991.
13. N. Arenas, "Análisis químico instrumental en la industria del cemento. Observación microscópica cualitativa del clínker del cemento portland " 1997.

- 
14. D. H. Campbell, "Microscopical examination and interpretation of portland cement and clinker," *PORTLAND CEMENT ASSOCIATION*, vol. 2, 1999.
  15. J. I. Tobón, "Caracterización petrográfica de algunos clinker Colombianos," 2001.
  16. A. C1356M-96, "Standard test method for quantitative of phases in Portland cement clinker by microscopical point-count procedure." vol. 4, 1996.
  17. J. M. García-Márquez, "Automatic quantification of phases and mechanical characterization of materials base don Pórtland clinker modified with silica and alumina additions," *Journal of Materials Processing Technology*, p. 4, 2003.
  18. D. Lawrence, "The constitution and specification of portland cements," in *Lea's chemistry of cement and concrete*, 4 ed, 1998, p. 64.
  19. H. F. W. Taylor, *Cement Chemistry*, 2 ed., 1997.
  20. A. Crumbie, "Where is the iron? Clinker microanálisis with XRD Rietveld, optical microscopy/point counting, Bogue and SEMEDS techniques.Cement and Concrete and Research," p. 6, 2006.
  21. E. Marciano, "Estudio comparativo entre dos resultados obtenidos microscópicamente e pelo metodo de bogue e suas implicacoes," Sao Paulo1983.
  22. H. F. W. Taylor, "Modification of the Bogue calculation," *Advances in Cement Research*, 1989.

## MODELOS DE REGRESIÓN LINEAL PARA DIÁMETRO Y ÁREA FOLIAR EN *FRAXINUS UHDEI* (WENZ.) EN RESPUESTA A CONCENTRACIONES DE PARTÍCULAS ATMOSFÉRICAS EN BOSQUES URBANOS.

Luz A. Sánchez Landero\*,\*\*; Sergio Hernández González\*, María Y. Zavaleta Sánchez\*  
&

Miguel Á. López López\*,

\*Facultad de Estadística e Informática, Universidad Veracruzana.

\*\*Colegio de Postgraduados, Campus Montecillo, México

### ABSTRACT

The damage in vegetation in urban environments is diverse for the effects that air pollution causes it. Multiple linear regression models were performed using the statistical package Minitab® with Stepwise procedure to evaluate the effects of surface deposits and intracellular concentrations of copper (Cu), zinc (Zn), lead (Pb), nickel (Ni), Chromium (Cr), cobalt (Co), cadmium (Cd), nitrogen (N), phosphorus (P), potassium (K) and the pH of the soil, on the following characteristics of the foliage: leaf area, specific leaf area (SLA), the dry weight to 100 sheets and increases in diameter at breast height (DBH) of *Fraxinus uhdei* specie, on three urban forests located in the metropolitan area of the Valley of Mexico (ZMVM) Naucalli park, Chapultepec and Tlalpan forests. For the interpretation of the obtained models it assumed that there was a causal relationship between the variables of the tree and contaminants, although this assumption may not be entirely valid. The results obtained in this work in the future could be important starting base for further studies.

The main results show that in the spring, autumn and winter are the seasons in which cause and effect relationships are presented. While in the forest Tlalpan in the fall it found that leaf area, specific leaf area, dry leaves and increased diameter at breast height weight are affected by variables potassium, cadmium, zinc and soil pH.

**KEYWORDS:** Multiple Linear regression, heavy metals, Distrito Federal, Fresno.

### RESUMEN

La vegetación en ambientes urbanos padece en mayor o menor grado los efectos que la contaminación atmosférica le ocasiona. Se realizaron modelos de regresión lineal múltiple a través del paquete estadístico Minitab® con el procedimiento Stepwise para evaluar los efectos que tienen los depósitos superficiales y las concentraciones intracelulares del Cobre (Cu), Zinc (Zn), Plomo (Pb), Níquel (Ni), Cromo (Cr), Cobalto (Co), Cadmio (Cd), Nitrógeno (N), fósforo (P), potasio (K) y el pH del suelo, sobre las siguientes características del follaje: área foliar, área foliar específica (AFE), el peso seco para 100 hojas y el incremento en diámetro a la altura del pecho (DAP) de árboles de la especie *Fraxinus uhdei* ubicados en tres bosques urbanos de la Zona metropolitana del Valle de México (ZMVM): Parque Naucalli, bosque de Chapultepec y bosque de Tlalpan. Para la interpretación de los modelos obtenidos se asumió que existía una relación causa-efecto entre las variables del árbol y los contaminantes, aunque dicha suposición puede no ser del todo válida. Los resultados obtenidos en este trabajo pudieran ser en un futuro una base de partida importante para otros estudios.

Entre los resultados más destacados, se encontró en general que en la primavera, otoño e invierno son las épocas del año en las que se presentan relaciones causa-efecto. Mientras que en el bosque de Tlalpan en el otoño se encontró que el área foliar, área foliar específica, peso seco de hojas e incremento en el diámetro a la altura del pecho se ven afectados por las variables potasio, cadmio, zinc y pH del suelo.

## 1. INTRODUCCIÓN

Las partículas o material particulado (PM) cuyo término proviene del inglés *Particulate Matter*, están compuestas de partículas finas y gruesas que se encuentran suspendidas y dispersas en el aire. Éstas varían en términos de su composición química, morfología y propiedades ópticas, siendo el diámetro de las partículas su propiedad más importante, ya que existen partículas nanométricas y submicrométricas, así como la aglomeración de otras partículas de mayor diámetro, denominadas finas (1-2.5  $\mu\text{m}$ ) y gruesas (> 2.5  $\mu\text{m}$ ) (Sosa *et al.*, 2012). Se denomina como PM<sub>10</sub> a las partículas con diámetros iguales o inferiores a 10  $\mu\text{m}$  (Campos *et al.*, 2007). Las PM<sub>10</sub> están formadas por polvo, líquidos y metales que llegan a ser tóxicos para las plantas, ya que, en exceso inhiben el crecimiento y producen toxicidad en las células vegetales llegando en algunos casos a provocar la muerte (vanAssche y Clijsters, 1986; Steffens, 1990).

La vegetación urbana reduce la concentración del polvo en el ambiente, actuando como sumidero de los contaminantes que se encuentran contenidos en el aire (Younis *et al.*, 2013). Las ramas, corteza y el follaje de las plantas contribuyen a la captura de las partículas ocasionando con ello daño a la morfología interna y estructura de las hojas además de oclusión de estomas y disminución de la tasa de fotosíntesis (Farooq, *et al.*, 2000, Pourkhabbaz *et al.*, 2010, Sukumaran, 2012), clorosis

---

ocasionada (Seyyednejad, *et al.*, 2011), exceso de producción de ácido ascórbico y disminución de la transpiración (Prajapati y Tripathi, 2008, Sharifi *et al.*, 1997), por lo que es necesario considerar especies resistentes o tolerantes a ambientes urbanos, ya que los depósitos de partículas en la superficie de las plantas se ve influida por una variedad de factores.

Aparte del diámetro y forma de las partículas, también los parámetros meteorológicos: humedad y la velocidad del viento, así como las fuentes emisoras que rodean a la vegetación, influyen en la retención de las partículas (Litschke y Kuttler, 2008).

La contaminación atmosférica en la Zona Metropolitana del Valle de México (ZMVM) es muy elevada durante el año (de Bauer *et al.*, 2006, Escobedo y Chacalo 2008) Se han realizado trabajos que reportan la influencia y el daño que ésta ocasiona a la vegetación (de Bauer *et al.*, 2000, Alvarado-Rosales y Hernández-Tejeda, 2002, Bravo-Álvarez y Torres-Jardón 2002, Vega *et al.*, 2010), por lo que el propósito de este estudio fue el determinar el comportamiento que tienen el área foliar, área foliar específica, peso seco de 100 hojas, e incremento en diámetro, en función de los depósitos superficiales e intracelulares de los elementos Cu, Zn, Pb, Ni, Cr, Co, Cd y la concentración de N, P y K, así como el pH del suelo, en follaje de árboles de *Fraxinus uhdei* (Wenz) durante el año 2012. Lo anterior se realizó con modelos de regresión múltiple.

La selección de esta especie se basó en que cerca del 80% del arbolado urbano en la Ciudad de México está compuesto por los géneros *Fraxinus* y *Ligustrum* (Benavides *et al.*, 2002). Sin embargo no existen datos sobre los efectos que los elementos bajo estudio puedan tener sobre las variables del follaje y el crecimiento de los árboles

## 2. MATERIALES Y MÉTODOS

### 2.1. Área de estudio

El estudio se realizó en tres bosques urbanos de la Zona Metropolitana del Valle de México (ZMVM). Parque Naucalli, localizado al norte en las coordenadas 19° 29' 27.6" N, 99° 14' 21.2" W; bosque de Chapultepec, ubicado al poniente de la Ciudad de México a los 19° 23' 40" y 19° 25' 45" de N y 99° 10' 40" y 99° 14' 15" W y el bosque de Tlalpan, localizado en el sur a los 19° 17' 38.2" N, 99° 11' 36.3" W (Figura 1).

Se colectaron hojas ubicadas a una altura de 7 metros sobre el nivel del suelo con ayuda de una garrocha adaptada con una tijera para podar, una vez por estación y durante un año, en los tres bosques. La recolección de hojas se hizo para 45 árboles de la especie *Fraxinus uhdei* (Wenz.) distribuidos equitativamente en los tres bosques urbanos. Se definió en un transecto por bosque, procurando un espaciamiento de 50 m entre árboles y se consideraron las condiciones particulares de cada bosque, tales como distribución de la especie, obstáculos y avenidas circundantes. Se procuró que la muestra de hojas fuera de toda la copa del árbol. Además se obtuvo la ubicación de cada árbol por bosque con ayuda de un Sistema de Posicionamiento Global (GPS) Garmin eTrex y, con una cinta diamétrica, se tomó el diámetro a la altura del pecho (DAP, 1.30 m) durante otoño de 2011, 2012 Y 2013. Se estimó el incremento de DAP como la diferencia entre los años 2013 y 2011. Se tomaron muestras de suelo a una profundidad de 30 cm en los tres bosques en el año 2013 en los meses de abril, junio y septiembre para determinar el pH con un medidor marca Hanna pH 210. Asimismo, se colectó una muestra foliar del dosel de los árboles antes de la caída de hojas en el otoño con la cual se obtuvo el área foliar para 100 hojas por árbol, con la ayuda un medidor de área foliar marca Li-cor modelo Li-3100C. Posteriormente las muestras se secaron en horno a 70°C hasta que tuvieron un peso constante y se pesaron para obtener el área foliar específica (AFE) de las 100 hojas a través de la siguiente fórmula:

$$AFE = \frac{AF}{W}$$

dónde:

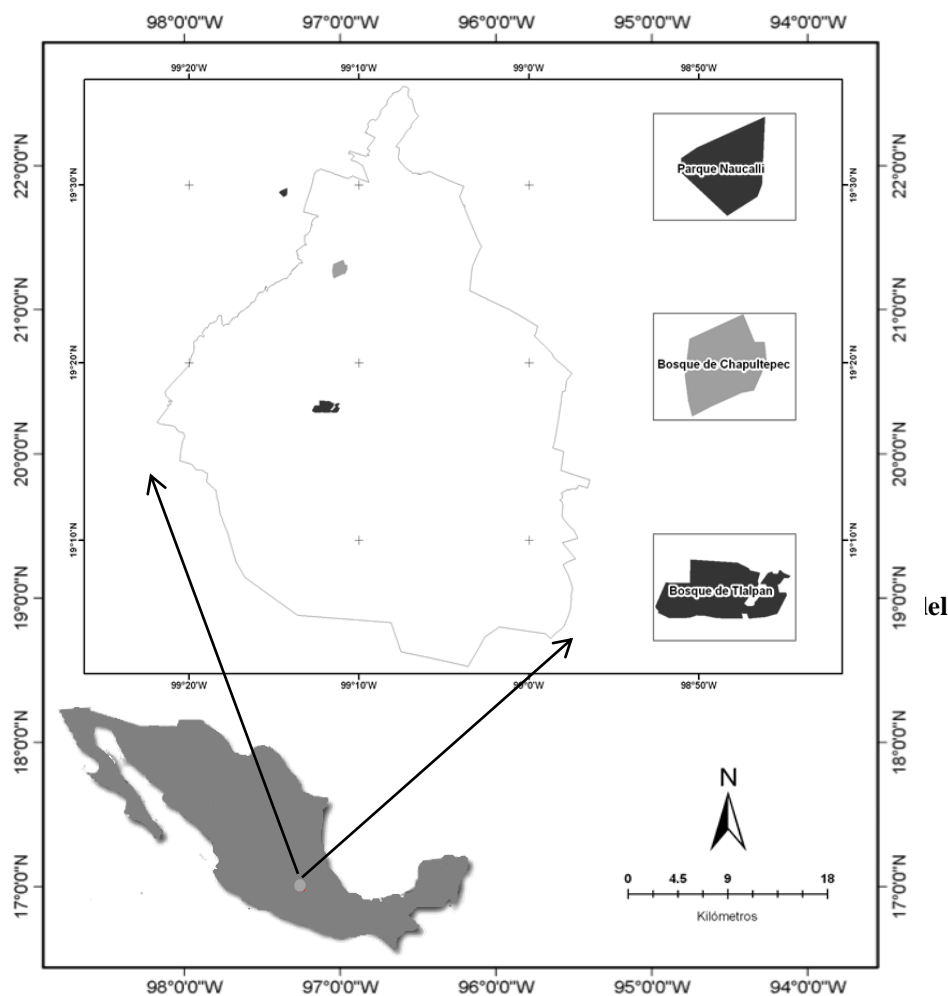
AFE= Área foliar específica de 100 hojas (cm<sup>2</sup>/g)

AF= Área foliar medida para 100 hojas (cm<sup>2</sup>)

W= peso seco de las 100 hojas (g)

Tanto el área foliar, área foliar específica y peso seco de hojas se obtuvieron a partir de una colecta antes del otoño, y éstas se correlacionaron con los metales y los macronutrientes.

Para la obtención de las concentraciones de las partículas de Cu, Zn, Pb, Ni, Cr, Co y Cd, se colectó una muestra por árbol en la primavera y el en el verano, la cual fue subdividida en dos submuestras. De la primera, se pesaron 20 g de follaje húmedo por árbol y éste fue sometido a un lavado con una solución a base de EDTA disódico a una relación molar Pb/EDTA=0.12, tal y como lo hizo Olguín *et al.* (2005), con la finalidad de cuantificar la cantidad de metales en la superficie de las hoja. Dado que el EDTA es una sustancia que tiene la capacidad de adsorber metales presentes en una solución o material, permite la cuantificación de los mismos. La segunda submuestra permaneció sin ningún tratamiento para cuantificar la concentración total de metales.



Posteriormente, ambas muestras se secaron en horno hasta que alcanzaron peso constante a una temperatura de 70°C y después se tritararon en un molino marca GI y se tamizaron con una malla de 2 mm. Se pesaron 0.25 g de la muestra molida y se sometieron a un proceso de digestión con 6 mL de HNO<sub>3</sub> en un digestor tipo micro ondas. Las submuestras fueron aforadas a 25 mL y filtradas con papel whatman sin cenizas del número 45.

En los muestreos de otoño e invierno se colectó una única muestra a la que se le aplicó un primer lavado con una solución de EDTA, de la misma forma que a las muestras de la primavera y el verano. Se realizó un segundo lavado con agua deionizada. Y las soluciones resultantes de los lavados se analizaron químicamente mientras que el follaje triturado fue digerido para su análisis.

La lectura de todas las muestras se realizó con un equipo ICP (Inductively Coupled Plasma) modelo 3000, con el cual se determinó la concentración de Cu, Zn, Pb, Ni, Cr, Co y Cd.

---

Para conocer las concentraciones superficiales de los elementos bajo estudio, que se depositan sobre las hojas, se realizaron dos tipos de manejo:

- 1) Una submuestra de hojas se lavó con EDTA y se digirió. Al lavar así, se desorbieron los metales superficiales y sólo se determinó la concentración intracelular del metal.
- 2) La Otra sub-muestra de hojas, se lavó con EDTA y la determinación se realizó al agua de lavado. En este caso, se determinaron directamente los elementos adheridos a la superficie.

$$C_t = C_i + C_s$$

Dónde:

$C_t$ = Concentración total de elementos (superficiales e intracelulares)

$C_i$ = Concentración intracelular (concentración en hojas lavadas con EDTA y digeridas)

$C_s$ = Concentración superficial obtenida por determinación directa al agua de lavado con EDTA

Para determinar P y K se realizaron los mismos procedimientos que en los metales y la lectura se llevó a cabo con un equipo ICP. Para la determinación del nitrógeno se utilizó la metodología de Kjeldahl.

Además, se consiguieron datos de precipitación, por época estacional (2011 y 2012), de las estaciones meteorológicas automáticas de la Comisión Nacional del agua (CONAGUA) más cercanas a los bosques (Presa madin, tezontle y ecoguardas) y de partículas de  $PM_{10}$  y Pb suspendidas en el aire que reporta el sistema de monitoreo atmosférico de la Ciudad de México (SIMAT) de las estaciones FES Acatlán (FAC), Merced (MER) y Pedregal (PED) para dilucidar si existe alguna relación de estas partículas con respecto a las concentraciones encontradas de Cu, Zn, Pb, Ni, Cr, Co y Cd.

Los análisis estadísticos se realizaron con el paquete MINITAB® versión 14. Se obtuvieron modelos de regresión lineal múltiple para conocer si existe algún efecto sobre el área foliar, área foliar específica y el peso seco de las 100 hojas así como el incremento del DAP ocasionado por las concentraciones foliares intracelulares y superficiales de los elementos Cu, Zn, Pb, Ni, Cr, Co, Cd, N, P, K y el pH del suelo. Lo anterior se logró a través del método *Stepwise* con *forward selection* y *backward elimination*, procurando seleccionar aquellos modelos que asemejaban un mejor comportamiento con los datos reales.

## 2. RESULTADOS Y DISCUSIÓN

Las concentraciones promedio superficiales e intracelulares de cobre (Cu), zinc (Zn), plomo (Pb), níquel (Ni), cromo (Cr) y cadmio (Cd) en el follaje de *Fraxinus uhdei* (Wenz.) Encontradas en el parque Naucalli y los bosques Chapultepec y Tlalpan, por época del año se reportan en el Cuadro 1. Dichas concentraciones cambian según el bosque urbano y la época del año (Sánchez-Landero *et al.*, 2013 y Sánchez-Landero *et al.*, 2015).

Para cobre y zinc en general se observó que las concentraciones promedio superficiales están dentro de los rangos normales para el género *Fraxinus* de acuerdo a lo reportado por Mills y Jones (1996) las cuales son de 7-20 mg Kg<sup>-1</sup> cobre y hasta 34 mg Kg<sup>-1</sup> para zinc y son más altas según lo que reporta Gutiérrez (2013) para *Fraxinus uhdei* exceptuando el Cu en primavera en el bosque los Colomos (44.42 mg Kg<sup>-1</sup>).

Plomo, níquel, cromo y cadmio presentaron concentraciones promedio superficiales e intracelulares bajas con respecto a las reportadas por Aksoy y Demerizen (2006) para un parque urbano sobre la especie *Fraxinus excelsior* (12.39 mg Kg<sup>-1</sup>, 20.01 mg Kg<sup>-1</sup>, 1.70 mg Kg<sup>-1</sup>, y 0.36 mg Kg<sup>-1</sup> respectivamente); así mismo también son menores solo para Pb comparadas con las reportadas por Gutiérrez (2013) para *Fraxinus uhdei* (9.34 mg Kg<sup>-1</sup> en hojas sin lavar y 8.6 mg Kg<sup>-1</sup> en lavadas). Sin embargo las concentraciones halladas en este estudio no llegan a exceder los niveles críticos tóxicos para plantas según Kabata- Pendias (2011), excepto níquel, dichas concentraciones son de 0.5-3 mg Kg<sup>-1</sup> para Pb, 0.5-5 mg Kg<sup>-1</sup> en Ni, 5-30 mg Kg<sup>-1</sup> Cr y 3-5 mg Kg<sup>-1</sup> Cd.

**Cuadro 1.** Concentraciones promedio de las partículas de Cu, Zn, Pb, Ni, Cr y Cd depositadas sobre hojas de *Fraxinus uhdei* (Wenz.) en las cuatro estaciones del año para el Parque Naucalli y los bosques Chapultepec y Tlalpan.

Bosque		Primavera	Verano	Otoño	Invierno					
Metal	Concentración									
(mg kg <sup>-1</sup> )	Superficial	intracelular	superficial	intracelular	superficial	intracelular	superficial	intracelular		
Naucalli	Cu	14.29 ± 6.09	13.64 ± 5.85	17.90 ± 7.78	16.75 ± 7.31	7.06±3.58	1	5.49±2.00	0.01±0.01	
	Zn	27.14 ± 9.42	20.92 ± 6.11	36.0 ±14.76	23.60 ± 8.14	8.98±2.35	0.01±0.01	10.30±2.79	0.02±0.01	
	Pb	0.24 ± 0.16	N/d	2.23 ± 2.25	0.04 ± 0.01	0.05±0.01	1	1.98±2.47	0.01±0.001	
	Ni	N/d	N/d	16.36±11.8	2.35 ± 0.58	0.01±0.01	1	0.002±0.00	0.01±0.001	
	Cr	0.52 ± 0.23	0.21 ± 0.17	3	0.84 ± 0.57	0.001±0.00	1	0.28±0.61	0.01±0.001	
	Cd	0.11 ± 0.04	0.06 ± 0.26	0.05 ± 0.05	0.02 ± 0.02	0.006±0.00	0.001±0.00	0.07±0.02	0.001±0.00	
Chapultepec	Cu	11.37±6.52	8.32±6.01	10.63±5.07	5.96±1.91	5.52±3.03	0.01±0.01	6.79±3.26	0.02±0.02	
	Zn	22.09±10.0	16.82±8.63	21.50±7.83	14.54±2.56	11.85±3.30	0.02±0.01	11.66±4.49	0.04±0.03	
	Pb	0.58±0.53	N/d	0.26±0.27	N/d	0.01±0.01	0.01±0.01	4.69±4.34	0.01±0.01	
	Ni	11.95±1.63	N/d	11.79±10.1	N/d	0.01±0.01	0.001±0.00	0.01±0.01	0.01±0.001	
	Cr	1.17±1.07	0.19±0.19	6	0.34±0.02	0.01±0.01	1	0.98±0.66	0.001±0.00	
	Cd	0.15±0.08	0.05±0.02	0.21±0.03	0.12±0.13	0.01±0.01	0.01±0.001	1.10±0.20	1	
Bosque	Metal	Primavera	Verano	Otoño	Invierno					
	(mg kg <sup>-1</sup> )	Concentración								
		superficial	intracelular	superficial	intracelular	superficial	intracelular	superficial	intracelular	
	Cu	7.59±3.79	7.59±3.79	11.82±6.76	7.75±3.97	4.17±1.05	0.02±0.01	4.77±1.96	0.01±0.001	
	Zn	25.07±7.44	25.07±7.44	29.10±6.74	19.35±5.30	10.19±2.66	0.01±0.01	9.74±4.10	0.02±0.01	
	Pb	N/d	N/d±	0.42±0.42	0.10±0.01	0.001±0.01	0.001±0.00	1.81±2.68	0.01±0.001	
	Ni	0.45±0.01	0.45±0.02	7.13±3.65	N/d±	0.01±0.01	0.01±0.001	0.001±0.01	0.001±0.00	
Cr	0.32±0.01	N/d±	0.57±0.52	N/d±	0.001±0.00	0.001±0.00	0.07±0.10	1		
Cd	0.06±0.03	0.06±0.03	0.14±0.05	0.09±0.04	1	1	0.18±0.01	0.01±0.001		

Notas: N/d= no determinado. ± es la desviación estándar.

Para la interpretación de los resultados obtenidos en este estudio, sobre los modelos de regresión lineal múltiple (Cuadros 2 - 5) se asume que hay una relación causa-efecto entre las variables del árbol: Área foliar (AF), área foliar específica (AFE) y el peso seco de 100 hojas (PS), así como el incremento de diámetro ( $\Delta D$ ) con respecto a las concentraciones superficiales e intracelulares de los elementos: Fósforo, potasio, cobre, zinc, plomo, níquel, cromo y cadmio; así como el pH del suelo y Nitrógeno (N) total. Aunque esta suposición no es del todo válida puesto que hay diversas variables que pueden influir y no fueron consideradas para los fines de este estudio. No obstante, los resultados de este trabajo pueden ser una base de partida importante para trabajos posteriores.

Se encontró que el AF, AFE, PS y  $\Delta D$ , se relacionan de manera más consistente con las concentraciones superficiales e intracelulares de los elementos potasio, cobre, zinc y cadmio; y en dos casos con el pH del suelo de otoño para el bosque de Tlalpan.



---

En el área foliar, la cual define la capacidad de la cubierta vegetal para realizar el proceso fotosintético, se encontró que el Cu y Zn se representaron de manera más constante en algunos modelos (Cuadro 2). De los modelos obtenidos para el AF llama la atención el primer caso en donde se obtuvo que para el parque Naucalli en el invierno, asumiendo valores constantes de  $Cu_{sup}$ , en promedio el AF disminuye  $118 \text{ cm}^2$  por unidad de  $Cu_{sup}$ . De la misma forma, asumiendo valores constantes de  $Zn_{sup}$ , en promedio el AF aumenta en  $144 \text{ cm}^2$  por cada unidad de  $Zn_{sup}$ , mientras que bajo un valor constante de  $Cu_{intra}$  en promedio el área foliar aumentaría  $28.66 \text{ cm}^2$ , mientras que la constante 994 es un factor de ajuste. La contradicción de  $Cu_{sup}$  y  $Cu_{intra}$ , quizás se deba a que la concentración en el ambiente de partículas suspendidas en el aire  $PM_{10}$  (entre las que se encuentra el Cu) en el invierno es en promedio de  $40 \mu\text{g}/\text{m}^3$  (Figura 2) ligeramente elevada, sin rebasar lo recomendado por la NOM-025-SSA1-1993 ( $50 \mu\text{g}/\text{m}^3$ ). El Cu puede estar depositándose sobre el follaje de *Fraxinus*, generando problemas de permeabilidad e interviniendo en procesos fotosintéticos (Reilly y Reilly, 1973 y Younis *et al.*, 2013) sin embargo, si penetra en la hoja no estará en forma asimilable mientras no se asocie con aminoácidos o péptidos (Reilly *et al.*, 1970). Por lo que en los modelos donde aparece el  $Cu_{sup}$  como un elemento con efectos negativos indican que en forma superficial en las hojas, éste tiene efectos negativos, contrariamente a lo que sucede cuando se encuentra el elemento en el interior del tejido foliar. Esto puede implicar que internamente se trata de un nutrimento que se encuentra no asimilable.

Por otro lado también se encontró que el plomo depositado superficialmente sobre el follaje, para el bosque de Chapultepec en el otoño y el invierno cuando las concentraciones promedio de las partículas suspendidas en el aire son altas de acuerdo a la NOM-025-SSA1-1993 en promedio de 48 y  $54 \mu\text{g}/\text{m}^3$  respectivamente (Figura 2), tiene un efecto negativo sobre el área foliar, por lo que en el caso del modelo otoño, este indica que asumiendo valores constantes de  $Pb_{sup}$  en promedio el AF disminuye  $120.13 \text{ cm}^2$  por cada unidad de  $Pb_{sup}$  que se aumente, mientras que en el invierno disminuye  $85.7 \text{ cm}^2$ . Lo anterior tiene sentido ya que este elemento es tóxico para la plantas y aun cuando no está muy definido el grado de toxicidad que presenta en ellas, se sabe que es inhibidor del proceso de respiración y fotosíntesis principalmente (Kabata-Pendias, 2011). Los coeficientes obtenidos son muy altos, lo que indica que dichas concentraciones son sumamente bajas (tabla 2.1) y no están siendo tóxicas para los árboles; es decir, no exceden los  $3\text{-}5 \text{ mg Kg}^{-1}$  en promedio que reporta Kabata-Pendias (2011) como tóxicos en la vegetación

Para los modelos de área foliar a nivel época estacional, los principales cambios se observan en el otoño y en el invierno (Cuadro 1), justo cuando las concentraciones de partículas suspendidas en el aire fueron altas (Figura 2) y en la temporada en la que las hojas caen y las plantas dejan de crecer y como bien se sabe, el área foliar es un parámetro característico por especie (Nouvellon *et al.*, 2010) que cambia según la edad de los individuos y la época del año (Delzon *et al.*, 2005).

El área foliar específica es definida como la razón entre la área foliar y peso de la hoja y además de eso un indicador del índice de espesor y densidad de la hoja (Hunt, 1990). Las principales causas-efecto de las partículas sobre esta variable se observan en el Cuadro 3. Se encontró que P, K, Zn y el pH del suelo son los principales elementos que generaron algún cambio sobre el área foliar. De los modelos obtenidos el del parque Naucalli en el invierno, destacan los efectos del  $P_{sup}$  y el  $P_{intra}$ . En estos casos se observa que la concentración de fósforo tiene un efecto negativo sobre el área foliar específica, lo que indica que asumiendo valores constantes de  $P_{sup}$ , en promedio el AFE disminuye  $16.2 \text{ cm}^2$  por unidad de  $P_{sup}$ . De la misma forma, asumiendo valores constantes de  $P_{intra}$ , en promedio el AFE disminuye en  $280 \text{ cm}^2$  por cada unidad de  $P_{intra}$ . La constante 135 es un factor de ajuste. En este caso las concentraciones promedio obtenidas para esta época son de  $5.3 \% P_{sup}$  y  $0.3 \% P_{intra}$ , por arriba de lo reportado como estándar por Mills y Jones (1996) para *Fraxinus*  $0.16\text{-}0.48\%$  lo que ocasiona estos efectos negativos sobre el AFE.

**Cuadro 2.** Cambios encontrados para el área foliar de 100 hojas por bosque y época estacional.

Localidad	Bosque	Época estacional	Modelo	Valor-p	V <sup>2</sup> (%)	F
Calli	Nauarero	Invierno	$AF = 994 - 118 Cu_{sup} + 144 Zn_{sup} + 28.66 Cu_{intra}$	.021	5	4
Pultepec	Chalco	Otoño	$AF = 8.52 - 23.98 Ni_{sup} - 2.16 N - 99.2 Cu_{sup} - 120.13 Pb_{sup}$	.001	2	6
		Invierno	$AF = 939 + 1.257 Co_{sup} - 85.7 Pb_{sup}$	.003	9	5
Tlalpan	Tlalco	Otoño	$AF = 5.240 - 673.6 Cd_{intra} - 151 Zn_{sup} + 678 pH_{septiembre}$	.002	6	6

$AF$ = área foliar;  $Cu_{sup}$ = Cobre superficial;  $Cu_{intra}$ = cobre intracelular;  $Zn_{sup}$ = Zinc superficial;  $Zn_{intra}$ = Zinc intracelular;  $Cd_{intra}$ = cadmio intracelular;  $pH$ = pH del suelo tomado según mes;  $N$ = nitrógeno;  $Pb_{sup}$ = plomo superficial;  $Ni_{sup}$ = níquel superficial;  $Co_{sup}$ = cobalto superficial.

Potasio también ocasionó cambios en el área foliar específica para el bosque de Tlalpan. En la primavera el modelo indica que asumiendo valores constantes de  $K_{sup}$ , en promedio el AFE aumenta 46.2 cm<sup>2</sup> por unidad de  $K_{sup}$ . De la misma forma en el otoño, asumiendo valores constantes de  $K_{intra}$ , en promedio el AFE aumenta 17.8 cm<sup>2</sup> por unidad de  $K_{intra}$ . Las concentraciones promedio obtenidas para estos periodos son de 3.6 y 3.2 % respectivamente y son bajas con respecto a Para el nitrógeno total que mostró efectos negativos (Cuadro 3) en bosque de Tlalpan en la primavera en donde el AFE disminuye en promedio 32.6 cm<sup>2</sup> por cada unidad de  $N$ , el efecto se deba quizás a que la zona estudiada, se reporta una saturación de este elemento como producto de la lluvia ácida, (Fenn *et al.*, 1999; Fenn *et al.*, 2002).

**Cuadro 3.** Relaciones entre el área foliar específica de 100 hojas por bosque y época, y metales estudiados superficial e intracelularmente.

Localidad	Bosque	Época estacional	Modelo	Valor-p	V <sup>2</sup> (%)
Calli	Nauarero	Invierno	$AFE = 135 - 16.2 P_{sup} - 280 P_{intra}$	.002	2
Tlalpan	Tlalco	Primavera	$AFE = 77.5 + 46.2 K_{sup} - 32.6 N - 0.646 Zn_{sup}$	.003	3
		Otoño	$AFE = 32.3 + 27.6 pH_{septiembre} - 3.60 Zn_{sup} + 17.8 K_{intra}$	.007	5

$AFE$ = área foliar específica;  $Cr_{intra}$ = cromo intracelular;  $N$ = nitrógeno;  $Cr_{sup}$ = cromo superficial;  $P_{sup}$ = fósforo superficial;  $P_{intra}$ = fósforo intracelular;  $K_{sup}$ = potasio superficial;  $K_{intra}$ = potasio intracelular;  $pH$  del suelo tomado según mes;  $Zn_{sup}$ = Zinc superficial;  $Ni_{sup}$ = níquel superficial.

**Cuadro 4.** Relaciones entre peso seco de 100 hojas por bosque y época, y metales estudiados superficial e intracelularmente.

Bosque	Época estacional	Modelo	Valor-p	R <sup>2</sup> (%)
Naucalli	Invierno	$PS = 11.9 + 4.81 Zn_{sup} - 38.62 Cd_{intra} + 922 Ni_{sup} + 5.81 K_{sup}$	0.002	71
Chapultepec	Primavera	$PS = 10.6 + 1.27 Zn_{sup} - 3.89 Cr_{sup} - 1.12 Cu_{intra}$	0.020	62
	Invierno	$PS = 49.5 - 12.7 K_{intra} - 1.03 Cu_{sup} - 1.41 Pb_{sup}$	0.002	64
Tlalpan	Otoño	$PS = 91.1 - 80.24 Cd_{intra} - 6.68 K_{intra}$	0.006	50

*PS*= peso seco de 100 hojas; *Zn<sub>sup</sub>*= Zinc superficial; *Cd<sub>intra</sub>*= cadmio intracelular; *Ni<sub>sup</sub>*= níquel superficial; *Zn<sub>sup</sub>*= Zinc superficial; *Cu<sub>intra</sub>*= cobre intracelular; *K<sub>intra</sub>*= potasio intracelular; *Cu<sub>sup</sub>*= Cobre superficial; *Pb<sub>sup</sub>*= plomo superficial.

Para el peso seco de 100 hojas (Cuadro 4), se observó que el Cd intracelular tuvo un efecto negativo para el parque Naucalli; en el invierno y en el otoño en el bosque de Tlalpan justo en épocas donde escasea la precipitación en la ZMVM y ocurren inversiones térmicas (Jáuregui, 1989). Este elemento puede llegar a ser tóxico para la vegetación en concentraciones de 3-5 mg kg<sup>-1</sup> (Kabata-Pendias, 2011) y se ha reportado que cuando se encuentra depositado en el suelo en concentraciones altas este tiende a ser absorbido por las plantas y a acumularse en el follaje (Migeon *et al.*, 2009). De este elemento se sabe que cuando está presente en áreas urbanas, es proveniente del tráfico vehicular (Nazzal *et al.*, 2014). Las concentraciones halladas en este estudio indican que asumiendo valores constantes de Cd para el Parque Naucalli en la primavera el PS en promedio disminuiría en 38.62 g, mientras que en Tlalpan en el otoño sería de 80.24 g.

Los principales efectos sobre el incremento en diámetro se encontraron en la primavera para el parque Naucalli, invierno para Chapultepec y otoño en el bosque de Tlalpan (Cuadro 5). Los elementos que generaron algún cambio sobre el incremento en diámetro fueron K, Cr, Ni, Cd y Zn.

Potasio se considera un nutrimento esencial para el desarrollo y crecimiento de las plantas, sin embargo en concentraciones altas (salinas) puede llegar a ser tóxico (Marschner's, 2012), quizá en ciertas épocas del año como se observó para el bosque de Chapultepec en el invierno (5 %), existan cambios en el suelo que esté siendo absorbido por la raíz, ocasionando cierto grado de toxicidad a los árboles de *Fraxinus*.

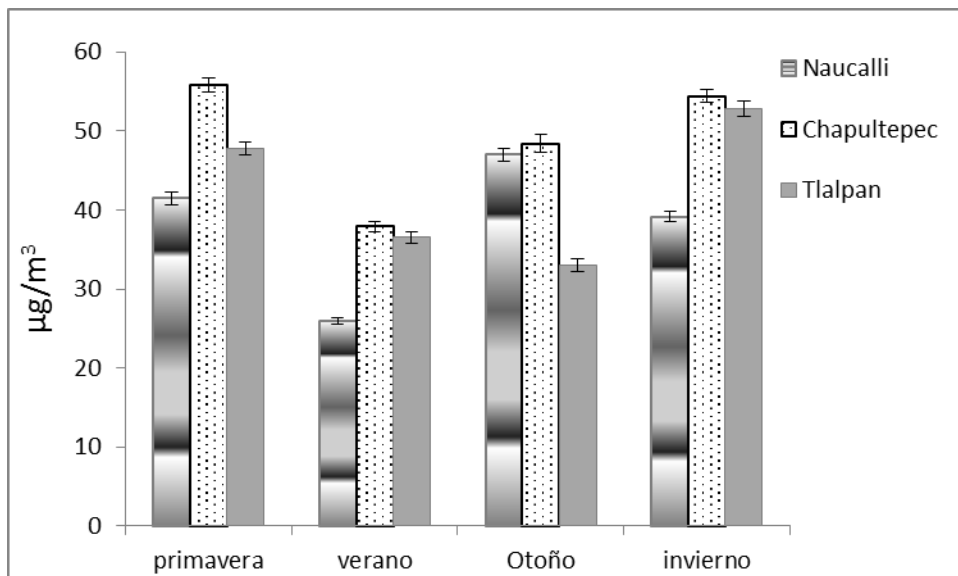
Por otro lado el Cd *intra* para el bosque de Chapultepec en el invierno mostró que el  $\Delta D$  considerando valores constantes de Cd se ve afectado en menos 0.66 cm por cada unidad de Cd *intra* que se aumente. De este elemento se sabe que en las plantas ocasiona disminución de la tasa de transpiración y la fotosíntesis así como limitación en el crecimiento de las plantas (Vassilev y Yordanov, 1997, Kos *et al.*, 2003).

**Cuadro 5.** Relaciones entre incremento en diámetro, por bosque y época, y metales estudiados superficial e intracelularmente.

Bosque	Época estacional	Modelo	Valor-p	R <sup>2</sup> (%)
Naucalli	Primavera	$\Delta D = 1.80 - 2.59 Cr_{intra}$	0.000	65.5
Chapultepec	Invierno	$\Delta D = 8.13 - 0.20 K_{intra} - 18.2 Ni_{intra} - 0.66 Cd_{sup} - 2.50 Zn_{intra}$	0.001	73.8
Tlalpan	Otoño	$\Delta D = 0.69 + 1.03 K_{sup} + 0.89 K_{intra}$	0.050	28.8

$\Delta D$ = incremento en diámetro;  $Cr_{intra}$ = cromo intracelular; pH del suelo tomado según mes;  $Cd_{intra}$ = cadmio intracelular;  $K_{intra}$ = potasio intracelular;  $Cd_{sup}$ = cadmio superficial;  $K_{sup}$ = potasio superficial;  $P_{intra}$ = fósforo intracelular;  $P_{sup}$ = fósforo superficial;  $Zn_{intra}$ = Zinc intracelular.

Gráficamente se observó que de acuerdo a lo reportado por las estaciones meteorológicas del SIMAT para partículas suspendidas en el aire PM<sub>10</sub> en el otoño y el invierno las concentraciones promedio son más altas con respecto a primavera y verano a excepción del bosque de Chapultepec el cual muestra sus valores promedios más altos en la primavera 56 µg/m<sup>3</sup> y el invierno 52 µg/m<sup>3</sup> valores que exceden lo recomendado por la NOM-025-SSA1-1993.



**Figura 2.** Concentraciones de Partículas suspendidas en el aire de tamaño PM<sub>10</sub> obtenidas de las estaciones meteorológicas del sistema de monitoreo atmosférico de la Ciudad de México cercanas al Parque Naucalli y los bosques Chapultepec y Tlalpan.

#### 4. CONCLUSIONES

Se observaron cambios en el área foliar, área foliar específica, peso seco de hojas y el incremento en diámetro principalmente en la primavera y el invierno, justo cuando los periodos de crecimiento de las plantas se ven afectados por efecto de la estación del año y cuando además en el invierno los niveles de contaminación en general aumentan por efecto de la inversión térmica.

---

El área foliar, área foliar específica, peso seco de hojas y el incremento en diámetro se ven afectados de manera más consistente por las concentraciones superficiales e intracelulares de algunos elementos como K, Cu y Zn.

## REFERENCIAS

1. Aksoy, A. & Demirezen, D. (2006). Fraxinus excelsior as a Biomonitor of Heavy Metal Pollution. *Polish Journal of Environmental Studies*, 15(1), 27-33.
2. Alvarado-Rosales, D., & Hernández-Tejeda, T. (2002). Decline of sacred fir in the Desierto de los Leones National Park. In *Urban Air Pollution and Forests* (pp. 243-260). Springer New York.
3. Beatriz, G. M. P. (2013). Acumulación de contaminantes en especies vegetales de la zona metropolitana de Guadalajara. Recuperado de [http://biblioteca.cucba.udg.mx:8080/xmlui/bitstream/handle/123456789/5561/Gutierrez\\_Martinez\\_Paulina\\_Beatriz.pdf?sequence=1](http://biblioteca.cucba.udg.mx:8080/xmlui/bitstream/handle/123456789/5561/Gutierrez_Martinez_Paulina_Beatriz.pdf?sequence=1)
4. Benavides, M. H.M., López, M. R. & Flores, H. J. (2002). Daños a banquetas por arbolado de alineación establecido en cepas en la delegación Coyoacán, Distrito Federal. *Revista Ciencia Forestal en México*, 27, 53-77.
5. Bravo-Álvarez, H., & Torres-Jardón, R. (2002). Air pollution levels and trends in the Mexico City metropolitan area. In *Urban Air Pollution and Forests* (pp. 121-159). Springer New York.
6. Campos, A., Alcaraz, G. I., Herrera, E. F., Sosa, M., Jiménez, J., Delgado, M., ... & Puga, S. (2007). Análisis temporal de las concentraciones, distribución de tamaño y morfología de partículas suspendidas menores a 10 micras en la ciudad de Chihuahua, México. *Revista Latinoamericana de Recursos Naturales*, 3(1), 44-5.
7. Delzon, S., Bosc, A., Cantet, L. & Loustau, D. (2005). Variation of the photosynthetic capacity across a chronosequence of maritime pine correlates with needle phosphorus concentration. *Annals of Forest Science*. 62(6), 537-543.
8. de la Federación, D. O. (2005). Modificación a la Norma Oficial Mexicana NOM-025-SSA1-1993, Salud ambiental. *Criterios para evaluar la calidad del aire ambiente, con respecto a material particulado*. Recuperado de <http://www.aire.df.gob.mx/descargas/monitoreo/normatividad/NOM-025-SSA1-1993.pdf>
9. Escobedo, F. & Chacalo, A. (2008). Estimación Preliminar de la descontaminación atmosférica por el arbolado urbano de la ciudad de México. *Interciencia*, 33, 29-33.
10. Farooq, M., Arya, K. R., Kumar, S., Gopal, K., Joshi, P. C., & Hans, R. K. (2000). Industrial pollutants mediated damage to mango (*Mangifera indica*) crop-a case study. *Journal of Environmental Biology*, 21(2), 165-167.
11. Fenn, M. E., De Bauer, L. I., Quevedo-Nolasco, A., & Rodríguez-Frausto, C. (1999). Nitrogen and sulfur deposition and forest nutrient status in the Valley of Mexico. *Water, Air, and Soil Pollution*, 113(1-4), 155-174.
12. Fenn, M. E., De Bauer, L. I., Zeller, K., Quevedo, A., Rodríguez, C., & Hernández-Tejeda, T. (2002). Nitrogen and sulfur deposition in the Mexico City Air Basin: Impacts on forest nutrient status and nitrate levels in drainage waters. In *Urban Air Pollution and Forests* (pp. 298-319). Springer New York.
13. Hunt, R. (1990). *Basic growth analysis*. Unwin Hyman Ltd.
14. Jáuregui, E. (1989). The dust storms of Mexico City. *International Journal of Climatology*, 9(2), 169-180.
15. Kabata-Pendias, A. (2011). *Trace elements in soils and plants*. FL. (USA): CRC Press.
16. Kos, B., Greman, H., & Lestan, D. (2003). Phytoextraction of lead, zinc and cadmium from soil by selected plants. *Plant Soil and Environment*, 49(12), 548-553.
17. Litschke, T., & Kuttler, W. (2008). On the reduction of urban particle concentration by vegetation—a review. *Meteorologische Zeitschrift*, 17(3), 229-240.
18. Migeon, A., Richaud, P., Guinet, F., Chalot, M., & Blaudez, D. (2009). Metal accumulation by woody species on contaminated sites in the north of France. *Water, air, and soil pollution*, 204(1-4), 89-101.
19. Mills, H. A., Jones, J. B., & Wolf, B. (1996). *Plant analysis handbook II: A practical sampling, preparation, analysis, and interpretation guide*. Athens, GA: MicroMacro Publishing.

- 
20. Nazzal, Y., Ghrefat, H., & Rosen, M. A. (2014). Heavy Metal Contamination of Roadside Dusts: a Case Study for Selected Highways of the Greater Toronto Area, Canada Involving Multivariate Geostatistics. *Research Journal of Environmental Sciences*, 8(5), 259-273.
  21. Nouvellon, Y., Laclau, J. P., Epron, D., Kinana, A., Mabiala, A., Roupsard, O., ... & Saint-Andre, L. (2010). Within-stand and seasonal variations of specific leaf area in a clonal Eucalyptus plantation in the Republic of Congo. *Forest ecology and management*, 259(9), 1796-1807.
  22. Olguín, E. J., Sánchez-Galván, G., Pérez-Pérez, T., & Pérez-Orozco, A. (2005). Surface adsorption, intracellular accumulation and compartmentalization of Pb (II) in batch-operated lagoons with *Salvinia minima* as affected by environmental conditions, EDTA and nutrients. *Journal of Industrial Microbiology and Biotechnology*, 32(11-12), 577-586.
  23. Pourkhabbaz, A., Rastin, N., Olbrich, A., Langenfeld-Heyser, R., & Polle, A. (2010). Influence of environmental pollution on leaf properties of urban plane trees, *Platanus orientalis* L. *Bulletin of environmental contamination and toxicology*, 85(3), 251-255.
  24. Prajapati, S. K., & Tripathi, B. D. (2008). Seasonal variation of leaf dust accumulation and pigment content in plant species exposed to urban particulates pollution. *Journal of environmental quality*, 37(3), 865-870.
  25. Reilly, A., & Reilly, C. (1973). Copper-induced chlorosis in *Becium Homblei* (De Wild.) Duvign. & Plancke. *Plant and Soil*, 38(3), 671-674.
  26. Sánchez-Landero, L.A., Hernández-González, S., de Bauer, L.I., López-López, M.A., Gómez-Guerrero, A., González-Avalos, E. & Olguín, E.J. (2013). Análisis de varianza de la Contaminación en hojas de *Fraxinus uhdei* en la Zona Metropolitana del Valle de México. En *Matemáticas Estadística y Medio Ambiente*. (pp. 1-9). Dirección de Fomento Editorial de la BUAP. Puebla, Pue. Recuperado de <http://rideca.cs.buap.mx/libros/Libro%20con%20ISBN2013.pdf>
  27. Sánchez-Landero, L.A., Hernández-González, S., Zavaleta-Sánchez, M.Y. & López-López, M.A. (2015). Análisis estadístico de la contaminación en hojas de *Fraxinus Uhdei* en la Zona Metropolitana del Valle de México. En *La metodología Estadística casos de aplicación* (pp. 93-107). Imaginaria Editores, México.
  28. Sharifi, M. R., Gibson, A. C., & Rundel, P. W. (1997). Surface dust impacts on gas exchange in Mojave Desert shrubs. *Journal of Applied Ecology*, 837-846.
  29. Seyyednejad, S. M., Niknejad, M., & Koochak, H. (2011). A review of some different effects of air pollution on plants. *Research Journal of Environmental Sciences*, 5(4), 302.
  30. Sukumaran, D. (2012). Effect of Particulate Pollution on various Tissue Systems of Tropical Plants. *Central Pollution Control Board (CPCB), Zonal Office, Kolkata, India*.
  31. Vega, E., Eidels, S., Ruiz, H., López-Veneroni, D., Sosa, G., Gonzalez, E., ... & Edgerton, S. A. (2010). Particulate air pollution in Mexico City: a detailed view. *Aerosol and Air Quality Research*, 10(3), 193-211.
  32. Van Assche, F. & Clijsters H. (1986). Inhibition of photosynthesis in zinc effects on electron transport and photophosphorilation. *Physiologia Plantarum*, 66(4), 717-721.
  33. Vassilev, A., & Yordanov, I. (1997). Reductive analysis of factors limiting growth of cadmium-treated plants: a review. *Bulg. J. Plant Physiol*, 23(3-4), 114-133.
  34. Younis, U., Bokhari, T. Z., Malik, S. A., Ahmad, S., & Raja, R. (2013). Variations in leaf dust accumulation, foliage and pigment attributes in fruiting plant species exposed to particulate pollution from Multan. *International Journal Agricultural Science and Research*, 3, 1-12.

## THE MULTINOMIAL SUPER POPULATION MODEL FOR BIODIVERSITY INDEXES

Carlos Bouza Herrera\*, Dante Covarrubias Melgar\*\* & Marja Li \*\*\*

\*Universidad de La Habana, Cuba

\*\*Universidad Autónoma de Guerrero, Mexico

\*\*\*Software development division, Institute of Computing Training, China

### ABSTRACT

This paper is concerned with the use of superpopulation models in the study of biodiversity indexes and considers the use of a multinomial one. Estimators are compared using simulated data.

**KEYWORDS:** biodiversity indexes, multinomial superpopulation, estimators.

### RESUMEN

Este trabajo esta referido al uso de modelos superpoblacionales para acometer el estudio de indices de biodiversidad y considera el uso de un modelo multinomial. Se comparan estimadores usando datos simulados.

## 1. INTRODUCTION

The study of diversity indexes may be tackled using different approaches. They are associated to some basic principles linked with data collection. The experimental approach is proactive when data collection is considered. The experimental approach is adaptive with respect to the management. It develops experiments under different treatments and levels of them and comparisons are developed. The use of models for inferring allows to disregard the randomness of the samples involved. Representative is granted by the model properties as it generates that variables or the measure of biodiversity.

It is well known that there exist collections of biodiversity measures. The great majority of them have been proposed with a basis on biological issue and a statistics frame work was not determined. Hence to analyze the results is considerably difficult due to nonlinearities and other aspects related with the structure of the measures. In any case ecologists, when performing quantitative studies, use measures for obtaining objective criteria. That is the case when studying biodiversity for evaluating the health of ecosystems. The simulation of the values of them indicates the stress level of the ecosystem studied.

The need of dealing with flora in large surfaces is very common. Kepton (2002) discussed how diversity measures must be used for elaborating decisions. Simpson (1949) proposed a diversity index which one of the most popular ones. Fager proposed another index which is based in a ranking criteria of the species. Bouza-Covarrubias (2005a y 2005b) developed a study of these indexes using as framework the theory of finite population sampling.

Section 2 is concerned with superpopulation models theory and motivate the use of the multinomial one. The third section are compared the behavior of the estimations of the indexes using simulated data.

## 2. SUPERPOPULATION MODELS

The basic model in finite population sampling is associated with the search for a probability measure that permits to implement the selection of the sample. It is called sampling model. The selection of sample  $s$  from a finite population  $U = \{1, \dots, N\}$  is made by assigning a probability  $d(s)$  to

each possible sample. The estimation is developed once an interest variable Y is defined. It may be measured for any  $i \in U$  determining a vector

$$Y(U) = (Y_1, \dots, Y_N)^T,$$

statistical population.

Consider that U is a population of sample sites and that sample s is selected. The selection of a simple sites is made performing independent Bernoulli trials. Let us define then the independent random variables:

$$I(j|s) = \begin{cases} 1 & \text{if } j \in s \\ 0 & \text{if } j \notin s \end{cases}$$

where  $E(I(j|i)) = \pi_i$ , for any  $i=1, \dots, K$  y  $j=1, \dots, n$ .

An approach that may be used is the Bayesian models implemented by a superpopulation model. It considers that the a priori  $\phi$  distribution of the population parameter Y (U) is generated by a random mechanism that belongs to a class  $\Phi = \{\phi\}$ . We are interested in determining structural relations between the expected value of functions of known variables. That allows working with broader subclasses of  $\Phi$  determined by properties of expectations and variances. Then simulations can be developed simply by using these superparameters.

A convenient superpopulation model, which is sufficiently general, see Chaudhuri-Stenger (1992) is

$$\zeta: E_{\zeta}(Y_i) = \mu_i; \text{Cov}_{\zeta}(Y_i, Y_j) = \begin{cases} \sigma_i^2 & \text{if } i = j \\ \rho\sigma_i\sigma_j & \text{if } j \neq i \end{cases}$$

The involved parameters are unknown generally. We may suppose an auxiliary variable X perfectly known exists. The superparameters are described by a model which depends of X. The simple regression superpopulation model is a good fit for characterizing the expectation of Y under the superpopulation model  $\zeta$ . Take it as

$$\mu_i = \alpha_i + \beta X_i + \varepsilon_i$$

$\alpha_i$  and  $\beta$  may be known or not but the values  $X_1, \dots, X_N$  are of common knowledge.

The biodiversity index of Simpson is:

$$\lambda_{Simp} = \sum_{i=1}^k \{1 - \pi_i\} \pi_i = 1 - \sum_{i=1}^k \pi_i^2$$

Using the plug-in-rule as the estimator  $\frac{n_i}{n} = p_i$  is unbiased for  $\pi_i$

$$\hat{\lambda}_{Simp} = 1 - \sum_{i=1}^k \left( \frac{n_i}{n} \right)^2$$

It is biased and its bias is given by:

$$\text{bias}(\hat{\lambda}_{Simp}) = - \sum_{i=1}^k \frac{\pi_i(1 - \pi_i)}{n}$$

A bias corrected estimator is

$$L_S = 1 - \sum_{i=1}^k \frac{n_i(n_i - 1)}{n(n - 1)}$$

And its variance is:



$$V\{L_S\} = \frac{2}{n(n-1)} \left[ \sum_{i=1}^k \pi_i^2 + 2(n-2)\pi_i^3 - (2n-3) \left( \sum_{i=1}^k \pi_i^2 \right)^2 \right]$$

See a proof in Bouza-Covarrubias (2005b).

Under the superpopulation model the vector  $\pi(U) = (\pi_1, \dots, \pi_K)^T$  is random.

Let us study the behavior of the superpopulation model

$$\begin{aligned} \varphi: E_{\varphi}(\pi_i) &= \theta_i, \quad V_{\varphi}(\pi_i) = \sigma_i^2, \quad \forall i, i' = 1, \dots, K; \\ \text{Cov}_{\varphi}((\pi_i, \pi_{i'})) &= \rho \sigma_i \sigma_{i'} = \sigma_{ii'} \quad \forall i \neq i', i, i' = 1, \dots, K. \end{aligned}$$

Note that this model allows using the Multinomial distribution in the analysis of the behavior of estimators. That is

Doing some algebraic work after doing the needed substitutions the expected random error under  $\varphi$  is

$$\begin{aligned} E_{\varphi} V\{L_S\} &= \frac{2}{n(n-1)} \left[ \sum_{i=1}^k \sigma_i^2 + \left( \frac{2n(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] + \\ &+ \frac{2}{n(n-1)} \left[ \left( 1 + \frac{4n-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 \right] + \\ &+ \frac{2}{n(n-1)} \left[ \frac{2n(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n-3)}{N^4} \Psi(4) \sum_{i=1}^k \theta_i^4 \right] - \\ &- \frac{2}{n(n-1)} \left[ \frac{(2n-3)}{N^4} \left( \Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) + \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right) \right] \end{aligned}$$

From this expression we may deduce easily the error under the use of a simple random sampling with replacement of  $m$  sample sites. Take  $n(j)$  as the number of individuals in the  $j$ -th sampled site, The estimate index in that site is

$$L_{Sj} = 1 - \sum_{i=1}^k \frac{n_{ij}(n_{ij}-1)}{n(j)(n(j)-1)}$$

We have two alternative predictors when  $m$  sampling sites are selected. One of them is the separate index

$$L_{SS} = \sum_{j=1}^m \frac{L_{Sj}}{m}$$

Consider that the same Multinomial models holds. We are working with the conditioning to fixed simple sizes  $n(j)$ ,  $j=1, \dots, m$  and  $L_{Sj}$  is the index of Simpson computed in the  $j$ -th sample. Hence, as we are working with a linear combination of the variances, the  $\varphi$ -expectation of  $V(L_{SS})$  is

$$V\{L_{SS}\} = \frac{1}{m^2} \sum_{j=1}^m V\{L_{Sj}\}$$

Its model expectation is

$$\begin{aligned} E_{\varphi} V\{L_{SS}\} &= \frac{1}{m^2} \sum_{j=1}^m \left\{ \left( \frac{2}{n(j)(n(j)-1)} \left[ \sum_{i=1}^k \sigma_i^2 + \left( \frac{2n(j)(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] \right) \right\} + \\ &+ \frac{1}{m^2} \sum_{j=1}^m \frac{2}{n(j)(n(j)-1)} \left( 1 + \frac{4n(j)-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \\ &+ \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[ \frac{2n(j)(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n(j)-3)}{N^4} \Psi(4) \sum_{i=1}^k \theta_i^4 \right] \right\} \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \frac{(2n(j)-3)}{N^4} \left( \Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right\} \\ &- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \frac{(2n(j)-3)}{N^4} \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right\} \end{aligned}$$

$$n^* = \sum_{i=1}^k \sum_{j=1}^m n_{ij}, \quad n_i^* = \sum_{j=1}^m n_{ij}, \quad p_i^* = \frac{n_i^*}{n^*}$$

Consider the combined estimator

$$L_{SC} = 1 - \sum_{i=1}^k \frac{n_i^* (n_i^* - 1)}{n^* (n^* - 1)}$$

Its unbiasedness has been derived in Bouza-Covarrubias (2005b) and its variance is

$$V\{L_{SC}\} = \frac{2}{n^* (n^* - 1)} \sum_{i=1}^k \pi_i^2 + 2(n^* - 2)\pi_i^3 - (2n^* - 3) \left( \sum_{i=1}^k \pi_i^2 \right)^2$$

Under  $\varphi$

$$\begin{aligned} E_{\varphi} V\{L_{SC}\} &= \frac{2}{n^* (n^* - 1)} \left[ \sum_{i=1}^k \sigma_i^2 + \left( \frac{2n^* (4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] + \\ &+ \frac{2}{n^* (n^* - 1)} \left[ \left( 1 + \frac{4n^* - N - 2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \frac{2n^* (N-1) - 4N + 3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 \right] - \\ &- \frac{2}{n^* (n^* - 1)} \left[ \frac{(2n^* - 3)}{N^4} \left( \Psi(4) \left( \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 - \sum_{i=1}^k \theta_i^4 \right) + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right] - \\ &- \frac{2}{n^* (n^* - 1)} \left( \frac{(2n^* - 3)}{N^4} \right) \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \end{aligned}$$

Under a Multinomial model

$$\varphi(B): E_{\varphi}(\pi_i) = \theta_i, \quad V_{\varphi}(\pi_i) = \theta_i(1-\theta_i)/N, \quad \forall i, = 1, \dots, K.$$

The expected errors have the particular expressions

$$\begin{aligned}
E_{\varphi} V\{L_{SS}|B\} &= \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[ \sum_{i=1}^k \frac{\theta_i + (N-1)\theta_i^2}{N} \right] \right\} + \\
&+ \frac{1}{m^2} \sum_{j=1}^m \frac{2}{n(j)(n(j)-1)} \left[ \left( \frac{4n(j)-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \left( \frac{2n(j)(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] \\
&+ \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[ \frac{2n(j)(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 \right] \right\} - \\
&- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left( \frac{(2n(j)-3)}{N^4} \right) \left( \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} - \Psi(4) \sum_{i=1}^k \theta_i^4 \right) \right\} - \\
&- \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left( \frac{(2n(j)-3)}{N^4} \right) \left( \Psi(4) \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 + \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) \right) \right\}
\end{aligned}$$

and

$$\begin{aligned}
E_{\varphi} V\{L_{SC}|B\} &= \frac{2}{n^*(n^*-1)} \left[ \sum_{i=1}^k \frac{\theta_i + (N-1)\theta_i^2}{N} \sigma_i^2 + \left( \frac{2n^*(4N-3)}{N^4} \right) \Psi(1) \sum_{i=1}^k \theta_i \right] + \\
&+ \frac{2}{n^*(n^*-1)} \left( \frac{4n^*-N-2}{N^4} \Psi(2) \right) \sum_{i=1}^k \theta_i^2 + \\
&+ \frac{2}{n^*(n^*-1)} \left[ \frac{2n^*(N-1)-4N+3}{N^4} \Psi(3) \sum_{i=1}^k \theta_i^3 - \frac{(2n^*-3)}{N^4} \Psi(4) \left( \sum_{i=1}^k \theta_i^4 - \sum_{i \neq i'} \theta_i^2 \theta_{i'}^2 \right) \right] \\
&- \frac{2}{n^*(n^*-1)} \left[ \frac{(2n^*-3)}{N^4} \left( \Psi(3) \sum_{i \neq i'} \theta_i \theta_{i'} (\theta_i + \theta_{i'}) + \Psi(2) \sum_{i \neq i'} \theta_i \theta_{i'} \right) \right]
\end{aligned}$$

When N is sufficiently large for accepting that  $Hn(j)/N \cong 0$  for  $H < 5$  as approximations:

$$E_{\varphi} V\{L_{SS}|B\} \cong \frac{1}{m^2} \sum_{j=1}^m \left\{ \frac{2}{n(j)(n(j)-1)} \left[ (2n(j)-4) \sum_{i=1}^k \theta_i^3 - (2n(j)-3) \left( \sum_{i=1}^k \theta_i^2 \right)^2 \right] \right\}$$

$$E_{\varphi} V\{L_{SC}|B\} \cong \left\{ \frac{2}{n^*(n^*-1)} \left[ (2n^*-4) \sum_{i=1}^k \theta_i^3 - (2n^*-3) \left( \sum_{i=1}^k \theta_i^2 \right)^2 \right] \right\}$$

Bouza-Schubert (2004) and Bouza-Covarrubias (2005 a) have studied the biodiversity index developed by Fager (1972)

$$\lambda^*_F = [N(K+1) - J(K-J)]/2 - \sum_{i=1}^K N_i R_i = \lambda^*_0 - \sum_{i=1}^K N_i R_i$$

where  $J \in [0, K]$  is an integer and  $R_1, \dots, R_K$  are ranks assigned to the species of interest in decreasing order in terms of the importance of them.

Fager (1972) assumed that  $R_i$  is fixed by the ecologist. Then  $\lambda_F$  is estimated unbiasedly by

$$L_F = \lambda_0 - \sum_{i=1}^K p_i R_i$$

Its variance is

$$V(L_F) = [\sum_{i=1}^K R_i^2 \pi_i - (\sum_{i=1}^K R_i \pi_i)^2] / n.$$

Under the superpopulation  $\varphi$ , as

$$V(L_F) = [\sum_{i=1}^K R_i^2 \pi_i - (\sum_{i=1}^K R_i^2 \pi_i^2 + \sum_{i \neq i'} R_i R_{i'} \pi_i \pi_{i'})] / n$$

the expected error is

$$E_{\phi} V(L_F) = [\sum_{i=1}^K R_i^2 \theta_i - (\sum_{i=1}^K R_i^2 (\theta_i^2 + \sigma_i^2) + \sum_{i \neq i'} R_i R_{i'} (\sigma_{ii'} + \pi_i \pi_{i'}))] / n =$$

$$= [\sum_{i=1}^K R_i^2 \theta_i (1 - \theta_i) - (\sum_{i \neq i'} R_i R_{i'} \sigma_{ii'} + \sum_{i=1}^K R_i^2 \sigma_i^2)] / n$$

Hence for an independent random sample of size m the separate estimator is

$$L_{FS} = \frac{1}{m} \sum_{j=1}^m L_{FS(j)} = \lambda_0 - \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^k R_i n_{ij} / n_i^*$$

Its sample error is

$$L_{FS} = \frac{1}{m^2} \sum_{j=1}^m V(L_{FS(j)}) = \frac{1}{m^2} \sum_{j=1}^m \frac{\sum_{i=1}^k R_i^2 \pi_i^2 + \sum_{i \neq i'} R_i R_{i'} \pi_i \pi_{i'}}{n(j)}$$

Combining the information for estimating  $\pi_i$

$$L_F(C) = \lambda_0 - m^{-1} \sum_{i=1}^K R_i p_i^*$$

The variance is given by

$$V(L_F(C)) = [\sum_{i=1}^k R_i^2 \pi_i - (\sum_{i=1}^k R_i^2 \pi_i^2 + \sum_{i \neq i'} R_i R_{i'} \pi_i \pi_{i'})] / n^*$$

Esta estructura ayuda a establecer que los errores esperados bajo  $\phi$  son

$$E_{\phi} V(L_F(S)) = m^{-2} \sum_{s(t) \in C} [\sum_{i=1}^k R_i^2 \theta_i (1 - \theta_i) - (\sum_{i \neq i'} R_i R_{i'} \sigma_{ii'} + \sum_{i=1}^k R_i^2 \sigma_i^2)] / n(t)$$

$$E_{\phi} V(L_F(C)) = [\sum_{i=1}^k R_i^2 \theta_i (1 - \theta_i) - (\sum_{i \neq i'} R_i R_{i'} \sigma_{ii'} + \sum_{i=1}^k R_i^2 \sigma_i^2)] / n^*$$

### 3. COMPARING THE ESTIMATORS

We use the data base of investigación Ordenamiento Ecológico Territorial desarrollada en la subcuenca del río San Juan, Guerrero, México, see Almazán et al. (2003). A simulation model was implemented for generating the population of the species with more than 9 units. The number of simple sites was fixed in 600 and 48 200 plants classified in 66 species. The generated populations allowed determining hypothetical values of the superpopulation parameters in each site. Were generated H=1000 samples of size 500. The results are presented in table 1 and we calculated the relative accuracy is

$$\tau(AB) = \frac{1}{H} \sum_{h=1}^H \frac{L_{AB(h)}}{\sqrt{E_{\phi} V\{L_A\}}}, A = S, F \quad B = C, S$$

**Table 1. Relative accuracy of the estimators under the superpopulation model**

Índex	Combined	Separate
Simpson	3,1668	6,4477
Fager	2,0381	4,7629

Note that Fager's is more accurate and the combined model the best criterion.

Each realization was generated by using the generator described by the superparameters

$$\vec{\tau} | \vec{n} = (n_1, \dots, n_K)$$

We took the distribution of Dirichlet  $\text{Dir}(\alpha)$ ,  $\alpha \in \mathcal{R}^K$ , as the conjugate of the Multinomial. Hence the involved probabilities were generated using the probability density function

$$f(\pi_1, \dots, \pi_K | \alpha_1, \dots, \alpha_K) = \begin{cases} \frac{\prod_{i=1}^K \pi_i^{\alpha_i - 1}}{\beta(\alpha_1, \dots, \alpha_K)} & \text{si para todo } i = 1, \dots, K \pi_i > 0 \text{ y } \sum_{i=1}^K \pi_i = 1 \\ 0 & \text{en otro caso} \end{cases}$$

where

$$\beta(\alpha_1, \dots, \alpha_K) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}$$

Then the superpopulation model is described by

$$E(\pi_i) = \frac{\alpha_i}{\sum_{i=1}^K \alpha_i} = \theta_i$$

$$V(\pi_i) = \frac{\alpha_i (\sum_{j \neq i} \alpha_j)}{(\sum_{i=1}^K \alpha_i)^2 (\sum_{i=1}^K \alpha_i + 1)} = \frac{\theta_i (1 - \theta_i)}{\sum_{i=1}^K \alpha_i + 1}$$

$$\text{Cov}(\pi_i, \pi_j) = \frac{-\alpha_i \alpha_j}{(\sum_{i=1}^K \alpha_i)^2 (\sum_{i=1}^K \alpha_i + 1)} = \rho \sqrt{\frac{\alpha_i (\sum_{j \neq i} \alpha_j)}{(\sum_{i=1}^K \alpha_i)^2 (\sum_{i=1}^K \alpha_i + 1)}} \sqrt{\frac{\alpha_j (\sum_{j \neq j} \alpha_j)}{(\sum_{j=1}^K \alpha_j)^2 (\sum_{j=1}^K \alpha_j + 1)}}$$

Then the hidden distribution of the species proportions are the superparameters,  $\pi_1, \dots, \pi_K$  of the distribution of the  $m$  samples given the priori  $\text{Dir}(\alpha) = \text{Dir}(\bar{\alpha} + \bar{r})$  where  $\bar{r}$  is the histogram computed from the samples.

For simplifying we consider only 2 species, the two higher frequencies, and the rest of the species in as a third one. We fixed  $n(j)=n(0)=10$ ,  $m=5$  and  $y \alpha=(6, 3, 1)$ . The results are given below in table 2.

**Table 2. Expected Variance under the MSP for  $n(0)=10$**

Estimator	$\alpha=(6,3,1)$	$\alpha=(4,4,2)$	$\alpha=(1,7,2)$
$L_{SS}$	0,033	0,021	0,118
$L_{SC}$	0,099	0,707	0,014
$L_{FS}$	0,052	0,027	0,052
$L_{FC}$	0,052	0,027	0,052
	0,052		

Note that if biodiversity is generated by  $\alpha=(6,3,1)$  o  $\alpha=(1,7,2)$  the best option is using  $L_{SC}$  while for  $\alpha=(4,4,2)$  is better using  $L_{SS}$ . Fager improve the accuracy. See table 3 where the relative errors are given.

Note that in the case  $\alpha=(6,3,1)$  or  $\alpha=(1,7,2)$   $L_{SC}$  is better while for  $\alpha=(4,4,2)$  that role is played by  $L_{SS}$ . Fager's estimators are the best alternatives.

**Table 3.** Expected Variation Coefficient under the MSP for  $n(0)=10$

Estimator	$\alpha=(6,3,1)$	$\alpha=(4,4,2)$	$\alpha=(1,7,2)$
$L_{SS}$	0,224	0,179	0,421
$L_{SC}$	0,113	0,945	0,133
$L_{FS}$	0,245	0,029	0,245
$L_{FC}$	0,245	0,029	0,245
	0,052		

**Acknowledgments:** This paper has been benefited by the project PNCB: Modelos Matemáticos para el Estudio de Medio Ambiente, Salud y Desarrollo Humano .

#### REFERENCES

1. ALMAZÁN, A. y otros (2003): “Ordenamiento Ecológico Territorial de la Subcuenca del Río San Juan del Estado de Guerrero”, Reporte Técnico de Investigación, SIBEJ-UAGro.
2. BOUZA C.N. D. COVARRUBIAS. (2005a): Estudio estadístico del índice de biodiversidad de Fager. Presentado en el “6th International Workshop on Operations Research”, La Habana.
3. **BOUZA, C.N. Y COVARRUBIAS, D.(2005b): Estimación del índice de diversidad de Simpson en m sitios de muestreo Inv. Operacional. 26, 187-195.**
4. BOUZA C.N. AND L. SCHUBERT (2004): The estimation of bio-diversity and the characterization of the dynamics: an application to the study of a pest. Rev. Matemática e Estadística. 21, 85-98.
5. CHEN, M-H., Q-M.SHIAO and J.G. IBRAHIM (2002): Monte Carlo Methods in Bayesian Computaion. Springer, N. York.
6. FAGER, E.W. (1972): Diversity: a sampling study. American Naturalist. 106, 293-310.
7. GELMAN A., J. B. CARLIN H. S. STERN and D. B. RUBIN (2003). Bayesian Data Analysis (2nd ed.). Chapman and Hall/CRC.
8. KEMPTON R. A. (2002): Species Diversity. In Encyclopedia of Environmetrics, 4, 2086-2092. (Eds Abbel H. El-Shaarawi and W. Piegorsh) John Wiley & Sons. N York
9. PATIL G. P. AND TAILLIE C. (1982): Diversity As A Concept And Its Measurement. Journal Of The American Statistical Association 77, 548-567.
10. SIMPSON, E. H. (1949). Measurement of diversity. Nature 163, 688.

# ALGORITMO Y PSEUDOCÓDIGO PARA CALCULAR EL VECTOR ELÉCTRICO INTEGRAL DE LA ACTIVIDAD VENTRICULAR CARDIACA EN EQUINO

Maikel Camejo Zamora, Alberto Pompa Núñez, Yasser Vázquez Alfonso  
Universidad Agraria de la Habana

## ABSTRACT

The electrocardiogram is a valuable tool in the study of cardiac electrical activity and the average comprehensive electric vector (IEV) QRS is an important parameter. Using a new algorithm and pseudo-analytical perspective is addressed IEV, graphical display allowing the solution to a problem. The objective of this work is to design the algorithm and the pseudo-code for determining the average QRS VEI. 35 horses were collected electrocardiograms and net QRS voltages were recorded in leads  $V_{1z}$ ,  $V_{2z}$  and  $V_{5z}$ . The calculations were performed manually and computationally having led the pseudocode to Java language through NetBeans IDE 8.0 environment. The results expressed no numerical difference for all procedures and are reliable in their entirety for determining QRS del VEI medium.

**KEYWORDS:** Electrocardiogram, algorithm, pseudocode, integral electric vector.

## RESUMEN

El electrocardiograma es una herramienta valiosa en el estudio de la actividad eléctrica cardíaca y el vector eléctrico integral (VEI) medio del QRS es un parámetro importante. Usando algoritmo y pseudocódigo se aborda una nueva perspectiva de análisis del VEI, permitiendo visualizar gráficamente la solución de un problema. El objetivo de este trabajo es diseñar el algoritmo y el pseudocódigo para la determinación del VEI medio del QRS. Se recolectaron 35 electrocardiogramas de equinos y se registraron los voltajes netos del QRS en las derivaciones  $V_{1z}$ ,  $V_{2z}$  y  $V_{5z}$ . Los cálculos se realizaron manualmente y computacionalmente después de haber llevado el pseudocódigo al lenguaje Java a través del entorno NetBeans IDE 8.0. Los resultados expresaron que no hay diferencia numérica por todos los procedimientos realizados y son confiables en su totalidad para la determinación del VEI medio del QRS.

## 1. INTRODUCCIÓN

El electrocardiograma de superficie de 12 derivaciones sigue siendo una herramienta valiosa en el estudio de la actividad eléctrica cardíaca, aún en los actuales momentos de alta tecnología y grandes descubrimientos. Sigue aportando información detallada sobre los procesos de despolarización y repolarización que ocurren en las células miocárdicas. Los avances en esta materia comienzan en 1903 con el asentamiento de los principios básicos de la electrocardiografía de W. Einthoven (Zerquera, 2012), sin olvidar los aportes de Lewis, Wiggers, Wilson (Rushmer, 1961; Laiken, 1976), entre otros.

Dentro de los parámetros de estudio, interesa de manera particular la determinación del vector eléctrico integral (VEI) usando diagramas de flujo y pseudocódigo, conceptos usuales en las ciencias de la computación. Los algoritmos y diagramas de flujo se vienen usando frecuentemente en Medicina (Jobe et al., 2012; Goldenberg et al., 2013).

Un diagrama de flujo según (Lanza, 1998; Vasconcelos, 2013; Lee *et al.*, 2013) es un esquema que nos permite recorrer visualmente los diferentes pasos y secuencias para resolver un problema de una manera clara, sencilla, discriminando los caminos a seguir de acuerdo a las condiciones planteadas. Para expresar la solución contenida en un gráfico y llevarla a una computadora se requiere transformar la información diagramática a través de un lenguaje que la máquina pueda entender. Existen diferentes lenguajes de alto nivel que hacen este cometido: Pascal, C, Visual Basic, Java, entre otros. Para facilitar esta tarea de una manera natural y progresiva, sin tener altos conocimientos de lenguajes de programación, se recurre al pseudocódigo. Consiste en utilizar el lenguaje cotidiano con algunos términos de los lenguajes de alto nivel. Este proceso entremezclado ayuda a entender el contenido de un diagrama de flujo hacia un lenguaje escrito.

El objetivo de este trabajo es diseñar el algoritmo y el pseudocódigo para la determinación del VEI medio del QRS de un electrocardiograma usando diagrama de flujo y pseudocódigo, para luego llevarlo a través de un lenguaje de alto nivel a fin de que la computadora pueda darnos información pertinente y compararlo con el que se realiza manualmente.

---

## 2. MATERIALES Y MÉTODOS

La investigación fue realizada en la “Finca Rancho Domingo García Frías” ubicada en la porción centro sur de las provincias actuales de La Habana y Mayabeque. Se trabajó con una población de 40 animales de las razas, pura sangre inglés (17), pura sangre español (14) y árabe (9); constituyendo el total de sementales en este centro. De ellos se tomó una muestra de 35 animales.

En la práctica médica se realiza el cálculo del VEI en humano de un electrocardiograma de superficie en el sentido de establecer el ángulo que tiene el vector medio o promedio del QRS con respecto a las derivaciones  $D_1$  y aVF, aunque se pueden considerar otras derivaciones perpendiculares si se utiliza un sistema de coordenadas cartesianas. De manera usual, se determina el voltaje neto del QRS en las derivaciones estándar  $D_1$  y aVF y se ubica el cuadrante. Existen diferentes métodos para calcular el ángulo que tiene el vector medio del QRS (Lanza, 2014).

El proceso de excitación del miocardio, se afecta por factores ambientales, nutricionales y de forma particular genéticos Sandersen *et. al.* (2006), Valladares *et. al.* (1975) plantearon, que la formación del impulso y su conducción genera corrientes eléctricas débiles que se diseminan por todo el cuerpo.

Para efectuar los registros electrocardiográficos los animales se mantuvieron, en un estado de correcto aplomo sobre sus extremidades. Se utilizó un electrocardiógrafo de marca Fukuda, modelo japonés Denshi FX-2111 de un solo canal, el cual se calibró con una señal de 1 mV para una deflexión de la aguja de 10 mm sobre el papel de calibración en la escala vertical. La velocidad del registro fue de 25 mm/s.

Las zonas o puntos donde se colocaron los electrodos se depilaron previamente y se frotaron suavemente con alcohol, para eliminar las sustancias que pudieran disminuir la conductividad eléctrica (grasas, partículas de polvo y otras), así como también se aplicó pasta hidrosoluble como elemento conductor. Se utilizaron electrodos de pinza (tipo caimán) para el registro de las derivaciones bipolares de hombro y placas metálicas para las derivaciones estándares de extremidades, con las que se obtuvo el cero de referencia para las derivaciones monopulares.

Los parámetros que fueron medidos en cada una de las derivaciones comprenden: Amplitud y duración de las ondas, y la duración de los segmentos e intervalos. La frecuencia cardíaca fue calculada dividiendo 1500 entre el número de cuadritos de 0.04 s comprendidos entre dos ondas R consecutivas, asumiéndose como divisor el promedio de cinco intervalos RR de forma similar a como ha sido planteado por otros autores (Matsui y Busano, 1987).

Se hicieron dos veces las mediciones a cada electrocardiograma y se tomaron como datos finales el promedio aritmético de cada uno para las derivaciones  $V_{1c}$ ,  $V_{5c}$  y  $V_6$ . Se define el cuadrante donde se ubica el eje eléctrico de acuerdo a los valores de  $V_{1c}$ ,  $V_{5c}$  y  $V_6$ .

Se recolectaron inicialmente 35 electrocardiogramas (realizados por un técnico entrenado) que estuvieran exentos de trastornos de conducción sobre el QRS. Este requisito se hace necesario debido a que la altura del QRS no determinaría el voltaje real en esa derivación y habría que hacer un cálculo del área bajo el complejo QRS, lo cual sería más complicado.

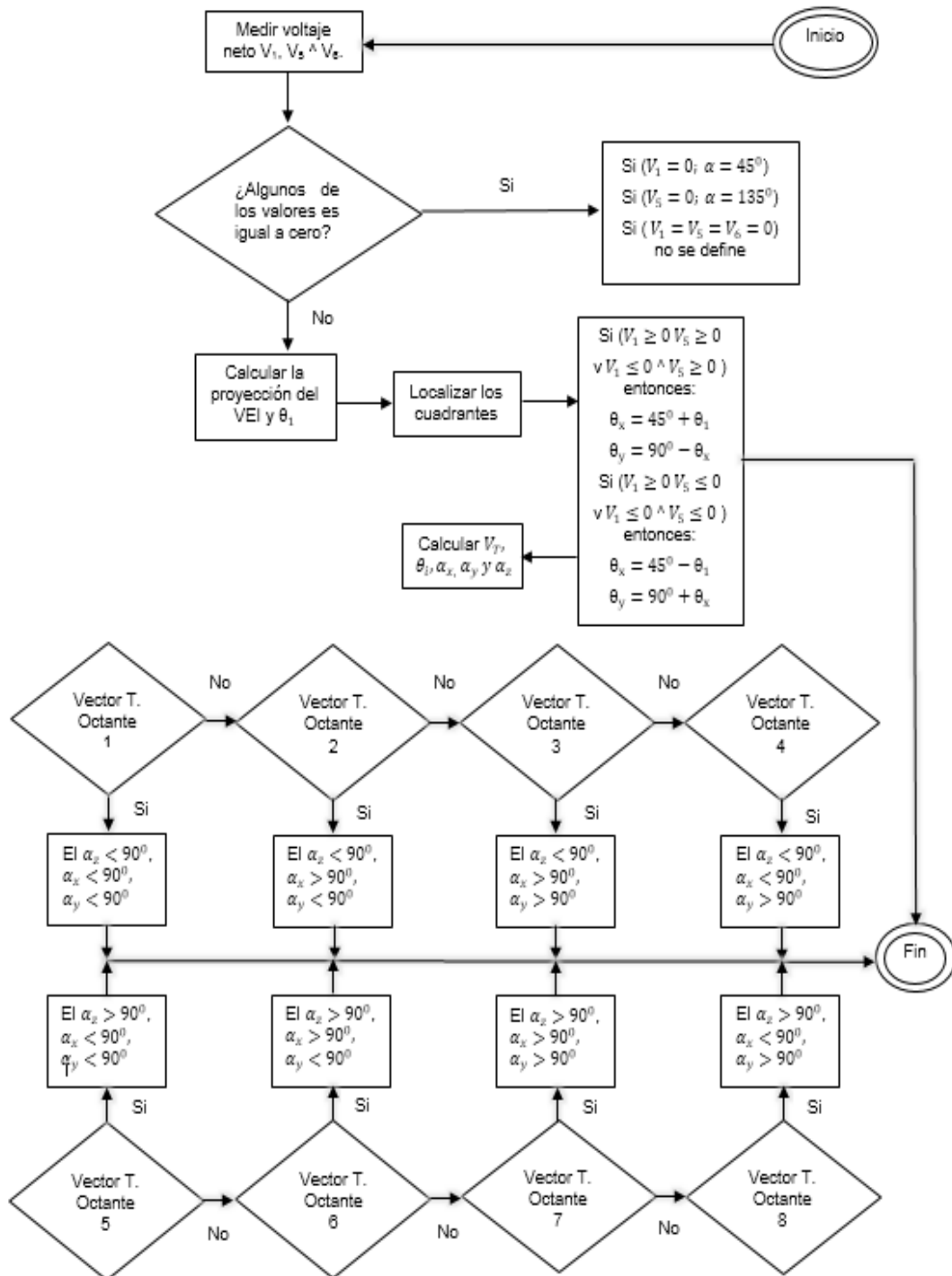
El modelo que se propuso, toma en consideración, todos estos factores y los principios que se consideraron anteriormente, y es una extensión del obtenido por Pompa (2000; 2003, 2004) donde a partir de tres derivaciones monopulares  $V_{1c}$ ,  $V_{5c}$  y  $V_6$  se determina la posición espacial del VEI de la actividad ventricular cardíaca.

## 3. RESULTADOS Y DISCUSIÓN

Para facilitar el cálculo del VEI se diseñó el algoritmo (Figura 1) que muestra los pasos y el pseudocódigo (Figura 2) que traduce los gráficos del algoritmo a través de un lenguaje natural o gramatical. La implementación del pseudocódigo a la computadora se realiza mediante el lenguaje de



programación Java a través del NetBeans IDE 8.0, creándose el software Cálculo de VEI (figura 3). Estos resultados complementan los estudios realizados por Lanza (2014) en humanos.



**Figura 1:** Algoritmo para cálculo del VEI medio.

**Fuente:** Elaboración propia.

Inicio-Programa

```

Leer voltaje: V1, V5 ^ V6
Si (voltaje V1 = 0 entonces α = 45°)
Si (voltaje V5 = 0 entonces α = 135°)
Sino si:
VT' = √(V1'² + V5'²)
Calcular:
θ1 = cos-1 (V1' / VT')
Si (V1 ≥ 0 ^ V5 ≥ 0 v V1 ≤ 0 ^ V5 ≥ 0) Entonces:
θx = 45° + θ1
θy = 90° - θ1
Si (V1 ≥ 0 ^ V5 ≤ 0 v V1 ≤ 0 ^ V5 ≤ 0) Entonces:
θx = 45° - θ1
θy = 90° + θ1
Calcular:
VT = √(V6'² + VT'²)
Calcular:
θi = cos-1 (VT' / VT)
Calcular:
αx = cos-1 [cos θx × cos θi]
αy = cos-1 [cos θy × cos θi]
αz = 90° - θi
Si αx < 90° ^ αx < 90° ^ αy < 90° Entonces 1er octante
Si αx < 90° ^ αx > 90° ^ αy < 90° Entonces 2do octante
Si αx < 90° ^ αx > 90° ^ αy > 90° Entonces 3er octante
Si αx > 90° ^ αx < 90° ^ αy > 90° Entonces 4to octante
Si αx > 90° ^ αx < 90° ^ αy < 90° Entonces 5to octante
Si αx > 90° ^ αx > 90° ^ αy < 90° Entonces 6to octante
Si αx > 90° ^ αx > 90° ^ αy > 90° Entonces 7mo octante
Si αx > 90° ^ αx < 90° ^ αy > 90° Entonces 8vo octante

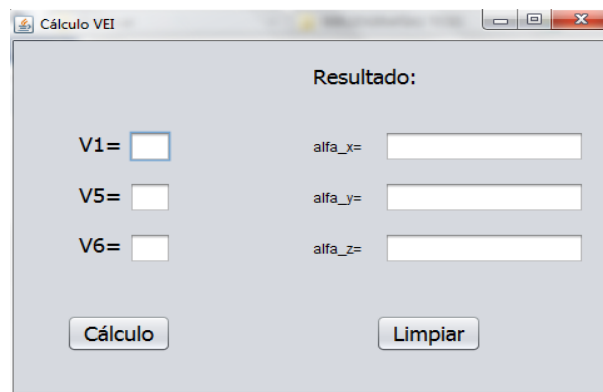
```

Fin-Programa

**Figura 2:** Seudocódigo para cálculo del VEI medio.

**Fuente:** Elaboración propia.

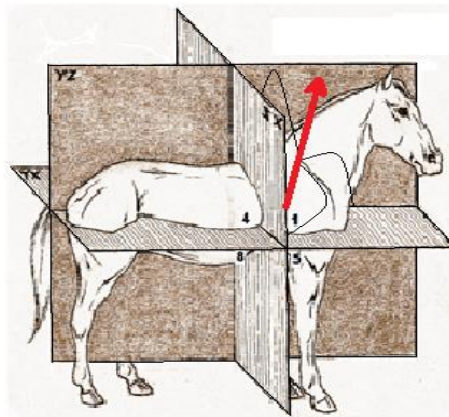
Los resultados de los cálculos del voltaje promedio ( $\pm$  desviación estándar) del QRS en  $V_{1c}$  fue de  $0,56 \pm 0,12$ , en  $V_{5c}$  de  $0,51 \pm 0,32$  y en  $V_6$   $0,94 \pm 0,22$  para  $n = 35$ . En relación a los ángulos obtenidos (en grados), entre el VEI y los ejes coordenados cartesianos X, Y, Z fue  $\alpha_x = 85 \pm 11$ ,  $\alpha_y = 52 \pm 8$  y  $\alpha_z = 39 \pm 8$ . Dos de los ángulos estaban en el octante 3; uno de los ángulos en el octante 2 y el resto en el octante 1. Esto expresa que el 91,4 % de los ángulos estaban en el primer octante, encontrándose el VEI en dirección cefálica. Este hecho es similar al predicho por Pompa (2003), Illera *et.al.* (1987), Naylor (2004), Hiraga y Sugano (2015), situación que se asemeja en el bovino por ser un animal perteneciente al grupo B, Hamlin y Smith (1960).



**Figura 3:** Software cálculo del VEI.

**Fuente:** Elaboración propia.

El análisis espacial de la posición del VEI se puede visualizar mediante la figura 4 que refleja la su ubicación en el primer octante, esto significa que el estado fisiológico del corazón equino se encuentra clínicamente sano.



**Figura 4:** Ubicación espacial del VEC medio en equino.

**Fuente:** Elaboración propia.

#### 4. CONCLUSIONES

El algoritmo presentado es confiable en cuanto a los pasos a seguir en la determinación del VEC medio de QRS en la forma establecida. Puede ser útil en el aprendizaje de la electrocardiografía al tener un algoritmo que guíe al especialista en los cálculos o que los mismos sean validados por una computadora. De la misma manera, podría ser incorporado a los equipos de electrocardiografía siempre y cuando se hagan las adaptaciones electrónicas correspondientes.

#### REFERENCIAS

1. Rushmer R (1961): Cardiovascular dynamics. 2ª edición. Saunders Co., Filadelfia..
2. Laiken S, Laiken N, O'Rourke RA, Karlner J. S. (1973): A rapid method for frontal plane axis determination in scalar electrocardiograms. *Am Heart J.* 85, 620-623.
3. Goldenberg G, Iakobishvili Z, Strasberg B. (2013): Electrocardiographic algorithm for the diagnosis of inadvertent implantation of permanent pacemaker lead in the left ventricular cavity. *Ann Noninvasive Electrocardiol*; 18, 294-296.
4. Lee G. A., Murray A, Bushnell R, Niggemeyer L. E. (2013): Challenges developing evidence-based algorithms for the trauma reception and resuscitation project. *IntEmergNurs*; 21,129-135.
5. Jobe J, Ghuysen A, D'Orío V. (2012): [Presentation of ELISA, A new triage algorithm suitable for emergency department]. *Rev Med Liege*, 67, 632-637.
6. Vasconcelos J. (2013): Manual de construcción de programas. [monografía en Internet]. México. [consultado 2013 Abr 12] Disponible en, <http://www.bibliodgca.unam.mx/manuales/manual.pdf>.
7. Lanza Tarricone G. (1998): Electrocardiograma, Modelo matemático del eje eléctrico y su uso en la determinación del voltaje del QRS en derivaciones estándar. *GacMédCaracas*; 106,509-514.
8. Lanza Tarricone G. (2014): Algoritmo y pseudocódigo, nueva perspectiva para calcular el eje eléctrico de un electrocardiograma. *Avances Cardiol*; 34,280-285.
9. Zerquera, F. M. (2012): Nociones de electrocardiografía. acimed, Editorial Ciencias Médicas, La Habana.
10. Sandersen, C. F., Detilleux, J. y Moffarts, B. (2006): Effect of atropine dobutamine stress test on left ventricular echocardiographic parameters in untrained warmblood horses. *J VetInternMed.*, 575-580.
11. Valladares, A. F., Céspedes, M. A., Morais, M. y Espinosa, J. (1975): Datos electrocardiográficos en bovino. 108. V Seminario Científico del SENIC, Editorial ORBE.
12. Pompa, A. (2003): Determinación de la orientación espacial del vector eléctrico integral de cada onda electrocardiográfica en cuadrúpedos, tomando como referencia al bovino. *Rev. Salud Anim*, 25 , 186-191.
13. Pompa, A. (2004): Biofísica, actividad eléctrica de algunos tejidos y órganos de los animales. potenciales de superficie. Tomo 1. Félix Varela, La Habana..

- 
14. Pompa, N.A. (2000), Determinación de las derivaciones electrográficas estables en bovinos Holstein. *Rev. Salud Anim.* 22, 27-34.
  15. Matsui, K. y Busano, S. (1987): Species differences in the changes in heart and T- wave amplitude after autonomic blockade in thoroughbredhorses, ponies, cows, pigs, goats and chickens. *Japan Vet. Sci.* 49 , 637-644.
  16. Illera, J. C., Illera, M. y Hamlin, R. L. (1987): Unipolar thoracic electrocardiograms that induces QRS complexis of relative uniformity from male horses. 48. 12 ed. American, *Am. J. Vet. Res.* 1700-1702.
  17. Naylor, J. M. (2004): *The Art of Equine Auscultation: An Interactive Guide.* Saskatoon.
  18. Hiraga, A. y Sugano, S. (2015): History of research in Japan on electrocardiography in the racehorse. *Journal of EquineScience* [en línea]. 26 , 1-13. Disponible en, 10.1294/jes.26.1 [Consulta, 10 junio 2015].
  19. Hamlin, R. L. y Smith, C. R. (1960): Categorization of common domestic animals base upon their ventricular activation process. *American, Ann. Acad. Sci.* 17, 195- 203.

## **PLATAFORMAS DE APRENDIZAJE: UN ENTORNO PARA LA MEJORA DE LA DOCENCIA EN EL GRADO DE ENFERMERÍA**

Emilio González-Jiménez\*, Miguel Ángel Montero-Alonso\*\*

\* Departamento de Enfermería. Facultad de Enfermería (Campus de Melilla). Universidad de Granada. C/ Santander 1, (52071) Melilla - (España).

\*\* Departamento de Estadística e I.O. Facultad de Ciencias Sociales (Campus de Melilla). Universidad de Granada. C/Santander 1, (52071) Melilla - (España).

### **ABSTRACT**

The implementation of degree in the Spanish university system requires new teaching methodologies. The objective was to determine the opinion of students in Nursing Degree Melilla (Spain) have about using the Teaching Board, SWAD and Moodle as tools to support teaching. El 47 % of students considered the most comprehensive platform Moodle. 63% used the platforms properly. 70% indicated that enrich learning platforms. Its use by teachers, 43% said they used the platforms properly. Whether the teacher properly handled platforms, 39% said yes. Whether its use should be mandatory or not between teachers, 82 % said yes. 4% answered do not know/no answer. On its usefulness for their learning, 70% of students said yes. For students, the new information technologies play an important role in its formation.

**KEYWORDS:** Virtual Platform; Teaching-learning; e-learning.

### **RESUMEN**

La implantación de los Grados en el sistema universitario español exige nuevas metodologías docentes. El objetivo fue conocer la opinión que los alumnos del Grado en Enfermería de Melilla (España) tienen acerca del uso del Tablón de Docencia, SWAD y Moodle como herramientas de apoyo a la docencia. El 47% de los alumnos consideraba Moodle la plataforma más completa. El 63% utilizaba adecuadamente las plataformas. El 70% indicó que las plataformas enriquecen el aprendizaje. Sobre su uso por profesores, el 43% afirmó que usaban correctamente las plataformas. Sobre si el profesorado manejaba adecuadamente las plataformas, el 39% manifestó que sí. Sobre si su uso debería ser o no obligatorio entre profesores, el 82% afirmó que sí. Un 4% respondió no sabe/no contesta. Relativo a su utilidad para su aprendizaje, el 70% de alumnos manifestó que sí. Para los alumnos, las nuevas tecnologías informáticas, ocupan un lugar importante en su formación.

## **1. INTRODUCCIÓN**

El creciente desarrollo tecnológico de la sociedad actual ha ocasionado cambios sustanciales a nivel de todas las profesiones, obligando ahora más que nunca a los profesionales a mantenerse actualizados y a modificar ciertos aspectos en el desempeño de sus trabajos [6]. En el ámbito educativo, profesores de todos los niveles educativos, pero especialmente del ámbito universitario, se han visto desbordados por el desarrollo de la informática y las telecomunicaciones. Si bien, este avance ha generado nuevos recursos y herramientas útiles para la docencia e investigación [1]. El impulso generado en el ámbito educativo derivado de la implantación de las TIC y sus potencialidades ha originado cambios en el rol de los profesionales de la enseñanza [3]. Asimismo, en el ámbito de la enseñanza superior, el alumno ha visto modificado igualmente su rol. Todo ello lleva al profesor a readaptar su discurso y sus estrategias docentes, dado que la universidad y el profesor dejan de ser las fuentes indiscutibles del conocimiento [4]. Así, la figura del profesor pasa a constituir una guía para el alumno facilitándole los recursos y herramientas que necesitan para explorar y elaborar nuevos conocimientos y destrezas [9].

El proceso político de convergencia europea ha contribuido hacia un cambio en la función docente. Así, se espera que el profesor universitario vaya más allá de la mera explicación o clase tradicional teórica o práctica y se convierta en una figura de ayuda y orientación en el aprendizaje para el estudiante, siendo éste último el verdadero protagonista en el proceso de aprendizaje ([11] y [12]).

Por su parte, las universidades españolas, en su proceso de actualización, cuentan con algún sistema de apoyo informático a la enseñanza tradicional, constituido en un LMS<sup>8</sup>, CMS<sup>9</sup> o plataforma

---

<sup>8</sup> LMS, del inglés Learning Management System, es un Sistema de Gestión de Aprendizaje.

---

educativa, y cada vez con más frecuencia las universidades presenciales desarrollan acciones formativas en formato b-learning<sup>10</sup>.

Múltiples estudios, sobre todo a nivel universitario, han analizado la incorporación de plataformas al proceso de enseñanza– aprendizaje, tratando aspectos como sus funcionalidades, la multiplicidad de tareas que pueden realizar, y aspectos tecnológicos del uso de las herramientas que las conforman. La formación virtual utilizando una plataforma de aprendizaje es una modalidad donde la relación entre profesor y alumno se produce en un entorno virtual y los actores no tienen necesidad de coincidir ni en espacio ni en tiempo. Este cambio por el cual el docente deja de ser fuente del conocimiento para adoptar funciones de guía, orientador y facilitador de recursos y herramientas de aprendizaje [2], lleva asociadas nuevas responsabilidades, debiendo estar al día en las innovaciones y avances en su campo de conocimiento, al mismo tiempo que a las posibles innovaciones en los procesos de enseñanza-aprendizaje derivados de la cada vez mayor implantación de las tecnologías de la información y la comunicación [5]. Si bien, es cierto que el simple uso de la tecnología en la educación no asegura la innovación ni la mejora del proceso de enseñanza-aprendizaje. Asimismo, el profesor en su rol de guía y facilitador de recursos deberá conseguir motivar a los alumnos a que participen en su propio proceso de aprendizaje. Para ello, el profesor deberá participar activamente y de forma continuada en el proceso de diseño, elaboración y actualización de los materiales de aprendizaje, organizando la enseñanza en función de las competencias que se deban adquirir [10]. Otro aspecto igualmente importante en la nueva labor del profesor será participar interactivamente en los procesos de intercambio de información y experiencias a través de las sesiones de tutorías con los alumnos, todo ello con el fin de potenciar la adquisición de herramientas de aprendizaje autónomo y permanente [13]. Además, nos va a permitir tener mucha más información del trabajo del alumno (número de accesos al sistema, tiempo empleado en realizar los cuestionarios, etc.) que en un momento dado puede ayudar para su evaluación.

En los últimos años ha comenzado a extenderse el uso de plataformas de E-learning las cuales proporcionan soporte a una infraestructura extensa de formación a través de internet. En esta línea, numerosas universidades utilizan plataformas propias creadas para satisfacer sus necesidades. Una de estas herramientas es SWAD, desarrollada en la Universidad de Granada y utilizada actualmente por 773 profesores y más de 31500 alumnos en 785 asignaturas de 137 titulaciones. SWAD constituye una plataforma web de teleformación docente que incluye a la vez funciones de apoyo al aprendizaje, docencia y gestión de estudiantes y profesores [7].

Otra plataforma igualmente utilizada y propia de la Universidad de Granada es el Tablón de docencia. Esta herramienta constituye un servicio centralizado de la Universidad de Granada que pretende facilitar el intercambio de información electrónica entre profesores y alumnos. En ella, el profesor puede escribir mensajes a sus alumnos con o sin contenido electrónico, de la misma forma que permite al alumnado escribir y enviar sus propios mensajes [8].

Si bien, a pesar de las numerosas ventajas derivadas del uso de las plataformas virtuales SWAD y Tablón de Docencia, es cierto que su uso no es generalizado entre el colectivo de alumnos ni entre el profesorado.

Moodle (Modular Object Oriented Developmental Learning Environment) es un Ambiente Educativo Virtual, sistema de gestión de cursos, de distribución libre, que ayuda a los educadores a crear comunidades de aprendizaje en línea. Este tipo de plataformas tecnológicas también se conoce como LMS (Learning Management Systems). Fue diseñado por Martin Dougiamas de Perth, (Australia), quien basó su diseño en las ideas del constructivismo en pedagogía, que afirman que el conocimiento se construye en la mente del estudiante en lugar de ser transmitido sin cambios a partir de libros o enseñanzas y en el aprendizaje colaborativo. Moodle se distribuye gratuitamente como Software Libre (Open Source), bajo Licencia pública GNU, lo que significa que Moodle tiene derechos de autor (copyright). Su primera versión apareció el 20 de agosto de 2002 y, a partir de ese momento, han

---

<sup>9</sup> CMS, en inglés Content Management System, sistema de gestión de contenidos, es un programa que permite crear una estructura de soporte (framework) para la creación y administración de contenidos, principalmente en páginas web, por parte de los participantes.

<sup>10</sup> b-learning, la formación combinada, del inglés *blended learning*, consiste en un proceso docente semipresencial; esto significa que un curso dictado en este formato incluirá tanto clases presenciales como actividades de e-learning.

---

aparecido nuevas versiones incorporando nuevos recursos, actividades y mejoras en muchos casos demandadas por su comunidad de usuarios. Actualmente, Moodle está traducido a 75 idiomas e incluye más de 27.000 lugares registrados en todo el mundo. En la actualidad, el volumen de usuarios registrados asciende a más de 260.000 al igual que el número de descargas, situándose muy cerca de las 60.000 mensuales.

Ahora bien, la metodología utilizada en un entorno semipresencial facilita el trabajo individual del alumno, ya que le permite trabajar desde cualquier lugar y a cualquier hora. Dicha metodología se ha desarrollado con la finalidad de incrementar tanto el rendimiento académico del alumno como su grado de satisfacción con el proceso de aprendizaje. Si bien, para hacer efectivo el proceso, es preciso que el profesor explicita las acciones educativas: mostrar la planificación, desde los objetivos hasta cada una de las actividades, así como un calendario de realización de las mismas.

interrelación entre todos los elementos que forman el proceso educativo: desde el diseño de los materiales didácticos, a las acciones docentes, ofrece un abanico de relaciones que es difícil comparar con el de la presencialidad. Por otro lado, se facilita el acceso a un espacio único, compartido por el profesor, estudiantes, material educativo, etc., donde las relaciones que se establecen entre estos son múltiples y facilitan procesos de construcción de aprendizajes.

## **2. OBJETIVO**

El objetivo de este estudio fue conocer la opinión que los alumnos del Grado en Enfermería de Melilla tienen acerca del uso del Tablón de Docencia, SWAD y Moodle como herramientas de apoyo a la docencia.

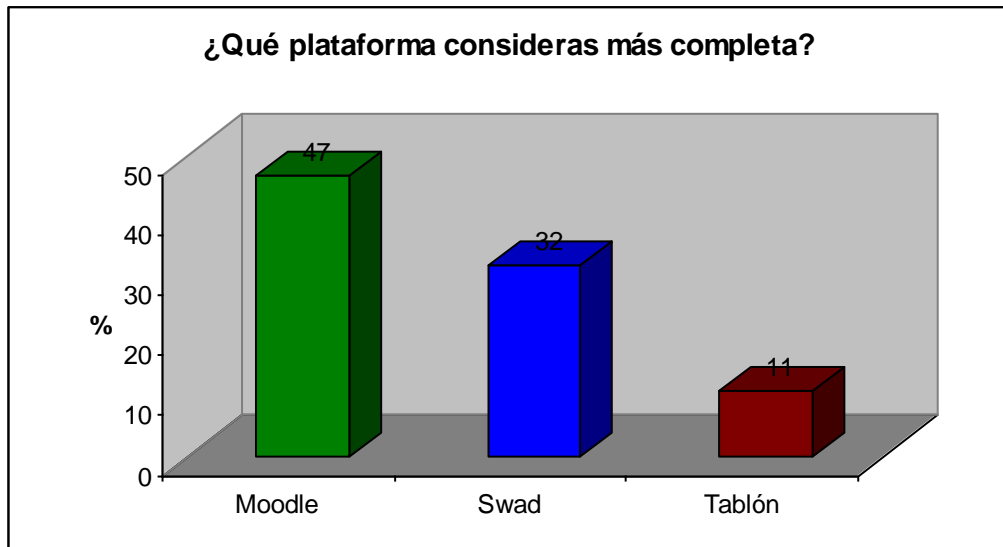
Además, se pretende fomentar el aprendizaje autónomo del estudiante, bien como método de estudio complementario al tradicional (basado en una actitud pasiva del alumnado a la hora de asistir a clase y tomar apuntes), o bien como método de aprendizaje sustitutivo para aquellos alumnos que no pueden asistir a clase por motivos diversos, trabajo, cargas familiares, discapacidad, etc. Se pretende, por tanto, favorecer que el alumno adquiera un papel activo y no se limite a recibir información sino que forme parte de su propia formación. Para ello, el alumno ha de disponer, además de la posibilidad de usar la herramienta de autoevaluación, de la documentación y material de apoyo necesario para facilitar el autoaprendizaje de las asignaturas y la comprensión de los conocimientos básicos que se exigen en las mismas.

## **3. MUESTRA Y METODOLOGÍA**

Estudio descriptivo realizado a partir de una muestra de 100 alumnos pertenecientes a primer y segundo curso del Grado de Enfermería de la Universidad de Granada en el Campus de Melilla. Se utilizó un cuestionario compuesto de seis preguntas cerradas, tres de ellas tendrían como posibles respuestas Sí o No, y las otras tres Sí, No y A veces. Los resultados se analizaron descriptivamente mediante el software estadístico SPSS 20.0.

## **3. RESULTADOS**

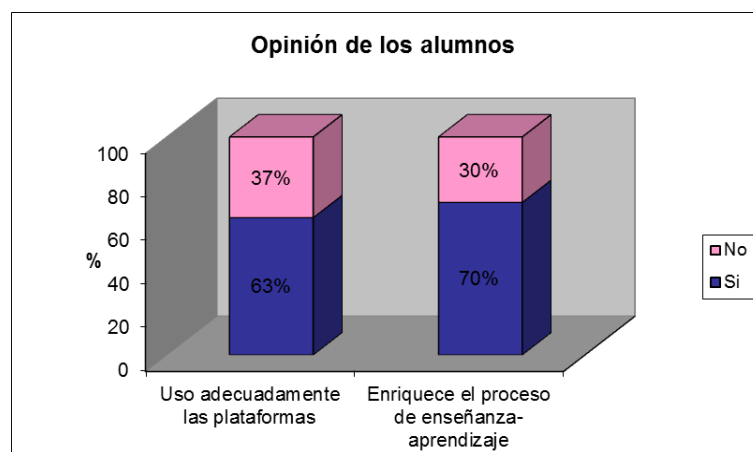
Los resultados obtenidos muestran que el 47% de los alumnos considera Moodle como la plataforma de apoyo a la docencia más completa, seguido de cerca por un 32% de alumnos para quienes la plataforma SWAD es la más completa. Tan sólo un 11% de alumnos consideró que el Tablón de Docencia era la plataforma docente más completa. Estos resultados se muestran más claramente en la Figura 1.



**Figura 1 – ¿Qué plataforma consideras más completa?**

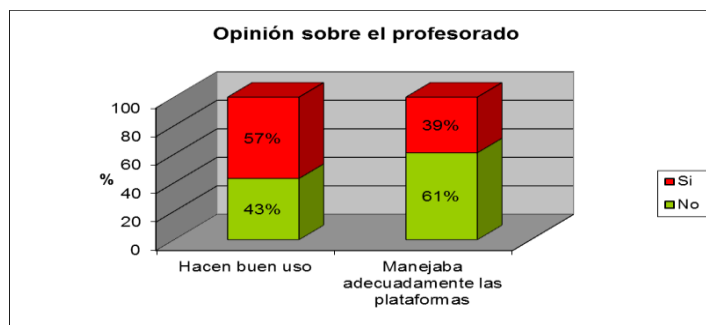
Relativo a si los alumnos usan o no adecuadamente las tres plataformas (Figura 2), un 63% manifestó utilizar adecuadamente las plataformas, seguido de otro 37% quienes indicaron no usarlas adecuadamente. Se realizó un análisis de contingencia entre el uso adecuado de las tres plataformas y el género de los estudiantes mediante la test de al Chi-cuadrado resultando un valor de 0,064, por lo que se concluye que el uso de las tres plataformas no está asociado con el sexo de los estudiantes con un p-valor de 0,8.

En relación a si su uso enriquece o no el proceso de enseñanza aprendizaje, el 70% de los alumnos indicó que usar dichas plataformas enriquece el proceso de enseñanza aprendizaje, frente a otro 30% de alumnos para quienes su uso no enriquecía. Igualmente, se realizó un análisis de contingencia entre si su uso enriquece el proceso de enseñanza aprendizaje de los estudiantes y el género de éstos mediante la test de al Chi-cuadrado resultando un valor de 0,002, por lo que se concluye que el uso de las plataformas enriquece el proceso de enseñanza aprendizaje de los estudiantes no está asociado con el género con un p-valor de 0,965.



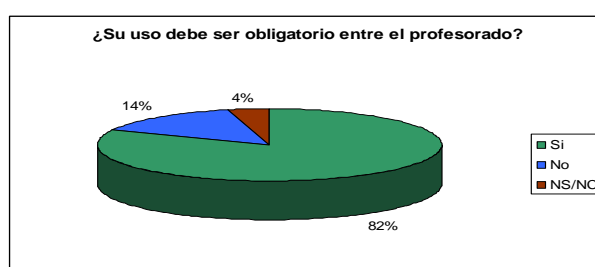
**Figura 2 – Opinión de los alumnos**





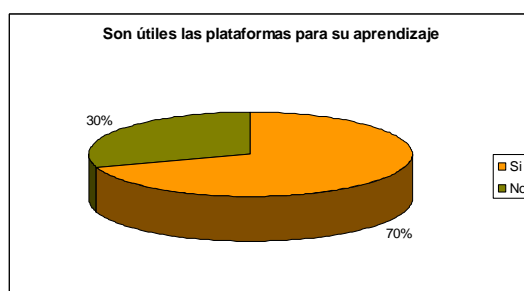
**Figura 3 – Opinión sobre el profesorado**

Con relación a si el uso de las plataformas debería ser o no obligatorio entre el profesorado (Figura 4), el 82% de los alumnos participantes afirmó que sí, frente a un 14% de alumnos quienes opinaban que no. Sólo un 4% de alumnos respondió como no sabe/no contesta.



**Figura 4 – ¿Su uso debe de ser obligatorio por el profesorado?**

Relativo a si son útiles o no las plataformas para su aprendizaje (Figura 5), el 70% de los alumnos manifestó que sí. Sólo un 30% de alumnos afirmó que su uso no había sido útil para su aprendizaje. De nuevo, se realizó un análisis de contingencia para analizar en función del sexo la utilidad de las plataformas para su aprendizaje mediante la test de al Chi-cuadrado resultando un valor de 0,002, por lo que se puede concluir que el la utilidad de éstas para el aprendizaje de los estudiantes es independiente del sexo con un pvalor de 0,965.



**Figura 5 – ¿Son útiles las plataformas para su aprendizaje?**

## 5. CONCLUSIONES

Los resultados obtenidos en este estudio muestran una alternativa que combina la metodología tradicional con actividades a realizar de forma virtual a través de una plataforma web. En ocasiones superar ciertas confusiones conceptuales necesitan de la intervención del profesor e incluso del grupo de alumnos ya que son difíciles de reconducir en un entorno puramente virtual. Ahora bien, sin perder dicha interacción presencial, estas metodologías ofrecen la mayoría de las ventajas de los cursos virtuales, como pueden ser la flexibilidad, el compromiso activo en las actividades de aprendizaje, un mayor número de oportunidades para practicar, etc. Nuestra experiencia se encuentra en esta línea:

---

clases presenciales combinadas con material y actividades virtuales junto con el apoyo de tutorías personalizadas.

La plataforma virtual proporciona un entorno en el cual los alumnos pueden aprender de forma individualizada y adaptada a sus necesidades horarias. Si bien, este proceso para completarse necesita ser evaluado. El hecho que los errores no sean penalizados y que el alumno sepa de forma inmediata si su respuesta ha sido correcta o errónea conlleva que el alumno considere los errores como una oportunidad de aprendizaje más que como un fracaso. A su vez, cada respuesta correcta refuerza el aprendizaje y permite al alumno valorar su progreso.

Las actividades a realizar por el alumno, deben estar correctamente diseñadas, en un entorno virtual que informe sobre cómo y cuándo los estudiantes realizan las actividades. Estos informes son una valiosa herramienta para identificar aquellos conceptos que suponen mayor dificultad a los alumnos y de esta forma nos ofrece la posibilidad de adaptar el ritmo del curso a los progresos de los estudiantes, haciendo especial hincapié en aquellos conceptos detectados como los más complejos. Así pues la principal función del profesor va a ser la motivación de los alumnos y las explicaciones conceptuales, dejando para la plataforma virtual la práctica necesaria para consolidar el conocimiento, así como el aprendizaje de algunos detalles de carácter fundamentalmente instrumental.

Los resultados obtenidos permiten concluir como para los alumnos, las nuevas tecnologías informáticas, ocupan actualmente, un lugar importante en el proceso de enseñanza-aprendizaje ya que facilitan el aprendizaje y la comunicación entre profesores y alumnos. Si bien, en aquellos casos en los que su uso no haya sido útil para el aprendizaje, probablemente estemos ante un problema de déficit de conocimientos en su manejo por parte del profesorado. En este sentido, cabría considerar la necesidad de formar adecuadamente al profesorado en el manejo óptimo de sendas plataformas.

Un uso adecuado de estas plataformas potencia y enriquece, sin duda, la docencia respecto del sistema tradicional, siendo esta opinión independiente del género de los estudiantes. La versatilidad y eficacia que brindan sendas herramientas promueve un aprendizaje integrado y carente de barreras, en donde el flujo de información e intercambio de actividades y tareas no cesa entre profesores y alumnos [14]. Si bien, el deficiente uso de ellas entre los colectivos de alumnos y profesores encontrado en este estudio muestra las dificultades con que este tipo de instrumentos académicos cuentan en la actualidad.

El uso de plataformas virtuales ha modificado el proceso tradicional de enseñanza-aprendizaje. En este sentido, resulta muy necesario conocer la opinión de todos los agentes involucrados (profesores y alumnos) para mejorar el proceso de enseñanza-aprendizaje.

## REFERENCIAS

1. AIMC. NAVEGANTES EN LA RED (2009). 11ª encuesta AIMC a usuarios de Internet. <http://www.aimc.es/03internet/macro2008.pdf> (Consulta: 08/02/2014).
2. Amorós, L. (2009). Weblogs para la enseñanza-aprendizaje. Pixel-Bit. Revista de Medios y Educación, 35, 61-71.
3. ANECA - Centro de Estudios de Gestión de la Educación Superior de la Universitat Politècnica de València (CEGES). (2007). El profesional flexible en la Sociedad de la Información.
4. Boyd, D. y Ellison, N. (2007). Social Network sites: Definition, History and Scholarship. *Journal of Computer-Mediated Communication*, 13 (1).
5. Cabero, J., López, E. y Ballesteros, C. (2009). Experiencias universitarias innovadoras con blogs para la mejora de la praxis educativa en el contexto europeo. *Revista de Universidad y Sociedad del Conocimiento*, 6 (2), 1-11.
6. Carretero, M. y Hermosilla, J.M. (2004). Gestión y Generación de conocimiento mediante el uso de plataformas digitales de formación. *Revista Complutense de Educación*, 15 (1), 139-164. [http://www.aneca.es/estudios/docs/InformejecutivoANECA\\_jornadasREFLEXV20.pdf](http://www.aneca.es/estudios/docs/InformejecutivoANECA_jornadasREFLEXV20.pdf) (Consulta: 09/01/2012) <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html> (Consulta: 5/02/2014).
7. CENTRO DE SERVICIOS DE INFORMÁTICA Y REDES DE COMUNICACIONES DE LA UNIVERSIDAD DE GRANADA (CSIRC). <http://www.ugr.es/informatica/> (Consulta: 12/02/2014).

- 
8. CENTRO DE ENSEÑANZAS VIRTUALES DE LA UNIVERSIDAD DE GRANADA (CEVUG) <http://cevug.ugr.es/> (Consulta: 10/02/2014).
  9. De Benito, B. y Salinas, J. (2008). Los entornos tecnológicos en la universidad. En Pixel-Bit. Revista de Medios y Educación, 32, 83-101.
  10. Domínguez, G. y Llorente, M.C. (2009). La Educación Social y la Web 2.0: nuevos espacios de innovación e interacción social en el Espacio Europeo de Educación Superior. Pixel-Bit. Revista de Medios y Educación, 35, 105-114.
  11. Guerrero, F. (1997). Tecnología de la información y comunicación en el proceso de enseñanza aprendizaje. <http://www.monografias.com/trabajos12/ltecdein/ltecdein.shtml> (Consulta: 7/02/2014).
  12. Guri-Rosenblit, S. (2005). Eight Paradoxes in the implementation process of E-Learning in higher education. Higher Education Policy, 18 (1), 5-29.
  13. Prendes, S. (2008). Portafolio electrónico: posibilidades para los docentes. Pixel-Bit. Revista de Medios y Educación, 32, 21-34.
  14. Marquès Graells, P. (2001). Algunas notas sobre el impacto de las TIC en la universidad. Educar, 28, 83-98.

# INCLUSIÓN FINANCIERA A TRAVÉS DEL CRÉDITO DE DESARROLLO HUMANO COMO HERRAMIENTA PARA EL FOMENTO DE CAPACIDADES Y EMPRENDIMIENTO

Eulalia-Elizabeth Salas\*, Viviana del Cisne Espinoza\*, Aurora Samaniego\*

&

Valentín-Alejandro Martínez-Fernández\* & \*\*

Dpto. Ciencias Empresariales., Universidad Técnica Particular de Loja. Ecuador.

Dpto. Ciencias Empresariales, Universidad de A Coruña, España.

## ABSTRACT

In recent years, the Ecuadorian economy has taken a new approach in its form of organization and with this has given rise to new trends that favor the most vulnerable sector, the popular and solidarity, whose current dimension displays a relevant social significance. This organizational initiative for social development is to reduce the poverty rate, with this, resolve the needs of a sector which for years has remained on the margins of the economy.

In this chapter we make an analysis of the impact caused in the financial sector of the country and the impact of the credit of Human Development Associative Credit (HDAC), as a key tool for social development, to be a financial service that offers the national government to the people who perceive the Bond of Human Development (BDH).

In the course of this study was carried out a documentary review on the incentives and subsidies that are given in other countries of Latin America, so as to be able to relate them to the government's policy of Ecuador.

The methodology used was diagnostic and verification of endeavors and was applied to a pilot sample of 78 individuals who have received the HDAC in the city of Loja; for this purpose, he carried out an entire process which consists of different phases that included the following tasks: a survey implementation; obtaining information of variables of general data such as: gender, marital status, age, educational level; and, achievement of information related to entrepreneurship, level of satisfaction, monthly income, technical support, and other related information.

**KEY WORDS:** Ecuador's financial system, Popular and Solidarity Economy, Financial Sector and People's Solidarity, Human Development Bond (BDH), Human Development Associative Credit (HDAC), entrepreneurship activities.

## RESUMEN

En los últimos años, la economía Ecuatoriana ha adoptado un nuevo enfoque en su forma de organización y con ello ha dado lugar a nuevas tendencias que favorecen al sector más vulnerable, el popular y solidario, cuya dimensión actual muestra una relevante significación social. Esta iniciativa de organización para el desarrollo social tiene por objetivo disminuir los índices de pobreza para, con ello, resolver las necesidades de un sector que por años ha permanecido al margen de la economía. En este trabajo se hace un análisis de las incidencias provocadas en el sector financiero del país y el impacto del Crédito de Desarrollo Humano Asociativo (CDHA), como herramienta clave de desarrollo social, al ser un servicio financiero que ofrece el gobierno nacional a las personas que perciben el Bono de Desarrollo Humano (BDH).

En la realización de este estudio se efectuó una revisión documental sobre los incentivos y subsidios que se dan en otros países de Latinoamérica, para así poder relacionarlos con la política gubernamental de Ecuador. La metodología utilizada fue de diagnóstico y verificación de emprendimientos y se aplicó a una muestra experimental de 78 personas que han recibido el CDHA en la ciudad de Loja; con tal finalidad se llevó a cabo todo un proceso consistente en diferentes fases que comprendieron las siguientes tareas: implementación de una encuesta; obtención de información de variables de datos generales como: género, estado civil, edad, nivel educativo; y, consecución de información relacionada con emprendimiento, nivel de satisfacción, ingresos mensuales, apoyo técnico, entre otras.

## 1. DESARROLLO SOCIAL EN AMERICA LATINA

El desarrollo social es la piedra angular que todo país necesita para su crecimiento, prosperidad y bienestar social. Todo ello depende de los modelos de Estado que existen en cada país, si bien varios de estos modelos han presentado incapacidad histórica, pues en su acción nuclear no promueven espacios de participación popular; a lo cual cabe añadir que algunos de estos se aplicaron en un marco democrático liberal, orientándose a eliminar formulaciones de desarrollo capitalista y la integración con los mercados externos, pero este hecho se mostró altamente ineficiente al llevar a que algunos sistemas democráticos excluyan de la vida política a las masas populares, privándolas de canales de participación social o de presión política (Max-Neef, Elizalde, & Hopenhayn, 2010).

---

A lo largo de los años, la economía mundial experimenta cíclicamente profundas crisis en el sistema capitalista. En el caso de la eclosionada en el 2008, si bien se ha centrado principalmente en Europa, en América Latina se ha sobrellevado con resultados favorables en el desarrollo social, evidenciándose una lucha paulatina contra la reducción de la pobreza.

Sin embargo, conforme a datos de la CEPAL, en el año 2014 América Latina presentó un estancamiento en sus procesos de reducción de la pobreza, ya que cerca del 28% de la población de la región se ubicó en los umbrales de pobreza, a lo cual ha de añadirse que la indigencia aumento del 11,3% al 12%, como consecuencia de la desaceleración económica que viven los países de la región.

A pesar de este panorama, cinco países destacan por presentar mejores resultados en reducción de la pobreza entre 2012 y 2013: Paraguay (49,6 a 40,7), El Salvador (45,3 a 40,9), Colombia (32,9 a 30,7), Perú (25,8 a 23,9) y Chile (10,9 a 7,8). En el caso de Ecuador la tasa de pobreza en el 2014 se situó en 22,5, con una disminución de 3 puntos respecto del año anterior, de acuerdo a los datos del Banco Central del Ecuador.

Dado que la pobreza es un problema social que afecta a todos los países, los gobiernos de América Latina y el Caribe, a partir de la década de 2000, realizaron importantes esfuerzos para incrementar su inversión social e introducir innovaciones en las políticas de protección, con el fin de quebrar así el esquema de política social residual. De acuerdo con Cechini et.al (2015), estas políticas se caracterizaron principalmente por abordar aspectos relacionados con la mejora de los sistemas de pensiones, ampliar el acceso a la salud y la formalización de los trabajadores.

Frente a estas reformas, es evidente que los países de América Latina han dejado atrás sus políticas de ajuste estructural, y se han centrado en crear e impulsar modelos de desarrollo sostenibles que permitan la generación de capacidad local y el fortalecimiento de los territorios, con una fuerte concepción social, donde los derechos de la población toman fuerza y son reconocidos como fundamento de la política pública.

En este sentido, la protección social tal y como señalan Cechini et.al (2005), se constituye en un instrumento esencial para acelerar el progreso, y como pieza clave de la política social, al considerar en éste la acción no solo del Estado sino también las estructuras operativas de los mercados, las familias, comunidades; es decir, se genera una visión holística donde todos los agentes de un territorio se erigen en piezas necesarias para conseguir el ansiado y esperado desarrollo. A este respecto, Hopenhayn (2007) estima que el desarrollo orientado por los derechos humanos supone que la articulación entre Estado y Sociedad se inspira en un contrato ético de ciudadanía.

En algunos países se pueden identificar varios ejemplos de políticas de protección social que, en sí mismos, responden a las necesidades de la población a lo largo de todo el ciclo de vida y en los cuales se incluyen programas de apoyo para niños, adolescentes y jóvenes, personas en etapa activa y reproductiva, así como la asistencia a personas en estado de vejez. Estas asistencias contemplan, entre otros, los siguientes aspectos: mayor acceso a la educación, salud infantil, seguridad alimentaria, educación sexual, regulación y protección laboral, seguridad social, políticas de empleo, jubilaciones y pensiones, políticas de acceso a vivienda, etc.

Si se toman como referencia los programas de transferencias condicionadas enfocadas a la infancia y niñez, se observa que, del 2000 al 2012, el porcentaje de apoyo se incrementa de un 5,7% al 21,1%. Sin embargo, es necesario reflexionar sobre las condicionalidades que dichos programas de ayuda implican, bajo el precepto de que los derechos humanos no dependen del cumplimiento de ninguna condición y son inherentes a la persona (Sepúlveda, 2014) y al ejercicio de la ciudadanía.

Por tanto, en aras a la coherencia y en concordancia con lo que plantean Cecchini et.al (2015) “se debe poner atención en la fijación de condiciones demasiado estrictas, ya que pueden generar una distinción poco afortunada entre los merecedores y no merecedores de la asistencia, lo cual choca con el principio de universalidad y violaría derechos humanos básicos con relación al aseguramiento de un nivel mínimo de vida”.

---

Al respecto, Sepúlveda (2014), plantea seis elementos que deben contener las políticas públicas y los programas de protección social basados en el enfoque de derechos; estos son: a) igualdad y no discriminación; b) integralidad; c) institucionalidad; d) participación; e) transparencia y acceso a la información; y, f) rendición de cuentas.

Se puede evidenciar que existen casos interesantes de apoyos condicionados. Brasil, por ejemplo, cuenta con el programa “Bolsa Familia”, cuya finalidad consiste en reforzar los derechos en educación y salud, así como fomentar el trabajo intersectorial y sin que ello implique medidas de castigo o punitivas.

Echecopar, Angelelli, Galleguillos, & Schorr (2006), señalan, en un estudio sobre la cuestión, toda una serie de iniciativas de apoyo a la creación de empresas con alto potencial de crecimiento y a microempresas de algunos países latinoamericanos, como es el caso de Chile con el apoyo a la generación de emprendimientos, vigente desde 1989, en donde se crea la normativa reguladora de la creación de fondos de inversión para el desarrollo empresarial y que, desde el año 2001, ha tomado un gran impulso a través de agentes del gobierno como son Corporación de Fomento de la Producción (CORFO). Entre los instrumentos con los que se cuenta, están el subsidio para la realización del Plan de Negocio y constituir la puesta en marcha de la empresa. El Comité INNOVA CHILE, mantiene también programas de financiamiento no reembolsables para el apoyo a las empresas, desde la incubación hasta la puesta en marcha e inicio de las actividades propiamente empresariales (Jiménez, 2008).

Para CL Group (2009), Chile cuenta con la presencia de un conjunto de instituciones que se han abierto a la atención micro-financiera hacia el sector micro-empresarial: la banca, las cooperativas de ahorro y crédito y diversas ONG, apoyadas por los diversos programas de las instituciones públicas.

Desde el 2001, en Argentina se da apoyo a la creación de empresas con alto potencial de crecimiento. La Subsecretaría de Pequeña y Mediana Empresa y Desarrollo Regional SEPYPME cuenta con un Fondo Nacional para el Desarrollo Pyme que tiene como objetivo otorgar financiación a empresas que ya están en funcionamiento.

En el caso de Colombia, las iniciativas públicas están orientadas al fortalecimiento de las microempresas. La principal iniciativa es el “Fondo Emprender” cuyo objetivo consiste en apoyar a personas con pocos conocimientos en el ámbito empresarial. Otra iniciativa es el “Banco de Oportunidades”, gestionado por la alcaldía de Medellín, el cual apoya con créditos dirigidos a la generación de microempresas y realiza concursos para otorgar capital semilla.

En El Salvador la iniciativa para el desarrollo de la microempresa está auspiciada por el “Programa Nacional de Emprendedores”, que lo implementa la Comisión Nacional de la Micro, Pequeña y Mediana Empresa, organismo público que tiene como finalidad la entrega de capital semilla no reembolsable a las pymes.

Si nos situamos en Ecuador cabe indicar que existen diferentes iniciativas públicas para orientar el desarrollo social. Así la Secretaría Nacional del Migrante Ecuatoriano SENAMI, creó el “Fondo Concursable Cucayo”, con el fin de apoyar a los migrantes ecuatorianos que deseaban retornar al país y emprender un negocio o mejorar uno ya existente. Otra iniciativa es la que mantiene el Ministerio de Cultura y Patrimonio, a través del Programa de Fondos Concursales “Emprende Patrimonio”, destinada a los gestores culturales ecuatorianos que estén dentro y fuera del país. Otra de las acciones con las que trabaja en gobierno ecuatoriano es la de desarrollo integral para la sociedad, competencia del Ministerio de Inclusión Económica y Social (MIES, 2014), reflejada en programas como: “Creciendo con nuestros hijos” y “Centro Infantil Buen Vivir (CIBV)”; programas que benefician a niños de 0-3 años. Otro programa de gran interés para el desarrollo social es “MIESpacios Juveniles”, dirigido a adolescentes de 12 a 18 años que pertenecen a la zona rural y urbana marginal. Igualmente se trabaja en programas de desarrollo integral enfocados a personas mayores de 65 años que requieren atención y cuidado integral; así como también en programas de atención a personas con discapacidad. Igualmente, entre las iniciativas públicas del país para orientar el desarrollo humano y social, es susceptible de destacar la Protección Especial que se presenta en la ejecución de programas como: “Acogimiento Familiar de niños y adolescentes en situaciones de vulneración de derechos o en situación de abandono”, “Programas de

---

Unidad de Apoyo Familiar”, “Erradicación de la Mendicidad, atención y protección a personas con discapacidad, a adultos mayores”, “Programas de Gestión de Riesgos para poblaciones damnificadas, afectadas por eventos adversos”, “Programas de Erradicación del Trabajo Infantil, Adopciones, Centros de Acogimiento”.

Dentro de los programas de aseguramiento no contributivo se trabaja con el Bono de Emergencia, destinado a personas y familias víctimas de desastres naturales o de conmoción social, con Redes de Descuento para las personas habilitadas al pago del bono de desarrollo humano, a los pensionistas y a las personas con discapacidad, las Pensiones para Adultos Mayores de 65 años que no tienen ningún aseguramiento.

Igualmente el Estado proporciona el servicio de pago de pensiones a personas con discapacidad igual o mayor al 40% que no estén aseguradas ni se beneficien de ningún otro bono.

Entre los programas de desarrollo social también se encuentra la entrega del Bono de Desarrollo Humano a núcleos de familias que están en condiciones de vulnerabilidad y extrema pobreza. Igualmente cabe subrayar que otro aspecto importante desarrollado en Ecuador como política pública es la Promoción y Movilidad Social instrumentado a través del Crédito de Desarrollo Humano, tanto el individual como el asociativo y destinado a usuarios del bono de desarrollo humano, pensionistas adultos mayores y personas con discapacidad (MIES, 2014).

## **2. ECONOMIA Y DESARROLLO SOCIAL EN ECUADOR**

El año 2007 se presenta en Ecuador como uno de los de mayor trascendencia en el ámbito económico, con lineamientos socialistas que abren espacios al desarrollo y a la acción social, al dejar de lado todas esas políticas económicas estrictas que se fijaron por años en el país y que enraizaron en la sociedad una economía capitalista materialista, de corte especulativo y no productivo; la misma que, a partir de los últimos años, está tomando un nuevo enfoque basado en el Sumak Kawsay, en donde se propone al ser humano y a la naturaleza como el eje central de la economía. Así, el objetivo principal es el Buen Vivir, que asegure de alguna manera la subsistencia de las futuras generaciones y fomente ante todo un entorno de armonía, en donde exista el principio de solidaridad y por ende no se base en la acumulación de capital para ciertos segmentos privilegiados, sino en el bien común que implica la igualdad y equidad para todos.

Para Ecuador, el nuevo modelo económico da gran prioridad a los procesos productivos del sector popular y solidario, determina el potencial de las micro-finanzas en el empoderamiento de las personas emprendedoras de este sector.

Si bien es cierto que en América Latina se han producido sustanciales cambios, en Ecuador en los últimos años se han experimentado modificaciones significativas en el desarrollo económico y social. A este respecto, ha de indicarse que la Constitución de la República de Ecuador, en su artículo 283, establece que el sistema económico es “social y solidario” e integrado por las siguientes formas de organización pública, privada, mixta y popular y solidaria.

Bajo esta premisa el sistema financiero ecuatoriano ha modificado de manera importante su organización, estructura y funcionamiento. Han sido muy diversos los cambios legislativos en este tema. Es así que, de acuerdo al artículo 309 de la nueva Constitución, el sistema financiero se compone de los siguientes sectores: público, privado y popular y solidario.

Si bien, no deja de ser cierto que el sistema financiero es uno de los factores más importantes dentro de la economía de un país, al canalizar el ahorro hacia la inversión, convirtiéndose en el eje central de la generación de la riqueza, los entes económicos se vienen apoyando en las diversas instituciones financieras para la obtención del capital de trabajo, para aumentar las actividades de producción, inversión, financiamiento, emprendimiento, etc. A este respecto, las entidades financieras se encargan de distribuir los fondos que reciben de los ahorradores y colocarlos en forma de créditos para los entes deficitarios que requieren de este capital al objeto de emprender nuevos retos empresariales,

---

negocios, etc., con lo que se mantiene de esta forma una economía saludable; pues bien, ante ello el objeto del sistema financiero consiste en canalizar el ahorro de prestamistas a prestatarios para realizar actividades económicas que conlleven al desarrollo económico y social del país. En Ecuador este proceso se realiza a través de las entidades financieras públicas, privadas y las del sector popular y solidario.

El sector financiero público está compuesto por bancos y corporaciones. El privado por los bancos múltiples y especializados, entidades de servicios financieros como los almacenes generales de depósitos, casas de cambio, corporaciones de desarrollo de mercado secundarios de depósitos, entidades de servicios auxiliares del sistema financiero. Por su parte, el sistema financiero popular y solidario está conformado por: cooperativas de ahorro y crédito, cajas centrales, entidades asociativas o solidarias, bancos comunales, cajas de ahorro, y las entidades de servicios auxiliares calificadas como tales por la Superintendencias de Economía Popular y Solidaria en el ámbito de su competencia. También son parte del sector financiero popular y solidario las asociaciones mutualistas de ahorro y crédito para la vivienda. (Código Orgánico Monetario y Financiero, 2014).

La economía popular y solidaria pretende romper grandes mitos que siempre trascendieron en la mente capitalista neoliberal, en donde se incluyen la aplicación de políticas monetaristas, orientadas a reducir la inflación y a mantener el balance fiscal; objetivos que en muchas ocasiones se lograban con la supresión de las regulaciones en el mercado laboral y con recortes en el sistema de bienestar y desarrollo social. Para Zurita (2005) quebrar este paradigma significa orientar la economía hacia el ámbito social, y situar la esperanza en que la sociedad a nivel mundial pueda crear un ambiente sostenible y adecuado para todos.

A criterio de Max-Neef, Elizalde, & Hopenhayn (2010) las dos grandes concepciones que han dominado el panorama de América Latina son el desarrollismo y el monetarismo neoliberal. Para dichos autores, ambas concepciones no han logrado cumplir su propuesta, Ante ello mencionan que no todo es negativo en un fracaso, los países que adoptaron estas concepciones y han fallado deben de reflexionar al objeto de mitigar los errores encontrados y superarlos mediante la adopción de las medidas necesarias para poder cumplir el fin al que todos deseamos apuntar: al bienestar y al desarrollo social que se puede lograr con la superación de los fracasos de la corriente neo-liberalista monetarista, los mismos que se pudieron haber producido al menos por las tres razones que citan: la primera es que el neo-liberalismo monetarista, a pesar de poder impulsar el crecimiento económico, no es generador de desarrollo; la segunda, porque es una concepción profundamente mecanicista e inadaptable, en donde las personas pobres y vulnerables que se encuentran en situación de miseria no pueden liberarse de un mercado al cual ni siquiera pertenecen; un mercado del que, de hecho, se encuentran totalmente marginados; la tercera razón es porque en mercados restringidos y oligopólicos, donde los grupos de poder económico no se enfrentan a fuerzas capaces de limitar su comportamiento, la actividad económica se orienta con sentido especulativo, y todo ello provoca resultados concentradores y mecanicistas.

En línea con la perspectiva que se acaba de exponer, actualmente en Ecuador se trata de incluir facilidades para que el grupo desprotegido, rural, pobre, etc., tenga acceso a servicios que los integren de manera solidaria en el entorno económico del país; de modo que se han puesto en marcha proyectos de iniciativa para la producción, en donde los pequeños, medianos productores, comerciantes, etc, puedan acceder a comercializar sus productos y a competir con grandes empresas que siempre los eclipsaron al no permitirles integrarse en el gran mundo competitivo. Por ello, el Estado promueve alternativas enfocadas a impulsar el desarrollo social equitativo, de forma que se incluyen programas como el de compras públicas inclusivas con la finalidad de hacer del Estado el demandante más importante de los productos de la economía popular, de modo que se ayude a impulsar el desarrollo económico y social de todos.

Una de las grandes críticas hacia la economía neoliberal es la falta de accesibilidad a productos y servicios financieros para los grupos más desprotegidos económicamente en el sector social, quienes no pueden acceder abiertamente a un crédito que les permita apuntar a su desarrollo personal y social.

Una alternativa de impulso económico consiste en facilitar productos y servicios a todos los segmentos económicos del país, especialmente a las pequeñas y medianas empresas en el campo del emprendimiento.



---

Sanhueza (2013) entiende que las pequeñas empresas ejercen un rol fundamental en la economía de un país, generando empleo, ingresos y un método de subsistencia para muchos hogares. No obstante, una serie de obstáculos complican la permanencia de las mismas, pues se enfrentan a grandes retos como el poder sostenerse en el tiempo, tener apertura a fuentes de financiamiento (De Asís, 2000; Foschiatto y Stumpo, 2006; García, 2004; Lacalle, 2008). Si bien ello puede no haber constituido siempre el principal escollo que han debido enfrentar; éste, sin duda, ha sido un gran problema ante la imposibilidad de ofrecer garantías que permitieran su acceso al mercado financiero formal, siendo sujeto de tasas de interés más altas (Herrera, 2003; Román, 2003).

Apuntar a una economía popular y solidaria, basada en los contornos del Plan Nacional del Buen Vivir, es una de las corrientes de reflexión más amplia que se ha presentado en América Latina en los últimos años, provocando con ello un cierto desarrollo, especialmente en el sector financiero popular y solidario. En Ecuador, desde hace varios años, existen diversas instituciones que aplican programas de micro finanzas para impulsar el desarrollo social y económico de diversos sectores; por su parte, las cooperativas de ahorro y crédito experimentan este tipo de economía y finanzas solidarias, impulsando la ejecución de proyectos de innovación, desarrollo, productividad todas ellas ancladas al objetivo principal de la economía popular y solidaria impulsar un desarrollo de vida justa y equitativa para todos.

Históricamente las finanzas populares y solidarias, según González, Prado & Miller (citado en Jácome & Cordovéz, 2003), promueven el desarrollo de microempresas, y de esta forma procuran contribuir a la generación de mayores ingresos, más numerosas fuentes de empleo y volver más dinámico el sector informal, para contribuir de esta forma al desarrollo económico de los países.

Las finanzas populares y solidarias se han convertido en un elemento catalizador de la lucha contra la pobreza; al generar oportunidades y libertades personales, estas se trasladan en el proceso de toma de decisiones; así el primer nivel es entendido como un indicador de autonomía que bajo el criterio de Nussbaum (2000), es la capacidad de decidir qué situación es mejor para la persona o comunidad y como segundo nivel está la libertad como el principal instrumento y objetivo del progreso, pues las libertades se desarrollan mutuamente y tienen efecto sinérgico sobre las oportunidades.

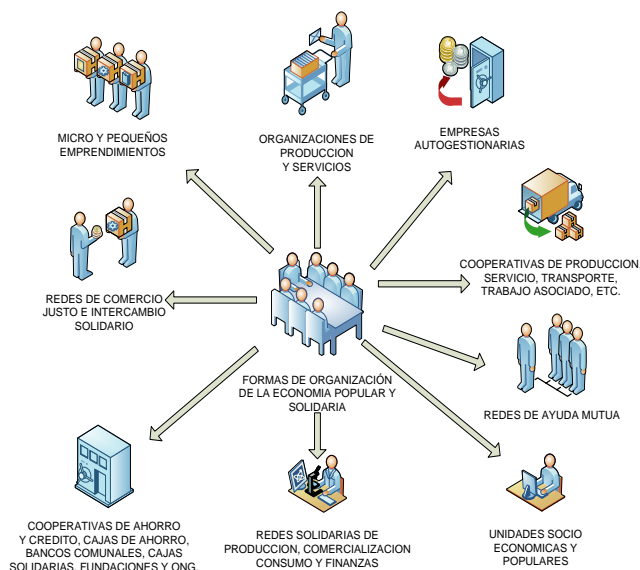
Para Sánchez (2013), la economía popular y solidaria tiene especial relevancia, toda vez que en este modelo de economía encontramos dos actores distintos, el popular y el solidario. Es la economía popular donde Coraggio (2011), citado por (Sánchez, 2013), sitúa lo que llama estrategia de supervivencia y movilidad del individuo, familias, unidades domésticas, donde se organizan y desarrollan procesos de producción, comercialización, financiamiento, etc., que les permita formar parte de una economía popular activa.

La misma autora menciona que la economía solidaria son unidades económicas organizadas que comparten principios: prevalencia del trabajo sobre el capital y de los intereses colectivos sobre los individuos, búsqueda del bien común, distribución equitativa y solidaria de excedentes. Ante ello (Dorkis, 2011) menciona que las finanzas solidarias buscan crear un modelo alternativo con entidades que generen confianza entre sus socios, enfocando su misión, visión y valores a atender las demandas de las personas más desfavorecidas, apoyando las iniciativas de cooperativismo y comercio justo, financiando actividades económicas que tengan un impacto social positivo y transformador.

Tonello & Hinojosa (2009), mencionan que las finanzas populares son consideradas las finanzas del pueblo. Ahora bien, en este escenario los actores principales son las personas con bajos recursos monetarios de manera que las finanzas populares son las finanzas del pueblo y para el pueblo.

No obstante, cabe señalar que para otros autores la economía popular y solidaria es una forma de organización económica donde los actores de este proceso se organizan y desarrollan procesos de producción, comercialización, financiamiento, intercambio y consumo de bienes y servicios que satisfacen de manera general las necesidades de la colectividad, basadas en relaciones de solidaridad, cooperación y reciprocidad al considerar de manera significativa las actividades económicas del ser

humano enfocadas en el Plan Nacional del Buen Vivir, las formas de organización de la economía popular y solidaria en Ecuador se presentan en la siguiente figura.



**Figura 1.** Formas de organización de la Economía Popular y Solidaria.

**Fuente:** Ley Orgánica de Economía Popular y Solidaria

**Elaboración:** Los Autores

En Ecuador el desarrollo social se refleja en la economía popular y solidaria que reconoce y da apertura a los grupos sociales vulnerables que hasta esos momentos estaban excluidos del sistema financiero tradicional, para con ello dar lugar a que formen parte de la nueva estructura financiera popular y solidaria. Para la nueva Constitución, el Sumak Kawsay implica mejorar la calidad de vida de la población, desarrollar sus capacidades y potencialidades; contar con un sistema económico que promueva la igualdad a través de la re-distribución social y territorial de los beneficios del desarrollo. (Art. 276).

En la economía social, el ser humano es el centro de la actividad económica y, por tanto, la economía debe de estar al servicio de la vida y no la vida en función de la economía. El objetivo de las finanzas populares se liga al desarrollo social desraizando la economía capitalista neoliberal y centra su prioridad en la equidad, la redistribución justa, la atención para todos y el combate a la pobreza.

Se trata entonces de apoyar las iniciativas económicas de la población desde la perspectiva del trabajo, con el fin de garantizar que la riqueza quede directamente en manos de los trabajadores y de las personas que quieren emprender y generar un motor incansable de desarrollo económico y social para el país.

### **3. EL CRÉDITO DESARROLLO HUMANO ASOCIATIVO EN LA ECONOMÍA SOCIAL DE ECUADOR.**

Para alcanzar el desarrollo humano y social hay que desafiar a que los políticos, los planificadores, los promotores, y sobre todo, los actores del desarrollo sean capaces de manejar el enfoque de las necesidades humanas (Max-Neef, Elizalde, & Hopenhayn, 2010) para orientar las acciones y aspiraciones al desarrollo y bienestar social.

Para Martínez (2008) las Instituciones Micro-Financieras (IMF) se han convertido en una de las herramientas favoritas para combatir la pobreza en los países en vías de desarrollo. Las microfinanzas se encargan de proveer servicios financieros, tales como depósitos, préstamos, transferencias de dinero y seguro a familias pobres (CGAP, 2008; Charitonenko y Campion, 2003; Rhyne, 2001; Ledgerwood 1999).

---

Martínez (2008) estima que los préstamos son utilizados para salud, educación, mejorar las condiciones de vivienda, para hacer frente a alguna emergencia familiar y en muchos de los casos para invertir en un micro- emprendimiento o en una microempresa, todo ello con la finalidad de incrementar el nivel de ingresos en las familias pobres. En esta idea abunda Sanhueza (2013) al puntualizar que las microempresas ejercen un rol fundamental dentro de la economía de un país y dentro de las localidades donde se encuentran, porque por lo general se convierten en pequeños motores generadores de empleo, ingresos y un método de subsistencia para muchas familias, constituyéndose como elementos dinamizadores de las localidades. Sin embargo, una serie de problemas atacan su permanencia, sostenibilidad y crecimiento en el tiempo, siendo uno de éstos el acceso a financiación (De Asís, 2000; Foschiato y Stumpo, 2006; García, 2004; Lacalle, 2008).

Como respuesta a esta situación, ha aparecido en las tres últimas décadas un grupo de servicios financieros caracterizados por una atención centrada en estas unidades económicas, implementando estrategias novedosas y originales. Es el caso de los programas micro-financieros, los cuales han surgido principalmente en economías en vías de desarrollo para atacar la pobreza, pero que a lo largo del tiempo se han ido expandiendo de tal forma, que en la actualidad atienden no solo a individuos pobres, sino que a micro-emprendedores que poco o nada tienen que ver con la pobreza, pero que en general se han visto excluidos del sistema financiero formal.

Entre sus servicios más conocidos de las micro-finanzas se encuentran los microcréditos, los cuales destacan por su alta demanda, oferta y difusión (Berezo, 2005; Lacalle, Rico, Márquez y Durán, 2006).

Por su parte, Gutiérrez (2005) considera que el microcrédito constituye una búsqueda de la extensión de los servicios de préstamo a las personas pobres que se han mantenido excluidos de los servicios financieros. La misma autora cita a Von Pischke (2002), quien afirma que el crédito moderno se inició a partir de las distintas organizaciones que ofrecían servicios de préstamo y, en distinto grado, de ahorro, a individuos en los márgenes de los mercados.

Berezo 2005, Lacalle, Rico, Márquez y Duran (2006), entienden el microcrédito como un instrumento financiero de apoyo a los sectores que quedan excluidos del ámbito financiero tradicional.

Ahora bien, el microcrédito es considerado una herramienta crucial para el crecimiento de economías domésticas que requieren del apoyo de una entidad financiera para superar los umbrales de pobreza, pero aún se registran casos en los que las familias en extrema pobreza no pueden acceder a estos servicios, muchas de las veces por falta de información o por el incumplimiento de requisitos que solicitan las entidades para colocar los microcréditos.

En el caso de Ecuador, debido a la crisis bancaria y a la eliminación de ciertos subsidios entre 1998 y 1999 como al gas, la electricidad y el combustible, lo que provocó que varias familias no puedan acceder al sistema financiero formal por la crisis que los abrumaba, varias familias, sobre todo del sector rural, se colocaron en situación de extrema pobreza. Por ello, el Estado se vio en la necesidad de crear un bono que compensara el impacto económico, fundamentalmente en los grupos sociales más desprotegidos y en situaciones críticas de pobreza y vulneración. Con el tiempo, este mecanismo de transferencia monetaria desencadenó muchas críticas que se enfocaban en creer que este subsidio se basa en el “padrinazgo” en una acción paternalista del Estado, llevando a un mayor estancamiento de actividades, desmotivando de manera clara el interés laboral de muchos ciudadanos, quienes en compañía de su grupo familiar conviven auspiciados por el bono de desarrollo humano. Bajo esta premisa, desde el 2007 el Instituto Nacional de Economía Popular y Solidaria (IEPS) implementa un nuevo programa de Gobierno con el cual los beneficiarios del BDH reciben el Crédito de Desarrollo Humano Asociativo (CDHA) con la finalidad de desarrollar, acrecentar y potenciar las actividades de producción, y micro-emprendimiento de los usuarios del Bono.(Coba Molina & Díaz Córdova, 2014).

En el estudio presentado por Coba & Córdova (2014), acerca del Crédito de Desarrollo Humano Asociativo en la Economía Social y Solidaria de la Provincia de Tungurahua – Ecuador, se presentan los siguientes resultados.

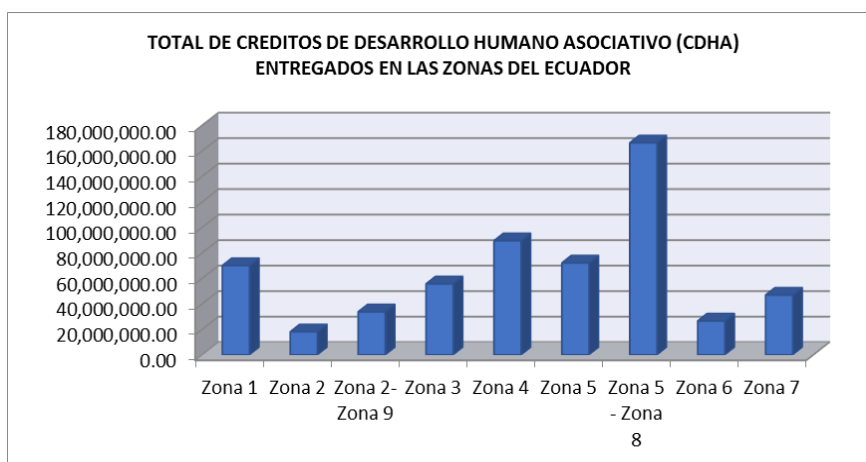
ZONAS	TOTAL ENTREGADO DESDE EL 2007 - 2012
Zona 1	70.101.216,00
Zona 2	18.115.791,00
Zona 2- Zona 9	33.784.587,00
Zona 3	55.728.785,00
Zona 4	89.587.053,00
Zona 5	72.158.232,00
Zona 5 - Zona 8	166.337.644,00
Zona 6	26.620.620,00
Zona 7	46.933.213,00
<b>TOTAL GENERAL</b>	<b>579.367.141,00</b>

**Tabla 1.** Montos de CDHA entregados por zonas de planificación 2007-2012

**Fuente:** IEPS (2013)

**Elaboración:** Los Autores

De los CDHA entregados desde el 2007 al 2012, como se aprecia en la siguiente figura la mayor colocación que se ha realizado es en la Zona 5 y Zona 8 integrada por las ciudades de Santa Elena, Bolívar, Guayas, Los Ríos y Galápagos, Guayaquil, Samborondón y Durán.



**Figura 2.** Monto de Créditos por zonas de Planificación 2007- 2013

**Fuente:** IEPS (2013) Análisis Histórico CDH 2007-2012

**Elaboración:** Autores

Como se puede apreciar en la figura, la inversión y apoyo social a las personas que solicitan el CDHA es considerable **579.367.141,00** dólares desde el 2007 al 2012 (Coba Molina & Díaz Córdova, 2014).

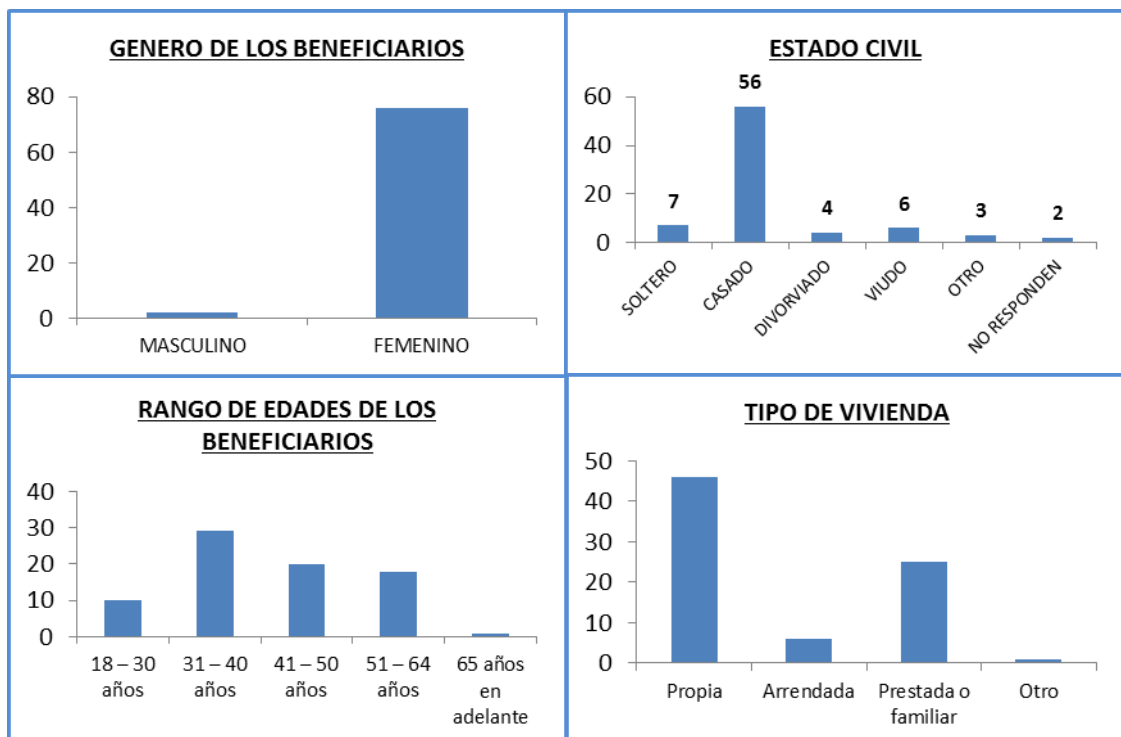


**Figura 3:** Barrios en los que se realizó la investigación.

**Fuente:** Aplicación de encuestas

**Elaborado:** Los autores.

Conceptualmente cabe precisar que el crédito de desarrollo humano es un servicio financiero que consiste en el pago adelantado del BDH, dependiendo de la naturaleza del crédito si el usuario solicita el individual se le anticipa el pago del bono de un año (600 dólares), recogiendo la solicitud y esperando el turno para la recepción del crédito. En el caso de solicitar el Crédito de Desarrollo Humano Asociativo a los usuarios se les anticipa (1200 dólares), además deben de participar de los talleres que imparte el MIES, con la finalidad de orientarlos a que modelen la idea de negocio en la que quieren emprender, para poder así cumplir con el objetivo principal de este crédito que se orienta a que los beneficiarios del CDHA puedan realizar actividades de micro-emprendimientos que les permita asociarse y desarrollar emprendimientos sustentables en el tiempo, para de este modo eliminar las barreras de pobreza, en las que por varios años han vivido atrapados.



**Figura 4:** Datos de los usuarios a los que se les realizó la investigación.

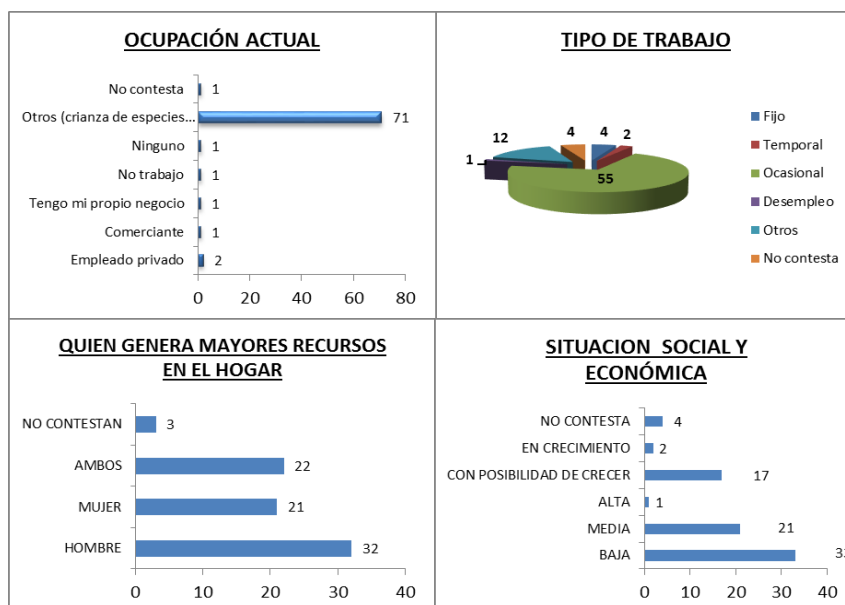
**Fuente:** Aplicación de encuestas

**Elaborado:** Los autores.

El Crédito de Desarrollo Humano es la primera puerta de acceso a los beneficiarios del Bono de Desarrollo Humano del sector popular y solidario a las micro-finanzas en el país, respondiendo con el desarrollo de sus micro-emprendimientos y con la generación de sus propios ingresos, antes que solamente recibir el subsidio asistencial de 50 dólares en la mayor parte de los casos para la alimentación y la supervivencia.

Los usuarios del bono que decidieron acceder a este beneficio de financiamiento promovido por el MIES, entregado a través del Banco de Fomento, forman parte del programa del Crédito de Desarrollo Humano, con el cual tienen acceso al pago adelantado de dos años del bono, a las capacitaciones y a las tres visitas técnicas que se les realiza durante la vigencia del crédito. (Ministerio de Inclusión Económica y Social, 2014).

Los CDHA tienen la finalidad de vincular a los beneficiarios del CDHA a ejes productivos agrícolas, ganaderos, porcinos, avícolas, producción de cuyes, comercio, turismo, artesanía, etc. Como puede apreciarse en la siguiente figura, varios barrios rurales de la ciudad de Loja también se han beneficiado de este servicio, de manera que la inversión y el apoyo social se extiende a todas las regiones de Ecuador.



**Figura 5:** Datos de los usuarios a los que se les realizó la investigación.

**Fuente:** Aplicación de encuestas

**Elaborado:** Los autores.

La presente investigación se desarrolló en la ciudad de Loja con 78 usuarios que recibieron el CDHA. Ha de indicarse que las familias encuestadas estaban ubicadas en lugares extremadamente alejados a la ciudad, dándose ocasiones en las que para encuestar y verificar la existencia de las actividades de emprendimiento se tuvo que caminar dos, tres o más horas, esto debido a que las viviendas son rurales y se encuentran bastante alejadas de la ciudad, en donde en ocasiones ni siquiera hay transporte. Con ello, se quiere subrayar el entorno de pobreza y exclusión.

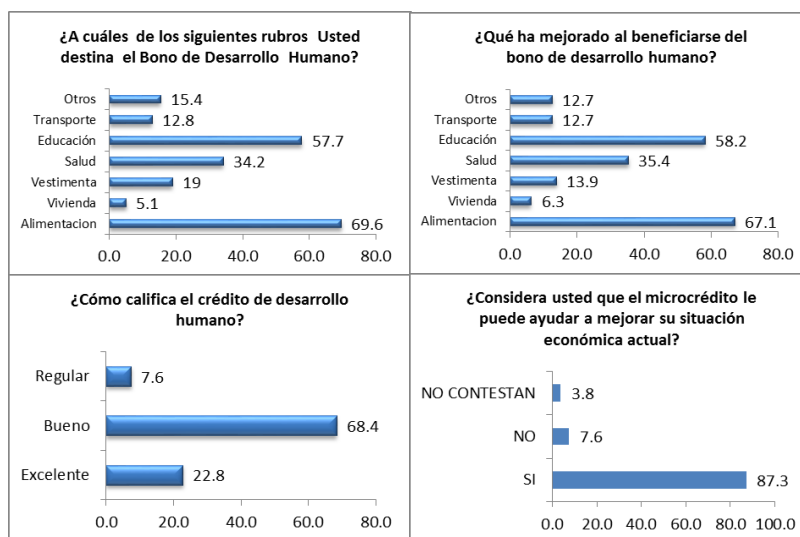
Con la recogida de la información facilitada por los beneficiarios, la encuesta permitió obtener datos como: género, edad, estado civil, tipo de vivienda, etc. Dichos datos se presentan en la figura 4.

Como se puede apreciar en la figura anterior la presencia de las mujeres en este programa es muy significativa. El 96,2% de las personas que han recibido el Crédito de Desarrollo Humano Asociativo (CDHA) en los barrios rurales de la ciudad de Loja analizados son mujeres que están entre los 31 y 40 años (58,2%); 56 de ellas son casadas, 7 madres solteras, 6 viudas y 4 divorciadas. Por otra parte, 3 de las mujeres encuestadas manifestaron que mantienen una relación de pareja de hecho y dos de ellas no respondieron la pregunta. El 58,2% de los encuestados manifestaron que vive en una casa humilde y pobre, pero propia porque los padres les proporcionaron una parcela de terreno en el que construyeron sus casas. En algunos casos (31,6%), la vivienda ha sido prestada o es de algún familiar. El 7,6% arrienda la vivienda.

#### 4. LOS MICROEMPRESARIOS COMO RESPUESTA A LA ASIGNACIÓN DEL CRÉDITO DE DESARROLLO HUMANO ASOCIATIVO (CDHA)

La finalidad del Crédito de Desarrollo Humano Asociativo es fortalecer las actividades productivas, de emprendimiento de comercialización, etc., con el objeto de procurar el desarrollo social, económico y personal de sus beneficiarios. Conseguir que se canalicen adecuadamente los CDHA al desarrollo de actividades de micro-empresarios es el objetivo primordial del Estado. Por ello se han adoptado las medidas necesarias y se actúa sobre los obstáculos que los micro-empresarios encuentran para el diseño, implementación, desarrollo y sostenimiento de su actividad. Es decir, los técnicos del MIES realizan seguimientos a estas actividades para verificar el crecimiento de los micro-empresarios y proporcionar apoyo, así como soporte técnico a las personas que lo necesitan.

Para conocer el punto de vista de los beneficiarios, en la investigación se realizó una indagación y verificación de los micro-empresarios puestos en marcha. En la figura, correspondiente se puede observar que el mayor porcentaje de personas beneficiarias del CDHA (89,9%) se dedica a la crianza de especies menores y a sembrados en el campo. Igualmente se aprecia la existencia de trabajo ocasional (69,6%) cuando llevan sus productos a vender a las plazas de mercado más cercanas; aunque las ventas no son periódicas se generan cuando tienen hortalizas, legumbres, animales de especies menores listos para ser comercializados, por regla general en mercados libres o de economía informal.



**Figura 6:** Destino de la aplicación del BDH,

**Fuente:** Aplicación de encuestas

**Elaborado:** Los autores.

En la generación de recursos de los hogares encuestados, el hombre es el que crea y aporta más recursos (40.5%). En algunas familias, el 26,6% de las mujeres es el que tiene protagonismo en la obtención de los recursos financieros necesarios para la subsistencia familiar. El 27,8% de los hogares encuestados

---

manifestó que los ingresos se generan en pareja; es decir, ambos trabajan para tener ingresos en el hogar. Solamente el 3,8% no respondió a la pregunta.

El 41,8% de los usuarios del bono considera que su situación económica es baja y tiene la esperanza de que con su trabajo y esfuerzo logrará salir de las condiciones de pobreza en la que actualmente vive. El 26,6% considera que su situación económica es media, porque gracias al CDHA genera la posibilidad de crecimiento económico en pequeños pasos; pero está en el camino de salir adelante. El 21,5% se ubica en una situación económica con posibilidades de crecer, de manera que confía y cree que su esfuerzo, así como la dedicación para mejorar, le permitirá tener una posibilidad de desarrollo social y económico. El 2,5% manifiesta que sus actividades de emprendimiento se han convertido en una herramienta de crecimiento y desarrollo familiar que los ubica en una economía en crecimiento.

Ha de resaltarse que en las visitas de campo realizadas se encontró a una mujer emprendedora Suquilanda (2015), quien manifestó que gracias a la ayuda del Crédito de Desarrollo Humano Asociativo, a sus ahorros y a las ganas de emprender, junto con su esposo, han alcanzado ya una situación económica alta. En concreto, ellos decidieron crear y poner en marcha una sastrería, toda vez que contaban con habilidades para la costura y adquirieron la pertinente formación. Con el CDHA compraron una aplanadora de telas, una máquina de coser y con los ahorros que habían generado pudieron implementar su pequeña sastrería y cuya marca es “Darío Vásquez”. Esta beneficiaria del Bono de Desarrollo Humano considera que, gracias a la ayuda que encontró en el CDHA, así como al apoyo de entidades del sector financiero popular y solidario que le permitieron a su esposo obtener un préstamo y a sus ahorros, ellos pueden decir que ya rompieron esa barrera de pobreza que por muchos años los mantuvo alejados del desarrollo y crecimiento económico, ubicándolos hoy en día en una situación económica estable y en crecimiento.

La generación de emprendimientos sostenibles en el tiempo, es el reto al que se enfrentan las políticas del Estado con los beneficiarios del CDHA, cuyo objetivo es mejorar los ingresos que mensualmente los beneficiarios obtenían con el bono.

De los 78 beneficiarios, el 69,6% destinaba el Bono de Desarrollo Humano a alimentación, educación (57,7%), salud (34,2%). Para un alto porcentaje (68,4%) de beneficiarios el CDHA es muy positivo, el 22,8% lo califica de excelente y el 7,6% valora el CDHA como regular. Para el 87,3% el microcrédito le puede ayudar a mejorar su situación económica actual.

El trabajo de campo llevado a cabo en la investigación, junto con el posterior análisis de los datos, ha permitido evidenciar el positivo impacto de los micro-emprendimientos en la creación de un tejido social abocado al crecimiento económico de economías excluidas. Dado que el objetivo de esta investigación experimental es diagnosticar y evaluar al CDHA de la ciudad de Loja- Ecuador, ha de concluirse que los micro-emprendimientos forman parte de un extracto productivo cuya dinámica debería estudiarse más a fondo para poder formular propuestas de políticas que generen beneficios de productividad y contribuyan al desarrollo de micro-emprendimientos sostenibles.

#### REFERENCIAS

1. Cecchini, S., Filgueira, F., Martínez, R., & Rossel, C. (2015). Instrumentos de protección social. Santiago: CEPAL.
2. Coba, E., & Díaz, J. (2014). El crédito de Desarrollo Humano Asociativo, en la economía social y solidaria de la provincia de Tungurahua - Ecuador. *Analitika, Revista de Analisis Estadístico*, 33-47.
3. Coraggio J (2011). Economía Social y Solidaria. El trabajo antes que el capital . Recuperado el 17 de agosto de 2015 de <http://www.coraggioeconomia.org/jlc/archivos%20para%20descargar/economiasocial.pdf>
4. Código Orgánico Monetario y Financiero (2014). Recuperado el 05 de agosto de 2015 de [http://www.ecuadorencifras.gob.ec/LOTAIP/2015/DIJU/enero/LA2\\_ENE\\_DIJU\\_CodOrgMonetario.pdf](http://www.ecuadorencifras.gob.ec/LOTAIP/2015/DIJU/enero/LA2_ENE_DIJU_CodOrgMonetario.pdf)



- 
5. Constitución de la República del Ecuador. (2008). <http://www.oas.org>. Recuperado el 01 de 07 de 2015, de <http://www.oas.org>: [http://www.oas.org/juridico/pdfs/mesicic4\\_ecu\\_const.pdf](http://www.oas.org/juridico/pdfs/mesicic4_ecu_const.pdf)
  6. Dorkis, S. (2011). Las monedas complementarias y la nueva realidad de las finanzas solidarias en países desarrollados y no desarrollados. *Revista Venezolana de Economía Social*, 24-30.
  7. Echeopar, G., Angelelli, P., Galleguillos, G., & Schorr, M. (2006). *Capital semilla para el financiamiento de las nuevas empresas: Avances y lecciones aprendidas en América Latina*. Washington.
  8. Foschiatto, P., & Stumpo, G. (2006). Políticas municipales de microcrédito: Un
  9. instrumento para la dinamización de los sistemas productivos locales. *Estudios de caso en América Latina*. Santiago: Cepal.
  10. Gutierrez, B. (2005). Antecedentes del microcrédito. Lecciones del pasado para las experiencias actuales. *Revista de Economía Pública, Social y Cooperativa*, 25-50.
  11. Jácome & Cordovéz (2003) *Microfinanzas en la economía ecuatoriana: Una alternativa para el desarrollo*. Recuperado el 17 de agosto de 2015 de <http://www.flacso.org.ec/docs/microec.pdf>
  12. Jiménez, L. (2008). *Capital de riesgo e innovación en América Latina*.
  13. Lacalle, M. (2008). *Microcréditos y pobreza*. Recuperado el 12 de agosto de 2014 de <http://www.redalyc.org/pdf/866/86601911.pdf>
  14. Martínez, A. (2008). El microcrédito como instrumento para el alivio de la pobreza: Ventajas y limitaciones. *Cuadernos de desarrollo rural*, 61.
  15. Max-Neef, M., Elizalde, A., & Hopenhayn, M. (marzo de 2010). *Desarrollo a escala humana. Opciones para el futuro*. Madrid, España: Biblioteca CF+S.
  16. Ministerio de Inclusión Económica y Social. (2014). Recuperado el 2 de Agosto de 2015, de <http://www.inclusion.gob.ec/>
  17. Nussbaum, M. (2000). *Women and Human Development the Capabilities Approach*. New York: Cambridge University Press
  18. Sánchez, J. (2013). *La necesaria integralidad de las políticas públicas para la economía social y solidaria*. Quito: Ministerio de Inclusión Económica y Social.
  19. Sanhueza, P. (2013). *Las micro-finanzas como instrumento de apoyo al sector micro-empresarial en Chile: Estado y Desafíos*. Recuperado el 14 de agosto de 2015, de <http://www.scielo.cl/pdf/jotmi/v8n2/art17.pdf>
  20. Suquilanda Johana (2015), entrevista personal a la beneficiaria del CDHA, 24 de julio de 2015.
  21. Tonello & Hinojosa (2009). *Finanzas Populares y Desarrollo Económico Local*. Módulo 4, Escuela de Formación Empresarial. "Mons. Cándido Rada Grupo Social FEPP".
  22. Zurita, J. (2005). *Globalización*. Recuperado el 2 de agosto de 2015, de <http://rcci.net/globalizacion/2005/fg574.htm>: <http://rcci.net/globalizacion/2005/fg574.htm>

## A RANKED SET SAMPLING ALTERNATIVE FOR SINGLE IMPUTATION IN A HOT DECK ENVIRONMENT

Sira Allende \*, Carlos Bouza\* and Iqbal Jeelani\*\*

\*Universidad de La Habana, Cuba

\*\* [Sher-e-Kashmir University of Agricultural Sciences and Technology, Agricultural Statistics, India](#)

### ABSTRACT

The objective of this paper is evaluating the convenience of using Ranked Set Sampling (RSS) instead of Simple Random Sampling with Replacement (SRAWR) for imputing item nonresponses, using simple imputation in a Hot-Deck methodology. A numerical study, for evaluating the accuracy of the alternative imputation procedures presented, is developed.

**KEYWORDS:** Ranked Set Sampling, Simple Random Sampling with Replacement, imputation, Hot-Deck .

### RESUMEN

Este trabajo tiene por objetivo evaluar la conveniencia de usar el muestro por rangos ordenados (Ranked Set Sampling) en ves del muestreo simple aleatorio con reemplazo para imputar no respuesta de ítem para una metodología de Hot-Deck. Se desarrolla un estudio numérico, para evaluar la precisión de los procedimientos de imputación propuestos.

## 1. INTRODUCTION

Most surveys have item the missing data become a cumbersome problem in the stage of dataset analysis. The researcher deals with item-level missing data and opts for one of the following strategies

- (i) Deleting complete cases whenever there are missing data for *any* variable being analyzed,
- (ii) Deleting cases but only as necessary for a particular family of estimates.
- (iii) Imputing (“fill in values for”) the missing data.

Methods (i) and (ii) are widely used but they are associated with serious violations and affects the analysis of the data. Method (i) results in a loss of cases, especially when many variables are being analyzed. It is well documented that the responses may not be representative of the value of the deleted cases, hence it generates a bias. Method (ii) has the same problems of method (i), tough in a lesser degree, but it may generate serious inconsistencies in the estimates. These inconsistencies will be misleading and result in a loss of confidence. Method (iii) allows using complete data methods and brings additional knowledge how was “solve” the missing data problem. Its drawbacks are the underestimation of the errors, as the amount of “real” data is smaller, and that the covariance estimates shrink toward zero. See a general discussion in Allison (2010), Kalton-Kish (1984).

- *Missing completely at random*—there are no systematic differences between the missing values and the observed values. For example, glycemi measurements may be missing because of breakdown of the glucometer.
- *Missing at random*—the systematic differences between the missing values and the observed values can be explained by the differences in the observed data. For example, missing glycemc measurements may be lower after a session of aerobics.
- *Missing not at random*—after the observed data are taken into account we observe systematic differences between the missing values and the observed ones. For example, old people with high blood pressure may be more likely to miss appointments with the physician for checking their glycemi.

If data are missing at random unbiased and statistically powerful analyses can be developed by including individuals with incomplete data. If that is the case is possible building some general model incorporating information on partially observed variables, for example by using some Bayesian methods

---

incorporating partially observed variables into a full statistical model. The analysis may use some weighting or using a maximum likelihood estimation.

If it is valid assuming the “Missing at random” model it is reasonable using a known predictive variable in the imputation model, but not if the variable is omitted from the model.

If the “Missing not at random” model is accepted it is not possible to account for systematic differences between the missing values and the observed values using the observed data. Then the bias in the analyses based on any imputation procedure may be very big, even bigger than using only the not-missing data.

Interested readers can consult Schafer (1997), Chen et al. (2000), Little- Rubin(2002), Baraldi-Enders (2010), Kadengye et al. (2012), Durrant (2015), for discussion on the potentialities, open problems and applications of interest.

Section 2 presents issues related to Hot-Deck methodologies and the third one of the use of single imputation. The objective of this paper is evaluating the convenience of using Ranked Set Sampling (RSS) instead of Simple Random Sampling with Replacement (SRAWR) for imputing item nonresponses using simple imputation in a Hot-Deck methodology. Section 4 is concerned with the use of RSS for imputation and section 5 presents a numerical study for evaluating the accuracy of the alternative imputation procedures presented.

## 2. HOT DECK METHODS

We seek methods for handling missing data that lead to statistically valid inferences, in the sense that satisfy

- (a) Approximately unbiased estimates of population
- (b) Interval estimates with at least their nominal
- (c) Tests of significance should reject at their close to the nominal level is highly desirable

Hot deck imputation replaces each missing value by the result of selecting at random a ‘donor pool’ consisting of observed values of that variable. The donor pools are determined by a set of units with complete information. The term comes from the ancient use of computer punch cards refers to the deck of cards for donors available for a non-respondent. The “hot” identifies that the cards were also in the physical sense hot, as the deck was being processed. They formed a “cold deck” if the cards of the donors were pre-processed data because they come a previous data set. A variation is to use categorical variables and use as donors units matching with the values of them in the missing units. See Little and Rubin (1987). Creating adjustment must take into account whether it is associated with the missing variable  $Y$ , and whether it is associated with the binary variable indicating whether  $Y$  is missing. Little & Vartivarian (2005) studied the effect of high or low levels of the associations on bias and variance of the estimated mean of  $Y$ .

There may be a gain in efficiency relative to using only the responses, since the information of the incomplete cases is being retained, and thus there is also a reduction in non-response bias, to the extent that there is an association between the variables defining imputation classes and both the propensity to respond and the variable to be imputed.

Hot deck imputation, when done properly, can also satisfy all three of the guidelines listed above for single imputation. This technique does not distort the distribution of the variable but it does not give a representation of uncertainty. The hot deck imputations does not use model fitting for imputing the variable values, hence they are potentially less sensitive to model misspecification than parametric model imputation. This imputation is invariant to transformations of the marginal distribution of  $Y$ . Hence, imputing  $Y$  yields the same imputations as imputing  $\log Y$ . Improvements may result from imputing a function of  $Y$  and  $x$ , rather than of  $Y$  directly. That is the case when  $R_{Y,S}$  is high as a result it seems that using the ratio  $R = Y/S$  produces that if for the donor  $r_j = \frac{y_j}{s_j}$  the imputation is  $\hat{y}_i = s_i r_j$  for  $y_i$ ,

Nowadays sequential hot deck methods are popular in some areas of applications. Recently complex matching metrics are in use.

The simplest hot deck procedure is using the entire sample of respondents as a single donor pool. The procedure generates consistent estimates only when the missingness is described as a MCAR, see Lanke (1983), Skinner, C. and Rao, J.N.K. (2002 and Little-Rubin (2002). In this case the hot deck estimate of the mean is equal to the respondent mean in expectation. If data are not MCAR, the imputation model approach (IM) and the non-response model approach (NM) allow establishing the consistence of the estimates, see Shao & Steel (1999) and Shao-Tang (2011). The conditions for accepting consistency depends on the used approaches.

The IM approach assumes a superpopulation model for the item to be imputed and inferences are based on the principle of repeated sampling. The response mechanism has not a role and only is considered that frame work of MAR. The response probability may depend on auxiliary variables only for creating the donor pools. Within adjustment cells leads to an unbiased estimator under a superpopulation model for the cell mean for nearest neighbor imputation, the estimates of sample means may be considered as asymptotically unbiased when a linear relationship between the item to be imputed and the auxiliary information is accepted.

Unconditional mean imputation, which replaces each is zero. Although (b) and (c) are accepted components of 'statistically valid frequentist. Unconditional mean imputation, which replaces each missing value with the mean of the observed values of that variable, generally meets none of the three guidelines listed above for imputation.

Nevertheless, a well-bussed theory has been not been developed as that of other imputation methods.

The lack of software in popular statistical packages deterred the applications of the hot deck in many applications. However, Hot-deck is commonly used by government statistics agencies and survey organizations to provide rectangular data sets for users. Examples of this fact are documented in Righi et al. (2014):

### 3. SINGLE IMPUTATION

Single imputation identifies the use of the imputation of one value for each missing datum. Singly-imputed data sets allows using standard complete data statistical methods. This is theoretically attractive. It is a commonly used and is available in data analysis software packages. It has the advantage of allowing the use of complete data analysis methods and is of easy implementation. A common single imputation methods is replacing missing values with the mean value of the responses. Little-Rubin (1987) pointed out that this method distorts the distribution of the variable and variance is underestimated.

There are a number of different approaches to imputation. One of them are the so called "Deductive *methods*". The imputation of a missing value is made using logical relations between variables and deriving a value for the missing item that is considered as "highly probable". The method of (*unconditional*) *mean imputation* imputes the mean of the respondents to each missing item within that variable. A variation of this method is to impute a class mean. Classes (imputation classes) may be defined based on some explanatory variables.

Denote by  $C(k)=C(k-r)+C(k-nr)$ ,  $k=1, \dots, q$ , an imputation class of size  $\|C(k)\| = n_k$ . It is plausible assuming that the set  $C(k-r)$  of the responding units for the item  $Y$  is a random subsample of all sampled units. Take  $\|C(k-r)\| = r_k \leq n_k$  the number of responding units and the set of the non-respondents  $C(k-nr)$ , such that  $\|C(k-nr)\| = m_k = n_k - r_k$ .  $C(k)$  is a post stratum and the best estimate of the mean of the item within  $C(k)$  is computed using:

$$\bar{Y}_{r_k} = \frac{1}{r_k} \sum_{j \in C(k-r)} Y_j$$

and the best estimate of the variance of the mean is

$$v(\bar{Y}_{r_k}) = \frac{1}{r_k(r_k - 1)} \sum_{j \in C(k-r)} (Y_j - \bar{Y}_{r_k})^2 = \frac{\sigma_k^2}{r_k}$$

If we impute the missing values by  $\bar{y}_{r_k}$  under limit conditions of  $r_k$  we have that  $\bar{Y}_{r_k} \cong \frac{1}{n} \sum_{j=1}^n Y_j$ . Hence

$$\frac{1}{n(n-1)} \sum_{j=1}^n (Y_j - \bar{Y}_{r_k})^2 \cong \frac{1}{n(n-1)} \sum_{j \in C(k-r)} (Y_j - \bar{Y}_{r_k})^2 \cong \frac{n_k(n_k-1)}{n(n-1)} v(\bar{Y}_{r_k})$$

If

$$\bar{Y}_{r_k} = \frac{1}{n_k} \sum_{j \in C(k-r)} Y_j \cong \frac{1}{m_k} \sum_{j \in C(k-nr)} Y_j = \bar{Y}_{m_k}$$

we have that

$$\frac{1}{n(n-1)} \sum_{j=1}^n (Y_j - \bar{Y}_{r_k})^2 \cong \frac{n-1}{n(n-1)(n_k-1)} \sum_{j \in C(k-r)} (Y_j - \bar{Y}_{r_k})^2 = \frac{n_k}{n} v(\bar{Y}_{r_k})$$

The estimator of the mean is

$$\bar{Y} = \frac{1}{n} \left( \sum_{k=1}^q \sum_{j \in C(k-r)} Y_j + m_k \bar{Y}_{r_k} \right)$$

and

$$V(\bar{Y}) = \frac{n_r \sigma^2}{n^2} + \left( \frac{n-n_r}{n} \right)^2 \sum_{k=1}^q \frac{\sigma_k^2}{n_k}, \quad n_r = \sum_{k=1}^q n_k$$

That is, the variance of the mean will be always underestimated even if  $\sigma_k^2 = \sigma^2$ . See a discussion in Cohen (2003):

The problem of underestimation of variances is one of the goals of the research for better imputation methods.

Classifying respondents and non-respondents into imputation classes is also called “adjustment cells, based on an auxiliary variable”, see Brick et al (2005). The imputation is made by selecting randomly a donor for each non-respondent within each cell. Groups of “similar” donors could be created empirically using classification algorithms. Optimal stratification by the propensity is another method for determining  $C(\vec{x}_i)$ . Sparseness of donors usually leads to over-using the same donor. Then is recommended to limit the number of times each donor is used for imputing. The optimal choice of this number  $d$  is a research theme for optimization.

A more general principle is to choose donor units considering their closeness to the non-respondent using some metric. Take the vector of auxiliary variables (covariates)  $\vec{x}_i = (x_{i1}, \dots, x_{iq})$ . They are used for determining the adjustment cell  $C(\vec{x}_i)$ . Matching the non-respondent  $i$  to donors in  $C(\vec{x}_i)$  is made once we fix a metric. Popular ones are

$$d(i, j) = \max_h |x_{ih} - x_{jh}|$$

where the variables have been previously scaled to make differences comparable;

$$d(i, j) = (\vec{x}_i - \vec{x}_j)^T V_i^{-1} (\vec{x}_i - \vec{x}_j)$$

where  $V_i$  is an estimate of the covariance matrix of  $\vec{x}_i$ ;

$$d(i, j) = \begin{cases} 0 & \text{if } j \in C(\vec{x}_i). \\ 1 & \text{if } j \notin C(\vec{x}_i). \end{cases}$$

Each distance determines a cluster. Hence general clustering algorithms may be used fixing the nonresponse as a centroid.

A simple method for defining the donor set for non-respondents is to use

$$C(\vec{x}_i) = \{j \text{ responded to } y \mid d(i, j) < \delta\},$$

for a pre-specified maximum distance  $\delta$ . A donor may be selected randomly or using the closest respondent.

Within each  $C(\vec{x}_i)$ , imputation class, is assumed that the responding units for item  $Y$  are a random subsample of all sampled units.

#### 4. RANKED SET SAMPLING PROCEDURE

Let us consider that instead of simple random sampling we use Ranked Set Sampling. To create ranked sets the selected first phase sample will be partitioned into sets of equal size, then choose a set size that is typically small. Call this set size  $m$ , where  $m$  is the number of sample units allocated to each set. The selection procedure works as follows:

- Step 1: randomly select  $q^2$  sample units from the population.
- Step 2: allocate the selected units as randomly as possible into  $q$  sets, each of size  $q$ .
- Step 3: without yet knowing any values for the variable of interest, rank the units within each set based on a perception of relative values for this variable. This may be based on personal judgment or done with measurements of a covariate that is correlated with the variable of interest.
- Step 4: choose a sample for actual analysis by including the smallest ranked unit in the first set, then the second smallest ranked unit in the second set, continuing in this fashion until the largest ranked unit is selected in the last set.
- Step 5: repeat steps 1 through 4 for  $c$  cycles until the desired sample size,  $n = qc$ , is obtained for analysis.

The procedure is a two-stage scheme. At the first stage, simple random samples are drawn and a certain ranking mechanism is employed to rank the units in each simple random sample. At the second stage, actual measurements of the variable of interest are made on the units selected based on the ranking information obtained at the first stage. The judgment ranking relating to the latent values of the variable of interest, as originally considered by McIntyre (1952), provides a ranking mechanism.

Let  $Y_1, \dots, Y_q$  be a sample selected using SRSWR from a probability density function  $f(y)$ , with mean  $\mu_Y$  and variance  $\sigma_Y^2$ . As we have selected  $q$  independent samples selected using a SRSWR design, each of size  $q$  each, we deal with  $Y_{11}, \dots, Y_{1q}, Y_{21}, \dots, Y_{2q}, \dots, Y_{q1}, \dots, Y_{qq}$ . They are ordered within each sample and obtain the ordered samples  $Y_{1(1)}, \dots, Y_{1(q)}, Y_{2(1)}, \dots, Y_{2(q)}, \dots, Y_{q(1)}, \dots, Y_{q(q)}$ . We measure  $Y$  only in the  $t$ -th order statistic of the  $t$ -th ordered sample. As we repeat the procedure  $c$  time independently we should have  $c$  observations of each order statistic. The non-response of a unit produces a missing observation of an order statistic.

It seems logical to take as imputation class the set of the  $k$ -th order statistics  $C_{(k)} = \{Y_{k(1)1}, \dots, Y_{k(1)c}\}$ ,  $k=1, \dots, q$ . Its size is  $\|C_{(k)}\| = c$ ,  $k = 1, \dots, q$ . Take  $C_{(k-r)}$  as the set of the responding units for the item  $Y$  in the  $c$  cycle and  $C_{(k-r)}$  with sizes  $c_k$  and  $m_k$ , the number of non-responding units. Note that  $c_k + m_k = c$ .

In this framework we have as estimate of the mean of the  $k$ -th order statistic

$$\bar{y}_{r(k)} = \frac{1}{c^{(k)}} \sum_{t \in C_{(k-r)}} y_{k(k)t}$$

The expectation of the estimator is

$$E(\bar{y}_{r(k)}) = \frac{1}{c^{(k)}} \sum_{t \in C_{(k-r)}} E(Y_{k(k)t}) = \bar{Y}_{(k)}$$

Imputing for the non-respondent  $h$  in  $C_{(k)}$

$$\hat{Y}_{k(k)h} = \frac{1}{\|s\|} \sum_{l \in s \subset C_{(k-r)}} Y_{k(k)l}$$

grants the unbiasedness of the estate of the overall mean as

$$E \left[ \frac{1}{q} \sum_{k=1}^q \left( \frac{1}{c} \left( \sum_{t \in C_{(k-r)}} Y_{k(k)t} + \sum_{h \in C_{(k-r)}} \hat{Y}_{k(k)h} \right) \right) \right] = \frac{1}{q} \sum_{k=1}^q \bar{Y}_{(k)} = \bar{Y}$$

From the variance of the mean of the order statistics we have that

$$V(\bar{Y}_{r(k)}) = \frac{\sigma_{(k)}^2}{c_{(k)}}, \quad V(\hat{Y}_{k(k)h}) = \frac{\sigma_{(k)}^2}{\|s\|}$$

Therefore

$$V \left[ \frac{1}{q} \sum_{k=1}^q \left( \frac{1}{c} \left( \sum_{t \in C_{(k-r)}} Y_{k(k)t} + \sum_{h \in C_{(k-r)}} \hat{Y}_{k(k)h} \right) \right) \right] = \frac{1}{cq^2} \sum_{k=1}^q \sigma_{(k)}^2 = \frac{\sigma^2}{n} - \frac{1}{nq} \sum_{k=1}^q \Delta_{(k)}^2$$

where  $\Delta_{(k)} = \bar{Y}_{(k)} - \bar{Y}$ .

It seem that the RSS alternative is more accurate than the use of SRSWR. We developed a numerical study for establishing if this fact is acceptable.

## 5. NUMERICAL STUDIES

The importance of imputing has motivated the development of different softwares for dealing with nonresponses. Popular softwares are the Generalized Edit and Imputation System (GEIS) of Statistics Canada, see Rancourt (1999); the Nearest-Neighbour Imputation Methodology (NIM) (Parshall et al., 2001; Rubin, 2003). SAS has the procedures in SAS PROC MI. MI algorithms are implemented in IVEware and MICE, see (Rässler et al., 2013).

There are good programs in the R's library. We used some of them for preparing a program for our numerical experiments.

The missing mechanism is simulated by generating an array of non-repeating random numbers using a random number generator based on a pre-specified missing rate. These random numbers are then used as keys to enter the complete data to determine which values to eliminate. Similar simulation procedures have been used or proposed by Drechsler (2011) and Nordholt (1998). The clustering algorithm used in Rancourt's proposal (1999) was considered.

**Table 1. Values of  $A_{IM,p}$  in a study of the production of citric fruits in farms**

		SRSWR		Ranked set Sampling
	Clustering based on $d(i,j) = \max_h  x_{ih} - x_{jh} $	Clustering based on $d(i,j) = (\vec{x}_i - \vec{x}_j)^T V_i^{-1} (\vec{x}_i - \vec{x}_j)$	Clustering based on Nearest neighbor	
Oranges	0,44(2)	0,81(4)	0,59(3)	0,29(1)
Pomegranate	0,11(2)	0,37(3)	0,39(4)	0,07(1)
Pompelmous	0,12(2)	0,19(3)	0,31(4)	0,09(1)
Lemmon	0,52(4)	0,41(3)	0,36(2)	0,10(1)
Lime-Lemmon	0,26(2)	0,37(4)	0,29(3)	0,16(1)
Consensus ranking	2	4	3	1

The evaluation of the experiments was based on the measure of accuracy

$$A_{IM,p} = \frac{1}{1000\bar{Y}} \sum_{h=1}^{1000} |\bar{Y}_{IM,p,h} - \bar{Y}|$$

In each sample was measured

$\bar{Y}_{IM,p,h}$   
 = estimated mean in the sample  $h$  selected by the sampling design  $p$  and imputation method  $IM$

The true mean was known. We fixed  $\delta/\bar{Y} = 0,25$  and  $n=N/10$ .

We used data from a census on the production of fruits in kilograms. There were 11 045 farms. The percent of cultivated area was the auxiliary variable  $X$ . It allowed determining clustering for imputing  $Y$  in the non-respondents and for ranking. 1000 samples were generated independently

These results established that clustering using order statistics is the best methods. It is followed by clustering using  $d(i, j) = \max_h |x_{ih} - x_{jh}|$ , but it is fairly worse for Lemmon, Lime-Lemmon and Oranges. Clustering based on the usual nearest neighbor procedure is ranked third. Clustering based on  $d(i, j) = (\vec{x}_i - \vec{x}_j)^T V_i^{-1} (\vec{x}_i - \vec{x}_j)$  was the worst method for organizing the imputation sets.

Another experiment was conducted using the data in Al-Omari et al. (2014) in a research on cancer. The variables were  $X$  = Outer mass of the tumor. For evaluating both a set of 351 patients was analyzed. 133 patients received double-chemo with an antiangiogenic agent. The results are given in the next table

**Table 2. Values of  $A_{IM,p}$  in a study in a lung cancer study**

		SRSWR		Ranked set Sampling
	Clustering based on $d(i, j) = \max_h  x_{ih} - x_{jh} $	Clustering based on $d(i, j) = (\vec{x}_i - \vec{x}_j)^T V_i^{-1} (\vec{x}_i - \vec{x}_j)$	Clustering based on Nearest neighbor	
Filling-in of cavitation	0,22(3)	0,21(2)	0,28(4)	0,12(1)
Outer mass measured using Computer Tomography	0,20(4)	0,19(3)	0,08(2)	0,03(1)
Maximum tumor diameter	0,60(4)	0,62(3)	0,38(2)	0,11(1)
Arch distance	0,18(3)	0,24(4)	0,15(2)	0,03(1)
Prediction of the tumor diameter using X-Ray	0,40(3)	0,38(2)	0,50(4)	0,19(1)
Consensus ranking	4	2	3	1

The study of lung cancer results also established that clustering using order statistics is the best method. It is followed by clustering using  $d(i, j) = (\vec{x}_i - \vec{x}_j)^T V_i^{-1} (\vec{x}_i - \vec{x}_j)$ . It may be considered as similar to clustering based using the distance  $d(i, j) = \max_h |x_{ih} - x_{jh}|$  though the third best is clustering based the nearest neighbor criteria.

## 6. CONCLUSIONS.



---

The theoretical models suggest that using RSS is the best decision if you expect needing to impute. The logic is supported that the fact that RSS generates a clustering on the order statistics. These stratum are no post-strata, as happens with a clustering generated by non responses.

Further studies in this line must provide more light on the validity of these arguments.

**Acknowledgements.** This paper was developed within the PNCB Modelos Matemáticos para el Estudio de Medio Ambiente, Salud y Desarrollo Humano of CITMA, Cuba.

## REFERENCES

1. Allison, P.D. (2001): *Missing Data*, Sage University Papers Series on Quantitative Applications in the Social Sciences, Series no. 07-136, Thousand Oaks.
2. Al-Omari, A. I., C. N. Bouza, A. Santiago and J. M. Sautto(2014): Studies of Cancer Problems Using Ranked Set Samples. In (Bouza Et Al Ed.) *Modelación Matemática De Fenómenos Del Medio Ambiente y la Salud Tomo 3*, 55-68
3. Cohen M. P. (2003): Imputation Allowing Standard Variance Formulas. *Journal of Data Science* 1, 275-292.
4. Baraldi, A. N., and Enders, C. K. (2010): An introduction to modern missing data analyses. *Journal of School Psychology*, 48, 5–37.
5. Blue-ETS (2013): Deliverable 6.1. Best practice recommendations on variance estimation and small area estimation in business surveys. Computer Codes. URL <http://www.blueets.istat.it/fileadmin/deliverables/Deliverable6.1.pdf>. (Last Consulted June 10, 2014)
6. Brick, J. M., Jones, M. E., Kalton, G. and Valliant, R. (2005): Variance Estimation with Hot Deck Imputation: A simulation Study of Three Methods. *Survey Methodology*, 31, 151-159.
7. Carpenter, J. and Goldstein, H. (2005): Multiple Imputation in MIwiN, available from [www.missingdata.org.uk](http://www.missingdata.org.uk). (Consulted Jul y 7, 2015)
8. Chen, J., Rao, J. N. K., and Sitter, R. R. (2000): Efficient random imputation for missing data in complex surveys. *Statistica Sinica*, 10, 1153–1170
9. Kadengye, D. T., W. Cools, E. Ceulemans and W. Van den Noortgate (2012): Simple imputation methods versus direct likelihood analysis for missing item scores in multilevel educational data. *Behav Res.*, 44, 516–531
10. Drechsler J. (2011): *Generating Multiply Imputed Synthetic Datasets: Theory and Implementation*. Lecture Notes in Statistics. Springer, N. York.
11. Durrant G. B. (2015): *Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review*. NCRM Methods Review Papers NCRM/002
12. Hox J. J. (1999): A Review of Current Software for Handling Missing Data, *Kwantitatieve Methoden*, 62, 123-138.
13. Kalton, G. and Kish, L. (1984): Some Efficient Random Imputation Methods, *Communications in Statistics, Part A, Theory and Methods*, 13, 1919-1939.
14. Little, R. J. A. and Rubin, D. B. (2002): *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics, Wiley, N. York.
15. Lanke, J. (1983): Hot deck imputation techniques that permit standard methods for assessing precision of estimates. *Statist. Review*, 21 105-110.
16. Nordholt, E.S. (1998): Imputation: Methods, Simulation, Experiments and Practical Examples, *International Statistical Review*, 66, 2, 157-180.
17. Parshall, C.G. (1998): Item development and pretesting in a computer-based testing environment. Paper presented at the ETS Sponsored Colloquium on Computer-Based Testing: Building the Foundation for Future Assessments, Philadelphia.
18. Parshall, C.G., Kromrey, J. D., Harmes, J.C., and Sentovich, C. (2001): Nearest neighbors, simple strata, and probabilistic parameters: An empirical comparison of methods for item exposure control in CATs. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle.
19. Rancourt, E. (1999): Estimation with Nearest Neighbour Imputation at Statistics Canada, in: *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 131-138.
20. Rässler, R. Donald B., Rubin and Elizabeth R. Zell (2013): Imputation. *WIREs Comput Stat.*, 5, 20–29.

- 
21. Righi, P., S. Falorsi and A. Fasulo (2014): Methods for variance estimation under random hot deck imputation in business surveys. *Rivista Di Statistica Ufficiale* N. 1-2, 45-64.
  22. Rubin, D.B. (1987): *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc., N. York.
  23. Rubin DB. (2003): Nested multiple imputation of NMES via partially incompatible MCMC. *Stat Neerl* , 57, 3–18.
  24. Schafer, J.L. (1997): *Analysis of Incomplete Multivariate Data*. Chapman and Hall., N. York.
  25. Schafer, J.L. (1999): NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT.  
<<http://www.stat.psu.edu/~jls/misoftwa.html>> (Last Accessed on May 2, 2015):
  26. Skinner, C. and Rao, J.N.K. (2002): Jackknife Variance Estimation for Multivariate Statistics under Hot Deck Imputation From Common Donors. *Journal of Statistical Planning and Inference*, 102, 421-422.
  27. Shao, J. and Steel, P. (1999): Variance Estimation for Survey Data with Composite Imputation and Non-negligible Sampling Fractions, *Journal of the American Statistical Association, Theory and Methods*, 94, 254-265.
  28. Shao, J. and Tang, Q. (2011): Random Group Variance Estimators for Survey Data with Random Hot Deck Imputation. *Journal of Official Statistics*, 27, 507-526.

## **REVISIÓN SISTEMÁTICA DE LOS MÉTODOS DE IMPUTACIÓN DE DATOS FALTANTES**

Carmen Viada\*, Carlos Bouza\*\*, Javier Ballesteros\*\*\*, Martha Fors\*, Mayteé Robaina\*\*\*  
&  
Rolando Uranga\*\*\*

\* Department of Clinical Information Management, Center of Molecular Immunology, Cuba

\*\*Universidad de LA Habana, Cuba.

\*\*\* University of the Basque Country, UPV/EHU Dept of Neuroscience, España

\*\*\*\*Centro Nacional Coordinador de Ensayos Clínicos, Cuba

### **ABSTRACT**

In practice, data analysts frequently deal with non-observed data. In this paper we compare properties of different statistical procedures for missing data with an arbitrary pattern for non-response. We study from heuristic methods to model-based methods, under different missingness mechanisms and considering heterogeneous variables. We analyze their effect on point estimates and on the coverage of confidence intervals. Finally, recommendations for practical data analysis are obtained.

**KEY WORDS:** missing data, multiple imputation, and statistical inference.

### **RESUMEN**

En la práctica estadística es frecuente encontrar muestras con datos que no han podido observarse. En este artículo se comparan las propiedades de distintas estrategias de inferencia a partir de muestras con datos faltantes según un patrón arbitrario. Se estudian desde métodos heurísticos hasta métodos basados en verosimilitudes, bajo distintos mecanismos para la no respuesta y con variables de características dispares. Se analiza el efecto sobre las estimaciones puntuales y la cobertura de los intervalos de confianza. Finalmente, se extraen conclusiones de utilidad para la práctica del análisis de datos.

### **1. INTRODUCCIÓN**

En el desarrollo teórico de la mayoría de las técnicas y modelos estadísticos no se tienen en cuenta algunas cuestiones que surgen en su aplicación práctica, en concreto, un problema al que con seguridad se ha enfrentado cualquier analista de datos es el de los datos faltantes, también denominados perdidos o incompletos.

Cuando se toma una muestra, en general con  $k$  variables, de tamaño  $n$  obtenemos una matriz de datos de dimensiones  $n \times k$ . Habitualmente esa matriz es incompleta en el sentido de que faltan datos sobre alguna o algunas de las variables para alguno o algunos de los casos, u observaciones, de la muestra. El estudio sistemático y la formalización de este problema desde un punto de vista probabilístico no se inicia hasta mediados de los años setenta, destacando principalmente el trabajo de Rubin (1976). Aún hoy, se tiende a infravalorar el efecto de eliminar de la matriz de datos aquellos casos con valores perdidos o a sustituirlos por valores que intuitivamente parecen razonables con el fin de eludir el problema y disfrutar de una nueva matriz completa sobre la cual aplicar los análisis pertinentes. De hecho, muchos de los programas informáticos de análisis de datos de uso generalizado incorporan dichas pseudo-soluciones en sus versiones estándar, de modo que son las empleadas por la mayor parte de los usuarios no especialistas.

Hasta hace relativamente poco, los únicos métodos generalmente utilizados para tratar el problema de los datos perdidos eran métodos como la eliminación del caso con valores perdidos, la sustitución/imputación de éstos por valores plausibles como la media de la variable o la predicción obtenida mediante regresión sobre las demás variables del vector, etc. Este tipo de métodos clásicos no

---

suelen tener una base teórica sólida y, aunque fáciles de implementar y adecuados en situaciones concretas, presentan en general importantes inconvenientes y carencias, especialmente en contextos multivariantes. Los principales problemas inferenciales asociados son ineficiencia, aparición de sesgos, distorsión de la estructura de covarianzas; además de no incorporar la incertidumbre asociada a los datos faltantes.

Frente a estos métodos clásicos, en los últimos años, y de forma paralela a la formalización del problema de los datos faltantes, se han ido desarrollando métodos con una base teórica más sólida. Así, en Dempster, Laird y Rubin (1977) se establece una formulación general y rigurosa para la inferencia en presencia de datos faltantes mediante el algoritmo EM. Por otro lado, Rubin (1987) desarrolla una nueva metodología de propósito general, flexible y fundamentada que denomina imputación múltiple, y que salva muchos de los inconvenientes asociados al tratamiento tradicional de los datos faltantes.

Nuestro propósito en este artículo es analizar y comparar el comportamiento de las principales estrategias existentes en la actualidad para el análisis estadístico de matrices de datos incompletos. Este trabajo se inspira fundamentalmente en el artículo de Schafer y Graham (2002) donde se analiza un problema bidimensional con un patrón de no respuesta univariante. Nuestro objetivo es ampliar dicho estudio planteando una situación más compleja y general, y por ende, más realista, con un patrón de no respuesta arbitrario, considerando variables de comportamiento y rango no homogéneos, con distintos niveles de variabilidad y distintos grados de interrelación entre ellas.

Lo que pretendemos es que el investigador aplicado disponga de una herramienta más para poder elegir qué es lo que más le interesa, dadas las características de sus datos, a la vista de las ventajas e inconvenientes de cada estrategia. Mostraremos cómo afectan los distintos métodos a las medidas de variabilidad y correlación, qué métodos son a priori más fiables, cuales se muestran más robustos ante distintos mecanismos de no respuesta, etc.

En las secciones siguientes iremos describiendo los procedimientos analizados y la forma en la que se han implementado. Se compararán los resultados obtenidos y se extraerán conclusiones útiles desde un punto de vista práctico.

### 3. MECANISMOS DE NO RESPUESTA

Cuando en una muestra aparecen valores perdidos por razones fuera del control del investigador, es necesario establecer unos supuestos sobre el proceso que los ha generado. Estos supuestos serán en general no verificables, y por ello deberán hacerse explícitos y analizar la sensibilidad del procedimiento general de estimación frente a desviaciones de los mismos. Si entendemos la presencia de valores perdidos como un fenómeno probabilístico necesitamos un mecanismo matemático que describa las leyes que rigen su aparición, y que capte aproximadamente las posibles relaciones entre la aparición de valores perdidos y los datos no observados en sí mismos.

Consideremos, con carácter general, un vector aleatorio  $X$   $k$ -dimensional que genera los datos y un vector  $R$ , también  $k$ -dimensional, formado por variables aleatorias binarias tomando valores 0 ó 1 para indicar valor observado o no observado. Llamaremos mecanismo de no respuesta a la distribución de probabilidad de  $R$ . Asociada a una muestra del vector  $X$  tendremos una muestra de  $R$ , cuya forma dependerá de la complejidad del patrón de no respuesta. Por ejemplo, si el patrón es univariante(1) tendremos una realización de una variable aleatoria binaria unidimensional indicando si un valor concreto es observado o perdido. Si el patrón es arbitrario(2), tendremos entonces una matriz de dimensiones  $n \times k$  con elementos  $r_{ij}$  tomando valor 0, si  $x_{ij}$  es observado, ó 1, si  $x_{ij}$  es no observado.

Si denotamos mediante  $X$  a una muestra multidimensional de  $X$  podemos hacer una partición de forma que  $X = (X_{\text{obs}}, X_{\text{per}})$ , donde  $X_{\text{obs}}$  y  $X_{\text{per}}$  denotan la parte observada y la parte no observada, o perdida, respectivamente. Se dice que los datos faltantes son de tipo MAR (missing at random) si la probabilidad de que un valor no se observe depende de los valores de los datos observados, pero no de los faltantes. Esto es, si  $P[R|X_{\text{obs}}, X_{\text{per}}, \xi] = P[R | X_{\text{obs}}, \xi]$ , siendo  $\xi$  un vector de parámetros desconocidos del mecanismo de no respuesta. Como indican los resultados de Rubin (1976), no es necesario que se satisfaga para todas las posibles realizaciones de  $R$ , basta con que se verifique en la muestra dada. Por

---

otro lado, se dice que los datos faltantes son de tipo MCAR si  $P[R|X_{obs}, X_{per}, \xi] = P[R | \xi]$ . La hipótesis MCAR (missing completely at random) es más restrictiva ya que implica que R y X son independientes, algo difícil de mantener en muchas situaciones prácticas. Por último, se dice que los datos faltantes son de tipo NMAR (not missing at random) si el mecanismo de no respuesta depende del verdadero valor del dato perdido (es decir, depende de  $X_{per}$ ), o de variables no observables. La hipótesis NMAR es la más general, pero al mismo tiempo es la más difícil de modelizar ya que exige la especificación de un modelo para R, por lo que es frecuente hablar de mecanismo de no respuesta no ignorable.

Sobre la hipótesis MAR descansan la mayoría de las técnicas actuales para datos faltantes, sin embargo no existen procedimientos generales para contrastarla sobre un conjunto de datos incompletos. Nos interesarán por lo tanto métodos que ofrezcan resultados robustos frente a posibles desviaciones. La sensibilidad de las respuestas obtenidas a partir de una muestra incompleta frente a supuestos débiles o injustificables es un problema básico asociado al análisis de datos incompletos, especialmente en el caso NMAR.

En muchas aplicaciones lo prudente será considerar distintos modelos plausibles para el mecanismo de no respuesta y realizar un análisis de sensibilidad de las estimaciones. Aún así, como destacan Molenberghs, 2001 y 2007, esta estrategia puede llevar a conclusiones equivocadas. Podemos encontrar una revisión de las técnicas de análisis de sensibilidad en Serrat (2001), en el contexto del análisis de supervivencia. Por último, destacar el trabajo de Troxel et al (2004), el que se presenta un índice de sensibilidad a la no ignorabilidad, una medida del potencial impacto de la no ignorabilidad en un análisis.

(1) *Se habla de patrón univariante cuando los valores perdidos sólo aparecen en una de las variables del vector aleatorio.*

(2) *Se habla de patrón arbitrario o general cuando los valores perdidos pueden aparecer en cualquier variable y observación de la muestra.*

## 5. PROBLEMAS GENERALES DE LA NO RESPUESTA

Cuando se aplica una encuesta, es muy probable que nos encontremos con espacios en blanco en el área de respuesta que denominamos “no respuesta”. Su origen puede estar en un “no lo sé” o en expresiones de valores o frases incongruentes con la pregunta planteada, debido posiblemente a inconvenientes que se presentan para lograr la medición en el momento preciso, o son mediciones muy difíciles de obtener, o porque la pregunta o medida lleva consigo ciertos compromisos que prefieren mantenerse sin responder. Esta variedad de situaciones llevan a la “no respuesta”, también conocida como datos faltantes.

La ausencia de datos plantea un gran problema que enfrenta el analista. En ocasiones, por ejemplo, con grandes conjuntos de datos y con poca proporción de pérdida, pudiera ignorarse la ausencia de datos, pero esto no es conveniente cuando se trata de pocas observaciones o de altas proporciones de pérdida.

La proporción de ausencia de ítems en un registro puede variar; dependiendo del estudio, de la dificultad de llenar el cuestionario o de la medición de una unidad, sin embargo, dependerá del investigador considerar el registro como pérdida parcial o como pérdida total cuando el número de ítems perdidos sea considerable. Otra situación cuestionada por el investigador es determinar hasta qué porcentaje de pérdida de ítems se considera tratable mediante imputación o considerarlo como una mala recolección de información, y por tanto, la base de datos obtenida es muy defectuosa que simplemente no debe ser usada. La respuesta a esto no es tan sencilla. En la práctica se habla de pérdidas máximas entre 1 y 20% de la data dependiendo de la exactitud del estudio y del área de investigación entre otros factores. Por ejemplo, en las ciencias médicas, la precisión es un factor muy importante en la obtención de resultados y estudios profundos, no pueden permitir la imputación de muchos valores que nunca serán reales, sólo para poderlos analizar, mientras que en las ciencias sociales permite porcentajes de ítems imputados más altos, siempre estará en manos del investigador esta decisión.

---

Cuando no se pueden ignorar los datos faltantes, la manera más adecuada de tratarlos es llenar esos espacios faltantes con valores plausibles; a este procedimiento es lo que denominamos imputación. Lohr (1999), refiriéndose a la importancia de los procedimientos de imputación señala que ésta no radica sólo en reducir el sesgo por las ausencias de respuestas, sino también para producir un conjunto de datos rectangulares y “limpios” sin datos faltantes.

Son muchas las técnicas de imputación que han surgido, sobre todo desde la década de los setenta, que emplean enfoques univariantes y multivariantes. Se han empleado enfoques basados en modelos como: funciones de verosimilitud, regresión y descomposición de matrices en valores singulares, entre otros. A pesar de estos avances, no se ha encontrado una metodología capaz de reproducir la data o que pueda resolver en forma totalmente satisfactoria el tratamiento de los datos faltantes, debido, generalmente, a problemas en cuanto a las alteraciones de la distribución de los datos, alteración en la relación de las variables, sesgo en las estimaciones, inflación de la varianza, entre otros, razón por la cual, aún se sigue investigando en busca de mejorar las técnicas existentes.

### 3.1. Tipos de no Respuesta

La no respuesta, aunque cuando no se quiera, siempre estará presente en toda investigación que involucre medición, de la cual, muchas veces no se obtendrán registros completos por diversas causas ajenas al investigador.

La situación idónea en una investigación es obtener una base de datos completa, de valores reales que permita llevar a cabo las tradicionales técnicas de análisis de datos.

La no respuesta, puede presentarse de dos maneras:

- La **no respuesta total**, es cuando falta todo el registro de una base de datos, por ausencia de la unidad a medir o por impedimento de efectuar un conjunto total de mediciones de variables en un determinado momento específico, es decir, no se recoge ningún dato de la unidad de la muestra. Por ejemplo, cuando se lleva a cabo la aplicación de encuesta en hogares y en algunas viviendas seleccionadas, y no se encuentran personas al momento de aplicar el instrumento, generándose una pérdida total de las respuestas del cuestionario que se le iba a aplicar a ese hogar o a esa persona.
- La **no respuesta parcial**, se presenta cuando hay ausencia de una o más variables, sin llegar a la ausencia completa de un registro, ejemplo; un individuo a encuestar se encuentra, pero no responde algunas preguntas del cuestionario o a una unidad no se le efectuaron algunas mediciones, por fallas en los equipos, sin embargo, otras mediciones si se llevaron a cabo.

La no respuesta parcial puede tener dos formas de presentarse; cuando las variables de un registro están ausentes, porque la data no está disponible, o cuando una variable produce una inconsistencia con el resto de las variables, entonces se dicen que están “pérdidas artificialmente” debido a que han sido eliminadas mediante un proceso de depuración, éstas últimas serán tratadas como la primera forma.

### 3.2. Patrones de Pérdida de Respuesta

Uno de los puntos a considerar en la no respuesta, es el patrón de pérdida de los datos faltantes, ya que estos pueden influir en la selección del método de imputación. Los patrones de pérdida pueden ser ignorables o no ignorables.

Los patrones de pérdida pueden ser ignorables si ocurren de manera completamente aleatoria (MCAR, Missing Completely At Random) o de manera aleatoria (MAR, Missing At Random). El primer caso (MCAR), ocurre cuando la ausencia de información depende de alguna variable presente en la matriz de datos ya sea  $x$  o  $y$ . Para el segundo caso (MAR), ocurre cuando la ausencia de los datos depende de variables presentes en la matriz de datos, excluyendo la variable perdida.

Los patrones de pérdida no ignorables (NMAR) son los que ocurren cuando la ausencia de los datos depende de la variable perdida, esto traería como consecuencia estudiar el patrón de pérdida de los datos ausentes para luego imputar tomando en cuenta dicho patrón.

---

### 3.3. Causas y Consecuencias de la no Respuesta

Existen múltiples causas que producen constantemente la no respuesta, ya sea total o parcial, las más comunes son; las unidades de las muestras son inaccesibles, ausencia o impedimento para contactar a los respondientes, no hay cooperación, rechazo o incapacidad para responder, material perdido, el encuestador, además del diseño, la extensión, el tema a tratar, el tipo y orden de las preguntas, la redacción y el vocabulario del cuestionario. Platek (1986) plantea: "... ¿qué es entonces lo que lleva a los encuestados a no responder? Hay tres razones: violación del derecho a la intimidad, carga de respuestas y hostilidad general del Gobierno. Al tratar con éstos problemas, es importante enfocarlos desde el lado de los encuestados, no desde el lado de las ideas preconcebidas del diseñador de encuestas".

En cuanto a las consecuencias que causa la falta de respuesta tenemos: resultados deficientes e incluso inválidos que puede llevar a una pérdida de toda la investigación, distorsión de las frecuencias marginales y/o conjuntas de las variables, sesgos en las estimaciones, disminución del tamaño de la muestra y todo lo que esto implica (aumento del error de muestreo, falta de representación en grupos o variables, estimaciones imposibles de obtener).

### 3.4. Tratamientos Generales de la no-Respuesta

Hay varias maneras de tratar la no respuesta, según Sande (1982), estas pueden ser:

- Eliminar todos los registros que tengan al menos un dato faltante, lo cual puede ser factible si tenemos un gran conjunto de datos y la pérdida es ignorable o si la proporción de registros a eliminar es muy pequeña, en caso contrario, se estaría perdiendo mucha información que puede ser importante.
- Crear una categoría como "no respuesta" en aquellas variables donde hay ausencia de información, pero esto no considera información parcial que puede servir para otros análisis o para otras variables.
- Ignorar los datos faltantes en cada caso, ponderando la variable convenientemente, pero se debe tener precaución, ya que se puede llegar a base de datos inconsistentes, es decir, valores en diferentes variables en un mismo registro que no tengan sentido que ocurran.
- Imputar los datos faltantes.

### 3.5. La Evolución de los Estudios de la no-Respuesta

Los primeros aportes en imputación se realizaron en 1932 por Wilks, quien propuso el reemplazo de los datos faltantes por la media de los datos presentes de la variable. Este método, puede ser aplicado cuando existen pocos datos faltantes, ya que tienden a distorsionar la distribución de las variables.

Con el avance tecnológico de los sistemas computacionales se iniciaron investigaciones sobre el particular en las décadas de los setenta y ochenta. Entre los principales autores de los últimos años que han hecho grandes investigaciones referentes a imputación figuran Kalton, Kasprzyk, Rubin, Little y Sande. Entre sus principales aportes se encuentran los siguientes:

Rubin (1976), hizo una distinción entre MAR (valores perdidos o faltantes de manera aleatoria) y MCAR (valores perdidos o faltantes de manera completamente aleatoria). En el caso de MAR, los datos perdidos dependen de los valores observados, pero no de las variables perdidas propiamente, es decir, la pérdida de información de la variable no depende de ella misma, mientras que en MCAR los datos perdidos no dependen de otros valores observados ni de otros datos perdidos. Ambas distinciones corresponden a patrones de pérdida ignorables, existiendo además de este, un segundo enfoque el cual corresponde a patrones de pérdida no ignorables, que como se mencionó anteriormente, no ocurren de manera aleatoria, sino que siguen un patrón sistémico específico, el cual hay que estudiar previamente a la imputación.

---

Kalton y Kasprzyk (1982), estableció las diferencias existentes entre las técnicas de ajuste ponderado, cuando hay pérdida de un registro completo llamado no respuesta total, y las técnicas de imputación para los casos de registros con pérdidas sólo de algunas variables, es decir, no respuesta parcial, estableciendo los principales efectos de ambos métodos en las respuestas, como; aumento de sesgo, principalmente en las estimaciones de ajuste ponderado y pérdida de relación entre variables al aplicar las técnicas de imputación.

Rubin (1983), expuso otros enfoques para el estudio de los datos faltantes, los cuales los clasificó como enfoque basado en la aleatorización, frecuentista, y enfoque basado en el modelo de superpoblaciones, enfoque Bayesiano. Rubin estableció algunas ventajas y desventajas entre los diferentes enfoques ante un conjunto determinado de datos, tales como; el enfoque aleatorio es más simple de computar y es más robusta, sin embargo, el enfoque Bayesiano obtiene más precisión del intervalo de probabilidad para la variable a imputar.

Little y Rubin (1987), desarrollan una nueva técnica llamada imputación múltiple, en la que los datos faltantes, son sustituidos por  $m > 1$  valores simulados. La imputación múltiple permitió hacer un uso eficiente de los datos, obtener estimadores no sesgados y reflejar adecuadamente la incertidumbre que la no-respuesta parcial introduce en la estimación de parámetros (Goicoechea, 2002).

Helmel (1987), aportó al tratamiento de la no respuesta un método llamado Listwise, el cual, es usado cuando se tiene un gran conjunto de datos y se puede eliminar la fila o columna donde se encuentra la data perdida, para obtener una base, aunque más pequeña, completa. Esta técnica es comúnmente usada hoy en día, pero no es recomendable para casos de pequeños conjuntos de datos, por la pérdida de información que ocasiona.

En la década de los 90, Todeschini (1990) propuso un k-vecino más cercano como método de estimación de valores perdidos; obteniéndose buenos resultados cuando se cuenta con información auxiliar.

Otras investigaciones han buscado maneras de mejorar las técnicas de imputación o crear nuevas, como las basadas en redes neuronales (Koikkalainen, 2002), análisis factorial (Genz y Li, 2003), en análisis de componentes principales (Gleason, y Staelin, 1975), o basada en la descomposición GH-Biplot (Vásquez, 1995), entre otras.

A partir del año 2000 se ha implementado el uso de árboles de clasificación como mejora de los procedimientos de imputación. Mesa, Tsai y Chambers (2000) realizaron un estudio de imputación del Censo del Reino Unido mediante el uso de árboles de clasificación, en el cual se llegó a las siguientes conclusiones:

1. Los árboles grandes no son garantía de obtener mejores imputaciones.
2. El software a usar tiene comparativamente poca importancia.
3. Los métodos de probabilidad más alta y la selección de una categoría aleatoria no mantienen la distribución de la variable,
4. Ninguno de los métodos investigados fueron satisfactorios desde el punto de vista de recuperar los datos perdidos actuales individuales.

Aún son muchos los estudios que se deben llevar a cabo para resolver los problemas mencionados anteriormente, los cuales se presentan en la mayoría de los métodos de imputación que existen hoy en día, haciendo distinción en los diferentes enfoques que puedan existir y dependiendo además del conjunto de datos a analizar. Bacallao, 2010 explicó el uso de los árboles de clasificación en la imputación múltiple.

#### **4. PRIMEROS MÉTODOS HEURÍSTICOS**

En esta sección nos ocuparemos de las soluciones habitualmente utilizadas en la práctica ante una matriz de datos con valores perdidos. Aunque intuitivamente pueden parecer soluciones razonables y cuando la cantidad de información perdida es pequeña, pueden funcionar relativamente bien en algunos casos, veremos en la sección 8 que no son procedimientos generalmente aceptables.

##### **Análisis de casos completos**



---

Dada una muestra  $X$  de  $k$  variables y  $n$  casos, supongamos que  $n_{\text{per}} < n$  de estos casos presentan al menos un valor perdido para alguna de las  $k$  variables. Como su propio nombre indica, mediante este método se descartan los  $n_{\text{per}}$  casos y sólo se aplica la técnica o análisis sobre aquellos con valores observados para todas las variables. De este modo se pasa a trabajar con una muestra de datos completa de tamaño  $n - n_{\text{per}}$ . Las consecuencias de esta medida dependerán fundamentalmente de la cantidad de información que se pierda al descartar los casos con faltantes, del mecanismo según el cual faltan los datos y de la relación entre los casos completos e incompletos. La pérdida de información relevante se traducirá en sesgo y falta de precisión de las estimaciones si los faltantes no son una muestra aleatoria de la muestra completa, es decir, si no se verifica la hipótesis MCAR. Como ventajas destacaremos su simplicidad y el hecho de que todos los estadísticos se calculan utilizando el mismo tamaño muestral, lo que permite su comparación.

### Análisis de casos disponibles

Si para el  $i$ -ésimo caso de una muestra se observan  $p$  de las  $k$  variables, al aplicar análisis de casos completos estamos perdiendo la información que sobre las  $k - p$  variables restantes contiene dicho caso. Una alternativa natural es utilizar para cada cálculo toda la información disponible en la muestra. Por ejemplo, a la hora de calcular la media o varianza de las variables  $X_i$  y  $X_j$  se utilizarán los  $n_i$  y  $n_j$  datos disponibles sobre cada una de ellas respectivamente. O bien, para calcular la covarianza entre las dos variables  $X_i$  y  $X_j$  se considerarán los casos  $h$  para los que los pares  $(x_{hi}, x_{hj})$  son observados. Es obvio que ello implicará en general trabajar con distintos tamaños muestrales e incluso combinarlos en el cálculo de un mismo estadístico. Como puede verse en Little y Rubin (2002) es posible entonces que resulten correlaciones fuera del intervalo  $[-1,1]$  o matrices de correlaciones no definidas positivas, condición requerida por diversas técnicas multivariantes.

El objetivo de la imputación es obtener un archivo de datos completos y consistentes para que puedan ser analizados mediante técnicas estadísticas tradicionales.

## 5. VENTAJAS Y DESVENTAJAS DE IMPUTAR

Cuando imputamos, logramos obtener una base de datos completa, la cual permitirá llevar a cabo metodologías de análisis de datos comunes y el uso de software tradicionales para su manejo. Si una imputación se lleva a cabo de manera adecuada, podría disminuir el sesgo, en caso de existir. Por otra parte, el investigador debe estar consciente que el uso de imputación también puede llevar a afectar las distribuciones conjuntas, o incluso, distribuciones marginales de las variables, aunque el problema es menor si la distribución de los casos ausentes es la misma que la de los casos completos (patrón de pérdida ignorable), como se mencionó anteriormente. Si la técnica no es la adecuada, posiblemente, aumenta el sesgo, subestima o sobrestima la varianza, se obtienen datos imputados inconsistentes produciendo una base de datos no confiables, llevando a la interpretación errónea de los resultados por parte de los usuarios.

## 6. DIFERENTES TIPOS DE TÉCNICAS DE IMPUTACIÓN

Varios estudios (Goicoechea, 2002; Platek 1986; y Government Statistical Service 1996), indican que las técnicas de imputación se pueden clasificar de la siguiente manera:

1. **Técnicas fundamentadas en información externa:** cuando son basadas en variables relacionadas con una encuesta perteneciente a otras bases de datos o reglas previas. Entre estas se encuentran:
  - a) **Métodos deductivos:** cuando los datos faltantes se deducen con cierto grado de certidumbre de otros registros completos del mismo caso, siguiendo algunas reglas específicas.
  - b) **Tablas Look-up:** cuando se hace uso de una tabla con información relacionada, como fuente de data externa para imputar los datos faltantes.
2. **Técnicas determinísticas:** cuando al repetir la imputación en varias unidades bajo las mismas condiciones, producirá las mismas respuestas.
  - a) **Imputación de la media o moda:** se llena el vacío del dato faltante de cada variable con la media de los registros no faltantes en caso de variables cuantitativas, o con la moda en caso de variables cualitativas. Tiene como desventaja la modificación de la distribución de la

variable haciéndose más estrecha ya que reduce su varianza, además, no conserva la relación entre variables y se debe asumir una MAR. Su ventaja es la facilidad de la aplicación del método.

Dada una variable  $X_i$  que presenta valores perdidos, mediante este método se reemplaza cada uno de ellos por  $x_i^{obs}$ , la media de los valores observados de  $X_i$ . Aunque esta estrategia es sencilla y puede resultar intuitivamente satisfactoria, presenta un importante defecto, y es que, como veremos en la sección 8, tiende a subestimar la variabilidad real de la muestra al sustituir los faltantes por valores centrales de la distribución. Bouza, 2011 construye el estimador de la mediana para datos incompletos.

b) **Imputación de media de clases:** las respuestas de cada variable son agrupadas en clases disjuntas con diferentes medias, y a cada registro faltante se le imputará con la media respectiva de su grupo. Tiene las mismas desventajas que el caso anterior, pero en menor proporción por estar agrupadas. Igualmente es de fácil aplicación.

c) **Imputación por regresión:** se ajusta un modelo lineal que describa a  $y$ , variable a imputar, para un conjunto  $X$  de variables auxiliares que se deben disponer. Resuelve el problema de la distorsión de la distribución de la variable a imputar, pero puede crear inconsistencias dentro de la base de datos, pues podría obtenerse valores “imposibles”, ya que el valor  $y$  es obtenido de variables auxiliares.

Consideremos una variable  $X_i$  que presenta  $n_{per}$  valores perdidos y  $n_i = n - n_{per}$  valores observados. Supongamos que las  $k-1$  restantes variables  $X_j$ , con  $j \neq i$ , no presentan valores perdidos. Con este método se estima la regresión de la variable  $X_i$  sobre las variables  $X_j$ ,  $\forall j \neq i$ , a partir de los  $n_i$  casos completos y se imputa cada valor perdido con la predicción dada por la ecuación de regresión estimada. Esto es, si para el caso  $l$  el valor  $x_{li}$  no se observa, entonces se imputa mediante:

$$x_{li}^{\wedge} = \beta_{0obs}^{\wedge} + \sum_{j \neq i} \beta_{jobs}^{\wedge} x_{lj} \quad [4.1]$$

donde  $\beta_{0obs}^{\wedge}$  y  $\beta_{jobs}^{\wedge}$ ,  $j \neq i$ , representan los coeficientes de la regresión de  $X_i$  sobre  $X_j$ ,  $\forall j \neq i$ , basada en las  $n_i$  observaciones completas. Frente a la imputación mediante la media, este método incorpora la información que sobre  $X_i$  contienen el resto de variables.

d) **Emparejamiento media:** se lleva a cabo el método (e) donde el valor de  $y$  (estimado) es comparado con casos completos, y el caso más cercano correspondiente provee el valor imputado  $y$ .

e) **Imputación por el vecino más cercano:** se identifica la distancia entre la variable a imputar  $y$ , y cada una de las unidades restantes ( $x$  o variables auxiliares) mediante alguna medida de distancia, entonces se determina la unidad más cercana a  $y$ , usando el valor de esta unidad cercana para imputar el faltante.

f) **Algoritmo EM (Expectation Maximization):** basada en la función de máxima verosimilitud, permite obtener estimaciones máximo verosímiles (MV) de los parámetros cuando hay datos incompletos con unas estructuras determinadas. Resuelve de forma iterativa el cálculo del estimador máximo verosímil mediante dos pasos en cada iteración (Little y Rubin, 1987). Este algoritmo tiene la ventaja de que puede resolver un amplio rango de problemas, incluyendo problemas no usuales que surgen de la pérdida o data incompleta, como lo es la estimación de los componentes de la varianza.

El algoritmo EM (Dempster, Laird y Rubin, 1977) es un algoritmo iterativo diseñado para la obtención de estimadores máximo-verosímiles (EMV) en problemas con muestras incompletas. Sea  $X = (X_{obs}, X_{per})$  una muestra incompleta de  $X \sim P[X; \theta]$  a partir de la cual se desea obtener el EMV de  $\theta$ . Podemos factorizar  $P[X; \theta]$  como

$$P[X; \theta] = P[X_{obs}; \theta] P[X_{per} | X_{obs}, \theta],$$

de donde es sencillo deducir que

$$\log L(\theta | X_{obs}) = \log L(\theta | X) - \log P[X_{per} | X_{obs}, \theta], \quad [4.2]$$

siendo  $\log L(\theta|X_{\text{obs}})$  la log-verosimilitud para los datos observados y  $\log L(\theta|X)$  la log-verosimilitud para los datos completos  $X$ . En presencia de datos faltantes, el objetivo es estimar  $\theta$  mediante la maximización de  $\log L(\theta|X_{\text{obs}})$  con respecto a  $\theta$  dada  $X_{\text{obs}}$ . Como vemos a continuación, el algoritmo EM relaciona el EMV de  $\theta$  a partir de  $\log L(\theta|X_{\text{obs}})$  con el EMV de  $\theta$  a partir de  $\log L(\theta|X)$ . Tomando esperanzas respecto a  $P[X_{\text{per}}|X_{\text{obs}}, \theta]$  a ambos lados de [4.2] y dado un estimador  $\theta^{(t)}$  de  $\theta$ , tenemos que

$$\log L(\theta|X_{\text{obs}}) = Q(\theta; \theta^{(t)}) - H(\theta; \theta^{(t)}),$$

donde  $Q(\theta; \theta^{(t)}) = \int \log L(\theta|X) P[X_{\text{per}}|X_{\text{obs}}, \theta^{(t)}] dX_{\text{per}}$  y  $H(\theta; \theta^{(t)}) = \int \log P[X_{\text{per}}|X_{\text{obs}}, \theta] P[X_{\text{per}}|X_{\text{obs}}, \theta^{(t)}] dX_{\text{per}}$

El paso E (expectation) del algoritmo EM calcula  $Q(\theta; \theta^{(t)})$ , reemplazando los valores perdidos, o una función de ellos, por su esperanza condicionada dados  $X_{\text{obs}}$  y  $\theta^{(t)}$ . El paso M (maximization) simplemente determina el EMV  $\theta^{(t+1)}$  que maximiza  $Q(\theta; \theta^{(t)})$  como si no hubiera datos perdidos. Los pasos E y M se repiten alternativamente generando una sucesión de estimadores  $\theta^{(t)}$ . La diferencia en el valor de la log-verosimilitud  $\log L(\theta|X_{\text{obs}})$  en dos iteraciones sucesivas viene dada por :

$$\log L(\theta^{(t+1)}|X_{\text{obs}}) - \log L(\theta^{(t)}|X_{\text{obs}}) = Q(\theta^{(t+1)}; \theta^{(t)}) - Q(\theta^{(t)}; \theta^{(t)}) + H(\theta^{(t)}; \theta^{(t)}) - H(\theta^{(t+1)}; \theta^{(t)})$$

Como el estimador  $\theta^{(t+1)}$  se escoge de manera que  $Q(\theta^{(t+1)}; \theta^{(t)}) \geq Q(\theta^{(t)}; \theta^{(t)})$ , y  $H(\theta^{(t)}; \theta^{(t)}) \geq H(\theta^{(t+1)}; \theta^{(t)})$ , lo cual se deduce de la desigualdad de Jensen y la concavidad de la función logarítmica, tenemos que  $\log L(\theta|X_{\text{obs}})$  se va incrementando en cada iteración con lo que se converge hacia el EMV de  $\theta$ .

En McLachlan y Krishnan (1996) o Little y Rubin (2002), pueden encontrarse resultados teóricos y condiciones acerca de la convergencia del algoritmo. Un criterio de convergencia habitual en la práctica consiste en detener el proceso cuando la diferencia entre dos estimaciones sucesivas de  $\theta$  sea suficientemente pequeña.

Es sencillo emplear el algoritmo EM como método de imputación. Una vez que se ha producido la convergencia, basta con dar un nuevo paso E y obtener las esperanzas matemáticas de los valores no observados condicionadas a los valores observados dado el EMV del vector de parámetros  $\theta$ .

Para profundizar en los detalles, la base teórica y extensiones del algoritmo EM nos remitimos a la monografía de McLachlan y Krishnan (1996).

- g) **Redes Neuronales:** son sistemas de información procesados, que reconocen patrones de los datos sin algún valor perdido para aplicarlo a la data a imputar. Estas redes son más usadas para variables cualitativas que cuantitativas, siendo más adecuadas cuando la distribución es no lineal. No es aconsejable cuando hay registros atípicos que distorsionan la red. Son costosos y requieren de capacitación del analista así como de “software” adecuado.
- h) **Modelos de series de tiempo:** se asume que la data perdida ocurre de tal forma, y en tal sistema, que el problema se reduce a una situación, en la cual, hay una serie de tiempo, donde una(s) serie(s) de observaciones están perdidas, haciendo óptimo el uso de interrelaciones entre sucesivas observaciones en cada serie de tiempo, mediante el uso de un modelo adecuado para estas series.
- i)
- 3. **Técnicas aleatorias o estocásticas:** son aquellas que cuando se repite el método de imputación bajo las mismas condiciones para una unidad, producen resultados diferentes.
  - a) **Imputación aleatoria de un caso seleccionado:** para cada caso con una celda faltante, se selecciona un donante aleatoriamente para ser asignado al dato faltante.
  - b) **Imputación aleatoria de un caso seleccionado entre clases:** se realiza de igual forma que para el caso (k) pero se lleva a cabo dentro de clases previamente creadas.
  - c) **Imputación secuencial Hot-Deck:** cada caso es procesado secuencialmente. Si el primer registro tiene un dato faltante, este es reemplazado por un valor inicial para imputar, pudiendo ser obtenido de información externa. Si el valor no está perdido, éste será al valor inicial y es usado para imputar el subsiguiente dato faltante. Entre las desventajas se encuentra que cuando el primer registro está perdido, se necesita de un valor inicial, (generalmente

obtenido de manera aleatoria), además cuando se necesitan imputar muchos registros se tiende a emplear el mismo registro donante, llevando esto a su vez la pérdida de precisión en las estimaciones.

d) **Imputación jerárquica Hot-Deck:** similar al método secuencial anterior. En esta se organizan dentro de clases haciendo uso de variables auxiliares en forma de una estructura jerárquica. Si el donante no es encontrado en un nivel de clasificación, las clases pueden ser colapsadas en grupos más anchos hasta que el donante sea encontrado.

e) **Imputación por regresión aleatoria o estocástica:** se hace primero un procedimiento de regresión (e), luego un término residual es adicionado para imputar los valores de  $y$ . Este término de error puede ser obtenido de diferentes maneras, una de ellas es a través de los residuos del modelo de regresión, generado con registros completos, eligiendo uno de éstos residuos aleatoriamente.

Al imputar mediante regresión se está reemplazando el valor perdido por una media condicionada, por lo que, como destacábamos en el caso de imputación mediante la media, se tiende sistemáticamente a subestimar la variabilidad. Una sencilla alternativa para atenuar este efecto consiste en añadir al valor predicho por la regresión una perturbación aleatoria, con lo que se obtiene una realización de la distribución predictiva de los valores perdidos condicionada a los valores observados. Esto es, en vez de imputar mediante [4.1], utilizar:

$$x_{li}^{\wedge} = \beta_{0obs}^{\wedge} + \sum_{j \neq i} \beta_{jobs}^{\wedge} x_{lj} + \varepsilon_{li} \quad [4.3]$$

donde  $\varepsilon_{li} \sim N(0, \sigma_{resid}^2)$ , siendo  $\sigma_{resid}^2$  la varianza residual de la regresión de  $X_i$  sobre  $X_j, \forall j \neq i$ .

f) **Imputación por regresión logística:** similar a la técnica anterior, pero para imputar variables binarias.

Molenberghs, 2004, Uranga, 2012 y Mallinckrodt, 2012 se refieren a la imputación de datos incompletos para el casos de datos longitudinales.

## 7. MÉTODOS DE IMPUTACIÓN SIMPLE

Los métodos de imputación pretenden solucionar el problema de los datos faltantes sustituyendo los mismos por valores estimados a partir de la información suministrada por la muestra. Con ello se consigue una matriz completa sobre la que realizar los análisis, salvando algunos de los problemas mencionados en la sección 3. Es precisamente la forma de estimar o predecir los valores perdidos lo que diferenciará unos métodos de otros. Nos centraremos en algunos de los más utilizados, concretamente en aquellos de aplicabilidad general y basada en un modelo estadístico explícito: imputación mediante la media, imputación mediante regresión e imputación mediante regresión estocástica.

## 8. MÉTODOS BASADOS EN VEROSIMILITUDES

En esta sección nos centramos en métodos que se basan en funciones de verosimilitud, y que son por lo tanto métodos bajo los que subyace un modelo probabilístico. A continuación revisaremos el marco formal, debido a Rubin (1976), que da soporte a estos métodos y que se mantiene en la actualidad.

### 8.1. Un marco formal para la inferencia basada en muestras incompletas

Consideremos un fenómeno multivariante real cuyo comportamiento viene descrito por un vector aleatorio  $k$ -dimensional  $X = (X_1, \dots, X_k) \in R^k$  con distribución de probabilidad  $P[X; \theta]$ , siendo  $\theta$  el vector de parámetros desconocidos.

Cuando se dispone de una muestra completa de  $X$ , una amplia clase de métodos de inferencia se justifican en la interpretación de  $P[X; \theta]$  como una función de verosimilitud que resume la evidencia que sobre  $\theta$  hay en los datos. Pero en presencia de valores perdidos sólo disponemos de  $X_{obs}$  cuya distribución se obtiene como

$$P[X_{obs}; \theta] = \int P[X; \theta] dX_{per} \quad [5.1]$$

Si pretendemos hacer inferencia sobre  $\theta$  a partir de la parte observada, es necesario comprobar que [5.1] es una verosimilitud adecuada. Rubin (1976) identifica las condiciones para que así sea, estableciendo que basta con que se verifique la hipótesis MAR como comprobamos a continuación.

Según se ha formalizado el problema de las muestras incompletas, es necesario especificar un modelo para  $X$ ,  $P[X; \theta]$ , y un modelo para la no respuesta,  $P[R|X_{obs}, X_{per}, \xi]$ . Mediante el producto  $P[R|X_{obs}, X_{per}, \xi] P[X; \theta]$  obtenemos la distribución conjunta  $P[X, R; \theta, \xi]$ . La verosimilitud basada en la parte observada puede expresarse como

$$P[X_{obs}, R; \theta, \xi] = \int P[X, R; \theta, \xi] dX_{per} = \int P[R|X_{obs}, X_{per}, \xi] P[X; \theta] dX_{per} \quad [5.2]$$

Bajo el supuesto MAR, [5.2] queda como

$$P[X_{obs}, R; \theta, \xi] = P[R|X_{obs}, \xi] \int P[X; \theta] dX_{per} = P[R|X_{obs}, \xi] P[X_{obs}; \theta] \quad [5.3]$$

De modo que la verosimilitud [5.2] bajo el supuesto MAR queda factorizada en dos partes, una relativa al vector  $\theta$  y otra relativa al vector  $\xi$ . Si además  $\theta$  y  $\xi$  son distinguibles(3), entonces las inferencias sobre  $\theta$  basadas en verosimilitudes no se verán afectadas por  $P[R|X_{obs}, \xi]$ , esto es, el mecanismo de no respuesta puede ser ignorado y la función de verosimilitud  $L$  de  $\theta$  será  $L(\theta | X_{obs}) \propto P[X_{obs}; \theta]$ . Este resultado pone de relieve que bajo ignorabilidad podemos realizar inferencias sobre el vector de parámetros  $\theta$  de la distribución de  $X$  a partir de la verosimilitud  $L(\theta | X_{obs})$ .

Por otro lado, desde una perspectiva bayesiana, todas las inferencias se basan en la distribución de probabilidad a posteriori de los parámetros desconocidos, que puede escribirse utilizando el Teorema de Bayes como

$$P[\theta, \xi | R, X_{obs}] = P[R, X_{obs} | \theta, \xi] \phi(\theta, \xi) / \iint P[R, X_{obs} | \theta, \xi] \phi(\theta, \xi) d\theta d\xi \quad [5.4]$$

donde  $\phi$  denota la distribución a priori de  $(\theta, \xi)$ . Bajo el supuesto MAR, podemos sustituir [5.3] en [5.4], obteniendo que  $P[\theta, \xi | R, X_{obs}]$  es proporcional a  $P[R | X_{obs}, \xi] P[X_{obs} | \theta] \phi(\theta, \xi)$ . Si además  $\theta$  y  $\xi$  son distinguibles, entonces la distribución marginal a posteriori de  $\theta$  queda como

$$P[\theta | X_{obs}, R] = \int P[\theta, \xi | R, X_{obs}] d\xi \propto P[X_{obs} | \theta] \phi_{\theta}(\theta) \int P[R | X_{obs}, \xi] \phi_{\xi}(\xi) d\xi \propto L(\theta | X_{obs}) \phi_{\theta}(\theta)$$

Por lo tanto, bajo la hipótesis de ignorabilidad, toda la información sobre  $\theta$  se recoge en la distribución a posteriori que ignora el mecanismo de no respuesta observado,  $P[\theta | X_{obs}] \propto L(\theta | X_{obs}) \phi_{\theta}(\theta)$ . La necesidad de especificar una distribución a priori para los parámetros puede resultar algo subjetivo o artificial desde un punto de vista no bayesiano. Sin embargo, como se destaca en Gelman et al (1995), conforme aumenta el tamaño muestral la verosimilitud domina a la distribución a priori, y las respuestas bayesianas y verosímiles tienden a converger. Además, en la mayoría de los problemas, la especificación de una distribución a priori no informativa que refleje el estado de ignorancia sobre los parámetros resulta adecuada, y esta es la opción tomada para el presente trabajo.

## 8.2. El método de imputación múltiple

Mediante imputación múltiple se reemplaza cada valor perdido por un conjunto de valores simulados con el fin de incorporar a la estimación la incertidumbre debida a la presencia de datos faltantes. La referencia básica sobre imputación múltiple es Rubin (1987), aunque podemos encontrar una variedad de trabajos relevantes como por ejemplo Rubin (1996), Schafer (1997), Little y Rubin (2002) o Zhang (2003).

Esta metodología ha permanecido durante algunos años en un segundo plano por su limitada aplicabilidad, debida principalmente a la inexistencia de herramientas computacionales adecuadas para poder crear las imputaciones. El desarrollo tecnológico de las últimas décadas ha permitido la implementación de algoritmos y procedimientos de cálculo computacionalmente intensivos necesarios para dar solución a problemas intratables analíticamente. En concreto, durante la década de los 90 se han popularizado los algoritmos MCMC (Markov Chain Monte Carlo) (véase p. ej. Gilks, Richardson y Spiegelhalter, 1996 ó Palarea, 2003) que permiten una modelización estadística más compleja al tiempo que realista.

En la práctica este supuesto implica que  $\varepsilon$  proporciona poca información sobre  $\theta$ , y viceversa. Este tipo de algoritmos también han encontrado su aplicación en el ámbito de los datos faltantes, en concreto, su incorporación al contexto de la imputación múltiple (Schafer, 1997) ha convertido este procedimiento en un destacado método para el análisis de datos incompletos.

El método consta de tres etapas:

**1. IMPUTACIÓN:** en esta etapa cada valor perdido se reemplaza por un conjunto de  $m$  valores simulados a partir de la distribución predictiva de  $X_{\text{per}}$  dado un modelo de probabilidad para  $X$  y una distribución a priori para  $\theta$ . Dicha distribución  $P[X_{\text{per}} | X_{\text{obs}}]$  puede obtenerse como

$$P[X_{\text{per}} | X_{\text{obs}}] = \int P[X_{\text{per}} | X_{\text{obs}}, \theta] P[\theta | X_{\text{obs}}] d\theta \quad [5.5]$$

En [5.5] se refleja tanto la incertidumbre sobre  $X_{\text{per}}$  dado el vector de parámetros  $\theta$ , como la propia incertidumbre asociada a  $\theta$ . Destacar que en las imputaciones así generadas no interviene  $R$ , se elude el mecanismo de no respuesta. En consecuencia, como estudiamos en la subsección 5.1, esta forma de proceder sólo será teóricamente apropiada bajo la hipótesis MAR.

En general,  $P[\theta | X_{\text{obs}}]$  y los cálculos donde interviene resultan intratables analíticamente, especialmente en contextos multidimensionales. Es aquí donde intervienen de forma natural los algoritmos MCMC dentro de esta metodología. En concreto, se utiliza el algoritmo de Aumento de Datos (Tanner y Wong, 1987) adaptado a este contexto para simular valores de [5.5] con los que realizar las imputaciones. El algoritmo de Aumento de Datos responde al siguiente esquema iterativo:

Dado  $\theta^n$   
 Repetir  
 Generar  $X_{\text{per}}^{n+1} \sim P[X_{\text{per}} | X_{\text{obs}}, \theta^n]$   
 $\theta^{n+1} \sim P[\theta | X_{\text{obs}}, X_{\text{per}}^{n+1}]$   
 Incrementar  $n$

Puede demostrarse (véase p. ej. Robert y Casella, 1999) que este algoritmo genera una cadena de Markov que converge hacia la distribución  $P[X_{\text{per}} | X_{\text{obs}}]$  y, en consecuencia, tras un número suficientemente grande de iteraciones se estarán generando valores de  $X_{\text{per}}$  condicionados a  $X_{\text{obs}}$  que se emplearán para reemplazar los valores perdidos.

**2. ANÁLISIS:** en esta etapa se aplica el análisis de datos deseado (regresión, análisis discriminante,...) sobre cada una de las matrices imputadas y se almacenan los resultados.

**3. COMBINACIÓN:** Finalmente, las estimaciones resultantes se combinan para obtener una única estimación global.

Sea  $Q$  un parámetro poblacional unidimensional de interés (ej. una media, una varianza, un coeficiente  $\beta$  de una regresión,...), sea  $\hat{Q}$  su estimación puntual si no hubiera datos perdidos, denotando con  $U$  la varianza estimada de  $\hat{Q}$ . Tras analizar los datos imputados tenemos  $m$  estimaciones  $\{\hat{Q}_1, \dots, \hat{Q}_m\}$  con varianzas estimadas asociadas  $\{U_1, \dots, U_m\}$ . Se derivan las siguientes reglas (Rubin, 1987) para combinar las  $m$  estimaciones: la estimación puntual para  $Q$  basada en imputación múltiple,  $Q_m$ , y su varianza asociada,  $T_m$ , vendrán dadas por

$$Q_m = 1/m \sum_{i=1}^m \hat{Q}_i \quad \text{y} \quad T_m = U_m + (1+1/m) B_m \quad [5.6]$$

donde  $U_m = \sum_{i=1}^m U_i/m$  representa la variabilidad intra-imputaciones y  $B_m = \sum_{i=1}^m (\hat{Q}_i - Q)(\hat{Q}_i - Q)/(m-1)$  la variabilidad entre-imputaciones. Sin datos perdidos, las estimaciones  $\hat{Q}_i$ ,  $i=1, \dots, m$ , serían idénticas, con lo que  $B = 0$  y  $T = U$ .

Pueden encontrarse resultados ampliados para combinar ratios de verosimilitudes o p-valores en Meng y Rubin (1992) y Schafer (1997), ajustes para  $m$  pequeña en Rubin (1987), y análisis sobre el comportamiento asintótico de los estimadores en Rubin (1996) y Robins y Wang (2000).

## 9. CÓMO SELECCIONAR LA TECNICA ADECUADA DE IMPUTACIÓN

---

Seleccionar un método de imputación adecuado es una decisión de gran importancia, ya que para un conjunto de datos determinado, algunas técnicas de imputación podrían dar mejores aproximaciones a los valores verdaderos que otras. Para la selección de la técnica de imputación adecuada, no hay reglas específicas, dependerá entonces del tipo del conjunto de datos, tamaños del archivo, tipo de no respuesta, patrón de pérdida de respuesta, de los objetivos de la investigación, características específicas de la población, características generales de la organización del estudio, software disponible (Entilge 1996), importancia de los valores agregados o de los valores puntuales (microdato), distribuciones de frecuencias de cada variable, marginal o conjunta, etc. Hay que tomar en cuenta que muchas veces la técnica de imputación seleccionada puede ser adecuada para algunas variables pero para otras no y será decisión del investigador seleccionar la técnica que menos afecte las estimaciones de las variables.

Fellegi y Holt (1971), plantean que: “La técnica de imputación seleccionada debe superar las reglas de validación, cambiando lo menos posible los registros, manteniendo la frecuencia de la estructura de la data”. Goicoechea (2002), resume los criterios a tomar en consideración para seleccionar la técnica adecuada para imputar.

1. **Tipo de variable a imputar:** si es continua, tomar en cuenta el intervalo para la cual se define, y si es cualitativa, tanto nominal como ordinal, las categorías de las variables.
2. **Parámetros que se desean estimar:** si deseamos conocer sólo valores agregados como la media y el total, se pueden aplicar métodos sencillos como imputación con la media o moda, sin embargo, puede haber subestimación de la varianza. En caso de que se requiera mantener la distribución de frecuencia de la variable y las asociaciones entre las distintas variables, se deben emplear métodos más elaborados aplicando imputación de todas las variables faltantes del registro.
3. **Tasas de no respuesta y exactitud necesaria:** cuando el porcentaje de no respuesta es alto en una base de datos, se considera que no hay confiabilidad en los resultados que se obtengan con el análisis de esta base.
4. **Información auxiliar disponible:** es bueno hacer uso de la información auxiliar disponible, ya que con ella podemos deducir información de los valores ausentes de una variable o hallar grupos homogéneos respecto a una variable auxiliar que se encuentre altamente correlacionada con la variable a imputar, y de esta manera encontrar un donante adecuado que sea similar al registro receptor.

Según Goicoechea (2002) los pasos que se llevan a cabo para realizar imputación son:

**Paso 1:** una vez que se dispone de un archivo con datos faltantes, se recopila y valida toda la información auxiliar disponible que pueda ser de ayuda para la imputación.

**Paso 2:** se estudia el patrón de pérdida de respuesta. Posteriormente se observa si hay un gran número de registros que simultáneamente tienen no respuesta en un conjunto de variables.

**Paso 3:** se seleccionan varios métodos de imputación posibles y se contrastan los resultados.

**Paso 4:** se calculan las varianzas para los distintos métodos de imputación seleccionados con el objetivo de obtener estimaciones con el mínimo sesgo y la mejor precisión.

**Paso 5:** se concluye a partir de los resultados obtenidos

## 10. EVALUACIÓN DE LAS TÉCNICAS

Existen muchas maneras de llevar a cabo la evaluación y validación de una o más técnicas de imputación que hayan sido aplicadas a una situación particular, la más común de ellas es el uso de simulación, que consiste en aplicar pérdidas artificiales de datos a una base original, a esta base de datos con pérdidas artificiales es llamada base de datos disponible. Posteriormente a esta base de datos se le suprimirán todos los registros que tengan al menos un dato faltante, obteniendo finalmente una base de datos más pequeña, pero completa. Burzykowski, 2010 junto a un grupo de expertos discuten los puntos a considerar y Robaina, 2013 abordan algunas consideraciones para abordar mejor los datos incompletos en ensayos clínicos y Robaina, 2014 hacen recomendaciones para la prevención y el tratamiento de los datos incompletos en ensayos clínicos.

---

Asumiendo que las relaciones y características de los datos perdidos es igual a la de los datos presentes (casos ignorables), se procede a aplicarle a la base de datos completa el mismo patrón de pérdida aplicado a la base de datos original.

Ante tales pérdidas, se aplica la técnica de imputación, y ésta se evalúa midiendo en ambas bases (completa e imputada) la desviación estándar, el sesgo de las estimaciones, las distribuciones de frecuencia conjuntas y marginales, entre otras magnitudes de los efectos de imputación, mediante el uso de tablas de contingencia, de distribuciones de frecuencia o prueba de hipótesis. Para ello se hará uso de estadísticos de conservación de la distribución, como el estadístico T para variables continuas y el estadístico Chi-cuadrado para variables categóricas.

En la evaluación de las diferentes técnicas también se considerará el porcentaje de pérdida de los datos, para saber hasta qué porcentaje la base de datos es confiable, observando ciertos criterios, tales como, cuántas veces ha sido usado un donante, cuantos intentos fueron requeridos para completar un registro, entre otros.

Goicoechea (2002) propone una serie de medidas deseables, al evaluar las técnicas de imputación, estas medidas son:

1. **Precisión en la predicción:** el valor imputado debe ser lo más cercano al valor verdadero. Para ello se hace uso de tablas de contingencia.
2. **Precisión en la distribución:** mantener en lo posible las distribuciones marginales y conjuntas.
3. **Precisión en la estimación:** producir parámetros insesgados e inferencias eficientes de la distribución de los valores reales.
4. **Imputación plausible:** valores aceptables al aplicarles el proceso de edición.

## 11. CONCLUSIONES

En este trabajo se plantea y ejecuta un experimento de simulación con el fin de contrastar el rendimiento de las principales estrategias frente a problemas de inferencia estadística con datos faltantes bajo distintos mecanismos de no respuesta. Se ha comprobado que, globalmente, los resultados van empeorando desde la situación MCAR a la NMAR, y van mejorando desde el análisis de casos completos hasta la imputación múltiple o el algoritmo EM. En general, los métodos para datos faltantes estudiados provocan la aparición de sesgos, subestimación de las varianzas, sobreestimación de las correlaciones y subcobertura de los intervalos de confianza. Estos efectos se acentúan de forma creciente al pasar a las situaciones MAR y NMAR. Es entonces cuando los métodos basados en verosimilitudes muestran un comportamiento más estable y robusto, especialmente el método de imputación múltiple.

El análisis de casos completos o casos disponibles, y la imputación mediante la media son procedimientos nada recomendables para la inferencia, con un comportamiento muy inestable y unas estimaciones sólo aceptables bajo la hipótesis MCAR, especialmente en lo que se refiere a la cobertura de los intervalos de confianza. Para variables con alto coeficiente de variación, las estimaciones de los distintos métodos resultan ser más dispares y las cotas de error mayores. Por otro lado, cuando los valores perdidos se sitúan principalmente en el centro de la distribución, sus efectos nocivos son bastante más débiles.

En resumen, ante un problema de datos faltantes hay que tener en cuenta lo que se pretende estimar, la cantidad de datos faltantes, el conocimiento sobre cómo y dónde faltan los datos, las características de las variables, y el esfuerzo que se esté dispuesto a realizar para obtener mejores estimaciones. En este trabajo se pone de manifiesto que las soluciones heurísticas habituales son garantía de una inferencia pobre e imprecisa y que los métodos basados en verosimilitudes ofrecen la mejor alternativa. En determinadas situaciones, este tipo de métodos pueden suponer para el analista un esfuerzo y un gasto computacional que no quede compensado por sus bondades para la inferencia. En futuros trabajos se profundizará en el análisis de estos procedimientos desde una perspectiva multivariante, y se abordará el estudio de métodos y modelos no sustentados en la hipótesis MAR con el fin de modelizar situaciones NMAR de forma más adecuada.



---

## REFERENCIAS

1. Agencia Europea de Medicamentos (2009): Guideline on missing data in confirmatory clinical trials. (CPMP/EWP/1776/99); 2009. Disponible en: [ww.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf](http://www.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf).
2. Agencia Europea de Medicamentos (2001): Points to consider on missing data; 2001. Disponible en: [www.ema.europa.eu/docs/en\\_GB/document.../09/WC500003641.pdf](http://www.ema.europa.eu/docs/en_GB/document.../09/WC500003641.pdf).
3. Andrew B. y Selamat A, (2012): Systematic Literature Review of Missing Data Imputation Techniques for Effort Prediction. IPCSIT vol.45 (2012) IACSIT Press, Singapore
4. Bacallao Guerra J y Bacallao Gallestey J, (2010): IMPUTACIÓN MÚLTIPLE EN VARIABLES CATEGÓRICAS USANDO DATA AUGMENTATION Y ÁRBOLES DE CLASIFICACIÓN. Revista Investigación Operacional. Vol. 31, No., 2, 133-139, 2010.
5. BATERHAM M.J., TAPSELL L.C. y CHARLTON K.E. (2013): 'Analyzing weight loss intervention studies with missing data: which method should be used?', Nutrition, vol. 29, no. 7-8, pp. 1024-1029.
6. Bouza Herrera C.N. y Al-Omari A.I. (2011): Ranked Set Estimation with Imputation of the Missing Observations: The Median Estimator. Revista Investigación Operacional., 32 , 1, 30-37.
7. **Burzykowski T.**, Carpenter J., Coens C., Evans D. et al. (2010) : Missing data: Discussion points from the PSI missing data expert group. Pharmaceut. Statist. 9: 288–297.
8. Carpenter Jr, Kenward M. G. (2008): Missing data in randomized controlled trials—a practical guide. Birmingham: National Institute for Health Research, Publication RM03/JH17/MK.. Disponible en: [http://www.haps.bham.ac.uk/publichealth/methodology/projects/RM03\\_JH17\\_MK.shtml](http://www.haps.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml)
9. Casella G.. y Berger R.L. (2001): Statistical Inference, 2ª edición, Duxbury.
10. Dempster A.P., Laird N.M. y Rubin D.B. (1977): Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion): J. of the Royal Statistical Society, 39, 1-38.
11. Donders A.R.T., Van Der Heijden J.M.G., Stijnen T., Moons K.G.M. (2006): Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology 59. 1087-1091.
12. Dong Y. y Joanne C.Y. (2013): Principled missing data methods for researchers Springer Plus 2013, 2:222
13. Eltinge J. (1996): Discussion of Imputation Papers. Recuperado 05, noviembre del 2004 en: [http://www.amstat.org/sections/srms/Proceedings/papers/1996\\_049.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1996_049.pdf).
14. Fellegi I. y Holt D. (1976): A systematic approach to automatic edit and imputation. Journal of the American Statistical Association. 71, 353. 17-35.
15. Fleming T. R (2011):. Addressing Missing Data in Clinical Trials. Ann Intern Med.,154,113-7.
16. Gelman, A., Carlin, J.B., Stern, H.S. y Rubin, D.B. (1995): Bayesian Data Analysis, Chapman y Hall.
17. Geng Z. y Li K. (2003): Factorization of posteriors and partial imputation algorithm for graphical models with missing data. Statistics and probability letters. 64, 369-379.
18. Gilks W.R., Richardson S. y Spiegelhalter D.J., (1996): Markov Chain Monte Carlo in Practice, Chapman y Hall.
19. Gleason T. y Staelin R. (1975): A proposal for handling missing data. Psychometrika. Vol 40, 2. 229-252.
20. Goicoechea P. (2002): Imputación basada en árboles de clasificación. EUSTAT.
21. Gómez J., Palarea J. y Martín J.A. (2006): Métodos de inferencia estadística con datos faltantes. Estudio de simulación sobre los efectos en las estimaciones. Estadística Española. Vol. 48, Núm. 162, págs. 241 a 270.
22. Government Statistical Service. (1996): Report of the Task Force on Imputation. GSS Methodology Serie No. 3. Reino Unido.
23. Hemel, J. et al. (1987): Stepwise deletion: a technique for missing data handling in multivariate analysis. Analytical Chemical Acta. 193 255-268.
24. Higgins Jpt, White Ir y Wood AM.(2008): Imputation methods for missing outcome data in meta-analysis of clinical trials. Clinical Trials, 5: 225–239

- 
25. International Conference on Harmonisation (1998): Statistical Principles for Clinical Trials. Tripartite Harmonised Guideline, ICH 1998. Disponible en: <http://www.ich.org/cache/compo/276-254-1.html>.
  26. Janssens M., Molenberghs G. y Kerstensa R (2012): Handling of missing data in long-term clinical trials: a case study. *Pharmaceut. Statist.*, 11 442–448.
  27. Kalton G y Kasprzyk D. (1982): Imputing for Missing Survey Responses, American Statistical Association. Proceeding of the Section on Survey Research Methods.
  28. Karahalios et al.(2012): A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures *BMC Medical Research Methodology* 2012, 12:96
  29. Koikkalainen P. (2002): Neural Network for editing and imputation. University of Jyväskylä Finland. Recuperado 05, noviembre del 2004 en: <http://erin.mit.jyu.fi/dataclean/abstracs/node33.html#SECTION00093000000000000000>.
  30. Lachin J.M. (2000): Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167-89.
  31. Little, R. y Rubin, D. (1987): *Statistical Analysis with Missing Data*. Series in Probability and Mathematical Statistics. John Wiley y Sons, Inc. New York.
  32. Little, R.J.A. Y Rubin, D.B. (2002): *Statistical Analysis with Missing Data*, Wiley y Sons.
  33. Little R. J. A., D'agostino R, Cohen MI, Dickersin K, Emerson Ss, Farrar JT, et al. (2012): The Prevention and Treatment of Missing Data in Clinical Trials. *N Engl J Med* 2012;367:1355-60.
  34. Lohr, S. H. (1999): *Muestro: Diseño y Análisis*. Editorial Thomson.
  35. Mallinckrodt C.H., Lin Q., and Molenbergh M. (2012): A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. *Pharmaceut. Statist.* John Wiley y Sons, Ltd. DOI: 10.1002/pst.1547
  36. Marshall A., Altman D.G., Holder R.L and Royston P. (2009): Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*, doi:10.1186/1471-2288-9-57
  37. Marshall A., Altman D.G., Royston P., Holder. R.L. (2010): Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Medical Research Methodology*, 2010 10:7
  38. McLachlan, G.J. Y Krishnan, T., (1996): *The EM algorithm and extensions*, Wiley.
  39. Meng, X.L. Y Rubin, D.B., (1992): Performing likelihood ratio tests with multiply imputed data sets, *Biometrika*, 79, 103-111.
  40. Mengersen, K. Y Robert, C.P., (1999): «MCMC convergence diagnostics: a review», en *Bayesian Statistics*, eds. J. Berger, J. Bernardo, A.P. Dawid y A.F.M. Smith, Oxford Sciences Publications.
  41. Mesa, D. (2004): *Imputación y Árboles de Decisión*. Caracas, Venezuela. Guía práctica. Postgrado en Estadística, Universidad Central de Venezuela.
  42. Mesa, D. Tsai, P. y Chambers, R. (2000): *Using Tree-Based Models For Missing Data Imputation: An Evaluation Using Uk Census Data*, Reporte Técnico. Proyecto AUTIMP. Recuperado 20, septiembre del 2004 en: [http://www.cbs.nl/en/service/autimp/CART-Dutch%20Data-\(AUTIMP\):pdf](http://www.cbs.nl/en/service/autimp/CART-Dutch%20Data-(AUTIMP):pdf).
  43. Molenberghs G., Kenward M.G. y Goetghebeur E., (2001): Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case, *Applied Statistics*, 50 (1): 15-29.
  44. Molenberghs G., Thijs H., Jansen I., Beunckens C. (2004) Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5,3, pp. 445–464. doi: 10.1093/biostatistics/kxh001
  45. Molenberghs G, Kenward M. (2007): *Missing Data in Clinical Studies*. New York: Springer-Verlag;
  46. National Research Council. (2010): *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies; 2010.
  47. Navarro J.B., Losilla J.M. (2000): Análisis de datos faltantes mediante redes neuronales artificiales. *Psicothema* 2000. Vol. 12, nº 3, pp. 503-510.

- 
48. O'Neill RT, Temple R. (2012): The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical Pharmacology y Therapeutics*, 91:550-4.
  49. Palarea, J., (2003): Algoritmos Monte Carlo basados en cadenas de Markov. Aplicación a la inferencia mediante imputación múltiple, Trabajo de investigación de tercer ciclo, Universidad de Murcia.
  50. Platek, R. (1986): Metodología y Tratamiento de la no respuesta. Seminario Internacional de Estadística en EUSKADI. Cuaderno 10.
  51. Robaina M., Uranga R., Fors M.M., Viada C. (2013): Algunas consideraciones para abordar mejor los datos incompletos en ensayos clínicos. *ICB 81*, diciembre de 2012-enero de 2013.
  52. Robaina M., Uranga R., Fors M.M., Viada C. (2014): Recomendaciones para la prevención y el tratamiento de los datos incompletos en ensayos clínicos. *VacciMonitor* 2014;23(1):32-36.
  53. Robert, C.P. y Casella, G., (1999): *Monte Carlo Statistical Methods*, Springer.
  54. Robins, J.M. y Wang, N., (2000): Inference for imputation estimators, *Biometrika*, 87, 113-124.
  55. Rubin D.B. (1996): Multiple imputation after 18+ years, *J. of the American Statistical Association*, 91, 473-489.
  56. Rubin D.B. (1983): Panel of incomplete data in sample surveys. En Madow, W.G., Olkin, I. y Rubin D.B. *Incomplete Data in Simple Surveys. Vol 2,12*, 123-145. Report and Case Studies. New York. Academic Press.
  57. Rubin D.B., (1976): Inference and missing data, *Biometrika*, 63, 581-592.
  58. Rubin D.B., (1987): *Multiple Imputation for Nonresponse in Surveys*, Wiley y Sons.
  59. Sande, I. (1982): Imputation in Surveys: Coping with reality. *The American Statistician*, Vol.36, 3:145-152.
  60. Schafer J.L. y Graham, J. W., (2002): Missing data: our view of the state of the art, *Psychological Methods*, 7, 147-177.
  61. Schafer, J.L., (1997): *Analysis of Incomplete Multivariate Data*, Chapman y Hall.
  62. Schlomer G.L., Bauman S., and Card N.A. (2010): Best Practices for Missing Data Management in Counseling Psychology. *Journal of Counseling Psychology*, Vol. 57, No. 1, 1-10.
  63. Serrat C., (2001): Study and validation of data structures with missing values. Application to survival analysis, doctoral thesis, Universitat Politècnica de Catalunya, Barcelona.
  64. Tanner M. y Wong W., (1987): The calculation of posterior distributions by data augmentation, *J. of the American Statistical Association*, 82, 528-550.
  65. Todeschini, R. (1990): Weighted k-nearest neighbour method for the calculation of missing values. *Chenometrics and Intelligent Laboratory Systems* 9.201-205.
  66. Troxel A.B., Ma, G. y Heitjan, D.F., (2004): An index of local sensitivity to nonignorability, *Statistica Sinica*, 14, 1221-1237.
  67. Uranga R. y Molenberghs G. (2012): Longitudinal conditional models with intermittent missingness: SAS code and applications, *Journal of Statistical Computation and Simulation*, DOI:10.1080/00949655.2012.725403
  68. Useche L. y Mesa D. (2006): Una Introducción A La Imputación De Valores Perdidos Terra. Vol. XXII, No. 31, , pp. 127-152.
  69. Vásquez, M. (1995): Aportación al Análisis Biplot: Un enfoque Algebraico. Tesis doctoral. Universidad de Salamanca. España.
  70. Vélez, R. Y García, A., (1993): *Principios de inferencia estadística*, UNED.
  71. Vergouwe Y., Royston P., Moons K.G.M., Altman D.G. (2010): Development and validation of a prediction model with missing predict data: a practical approach. *Journal of Clinical Epidemiology* 63 (2010) 205e214.
  72. White Ir, Carpenter J, Nicholas, HJ. (2012): Including all individuals is not enough: Lessons for intention-to-treat analysis. *Clin Trials* 2012;9:396-407.
  73. Wilks, S. (1932): Moments and distributions of estimates of population parameters from fragmentary simple, *Annals of Mathematical Statistics*, B, 163-195.
  74. Wood M. A, White R. I., Thompson G. S. (2004): Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*, 1,368-76.
  75. Zhang, P., (2003): Multiple imputation: theory and method, *International Statistical Review*, 71, 3, 581-592.

## APLICACIÓN DEL MODELO DE HORVITZ DE RESPUESTA ALEATORIZADA AL CONSUMO DE DROGAS

B. Cobo

Departamento de Estadística e Investigación Operativa, Universidad de Granada, España.

### ABSTRACT

A survey is a research method that is based on questioning a sample of individuals. The interest in sample surveys studies often focuses on sensitive or confidential aspects to the interviewees. Because of this, the typical problem that arises is social desirability, which is defined as the tendency of respondents to answer based on what is socially acceptable. For this reason, many respondents refuse to participate in the survey or provide false answers or conditioned answers, altering the accuracy and reliability of the estimations in a major way. Randomized Response Technique introduced by Warner is a possible solution for protecting the anonymity of the respondent and is used to reduce the risk of escape or no response to sensitive questions.

**KEY WORDS:** Randomized response, Sampling, Confidentiality, Horvitz model

### RESUMEN

Una encuesta es un método de investigación que se basa en el cuestionamiento de una muestra de individuos. El interés en estudios de encuestas por muestreo a menudo se centra en aspectos sensibles o confidenciales a los entrevistados. Debido a esto, el problema típico que se plantea es la deseabilidad social, que se define como la tendencia de los encuestados a responder en función de lo que es socialmente aceptable. Por esta razón, muchos de los encuestados se niegan a participar en la encuesta o proporcionar respuestas falsas o respuestas condicionadas, que alteran la exactitud y fiabilidad de las estimaciones de una manera importante. Las técnicas de respuesta aleatorizadas son una posible solución para proteger el anonimato del encuestado y reducir el riesgo de evasión o falta de respuesta a las preguntas sensibles.

## 1. INTRODUCCIÓN

Las técnicas de respuesta aleatoria fueron introducidas por Warner en 1965 con el objetivo de proteger la privacidad de los encuestados en el estudio de temas sensibles y de este modo evitar el sesgo de respuesta.

Esta técnica consiste en la utilización de un mecanismo aleatorio por medio del cual se selecciona una de dos preguntas complementarias: ¿pertenece al grupo con la característica  $A$ ? o ¿pertenece al grupo que no tiene la característica  $A$ ?, donde  $A$  es la característica sensible de interés. El entrevistado contestará sí o no y el entrevistador no tiene la posibilidad de saber qué pregunta contestó el entrevistado.

Mediante esta técnica se obtiene una cooperación mayor por parte de los encuestados y reduce su motivación para falsear las respuestas dando lugar a estimadores que son más precisos en comparación a respuesta directa.

Se puede ver la mejora en gran cantidad de encuestas, como por ejemplo, una encuesta de respuesta aleatoria en el comportamiento de los estudiantes universitarios reveló un aumento sustancial de la copia con medios tecnológicos, bulling y consumo de drogas en los campus universitarios.

En este capítulo se verá en profundidad el modelo de Horvitz cuando la proporción de personas que tienen la característica inocua es desconocida, concretamente su definición, el tratamiento con software, y la aplicación a un caso real llevado a cabo en una universidad con el objetivo de estimar el consumo de drogas.

## 2. MODELO HORVITZ CUANDO LA PROPORCIÓN DE PERSONAS QUE TIENEN LA CARACTERÍSTICA INOCUA ES DESCONOCIDA

Horvitz et al., (1967) seguido por Greenberg et al., (1969) recomendaron una alternativa a la técnica de respuesta aleatoria de Warner (1965).

Ellos sugieren que al igual que la característica  $A$  es sensible,  $A^c$  también lo puede ser. Así, las personas pueden estar incómodas al dar a conocer sus características  $A$  y  $A^c$ , a pesar de que mediante el procedimiento aleatorio no divulgan al investigador el resultado observado en la técnica de respuesta aleatoria.

Junto con el atributo sensible bajo investigación  $A$ , conceptualizan uno aparentemente no relacionado, por ejemplo,  $B$ . Por lo tanto, mientras tratan  $y$  para denotar la pertenencia a  $A$ , se introduce otra variable  $x$  para denotar la pertenencia a  $B$ , en lugar de su complementario  $A^c$ .

Por tanto,

$$y_i = \begin{cases} 1, & \text{si } i \text{ tiene } A \\ 0, & \text{si } i \text{ tiene } A^c \end{cases}$$

y

$$x_i = \begin{cases} 1, & \text{si } i \text{ tiene } B \\ 0, & \text{si } i \text{ tiene } B^c \end{cases}$$

La proporción de personas que tienen la característica inocua puede ser conocida o desconocida. Aquí se va a abordar el caso en que  $\beta$  es desconocida (Chaudhuri, 2011), ya que es una situación más realista.

En el modelo Horvitz, cuando la proporción poblacional  $\beta$  es desconocida, se toman dos muestras independientes.

Se llenan dos cajas con un gran número de cartas similares excepto que en la primera caja se marcan una proporción  $p_1$  ( $0 < p_1 < 1$ ) de ellas con  $A$  y la proporción complementaria  $(1 - p_1)$  con  $B$ , mientras que en la segunda caja estas proporciones son  $p_2$  y  $(1 - p_2)$ , siendo  $p_2$  distinto de  $p_1$ .

Se elige una muestra y a cada persona muestreada se le pide que extraiga una carta aleatoriamente de la primera caja y repita esto independientemente con la segunda caja.

En el primer caso, debería dar la respuesta aleatoria como:

$$I_i = \begin{cases} 1, & \text{si el tipo de carta extraído coincide con la característica sensible } A \text{ o la inocua } B \\ 0, & \text{si no coincide} \end{cases}$$

y en el segundo caso, dará una respuesta aleatoria como

$$J_i = \begin{cases} 1, & \text{si hay coincidencia para la segunda caja} \\ 0, & \text{si no hay coincidencia para la segunda caja} \end{cases}$$

La variable transformada es

$$\eta_i = \frac{(1 - p_2)I_i - (1 - p_1)J_i}{p_1 - p_2}$$

y la varianza estimada es

$$\hat{V}(\eta_i) = \eta_i(\eta_i - 1)$$

Nuestro objetivo es estimar el total de la población  $Y = \sum_{i=1}^N y_i$  de la variable de interés  $y$ . Entonces, un estimador insesgado para la población total viene dado por

$$\hat{Y} = \sum_{i \in S} \frac{\eta_i}{\pi_i}$$

siendo  $\pi_i$  la probabilidad de inclusión de primer orden.

La varianza del estimador viene dada por

$$V(\hat{Y}) = \left[ \frac{1}{2} \sum_{i \neq j \in U} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_{i \in U} \frac{V(r_i)}{\pi_i} \right] = V_{HT} + \sum_{i \in U} \frac{V(r_i)}{\pi_i}$$

donde  $V(r_i)$  es la varianza de  $r_i$  bajo el mecanismo de aleatorización, y  $V_{HT}$  es la varianza del diseño del estimador de Horvitz-Thompson.

Esta varianza es estimada mediante:

$$\hat{V}(\hat{Y}) = \left[ \frac{1}{2} \sum_{i \neq j \in S} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left( \frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2 + \sum_{i \in S} \frac{\hat{V}(r_i)}{\pi_i} \right] = \hat{V}_{HT} + \sum_{i \in S} \frac{\hat{V}(r_i)}{\pi_i}$$

De manera similar, un estimador insesgado para la media de la población  $\bar{Y}$  es dado por

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i}$$

Y un estimador insesgado para su varianza es calculado como:

$$v(\hat{\bar{Y}}) = \frac{1}{N^2} \left[ \frac{1}{2} \sum_{i \neq j \in S} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left( \frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2 + \sum_{i \in S} \frac{\hat{V}(r_i)}{\pi_i} \right] = \frac{1}{N^2} \left[ \hat{V}_{HT} + \sum_{i \in S} \frac{\hat{V}(r_i)}{\pi_i} \right]$$

### 3. TRATAMIENTO CON SOFTWARE

El software estadístico R es uno de los más flexibles, potentes y profesionales que existen actualmente para realizar tareas estadísticas de todo tipo, desde las más elementales, hasta las más avanzadas.

Este software se compone de paquetes, que son un conjunto de funciones con una temática común.

El término R puede hacer referencia al entorno computacional para el análisis estadístico y la generación de gráficos y al lenguaje de programación propio que el entorno anterior incorpora.

En nuestro caso hemos creado el paquete RRTCS (Randomized Response Techniques for Complex Surveys) con el objetivo de calcular los estimadores puntuales y por intervalos de parámetros lineales con datos obtenidos de muestras complejas cuando utilizamos técnicas de aleatorización.

El paquete implementa estimadores y varianzas para métodos de respuesta aleatoria para variables cualitativas y cuantitativas. La varianza también puede ser estimada utilizando métodos de remuestreo. Además, incluye algunos conjuntos de datos de encuestas con métodos de aleatorización.

En el caso del modelo de Horvitz cuando la proporción de personas que tienen la característica inocua es desconocida, tiene implementada la función

```
HorvitzUB(I, J, p1, p2, pi, type=c("total", "mean"), cl, N=NULL, pij=NULL)
```

En la ayuda de esta función se explica cada uno de los parámetros:

- I primer vector de la variable observada; su longitud es igual a  $n$  (el tamaño de la muestra)
- J segundo vector de la variable observada; su longitud es igual a  $n$  (el tamaño de la muestra)
- p1 proporción de cartas marcadas con la característica sensible en la primera caja
- p2 proporción de cartas marcadas con la característica sensible en la segunda caja
- p<sub>i</sub> vector de probabilidades de inclusión de primer orden
- type tipo del estimador: total o mean
- cl nivel de confianza
- N tamaño de la población. Por defecto es NULL
- p<sub>i,j</sub> matriz de probabilidades de inclusión de segundo orden. Por defecto es NULL

En la descripción de la función se indica que calcula la estimación de la respuesta aleatoria, la estimación de la varianza y el intervalo de confianza. Además la función también puede devolver la variable transformada  $\eta$ .

En el apartado detalles viene explicado el modelo, e incluye las fórmulas de la variable transformada y la varianza de ésta.

También viene un ejemplo desarrollado, el cual utiliza uno de los conjuntos de datos que tiene el paquete, para que el usuario lo pueda reproducir en el momento que estime oportuno.

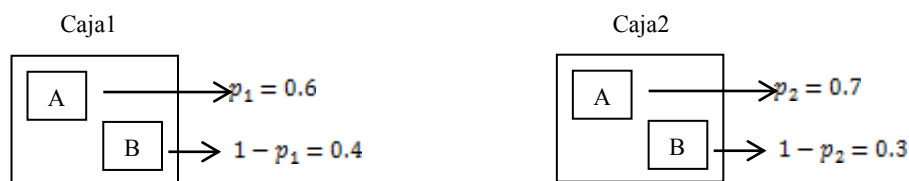
#### 4. APLICACIÓN A UNA ENCUESTA UNIVERSITARIA SOBRE EL CONSUMO DE DROGAS

Para la aplicación de la estrategia de respuesta aleatoria, concretamente para el modelo de Horvitz cuando la proporción de personas que tienen la característica inocua es desconocida, se ha realizado una encuesta a una muestra representativa de la población de estudiantes de la universidad para investigar el consumo de drogas.

El mecanismo para llevar a cabo la encuesta es el siguiente:

La técnica de respuesta aleatoria utilizada es el modelo de Horvitz con  $\beta$  desconocida con parámetros  $p_1 = 0.6$  y  $p_2 = 0.7$ , es decir,

- Se llena una primera caja con cartas marcadas en proporción 0.6 con **A** y 0.4 con **B**.
- La segunda caja en proporción 0.7 con **A** y 0.3 con **B**.



Dependiendo de la letra obtenida el encuestado contestará a una de las siguientes preguntas:

- La letra **A**, corresponde a la pregunta sensible: ¿Has consumido drogas alguna vez?
- La letra **B**, corresponde a la pregunta inocua: ¿Prefiere el fútbol al tenis?

Se elige una muestra, en este caso, la muestra es extraída mediante muestreo por conglomerados con probabilidades proporcionales al tamaño de la muestra, resultando una muestra de 188 observaciones de una población de 802 estudiantes divididos en cuatro conglomerados.

A cada persona muestreada se le pide que extraiga una carta aleatoriamente de la primera caja y responda

$$1 = \begin{cases} \text{si ha consumido drogas} \\ \text{si prefiere el fútbol al tenis} \end{cases} \quad 0 = \begin{cases} \text{si no ha consumido drogas} \\ \text{si no prefiere el fútbol al tenis} \end{cases}$$

y extraiga otra carta de la segunda caja. En este caso, al igual que en la primera caja, debe dar una respuesta aleatoria como

---

$$1 = \begin{cases} \text{si ha consumido drogas} \\ \text{si prefiere el fútbol al tenis} \end{cases}$$

$$0 = \begin{cases} \text{si no ha consumido drogas} \\ \text{si no prefiere el fútbol al tenis} \end{cases}$$



Una vez realizada la encuesta, se obtiene un conjunto de datos que contiene observaciones de una encuesta de respuesta aleatoria llevada a cabo en la universidad.

Este conjunto de datos se encuentra en el paquete RRTCS con el nombre `HorvitzUBData`. Concretamente el conjunto de datos es un data frame que contiene

- `ID`: identificador del estudiante
- `CL`: identificador del clonglomerado
- `I`: la primera respuesta aleatoria a la pregunta: ¿Ha consumido drogas alguna vez?
- `J`: la segunda respuesta aleatoria a la pregunta: ¿Ha consumido drogas alguna vez?
- `Pi`: probabilidad de inclusión de primer orden

Para obtener los resultados de esta encuesta se utiliza la función explicada en el apartado anterior `HorvitzUB(I, J, p1, p2, pi, type=c("total", "mean"), cl, N=NULL, pij=NULL)` donde para obtener los parámetros se tiene que recurrir al conjunto de datos `HorvitzUBData`.

Para poder utilizar esta función antes se tienen que ejecutar una serie de instrucciones:

En primer lugar se define el tamaño de la población, ya que se quiere estimar la media de universitarios que consumen drogas.

```
N=802
```

Se cargan los datos del conjunto de datos mediante la función `data`.

```
data(HorvitzUBData)
```

Con la función `with` se obtiene un nuevo data frame llamado `dat` que contiene solo las variables `I, J, Pi`, que son las que se utilizan posteriormente en la llamada a la función.

```
dat=with(HorvitzUBData, data.frame(I, J, Pi))
```

Se declaran los valores de las variables `p1, p2` y `cl`

```
p1=0.6
```

```
p2=0.7
```

```
cl=0.95
```

Se llama a la función pasándole como parámetros la primera y segunda respuesta aleatoria, la proporción de cartas marcadas con la característica sensible en la primera y segunda caja, el vector de probabilidades de inclusión de primer orden, le indico que quiero estimar la media, el nivel de confianza y el tamaño poblacional.

```
HorvitzUB(dat$I, dat$J, p1, p2, dat$Pi, "mean", cl, N)
```

Una vez ejecutada la función, se obtiene la siguiente salida

Call:

```
HorvitzUB(I = dat$I, J = dat$J, p1 = p1, p2 = p2, pi = dat$Pi,
  type = "mean", cl = cl, N = N)
```

```
Qualitative model
```

```
Horvitz unknown B model for the mean estimator
```

```
Parameters: p1=0.6; p2=0.7
```

```
Estimation: 0.1213169
```

```
Variance: 0.00758909
```

```
Confidence interval (95%)
```

```
  Lower bound: 0.1088642
```

```
  Upper bound: 0.1337696
```

que muestra la llamada a la función, indica que nos encontramos ante un modelo cualitativo, concretamente el modelo de Horvitz con  $\beta$  desconocida para el estimador de la media y el valor de los parámetros.

En las siguientes líneas ofrece la estimación de la media, la varianza y el intervalo de confianza (al nivel de confianza indicado).

En caso de querer estimar el total poblacional, habría que modificar el parámetro `type` e indicar que calcule "total" en lugar de "mean"

```
HorvitzUB (dat$I, dat$J, p1, p2, dat$Pi, "total", cl)
```

### REFERENCIAS

- [1] CHAUDHURI, A. (2011). Randomized response and indirect questioning techniques in surveys. **Boca Raton: Chapman and Hall, CRC Press.**
- [2] COBO, B., RUEDA, M., ARCOS, A. (2015). Package 'RRTCS': Randomized Response Techniques for Complex Surveys. URL = <http://cran.r-project.org/web/packages/RRTCS/>
- [3] GREENBERG, B.G., ABUL-ELA, A.L., SIMMONS, W.R., HORVITZ, D.G. (1969). The unrelated question RR model: Theoretical framework. **Journal of the American Statistical Association**, 64, 520-539.
- [4] HORVITZ, D.G., SHAH, B.V., SIMMONS, W.R. (1967). The unrelated question RR model. **Proceedings of the Social Statistics Section of the American Statistical Association**. 65-72. Alexandria, VA: ASA.
- [5] WARNER, S.L. (1965). Randomized Response: a survey technique for eliminating evasive answer bias. **Journal of the American Statistical Association** 60, 63-69.

## Capítulo 14

pp 139-150

# PANAMÁ: MODELACIÓN DE LA DEMANDA AGREGADA COMO ARISTA DEL DESARROLLO HUMANO

René Bracho

Universidad de Panamá, Panamá, Ciudad de Panamá

### ABSTRACT

The Aggregate demand is the economic infrastructure on which supported a category of human development of a countries. In the case of Panama, the aggregate demand has grown steadily in the past 12 years and is therefore useful the analytical description of this growth. In contrast derived from the explanation of the operation of macro-economic growth of the country, it is necessary to suggest some ideas on what must do Panama to advance to a higher level of human development, and this is to promote a diversification of the production model incorporating technological innovation and investing in human capital, education and health services with higher quality.

**KEYWORDS:** Human Development, aggregate demand, national income.

### RESUMEN

La demanda agregada constituye la infraestructura económica sobre la que se soporta la categoría de desarrollo humano de un país. En el caso de Panamá, la demanda agregada ha crecido constantemente en los últimos 12 años y por tal motivo es útil una descripción analítica de dicho crecimiento. Como contraste derivado de la explicación del funcionamiento del crecimiento macroeconómico del país es necesario sugerir algunas nociones sobre lo que debe hacer Panamá para avanzar a un nivel más alto de desarrollo humano, y esto consiste en impulsar una diversificación del modelo productivo incorporando innovación tecnológica e invirtiendo en capital humano, en educación y salud de calidad.

## 1. INTRODUCCIÓN: DESARROLLO HUMANO Y DEMANDA AGREGADA

El desarrollo humano está representado por la mejora de las condiciones de vida de las personas en una determinada sociedad; constituyendo un proceso caracterizado por el aumento de los bienes y servicios de los cuales disponen para satisfacer sus necesidades básicas y suplementarias. Este fenómeno de incremento de riqueza social, generalmente, va acompañado de un entorno de respeto por los derechos humanos de la colectividad.

Desde una perspectiva de historia de las ideas, el desarrollo humano como concepto posee orígenes clásicos grecolatinos. Ya en el siglo IV A.C., Aristóteles vinculaba esta noción con el alcance pleno del florecimiento de las capacidades humanas.

Lo real en el funcionamiento de las sociedades humanas es el cambio, la evolución y el movimiento, tanto en el plano tecnológico como en las demás dimensiones del desarrollo social. Es por ello que el concepto de desarrollo humano se ha ido nutriendo de otros aspectos diferentes al ámbito estrictamente económico para anexas otros factores igualmente importantes para la vida; es decir: la cultura, la educación, la salud y el bienestar en general, etc. que también desempeñan un rol para alcanzar el desarrollo.

Un efecto práctico del uso del concepto de Desarrollo Humano consiste en que es una forma de evaluar cuantitativamente la calidad de vida de las personas en el ámbito de su desenvolvimiento geográfico, nacional o regional.

La construcción de medidas de desarrollo humano como el Índice de Desarrollo Humano (IDH) son el resultado de la necesidad de plantear los parámetros para ubicar a los países en desarrollados o en

desarrollo. Proviene de ideas orientadas a catalogar a los países en función de variables distintas a las utilizadas tradicionalmente en la teoría económica y sociocultural. Así, este índice integró elementos económicos (PIB per cápita), con componentes de salud (esperanza de vida) y educación (años promedio de escolaridad). El IDH como instrumento es calculado desde 1990 por el Programa de Naciones Unidas para el Desarrollo (PNUD) y fue construido en base a las formulaciones teóricas y técnicas de economistas como Mahbub ul Haq y Amartya Sen.

El proceso de construcción teórica del término desarrollo humano también fue el fruto de una evolución sistemática. Iniciando por el momento en que se percibía como una categoría idéntica a la noción de crecimiento económico, seguido por una reformulación en donde se le definía como el crecimiento económico acompañado de una distribución eficaz de la riqueza, hasta alcanzar la elaboración de una concepción positiva que incorpora los indicadores numéricos del desarrollo humano. Categoría en la que coinciden índices que miden la existencia y profundidad de instituciones económicas y políticas inclusivas.

Así, distintas variables macroeconómicas y contables como: el nivel del ingreso nacional y sus formas de distribución, el volumen de la demanda agregada, el acceso a financiamiento, la rentabilidad del capital, la innovación tecnológica, la absorción de conocimientos aplicados a la producción, etc., se fusionaron con variables demográficas como: los ritmos de fecundidad y mortalidad, junto a elementos culturales: alfabetización, años de escolaridad, instituciones ligadas a los derechos humanos, etc. Dando como resultado los factores y componentes del IDH. Convirtiéndose en los elementos que, por la heterogeneidad mundial, definen las fronteras entre el desarrollo y el subdesarrollo. Siendo a la vez, el producto del devenir histórico, las diferencias institucionales, los patrones y pautas de dinámica socioeconómica específica para cada país en concreto.

El objeto de análisis de este trabajo es una descripción analítica de la demanda agregada de Panamá dado que es un factor del concepto de desarrollo humano.

## 2. DEMANDA AGREGADA ENFOCADA DESDE LA NUEVA ECONOMÍA KEYNESIANA Y LA MACROECONOMÍA POSTKEYNESIANA (NEOKALECKIANA)

Ambas tradiciones del pensamiento económico contemporáneo comparten tanto convergencias como divergencias conceptuales al momento de tratar la demanda agregada. Para ilustrar las afinidades tenemos el siguiente cuadro:

**Cuadro #1: Convergencias Conceptuales entre NEK Y MPK**

<b>Coincidencias entre Nuevos Keynesianos (NEK) y PostKeynesianos (MPK)</b>
1. No vaciado continuo de los mercados: rigidez de precios y salarios.
2. Mercado de bienes: competencia imperfecta
3. Mercado de trabajo: desempleo involuntario
4. Efectividad de las políticas de expansión de la demanda efectiva

**Fuente: (Carballo, Galindo y Usabiaga, 2000:17)**

El tratamiento que ambas escuelas le dan al concepto de demanda agregada parte de la existencia de estructuras de mercado no competitivas (monopolios, oligopolios, etc.) y de fallas de

mercado (externalidades, información asimétrica, bienes públicos, etc.) en el contexto de mercados sectoriales de los sistemas económicos y las economías globales. Las perspectivas de la NEK y la MPK poseen como elementos conceptuales de afinidad y convención las fundamentaciones teóricas que demuestran la rigidez de precios y salarios. Convirtiendo los vaciados de mercados en la excepción y no la regla común, toda vez que estas situaciones al ser los máximos posibles de los excedente de consumidores y productores, simultáneamente, son escenarios casi inexistentes en medio de la presencia de poderes corporativos en las economías de mercado contemporáneas.

En cuanto al mercado de trabajo, se aprecia la preocupación por ambas tradiciones entorno al desempleo involuntario. También, otro punto de proximidad entre NEK y MPK consiste en la importancia que brindan a los factores, convenciones e instituciones sociales en la determinación del salario real que es el elemento clave en la explicación de la demanda y la oferta de trabajo según ambas escuelas.

Por último, dichas doctrinas macroeconómicas convienen en aceptar la efectividad de las políticas de expansión de la demanda agregada. Para los teóricos de la NEK estas medidas tanto en el plano monetario como en el fiscal, tomando en cuenta un escenario de competencia imperfecta, son eficaces para aproximar la demanda en el sentido del pleno empleo. Y para los neokaleckianos la demanda efectiva posee un rol central en la configuración del pleno empleo. Y dado el carácter exógeno de las decisiones de inversión es acertado impulsar políticas de expansión de la demanda.

La nueva economía keynesiana indaga en los factores microeconómicos que pueden incidir sobre la dinámica económica generando efectos a nivel macro. Los modelos de esta corriente describen los problemas y decisiones que confrontan a los distintos agentes económicos, tanto hogares como empresas monopolísticas, gobiernos, autoridades monetarias y otros actores.

**Cuadro #2: Divergencias Metodológicas entre NEK y MPK**

NEK	MPK
METODOLOGÍA: Prediccionismo	METODOLOGÍA: Realismo
SUPUESTOS:	SUPUESTOS:
1. Enfoque agregado con enlace a la acción de los agentes económicos individuales (problema de los microfundamentos)	1. Enfoque holístico, grupo, agregado.
2. Agentes optimizadores	2. Racionalidad limitada
3. Expectativas racionales	3. Incertidumbre
4. Oferta monetaria exógena. (Aceptación de la dicotomía clásica: variables reales y nominales)	4. Oferta monetaria endógena. (Rechazo de la dicotomía clásica: variables reales y nominales)
5. Confianza en las soluciones mercados	5. Desconfianza en las soluciones de mercados o inexistencia

**Fuente:** Caraballo, Galindo y Usabiaga: 2000

Así, los problemas que enfrenten los agentes por motivo de rigideces en precios y/o salarios al momento de hacer ajustes en dichos rubros, debe tenerse en cuenta que tenderán a perdurar más allá de lo que el agente económico considera conveniente. Es decir, la demanda agregada depende de la relación

salario - precios. Eso indica que existe la tendencia que la creciente expansión industrial depende de factores tecnológicos.

Por el lado de los postkeynesianos, se enfatiza que tanto a corto como a largo plazo la economía está dirigida por la demanda agregada. Es decir, es fundamental el principio de demanda efectiva: “la producción se ajusta a la demanda”.

No obstante, se pueden identificar divergencias entre estas tradiciones. En el cuadro 2 las exponemos de forma simplificada:

### 3. DEMANDA AGREGADA Y LOS OBSTÁCULOS AL DESARROLLO HUMANO EN PANAMÁ

El modelo econométrico que hemos desarrollado se enfoca en generar una descripción analítica de una de las caras del desarrollo humano en su dimensión macroeconómica: la demanda agregada. Se pretenden estimar los coeficientes que relacionan los valores del consumo y la inversión fija de la contabilidad nacional con las expansiones registradas de la demanda agregada para ayudar a entender los aumentos en la renta nacional. Esto adquiere importancia, dados los resultados plausibles de Panamá en materia de crecimiento y acumulación de capital durante el periodo (2003-2013); que son el fruto de la política macroeconómica de corte inicialmente anticíclica y, posteriormente, expansiva de la demanda efectiva que llevaron a cabo los agentes económicos (trabajadores, empresarios locales, profesionales, policy-makers, banqueros y operadores del sector bursátil, inversionistas del exterior y consumidores) que concurren en el país.

El análisis de la demanda agregada panameña presentado a continuación junto a la observación de sus cuentas nacionales nos permite entender algunos datos relevantes para configurar un diagnóstico sobre la salud del crecimiento registrado por Panamá. Pero a la vez, invita a reorientar la indagación sobre las vías que debe transitar el país para alcanzar el desarrollo humano. Al parecer la mirada debe dirigirse hacia problemas distintos a los relacionados con la estabilidad económica. Reafirmar analíticamente que el problema de Panamá para avanzar en la ruta al desarrollo no es estrictamente macroeconómico obliga a contraponer los datos frente a las condiciones necesarias y suficientes para lograr el desarrollo humano en los países. Así, en el caso panameño se evidencia que la reasignación de recursos ineficaz por parte del estado se refleja en bajos resultados en materia de calidad educativa y salud pública; ello afecta la productividad de la población económicamente activa limitando la absorción de conocimientos tecnológicos, despotenciando la incorporación de innovaciones técnicas a algunos aspectos de la producción nacional. Ello crea un círculo vicioso que restringe la sostenibilidad de las potencialidades de creación de riqueza lo que a su vez se expresa en una distribución inequitativa del ingreso. Así, parece configurarse el cuadro de los aspectos centrales que obstaculizan el paso hacia niveles más altos de desarrollo humano.

Panamá no logra trascender el parámetro medio regional en materia de redistribución de recursos por parte del estado luego de impuestos y transferencias. Un dato ilustrador es el coeficiente Gini, que refleja para América Latina un promedio de 0.52 antes de impuestos y transferencias y 0.50 después de impuestos y transferencias, versus el promedio de 0.46 antes de impuestos y transferencias y 0.31 después que reflejan 15 países de la Unión Europea.

**Cuadro #3: Coeficiente Gini Promedio antes y después de Impuestos y Transferencias**

REGIÓN	ANTES	DESPUÉS
América Latina	0.52	0.50
Unión Europea	0.46	0.31

**Fuente:** Novelo, Federico. El desarrollo económico y social en América Latina: El doble atraso. CEPAL. México 2014.

Lo anterior sugiere que el círculo vicioso que puede estar posicionando a la inequidad como obstáculo al desarrollo humano en Panamá indica la importancia de impulsar una política económica que estimule la diversificación de la economía y la ampliación de la participación en la estructura productiva de sectores con alta intensidad en conocimientos. Efecto que únicamente puede alcanzarse mediante iniciativas tendientes a promover transformaciones estructurales progresivas que fomenten la producción de mediana y alta tecnología; es decir, estrategias industriales que produzcan el nacimiento de nuevos sectores fundamentados en la eficiencia schumpeteriana y con una significativa tasa de crecimiento de la demanda efectiva. Todo lo anterior acompañado de una fuerte inversión en capital humano: salud y educación de alta calidad.

Sólo esto puede complementar estratégicamente la bonanza macroeconómica registrada en Panamá recientemente, para llevar al país a mayor desarrollo humano. Ya que Panamá muestra en sus cifras oficiales una situación en su mercado laboral próxima al pleno empleo. La tasa de desempleo total es de 4% para 2013 según el Instituto de Estadística y Censo (INEC) de la Contraloría General de la República. Es decir, el 96% de la fuerza laboral nacional está ocupada. Si obviamos la discusión sobre la calidad del empleo en Panamá e ignoramos el 38.6% de la población económicamente activa que se encuentra en empleo informal, podemos asumir la ocupación plena como parte del contexto de nuestra estimación de la demanda agregada.

Destaca el significativo peso del sector terciario en la composición de la renta y el efecto multiplicador de las exportaciones provenientes del llamado Conglomerado de la Zona de Tránsito en función de la expansión de la demanda agregada. Las divisas generadas por las exportaciones de servicios, en el contexto de ser una economía dolarizada sin banco central, aumentan la demanda doméstica por bienes y servicios y dinamizando las posibilidades de financiamiento local que genera el Centro Bancario y crecen los depósitos bancarios. No obstante, el rubro de exportación de frutas y vegetales se ha visto beneficiado por la reciente mejora en los términos de intercambio de algunos alimentos y materias primas, explicados por la demanda proveniente de China Continental. Sin embargo, mientras la región atraviesa una reprimarización de la producción y de las exportaciones, ya que el total de la estructura de las exportaciones de América Latina posee un 60% del total representado por materias primas (39,1%) y por manufacturas basadas en recursos naturales (20,9%), mientras sólo un 9% a manufacturas de alta tecnología; Panamá durante el mismo periodo de tiempo posee cerca de un 75% de las exportaciones constituidas por servicios (tránsito por el canal, reexportaciones ZLC, turismo, puertos, servicios bancarios y transporte aéreo; mientras que la proporción restante estuvo vinculada a productos pesqueros, agropecuarios e industriales.

Las exportaciones tienen un efecto multiplicador en la actividad económica local. En el caso del Conglomerado del Canal, este fue estimado en 1.27, o sea, cada dólar de divisas que entra a la economía genera \$1.27 de PIB". (Barletta; 2011: 341).

Otro aspecto a resaltar denota las debilidades del crecimiento económico en el mediano y largo plazo. Este enfoque se desarrolla por el lado de la oferta incorporando a la variable tiempo. Dentro de esta perspectiva, se destaca un valioso dato sobre la productividad en Panamá. "En la última década la productividad total de factores, es decir el aumento en la producción real por persona ocupada descontado el efecto de la inflación, y tomando en cuenta la cantidad de capital utilizada por persona y el efecto de la tecnología aumentó apenas 0.7% por año. Este coeficiente es bajo ya que idealmente debiera ser del orden de 2.5 o aún 3.0% anual... Ello además nos dice que tenemos un modelo de crecimiento no sostenible que se asemeja al de la antigua URSS, Japón o Singapur, masivo en el uso de capital y no en la productividad". (Chapman; 2012: 9-10).

#### 4. MODELO ESTIMADO DE DEMANDA AGREGADA (INTERPRETACIÓN MACROECONÓMICA)

$$PIB_t = \alpha_1 + \beta_1 ConF_t + \beta_2 FBKF_t + \beta_3 ExpNetas_t + \beta_4 ExcNetoExp + \mu_t$$

$$PIB_t = 707.7 + 0.42ConF_t + 0.92FBKF_t + 0.35ExpNetas_t + 1.21ExcNetoExp + \mu_t$$

La ecuación de demanda agregada que hemos seleccionado está precisada a partir de una función del Producto Interno Bruto definido por tres variables que contablemente pertenecen a la vía del gasto: consumo, la inversión (formación bruta de capital fijo) y las exportaciones netas; y una variable perteneciente a la vía del producto: el excedente neto de explotación. Al margen de los problemas procedimentales que pueden aparecer por generar las estimaciones en base a un modelo que combina variables de dos enfoques contables diferentes se procederá a interpretar los resultados del modelo.

Los cálculos para la economía panameña de los coeficientes de los componentes de la demanda agregada son los siguientes:

##### 4.1. Coeficiente de Regresión Parcial del Gasto de Consumo Final ( $\beta_1$ ): $\beta_1=0.42$

Las estimaciones indican que por cada \$1 dólar de aumento en el consumo final agregado, el producto interno bruto nominal (PIB) aumentará \$0.42 centésimos.

Este coeficiente parcial de pendiente del consumo refleja la variación que se registra cuando en un lapso temporal anterior, un aumento del ingreso nacional disponible ha precedido a un alza del gasto de consumo, provocando alza en la demanda agregada que según los enfoques teóricos mencionados anteriormente NEK y MPK es el motor de la producción y el crecimiento económico. Es decir, por antonomasia es uno de los pilares del desarrollo humano. Por ende, se evidencia una relación directa o positiva entre este coeficiente y el PIB nominal. Se debe tomar en cuenta que, cuando aumenta el ingreso nacional disponible de un periodo anterior, el consumo no aumenta en la misma proporción, sino que aumenta en una proporción menor. Ya que una porción de dicho incremento del ingreso nacional disponible es destinado al ahorro o puede ser atesorado.

##### 4.2. Coeficiente de Regresión Parcial del Gasto de Inversión Fija : $\beta_2=0.92$

El gasto de inversión en nuestro modelo está representado por la formación bruta de capital fijo (FBKF), los cálculos para este parámetro sugieren que por cada incremento de \$1 dólar en la inversión fija, el producto interno bruto nominal está creciendo \$0.92 centésimos.

Este segundo componente de la demanda agregada, también puede ser entendido como: la demanda de inversión. Debido que las decisiones de inversión son tomadas por los empresarios en función de sus expectativas y objetivos de rentabilidad, la factibilidad de sus proyectos, del stock de tecnología y las posibilidades de financiamiento y crédito; la demanda de inversión puede considerarse como una variable exógena e independiente del ingreso disponible que se genere en el país

Los gastos de inversión son la clave del crecimiento futuro. El stock de capital dado por las inversiones realizadas en el pasado y que continúan operando actualmente (infraestructura, maquinaria, bienes inmuebles, elementos de logística, dispositivos tecnológicos e informáticos, etc.) conforman la capacidad instalada del país, condicionan su producción potencial. En consecuencia, las inversiones se



efectúan para aumentar estas existencias de capital o para reemplazarlas por motivo de su depreciación o desfase.

Para las cuentas nacionales, la formación bruta de capital (inversión) está conformada por:

- Las inversiones en infraestructura, planta y equipos.
- La construcción de bienes inmuebles para uso de las unidades familiares.

#### 4.3. Coeficiente de Regresión Parcial del Gasto en Exportaciones Netas ( $\beta_3$ ): $\beta_2=0.35$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	707.6578	393.3444	1.799080	0.0972
CONSFINAL	0.421683	0.147020	2.868202	0.0141
FBKFIJO	0.917354	0.223724	4.100388	0.0015
EXPNETAS	0.353479	0.131696	2.684063	0.0199
EXCNETOEXP	1.210280	0.331234	3.653848	0.0033

R-squared	0.999037	Mean dependent var	18108.82
Adjusted R-squared	0.998716	S.D. dependent var	8844.606
S.E. of regression	316.9838	Akaike info criterion	14.59551
Sum squared resid	1205745.	Schwarz criterion	14.84057
Log likelihood	-119.0618	Hannan-Quinn criter.	14.61987
F-statistic	3111.174	Durbin-Watson stat	1.787843

Las exportaciones netas constituyen un factor fundamental para establecer la renta y/o demanda agregada en una economía abierta. Representan el gasto de los extranjeros en bienes producidos en el interior menos el gasto de los residentes interiores en bienes extranjeros. En el caso de nuestro modelo, las estimaciones evidencian que cuando las exportaciones netas se elevan en \$1 dólar, el PIB sube \$0.35 centésimos.

Analíticamente, las exportaciones netas pueden ser deducidas así:

$$XN = X - M$$

$$XN = (\bar{X} + xY^* - x_r \varepsilon) - (\bar{M} + mY + m_r \varepsilon)$$

Reordenando:

$$\text{Siendo: } XN = \bar{XN} + xY^* - mY - \phi \varepsilon$$

XN: Exportaciones Netas

X: Exportaciones

M: Importaciones

Y\*: Renta exterior

Y: Renta interior

$x_r + m_r = \phi$ ; sensibilidad de las XN ante cambios en el tipo de cambio real ( $\varepsilon$ );  $\phi > 0$ .

#### 4.4. Coeficiente de regresión parcial del excedente neto de explotación ( $\beta_4$ ): $\beta_2=1.21$

En cuanto al excedente neto de explotación o remuneraciones al capital, tenemos que cuando éste rubro se alza en \$1 dólar, el PIB se amplía en \$1 dólar con \$0.21 centésimos. Al respecto, es ilustrador contrastar los resultados de la estimación de este coeficiente con el descubrimiento sobre la dinámica estancacionista del patrón de utilización de la capacidad instalada de la economía panameña (Jované, 2014). Atendiendo esto, tenemos que la evidencia econométrica demuestra que mientras mayor es la participación de las remuneraciones al capital en el ingreso nacional mayor es la tendencia a disminuir la utilización de la capacidad productiva; con el efecto desacelerador en el crecimiento de la demanda agregada que se deriva de esta tendencia en determinado lapso temporal. Sin embargo, las mediciones contables indican que esta tendencia no se ha materializado recientemente en el caso panameño. Esto puede deberse a que los efectos ralentizadores sobre la demanda agregada por motivo de la distribución regresiva del ingreso nacional en el periodo estudiado han sido atenuados por las exportaciones netas y la inversión fija gubernamental. Ello ha sido posible debido que el régimen de utilización de la capacidad productiva nacional es guiado por los salarios, pero también, por las exportaciones y la inversión pública.

### 5. LA EVALUACIÓN DE LA ESTIMACIÓN ECONOMETRICA

La evaluación individual de la significancia de los coeficientes indica que todos ( $\alpha_1$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  y  $\beta_4$ ) son significativos. El modelo presenta una elevada Bondad de Ajuste (0.99).

Al verificar el estadístico  $d$  Durbin Watson tenemos que cae en la región de indecisión:

$$d_L = 0.779 \leq d_{calc} = 1.79 \leq d_u = 1.90$$

Por tal motivo se procedió a efectuar la prueba  $d$  modificada:

“H0:  $\rho = 0$                       Zona de Indecisión                      H1:  $\rho > 0$ ”

Si el valor estimado  $d < d_u$ , recházese H0 al nivel  $\alpha$ ; es decir, hay correlación positiva estadísticamente significativa.

$$H0: \rho = 0$$

$$H1: \rho < 0$$

Si el valor estimado  $(4 - d) < du$ , recházese  $H_0$  al nivel  $\alpha$ ; hay evidencia estadísticamente significativa de autocorrelación negativa.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Si el valor estimado  $d < du$ , o  $(4 - d) < du$ , recházese  $H_0$  al nivel  $2\alpha$ ; estadísticamente, hay evidencia significativa de autocorrelación positiva o negativa". (Gujarati; 2007: 453).

Empleando este test obtenemos que: dado que  $d < du$  ( $1.79 < 1.90$ ); hay correlación estadísticamente significativa. No obstante, debido que  $(4 - d) > du$  ( $2.21 > 1.90$ ), se acepta  $H_0$ . No hay evidencia estadísticamente significativa de autocorrelación negativa o positiva.

Debido que la prueba  $d$  modificada indica la existencia de correlación estadísticamente significativa procederemos a calcular el coeficiente de autocorrelación ( $\rho$ ) para ver si dado que la muestra es pequeña (17 observaciones) se puede aplicar el criterio de Griliches y Rao. En donde se plantean las condiciones en las que los coeficientes estimados por Mínimos Cuadrados Ordinarios son aceptables en presencia de autocorrelación.

Coeficiente de autocorrelación estimado:

$$\hat{\rho} \approx 1 - \frac{d}{2}$$

Como el coeficiente de autocorrelación estimado es menor que 0.3 y la muestra es pequeña se pueden mantener sin corrección por mínimos cuadrados  $0.105 < 0.3$  generalizados factibles los coeficientes estimados previamente por mínimos cuadrados ordinarios. (Gujarati; 2007).

## 6. POSIBLES EFECTOS CONFUSIÓN E INTERACCIÓN EN EL MODELO UNIECUACIONAL DE DEMANDA AGREGADA

### 6.1. Conceptos y Resultados de la Estimación del Modelo

Los efectos de *confusión* e *interacción* son dos aspectos fundamentales a tomar en cuenta cuando se usan los modelos de regresión múltiple con una finalidad estimativa. Ya que, en esta clase de modelos posee una gran importancia la interferencia que una o varias variables pueden realizar en la asociación con otras.

El modelo de regresión múltiple generalmente supone la independencia de efectos de las distintas variables independientes. En este sentido, es un modelo aditivo - modelo de efectos principales- en el que cada variable ejerce efectos por separado sobre los criterios, y el conjunto de estas variables suman sus efectos para predecir dicho resultado. No existe pues, una variable ajena a las contempladas que module o mediatice esta serie de efectos sumados, es decir, que condicione sus influencias. En los modelos explicativos en los que existen efectos de interacción es necesario contemplar coeficientes que representen específicamente tales efectos. Así, en presencia de este fenómeno, además de los efectos aditivos (principales) de las variables predictoras escogidas inicialmente, habría que incluir un término anexo que represente la interacción entre las variables que se supone que interactúan. Así, cuando la interacción está presente la asociación entre dos variables varía según los diferentes niveles de otra u otras variables.

Por otro lado, hay presencia de confusión cuando la asociación entre dos variables difiere significativamente según que se considere, o no, otra variable. A esta última variable se le denomina variable de confusión para la asociación.

A continuación procederemos a indagar estadísticamente si existe alguno de estos efectos en las variables que configuran la dinámica de la demanda agregada de Panamá.

**Cuadro #4: Correlación entre las Variables**

Fuente: Datos del INEC y elaboración Propia

	<i>PIBNOMINAL</i>	<i>CONSFINAL</i>	<i>FBKFIJO</i>	<i>EXPNETAS</i>	<i>ExcNetoExp</i>	<i>Interacción</i>
<i>PIBNOMINAL</i>	↑ 1					
<i>CONSFINAL</i>	↑ 0.9834680	↑ 1				
<i>FBKFIJO</i>	↑ 0.9864536	↑ 0.95987	↑ 1			
<i>ExpNetas</i>	↓ -0.121945	↓ -0.24491	↓ -0.1691	↑ 1		
<i>ExcNetoExp</i>	↑ 0.9988312	↑ 0.98378	↑ 0.98276	↓ -0.1196	↑ 1	
<i>Interacción</i>	↓ -0.142863	↓ -0.27347	↓ -0.1726	↑ 0.9847	↓ -0.1422	↑ 1

Dependent Variable: PIBNOMINAL  
 Method: Least Squares  
 Date: 08/03/15 Time: 12:29  
 Sample (adjusted): 1996 2012  
 Included observations: 17 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	707.6578	393.3444	1.799080	0.0972
CONSFINAL	0.421683	0.147020	2.868202	0.0141
FBKFIJO	0.917354	0.223724	4.100388	0.0015
EXPNETAS	0.353479	0.131696	2.684063	0.0199
EXCNETOEXP	1.210280	0.331234	3.653848	0.0033
R-squared	0.999037	Mean dependent var		18108.82
Adjusted R-squared	0.998716	S.D. dependent var		8844.606
S.E. of regression	316.9838	Akaike info criterion		14.59551
Sum squared resid	1205745.	Schwarz criterion		14.84057
Log likelihood	-119.0618	Hannan-Quinn criter.		14.61987
F-statistic	3111.174	Durbin-Watson stat		1.787843

Las salidas del modelo uniecuacional estudiado (Con Interacción y sin interacción) son:

En nuestro análisis de la demanda agregada, las exportaciones netas (ExpNetas) no presentan una correlación significativa con el nivel del PIB Nominal cuando se considera la *supuesta interacción* entre las exportaciones netas y el excedente neto de explotación (ExcNetoExp); mientras que cuando no se considera dicha supuesta interacción, sí lo presenta. En este caso, la supuesta interacción entre las exportaciones netas y el excedente neto de explotación (remuneraciones al capital), constituye una variable de confusión para la asociación entre el PIB Nominal y las exportaciones netas.

Además, el hecho que la interacción entre las exportaciones netas y el excedente neto de explotación registre una valor p superior a  $\alpha$  ( $0.55 > 0.05$ ), indica que la variable interacción no arroja un coeficiente significativo en relación con el PIB Nominal. Por lo tanto, la suposición de interacción no tiene fundamento. No existe dicha interacción.

Al no comprobarse formalmente la existencia de interacción entre ExpNetas y ExcNetoExp, sí puede hablarse, por lo tanto, de un efecto particular de las exportaciones netas y otro efecto distinto del excedente neto de explotación; sin que ambos efectos interafecten combinadamente al PIB Nominal o demanda agregada del país.

Por el contrario, si la asociación entre las exportaciones netas y el PIB Nominal fuera diferente en función del nivel de remuneraciones al capital (ExcNetoExp), se diría que, para determinado nivel de PIB Nominal, existe interacción entre las exportaciones netas y las remuneraciones al capital efectuadas por la economía (ExpNetas X ExcNetoExp). No obstante, se ha demostrado formalmente, que en el caso analizado de la economía panameña durante el periodo 1996-2012 no existen varias estimaciones del coeficiente de la variable de interés, sino que habría una estimación independientemente del nivel de la otra variable. Es decir, no existe una estimación de la relación entre el PIB nominal y las exportaciones netas que esté influenciada por los niveles de remuneración al capital.

En base a esto se concluyó que el modelo uniecuacional más apto es aquel en el cual no se incluye la interacción entre las exportaciones netas (ExpNetas) y las remuneraciones al capital (ExcNetoExp), por eso fue el escogido para hacer la descripción analítica de la demanda agregada de Panamá:

$$PIB_t = \alpha_1 + \beta_1 ConF_t + \beta_2 FBKF_t + \beta_3 ExpNetas_t + \beta_4 ExcNetoExp + \mu_t$$

$$PIB_t = 707.7 + 0.42 ConF_t + 0.92 FBKF_t + 0.35 ExpNetas_t + 1.21 ExcNetoExp + \mu_t$$

#### REFERENCIAS

1. Barletta, N. (2011). Estrategia para el Desarrollo Nacional. Centro Nacional de Competitividad. Panamá.
2. Castaño Salas, H. (2008). Entender la Economía. Una Perspectiva Epistemológica y Metodológica. Editorial Félix Varela. Cuba.
3. Chapman, F. (2014). Declaraciones a Telemetro Reporta del Presidente de la Bolsa de Valores de Panamá en el marco del XV Foro de Inversionistas. Panamá.
4. Chapman, G. (2012). Conferencia Magistral en ocasión del 51 Aniversario de la fundación de la Escuela de Economía. Universidad de Panamá, Facultad de Economía. Panamá.
5. Contraloría General de la República. Instituto de Estadística y Censo (INEC). Panamá en Cifras 2011-2012. Panamá .
6. De Lima, F. (2013). Ponencia en el 8vo. Foro Nacional para la Competitividad. Ministerio de Economía y Finanzas. Panamá.
7. Furtado, C. (1984). Teoría y Política del Desarrollo Económico. Editorial Siglo XXI. México.
8. Gujarati, D. N. (2007). Econometría. Editorial Mc Graw Hill. Cuarta Edición. México.
9. Haberler, G. (1944). La Teoría del Multiplicador de Keynes: Una Crítica Metodológica. Capítulo de Ensayos Sobre El Ciclo Económico. Fondo de Cultura Económica. México.
10. Jované, J. (2014). Régimen de Utilización de la Capacidad Instalada y de Acumulación en Panamá: 2000-2012. Carta Informativa. Universidad de Panamá..
11. Keynes, M. J. (1936). Teoría General de La Ocupación, El Interés y El Dinero. Fondo de Cultura Económica. Primera Reimpresión, 2005. Argentina..
12. Machlup, F. (1939). Análisis del Tiempo y la Teoría del Multiplicador. Capítulo en: Ensayos sobre el Ciclo Económico. Fondo de Cultura Económica. México .
13. Patinkin, D. (1951). Flexibilidad de Precios y Pleno Empleo. Readings in Monetary Theory. American Economic Association. The Blakiston Company. Estados Unidos.
14. Pérez, C. (2004). Técnicas de Análisis Multivariante de Datos. Pearson Prentice Hall, Educación, S.A., Madrid, España.

15. Samuelson, P. (1948). La Matemática Elemental de la Determinación de la Renta. *Income, Employment and Public Policy.* 133-135.
16. Schwarz, G. (1978). Estimando la Dimensión de un Modelo. *Anales de Estadística*, número 6, 461-464.
17. Urbisaia, H. y Brufman, J. (2002). Evolución Metodológica de la Econometría, Visión Histórica. Universidad de Buenos Aires, Facultad de Ciencias Económicas. Argentina..

## Capítulo 15

pp 151-167

# CARACTERIZACIÓN DE LOS NIVELES DE DESARROLLO ECONÓMICO DE LAS REGIONES DEL CENTRO SUR DE MÉXICO

Ignacio Caamal Cauich\*, José Félix García Rodríguez\*\*, Verna Gricel Pat Fernández\*,  
David Martínez Luis\*\*\* & Felipe Jerónimo Ascencio\*

\* Universidad Autónoma Chapingo. México

\*\*Universidad Juárez Autónoma de Tabasco, Mexico

\*\*\* Universidad Autónoma del Carmen, Campeche. México

### ABSTRACT

The level of economic development of the Central South of Mexico is lower than the national average, the national average rural population is lower than the average of the region; the national average economically active population occupied in the primary sector is less than the average of the region; the national average primary gross domestic product is slightly greater than the regional; the national average gross domestic product per capita of the economically active population occupied in the primary sector is greater than the average of the region; the national average gross domestic product per capita in the primary sector is higher than the regional average. The results reflect a high weight of the rural population and the economically active population occupied in the primary sector; low levels of productivity in the economically active population and low levels of gross domestic product per capita of the rural population, which explains the low income levels and high levels of poverty of the rural population; and in general, the low levels of development of the Central South Region, compared to the national average.

**KEYWORDS:** rural population, agricultural population, agricultural gross domestic product.

### RESUMEN

El nivel de desarrollo económico del Centro Sur de México es inferior a la media nacional, la población rural promedio nacional es menor a la media de la región; la población económicamente activa ocupada en el sector primario promedio nacional es menor a la media de la región; el producto interno bruto primario promedio nacional es ligeramente mayor al regional; el producto interno bruto per cápita de la población económicamente activa ocupada en el sector primario promedio nacional es mayor a la media regional. Los resultados obtenidos reflejan un alto peso de la población rural y de la población económicamente activa ocupada en el sector primario; bajos niveles de productividad de la población económicamente activa y bajos niveles de producto interno bruto per cápita de la población rural, lo que explica los bajos niveles de ingreso y altos niveles de pobreza de la población rural; y en general, de los bajos niveles de desarrollo de la Región Centro Sur, en comparación con la media nacional.

## 1. INTRODUCCIÓN

El desarrollo económico, de acuerdo con la historia económica, las teorías económicas, la experiencia mundial y en términos generales, siempre se ha caracterizado por una transformación profunda y de gran magnitud de la estructura económica, social y política de la sociedad, de la organización, de la producción, de la distribución y del consumo. Así mismo, el desarrollo económico ha sido impulsado por clases y grupos sociales que están interesados en mejorar o cambiar el orden económico vigente con la finalidad de mejorar los niveles de bienestar de la población, especialmente de las capas más desprotegidas, pobres y marginadas de la sociedad (Baran, 1975). El desarrollo económico está estrechamente relacionado con el crecimiento económico, el cual se refiere al incremento del volumen y valor de la riqueza generada en un determinado periodo.

Las transformaciones de la estructura social y económica se encuentran directamente relacionados con los cambios en la composición y distribución de la población en general, de las clases y grupos sociales, de los productores y trabajadores (Lenin, 1974). Así mismo, están directamente relacionados con la composición y distribución de la economía en general, de la producción, de la distribución y del consumo de bienes en la sociedad. Así como de la composición y distribución de la riqueza generada en la sociedad.

## **2. CAMBIOS ESTRUCTURALES Y DESEQUILIBRIOS EN EL DESARROLLO AGRÍCOLA**

El nivel de desarrollo, en términos específicos, alcanzado por un país o una región, de acuerdo con las teorías económicas se manifiesta en la magnitud de la población que vive en el campo (rural) y en la que vive en la ciudad (urbana); en la que trabaja en el campo (PEA agrícola) y en la que trabaja en los otros sectores económicos (PEA industrial y servicios); en el valor generado por persona ocupada en los diferentes sectores económicos y en el valor generado en los diferentes sectores económicos (producto interno bruto primario, secundario y terciario); y en el valor que le corresponde a cada habitante (producto interno bruto per cápita); entre otros.

Las teorías y las experiencias a nivel mundial, especialmente en los países desarrollados, del desarrollo económico en el campo reflejan un proceso de disminución relativa del peso de la agricultura y un incremento de la participación absoluta y relativa de los sectores industrial y de servicios en la economía (Lenin, 1974). La disminución del papel de la agricultura en la economía se manifiesta en la reducción de la población rural y el aumento de la población urbana; la disminución de la población económicamente activa ocupada en el sector agropecuario y forestal, y el aumento de la población económicamente activa ocupada en la industria y servicios; la disminución del producto interno bruto del sector primario y el aumento del producto interno bruto de los otros sectores; entre otros.

El desarrollo económico en la inmensa mayoría de los países, especialmente en los que se encuentran en vías de desarrollo, es un problema muy fuerte, el cual se manifiesta en desequilibrios regionales, desequilibrios sectoriales, altos índices de desempleo y subempleo, bajos niveles de productividad de la mano de obra, bajos niveles de valor agregado por trabajador, bajos niveles de ingreso, altos niveles de pobreza, entre otros. Este proceso de desarrollo económico, actualmente está siendo directamente impactado por las grandes transformaciones estructurales y procesos de integración económica que se vienen desarrollando en todo el mundo.

En México existen graves problemas de desarrollo económico, tales como rezagos económicos y sociales, el desarrollo desigual de los diferentes sectores económicos, los niveles heterogéneos de desarrollo de las diferentes regiones que componen el país, entre otros. En específico se observa heterogeneidad en las proporciones de la población rural, de la población económicamente activa ocupada en el campo, del producto interno bruto del sector primario y del producto interno bruto per cápita en los diferentes estados y regiones de México. Así mismo, se observa una gran desproporción entre el peso del producto interno bruto primario en el conjunto de la economía y el peso de la población económicamente activa ocupada en el sector primario y la población rural.

La desproporción entre la población económicamente activa ocupada y la magnitud y proporción del producto interno bruto producido en el sector primario genera problemas de baja productividad, bajos niveles de valor generado por trabajador, bajos niveles de riqueza generada por habitante y altos niveles de pobreza en el campo mexicano, especialmente en el Centro Sur de México.

En este marco, en los siguientes apartados se realiza la caracterización de la estructura económica de México, del Centro Sur de México y de las zonas económicas del Centro Sur de México, con base en las principales variables relacionadas con el desarrollo económico, en el contexto de las teorías económicas del desarrollo y con los indicadores de desarrollo económico.

## **3. VARIABLES DE DESARROLLO ECONÓMICO Y REGIONALIZACIÓN DEL CENTRO SUR**

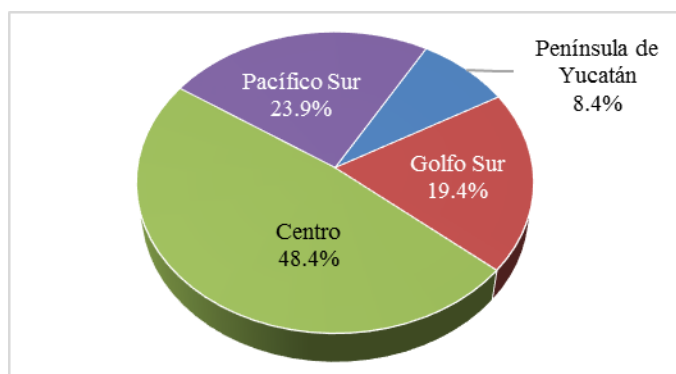


Las principales variables relacionadas con el desarrollo económico son: población total y rural; población económicamente activa total y del sector primario; producto interno bruto total y del sector primario; producto interno bruto global por trabajador ocupado y por trabajador ocupado en el sector primario; y producto interno bruto per cápita total y producto interno bruto per cápita en el sector rural. Así mismo, las principales variables señaladas permiten regionalizar al Centro Sur de México en cuatro grandes zonas, las cuales son: Pacífico Sur, Golfo Sur, Península de Yucatán y Centro. La zona Pacífico Sur está formada por los estados de Chiapas, Guerrero y Oaxaca; la Golfo Sur por los estados de Tabasco y Veracruz; la Península por los estados de Campeche, Quintana Roo y Yucatán; y, finalmente, la Centro por los estados de Puebla, Morelos, México y Tlaxcala.

### 3.1. Población Total

La población se refiere al conjunto de personas que habitan un territorio en un lugar y tiempo determinados (CESOP, 2006). La población es el resultado o balance, en un momento dado y en un determinado territorio, del conjunto de personas que ahí se encuentran. Económicamente, población se refiere al número de habitantes y sus características demográficas, en su calidad de fuente de fuerza de trabajo.

La población total de la Región Centro Sur de México representa un alto porcentaje (44.6%) de la población total de México. Las zonas con mayor proporción de población total en la Región Centro Sur son Centro (48.4%), Pacífico Sur (23.9%) y Golfo Sur (19.4%). La zona con menor proporción de población total es la Península de Yucatán (8.4%).



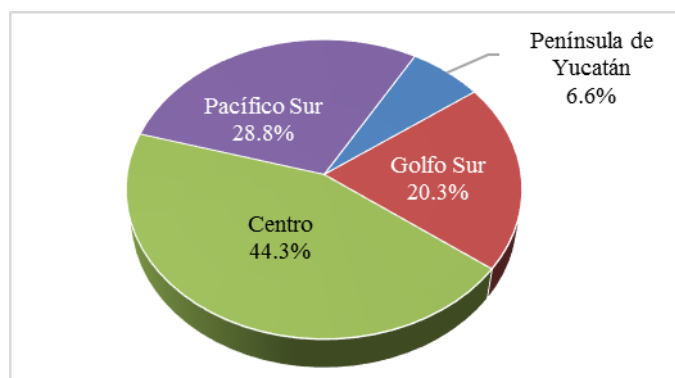
Gráfica 1. Distribución de la población total por zonas de la Región Centro Sur.

Fuente: Elaboración propia con datos de la Presidencia de la República (2015).

### 3.2. Población Rural

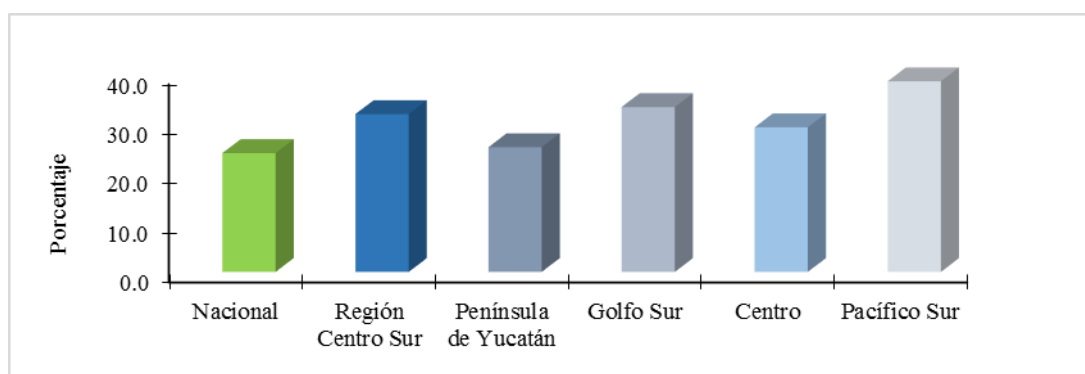
De acuerdo con el INEGI (2015c), una población se considera rural cuando tiene menos de 2,500 habitantes, mientras que la población urbana es aquella donde viven 2,500 o más personas.

La población rural de la Región Centro Sur de México representa un alto porcentaje (59.3%) de la población rural de México. Las zonas que concentran mayor población rural de la Región Centro Sur son Centro (44.3%), Pacífico Sur (28.8%) y Golfo Sur (20.3%). La zona que concentra menor proporción de población rural es la Península de Yucatán (6.6%).



Gráfica 2. Distribución de la población rural por zonas de la Región Centro Sur.  
Fuente: Elaboración propia con datos de la Presidencia de la República (2015).

La proporción de población rural respecto a la población total promedio de la Región Centro Sur de México (31.9%) es mayor a la promedio nacional (24.0%). Las zonas con mayor proporción de población rural en la Región Centro Sur son Pacífico Sur (38.5%) y Golfo Sur (33.3%), mayores a la media nacional y regional. Las zonas con menor proporción de población rural son Centro (29.2%) y Península de Yucatán (25.2%), mayores a la media nacional pero menores a la media regional.

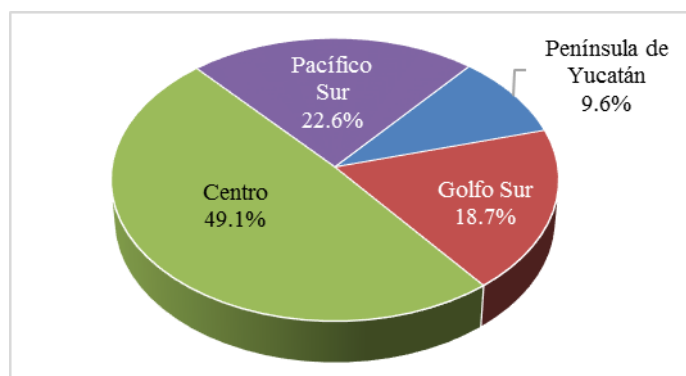


Gráfica 3. Proporción de la población rural por zonas de la Región Centro Sur.  
Fuente: Elaboración propia con datos de la Presidencia de la República (2015).

### 3.3. Población Económicamente Activa Ocupada

La población económicamente activa ocupada (PEAO) es la población de 12 y más años que participa en alguna actividad productiva, cumpliendo un horario definido y recibiendo un sueldo o salario (INEGI, 2010). Es la población que trabaja.

La PEAO de la Región Centro Sur de México representa un alto porcentaje (43.6%) de la PEAO de México. Las zonas de la Región Centro Sur con mayor proporción de PEAO son Centro (49.1%), Pacífico Sur (22.6%) y Golfo Sur (18.7%). La zona con menor proporción de PEAO en la Región Centro Sur es la Península de Yucatán (9.6%).

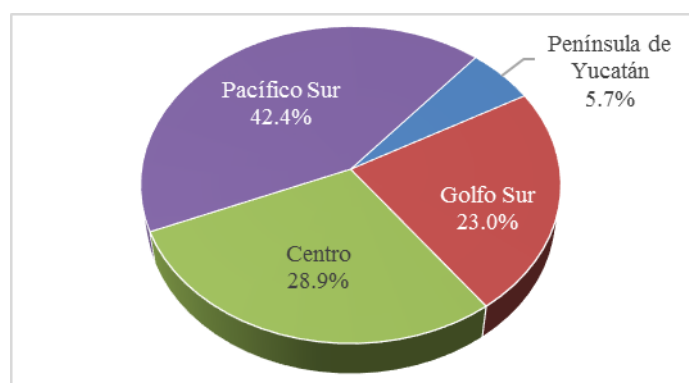


Gráfica 4. Distribución de la PEAO por zonas de la Región Centro Sur.  
Fuente: Elaboración propia con datos de la ENOE (INEGI, 2015b).

### 3.4. Población Económicamente Activa Ocupada En El Sector Primario

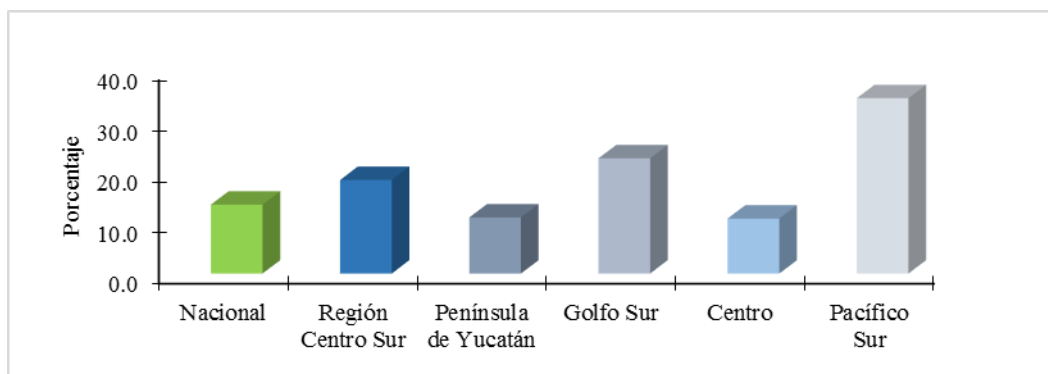
La población económicamente activa ocupada del sector primario (PEAOSI) es el conjunto de la PEAO localizada únicamente en el sector primario. El sector primario, llamado también sector agropecuario; se encuentra integrado por la agricultura, ganadería, silvicultura, caza y pesca; el sector secundario o sector industrial de la economía, se divide en dos subsectores, industria extractiva e industria de la transformación. La industria extractiva se integra por la extracción de petróleo y la minería, mientras que la industria de transformación incluye todas las demás ramas industriales como el envasado de frutas y legumbres, refrescos embotellados, abonos y fertilizantes, vehículos, cemento, aparatos electrodomésticos, entre otros. El sector terciario o también conocido como sector servicios de la economía, incluye todas aquellas actividades no productivas pero necesarias para el funcionamiento de la economía. Algunas ramas del sector servicios son el comercio, restaurantes, hoteles, transporte, servicios de educación, gobierno, entre otros.

La PEAOSI de la Región Centro Sur de México representa un alto porcentaje (59.2%) de la PEAOSI de México. Las zonas con mayor proporción de PEAOSI de la Región Centro Sur son Pacífico Sur (42.4%), Centro (28.9%) y Golfo Sur (23.0%). La zona con menor proporción de PEAOSI es la Península de Yucatán (5.7%).



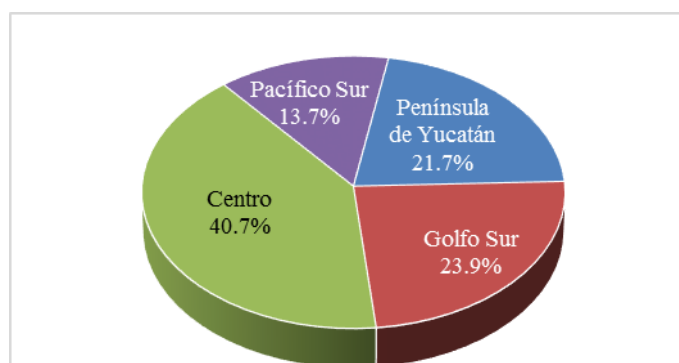
Gráfica 5. Distribución de la PEAOSI por zonas de la Región Centro Sur, 2013.  
Fuente: Elaboración propia con datos de la ENOE (INEGI, 2015b).

La PEAOSI promedio, respecto a la PEAO total, en la Región Centro Sur de México es mayor (18.4%) a la media nacional (13.5%). Las zonas con mayor proporción de PEAOSI de la región son Pacífico Sur (34.5%) y Golfo Sur (22.6%), mayores a la media nacional y regional. Las zonas con menor proporción de PEAOSI son Península de Yucatán (11.0%) y Centro (10.8%), menores a la media nacional y regional.



Gráfica 6. Proporción de la PEAOSI respecto a la PEAO por zonas de la Región Centro Sur. Fuente: Elaboración propia con datos de INEGI (2015b).

### 3.5. Producto Interno Bruto Total



Gráfica 7. Distribución del PIB por zonas de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

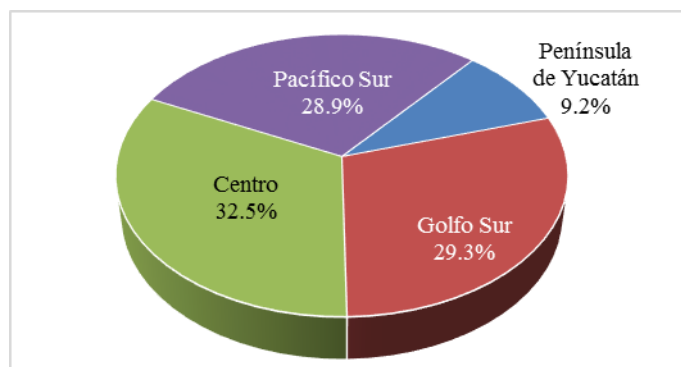
El producto interno bruto (PIB) es el valor monetario de todos los bienes y servicios finales producidos en un país durante un determinado periodo (Dornbusch et al., 2004).

El PIB total de la Región Centro Sur de México representa un bajo porcentaje (35.2%) del PIB total de México. Las zonas con mayor participación del PIB total de la Región Centro Sur son Centro (40.7%), Golfo Sur (23.9%) y Península de Yucatán (21.7%). La zona con menor participación en el PIB total de la región es Pacífico Sur (13.7%).

### 3.6. Producto Interno Bruto Del Sector Primario

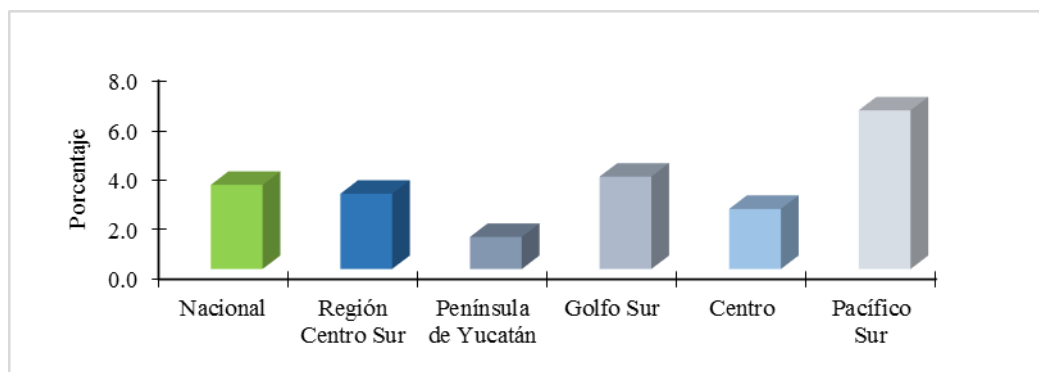
El producto interno bruto del sector primario (PIBSI) es el PIB generado exclusivamente por el sector primario. Incluye todas aquellas actividades donde los recursos naturales se aprovechan tal como se obtienen de la naturaleza, ya sea para alimento o para generar materias primas, las cuales corresponden a la agricultura, la silvicultura, la ganadería, la minería y la pesca (INEGI, 2015a).

El PIBSI de la Región Centro Sur representa un bajo porcentaje (31.5%) del PIBSI de México. Las zonas con mayor participación en el PIBSI de la Región Centro Sur son Centro (32.5%), Golfo Sur (29.3%) y Pacífico Sur (28.9%). La zona con menor participación en el PIBSI de la Región Centro Sur es la Península de Yucatán (9.2%).



Gráfica 8. Distribución del PIB del sector primario por zona de la Región Centro Sur, 2013.  
Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

El PIBSI respecto al PIB total en la Región Centro Sur de México es ligeramente menor (3.0%) al promedio nacional (3.4%). Las zonas de la Región Centro Sur con mayor proporción del PIBSI en el conjunto de la economía son Pacífico Sur (6.4%) y Golfo Sur (3.7%), mayores a la media nacional y regional. Las zonas con menor proporción de PIBSI son Península de Yucatán (1.3%) y Centro (2.4%), menores a la media nacional y regional.



Gráfica 9. Proporción del PIB del sector primario respecto al PIB total, por zonas de la Región Centro Sur, 2013.

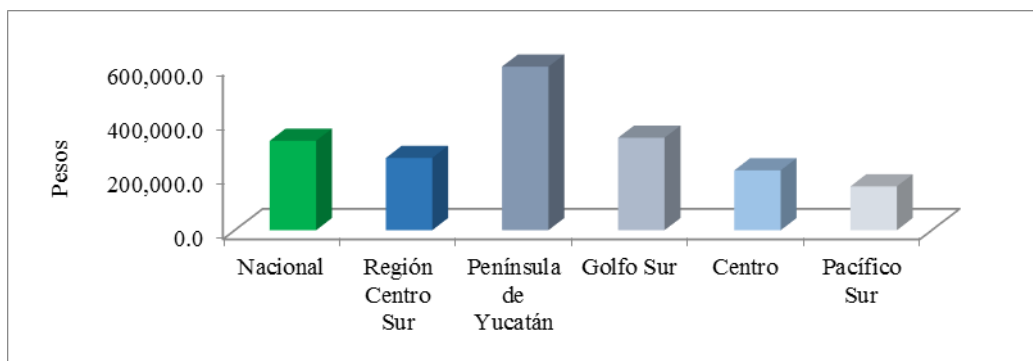
Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

### 3.7. Producto Interno Bruto Total Per Cápita De La Población Económicamente Activa Ocupada

El producto interno bruto per cápita de la población económicamente activa ocupada (PIBPPEAO) es el producto interno bruto total, dividido por la población económicamente activa ocupada total. Es el producto interno bruto per cápita de la población que trabaja.

El PIBPPEAO promedio en la Región Centro Sur de México (264,714.8 pesos) es menor al promedio nacional (327,489.8 pesos). Las zonas con menores niveles de PIBPPEAO son Pacífico Sur

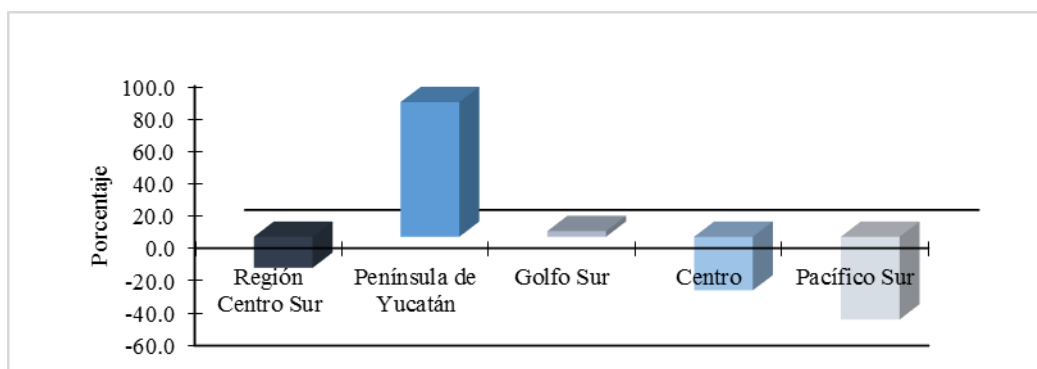
(160,490.6 pesos) y Centro (219,343.2 pesos), menores a la media nacional y regional. Las zonas con mayor nivel de producto interno bruto promedio generado por la población económicamente activa ocupada son la Península de Yucatán (599,263.9 pesos) y Golfo Sur (338,890.0 pesos) mayores a la media nacional y regional, lo anterior se explica principalmente, por el impacto de la producción de petróleo de los estados de Campeche y Tabasco y a la actividad turística del estado de Quintana Roo.



Gráfica 10. PIBPPEAO por zonas de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

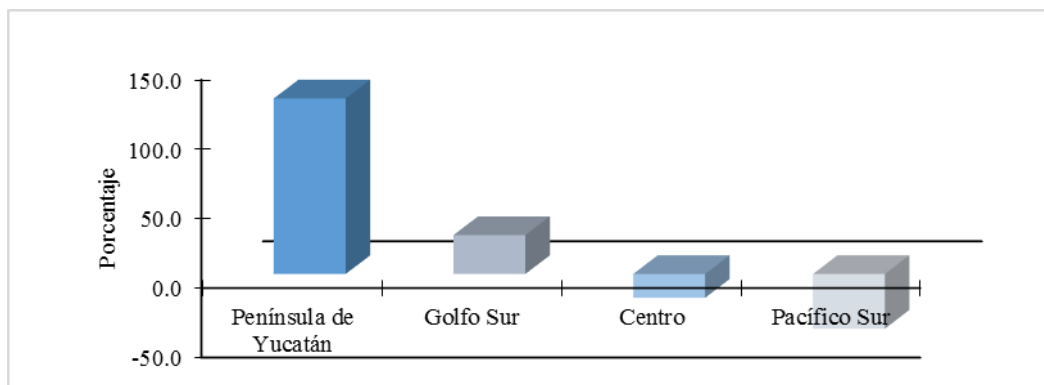
Con relación al PIBPPEAO promedio nacional, el de la Región Centro Sur es inferior en 19.2% a la media nacional. Las zonas con PIBPPEAO promedio menores a la media nacional son Pacifico Sur en 51.0% y Centro en 33.0%, mientras que las zonas que se encuentra por arriba de la media nacional en 83.0 y 3.5% son Península de Yucatán y Golfo Sur, respectivamente.



Gráfica 11. Diferencia del PIBPPEAO promedio de la Región Centro Sur, respecto al promedio nacional, 2013.

Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

Tomando como base la Región Centro Sur, se observa que el PIBPPEAO promedio de las zonas Península de Yucatán y Golfo Sur se encuentran por arriba de la media regional en 126.4 y 28.0%, mientras que las zonas Centro y Pacifico Sur se encuentran por debajo de la media regional en 17.1 y 39.4%, respectivamente.



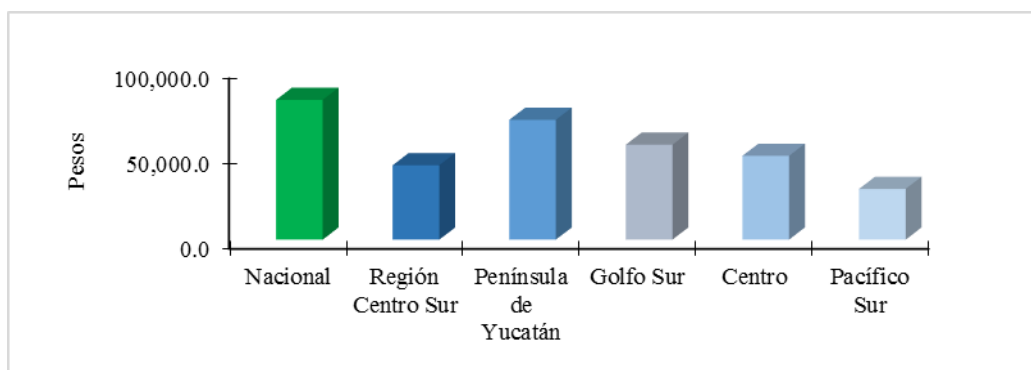
Gráfica 12. Diferencia del PIBPPEAO promedio por zonas, respecto al promedio de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

### 3.8. Producto Interno Bruto Per Cápita De La Población Económicamente Activa En El Sector Primario

El producto interno bruto per cápita de la población económicamente activa ocupada del sector primario (PIBPPEAOSI) es el producto interno bruto total del sector primario dividido por la PEAO del sector primario. Es el producto interno bruto per cápita de la población que trabaja en el sector primario.

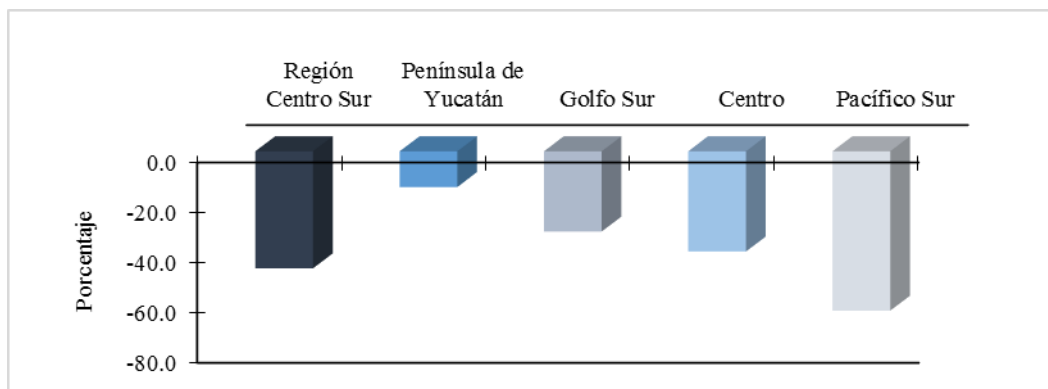
El PIBPPEAOSI en la Región Centro Sur de México es inferior (43,601.1 pesos) al promedio nacional (81,899.3 pesos). Todas las zonas de la Región Centro Sur presentan niveles de PIBPPEAOSI promedio, inferiores a la media nacional, mientras que las zonas de la Región Centro Sur que presentan niveles de PIBPPEAOSI promedio, superiores a la media regional son Península de Yucatán (70,191.5 pesos), Golfo Sur (55,633.7 pesos) y Centro (49,096.0 pesos). La zona que presenta un PIBPPEAOSI promedio inferior a la media regional y nacional es Pacífico Sur (29,733.7 pesos).



Gráfica 13. PIBPPEAOSI por zonas de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

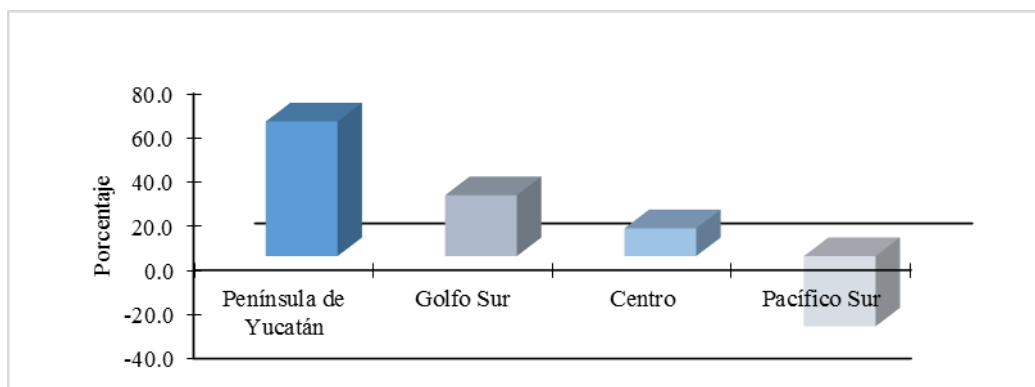
El PIBPPEAOSI promedio de la Región Centro Sur es inferior a la media nacional en 46.8%. El PIBPPEAOSI de las zonas es inferior a la media nacional en 63.7% en Pacífico Sur, en 40.1% en Centro, en 32.1% en Golfo Sur y en 14.3% en Península de Yucatán.



Gráfica 14. Diferencia del PIBPPEAOSI promedio de la Región Centro Sur respecto al promedio nacional, 2013.

Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

El PIBPPEAOSI promedio de las zonas Península de Yucatán, Golfo Sur y Centro, superan a la media regional en 61.0, 27.6 y 12.6%, respectivamente, mientras que el de la zona Pacífico Sur se encuentra por debajo de la media regional en 31.8%.



Gráfica 15. Diferencia del PIBPPEAOSI promedio por zonas, respecto al promedio de la Región Centro Sur, 2013.

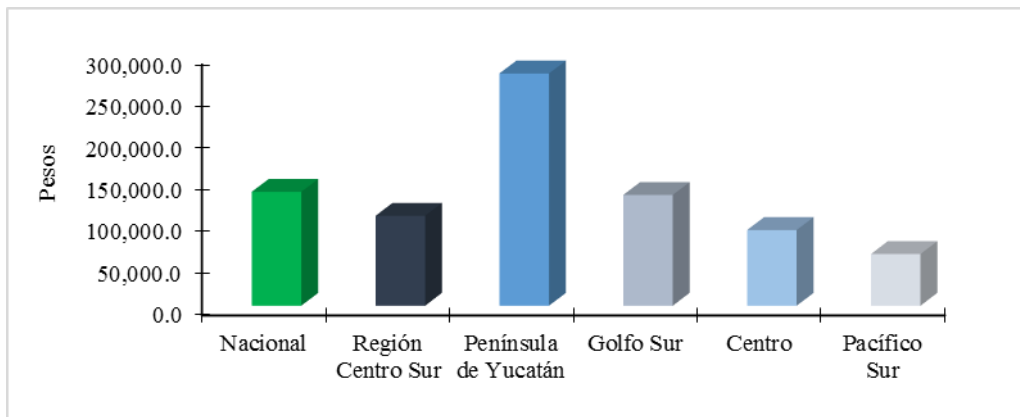
Fuente: Elaboración propia con datos de INEGI (2015b, 2015d) y Banco Mundial (2015a).

### 3.9. PRODUCTO INTERNO BRUTO PER CÁPITA GLOBAL

El producto interno bruto per cápita (PIB per cápita) es el producto interno bruto total dividido por la población total a mitad o a final del año (Banco Mundial, 2015b).

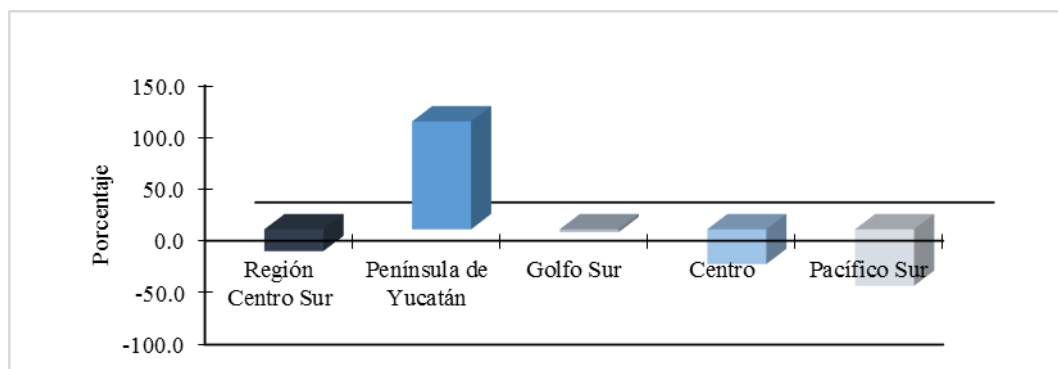
El PIB per cápita en la Región Centro Sur de México (107,501.9 pesos) es menor a la media nacional (136,166.5 pesos). Las zonas con menores niveles de PIB per cápita promedio con respecto a la media nacional y regional son Pacífico Sur (61,804.1 pesos) y Centro (90,511.8 pesos), mientras que en la zona Golfo Sur el PIB per cápita (132,459.2 pesos) es menor a la media nacional, pero mayor a la media regional. La zona de la Región Centro Sur que presenta el mayor PIB per cápita, superior a la media nacional y regional, es Península de Yucatán con 277,714.3 pesos, ésto se debe a que incluye al estado de Campeche, que es el primer productor de petróleo de México, siendo esa la principal actividad económica del estado, además de que en el 2013, fue el que presentó el mayor PIB per cápita del país. Sin considerar a este estado, el PIB per cápita de la región se reduce a 95,058.56 pesos, menor a la media nacional y regional.





Gráfica 16. PIB per cápita global promedio por zonas de la Región Centro Sur, 2013.  
Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

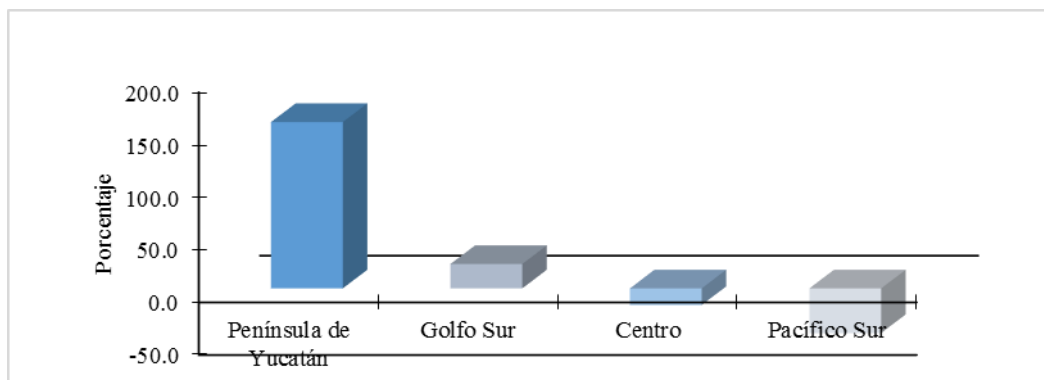
El PIB per cápita promedio de la Región Centro Sur es inferior a la media nacional en 21.1%. En las zonas Pacífico Sur, Centro y Golfo Sur el PIB per cápita es inferior a la media nacional en 54.6, 33.5 y 2.7%, respectivamente, mientras que en la zona Península de Yucatán es superior en 104.0%, considerando al estado de Campeche.



Gráfica 17. Diferencia del PIB per cápita promedio de la Región Centro Sur respecto al promedio nacional, 2013.

Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

Tomando como referencia el PIB per cápita promedio regional, se tiene que en las zonas Pacífico Sur y Centro es inferior a la media regional en 42.5 y 15.8%, por otro lado, las zonas Península de Yucatán y Golfo Sur superan a la media regional en 158.3 y 23.2%, respectivamente.



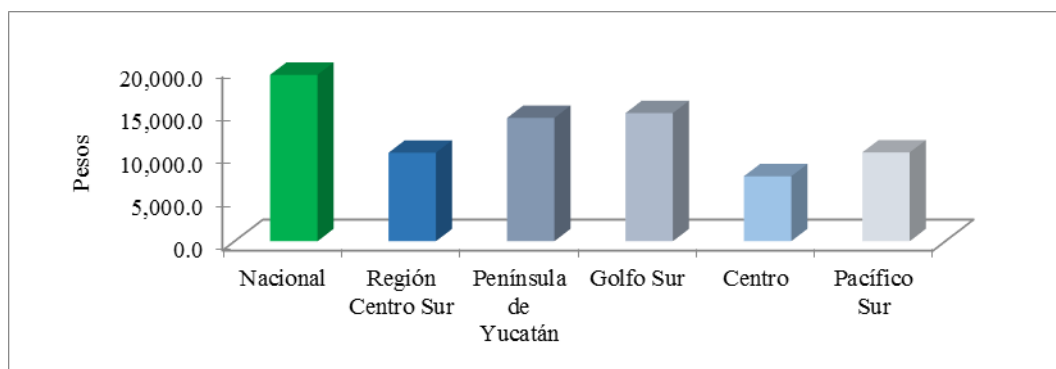
Gráfica 18. Diferencia del PIB per cápita promedio por zonas, respecto al promedio de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

### 3.9.1. Producto Interno Bruto Per Cápita De La Población Rural

El producto interno bruto per cápita de la población rural (PIB per cápita rural) es el producto interno bruto total del sector primario dividido por la población rural. Es el producto interno bruto per cápita de la población del sector rural.

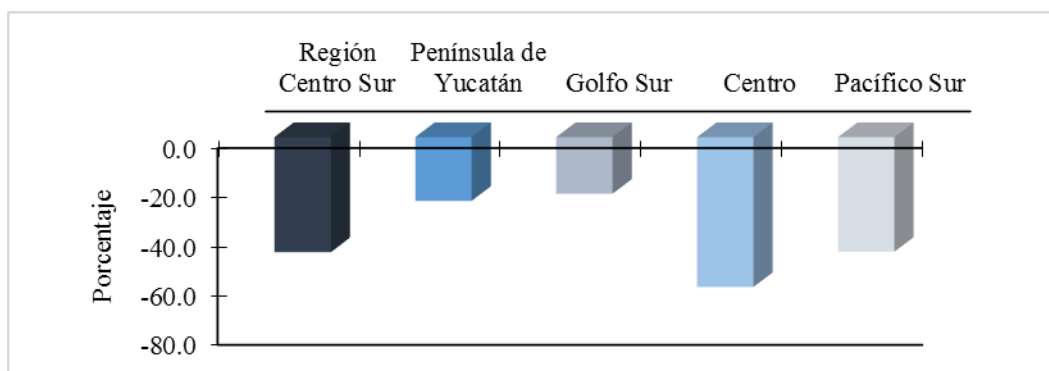
El PIB per cápita rural promedio de la Región Centro Sur de México (10,213.4 pesos) es menor al promedio nacional (19,204.9 pesos). Todas las zonas de la Región Centro Sur presentan un PIB per cápita rural promedio inferior a la media nacional. Las zonas con los más altos niveles de PIB per cápita rural promedio son Golfo Sur (14,785.5 pesos), Península de Yucatán (14,224.4 pesos), Pacífico Sur (10,249.8 pesos) y Centro (7,497.6 pesos). Las zonas que tienen el PIB per cápita rural promedio superior a la media regional son Golfo Sur y Península de Yucatán, mientras las que se encuentran por debajo de la media regional son las zonas Centro y Pacífico Sur.



Gráfica 19. PIB Per Cápita Rural Promedio De Las Zonas De La Región Centro Sur, 2013.

Fuente: Elaboración Propia Con Datos De INEGI (2015d) Y Banco Mundial (2015a).

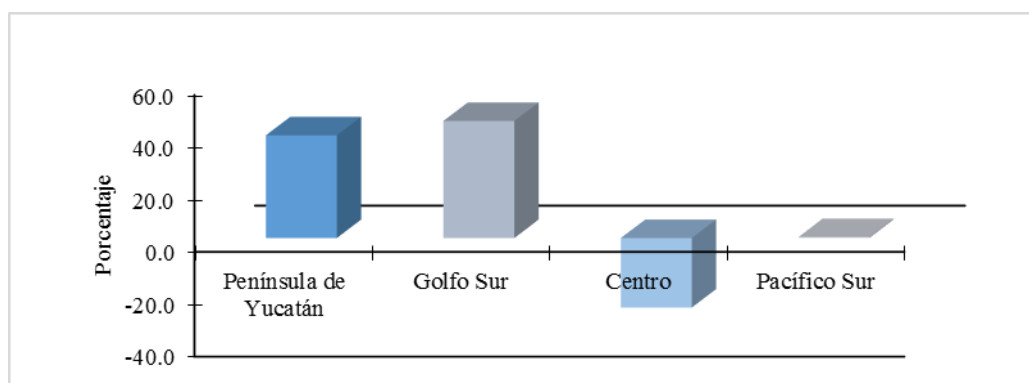
El PIB per cápita rural promedio de la Región Centro Sur es menor a la media nacional en 46.8%. El PIB per cápita rural promedio en todas las zonas es inferior a la media nacional, 61.0% en Centro, 46.6% en Pacífico Sur, 25.9% en Península de Yucatán y 23.0% en Golfo Sur.



Gráfica 20. Diferencia del PIB per cápita rural promedio de la Región Centro Sur respecto al promedio nacional, 2013.

Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

Tomando en cuenta el PIB per cápita rural promedio de la Región Centro Sur, las zonas que superan ese valor son Golfo Sur en 44.8% y Península de Yucatán en 39.3%, mientras que las zonas Centro y Pacífico Sur se encuentran por debajo de la media regional en 26.6 y 0.4%, respectivamente.



Gráfica 21. Diferencia del PIB per cápita rural promedio por zonas, respecto al promedio de la Región Centro Sur, 2013.

Fuente: Elaboración propia con datos de INEGI (2015d) y Banco Mundial (2015a).

#### 4. CONCLUSIONES

La alta proporción de población rural, de población económicamente activa ocupada en el sector agropecuario y de producto interno bruto del sector primario de la Región Centro Sur, reflejan el bajo nivel de desarrollo de esta región, especialmente de las zonas Pacífico Sur y Golfo Sur, y un gran desequilibrio entre los diferentes sectores económicos, el cual se observa en la gran cantidad de la población económicamente activa ocupada en el sector primario, agricultura, y la baja proporción del producto interno bruto generado en ese sector con relación al conjunto de la economía, lo que refleja la baja productividad. En ese contexto, también, los bajos niveles del producto interno bruto per cápita de la población económicamente activa en el sector primario, la alta proporción de la población rural y el alto porcentaje de la población económicamente activa ocupada en el campo de la Región Centro Sur de México reflejan los bajos niveles de producción, de productividad y de ingreso, y explican los altos niveles de pobreza en esta región, especialmente de las zonas Pacífico Sur y Golfo Sur.

Así mismo, los datos reflejan que el nivel de desarrollo económico de la Región Centro Sur es inferior a la media nacional y las zonas con menor nivel de desarrollo son Pacífico Sur y Golfo Sur, las

cuales se encuentran por debajo de la media regional. La situación descrita demanda la creación de programas de desarrollo económico que permita un mayor desarrollo de los sectores no agrícolas, el otorgamiento de facilidades para la canalización de inversiones, así como de programas de desarrollo agrícola que permitan el crecimiento y desarrollo del sector agrícola, y se incremente el ingreso per cápita y el nivel de bienestar del sector rural de la Región Centro Sur de México.

## REFERENCIAS

1. Banco Mundial. (2015a). *Datos de México*. [En línea] Disponible en: <<http://datos.bancomundial.org/pais/mexico>>
2. Banco Mundial. (2015b). *PIB per cápita (US\$ a precios actuales)*. Datos del Banco Mundial. [En línea] Consultado el 23 de mayo de 2015. Disponible en: <<http://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD>>
3. Baran, P. (1975). *Economía Política del Crecimiento*. México, FCE.
4. Centro de Estudios Sociales y de Opinión Pública (CESOP). (2006). *Definición, en Migración, Frontera y Población*. México, H. Cámara de Diputados. [En línea] Consultado el 07 de mayo de 2015. Disponible en: <[http://archivos.diputados.gob.mx/Centros\\_Estudio/Cesop/Comisiones/d\\_poblacion.htm](http://archivos.diputados.gob.mx/Centros_Estudio/Cesop/Comisiones/d_poblacion.htm)>
5. Dornbush, R., et al. (2004). *Macroeconomía*, novena edición, Colombia, McGraw-Hill.
6. INEGI. (2010). *Población ocupada*. Glosario del Censo de Población y Vivienda 2010. [En línea] Consultado el 13 mayo de 2015. Disponible en: <<http://www3.inegi.org.mx/sistemas/glosario/default.aspx?clvglo=cpv2010&s=est&c=27432>>
7. INEGI. (2015a). *Economía de México. Sector Primario*. [En línea] Consultado el 03 de mayo de 2015. Disponible en: <<http://cuentame.inegi.org.mx/economia/default.aspx?tema=E>>
8. INEGI. (2015b). *Encuesta Nacional de Ocupación y Empleo (ENOE)*, Consulta interactiva de indicadores estratégicos. México. [En línea] Disponible en: <[http://www3.inegi.org.mx/Sistemas/infoenoe/Default\\_CONAPO.aspx?s=est&c=27736](http://www3.inegi.org.mx/Sistemas/infoenoe/Default_CONAPO.aspx?s=est&c=27736)>
9. INEGI. (2015c). *Población rural y urbana*. [En línea] Consultado el 03 de mayo de 2015. Disponible en: <[http://cuentame.inegi.org.mx/poblacion/rur\\_urb.aspx?tema=P](http://cuentame.inegi.org.mx/poblacion/rur_urb.aspx?tema=P)>
10. INEGI. (2015d). *Producto interno bruto por entidad federativa*. México. Banco de información económica (BIE). [En línea] Disponible en: <<http://www.inegi.org.mx/sistemas/bie/>>
11. Lenin, V. I. (1974). *El Desarrollo del Capitalismo en Rusia*. Moscú, Editorial Progreso.
12. Presidencia de la República. (2015). *Segundo informe de gobierno, 2013-2014*. México. [En línea] Disponible en: <<http://www.presidencia.gob.mx/segundoinforme/>>

## ANEXOS

### Anexo 1. Procedimientos de cálculo

Valores totales. Los valores totales se refieren a la suma de los valores parciales, los cuales se obtienen de la siguiente manera:

$$VT = \sum VP$$

Donde: VT=Valor total; VP=Valor parcial.

Valores parciales. Los valores parciales se refieren a los diferentes valores obtenidos en cada fase, aspecto o área del objeto estudiado, como pueden ser los costos de producción obtenidos en cada fase de la actividad agrícola, las diferentes características de la población, entre otros.

Proporción. La proporción se refiere al valor que representa la participación de un valor parcial con respecto de un total. El procedimiento de cálculo es:

$$\% = (VP / VT) * 100$$

Donde: %=Participación porcentual; VP=Valor parcial; VT=Valor total.

Valor nominal. El valor nominal se refiere a cualquier valor económico a precios de mercado, de cada momento: del día, de la semana, del mes, del semestre o del año. El valor nominal se puede obtener a partir de un valor base, en este caso el procedimiento de cálculo es:

$$VN = VB (1 + J_1)(1 + J_2) \dots (1 + J_n)$$

Donde: 1=Factor de arrastre; VN=Valor nominal, valor de cada momento; VB=Valor inicial o valor del momento base; J=Tasa de inflación en unidades; J<sub>1</sub>=tasa de inflación en el período 1, en unidades; J<sub>2</sub>=Tasa de inflación en el periodo 2, en unidades; J<sub>n</sub>=Tasa de inflación en el periodo n, en unidades.

Valor per cápita. El valor per cápita se refiere a un valor global (PIB total, PIB del sector primario, entre otros) dividido entre una determinada cantidad de población (población total, población económicamente activa total, población rural, población económicamente activa del sector primario, entre otros), con la finalidad de conocer la proporción o parte que le corresponde a cada miembro de la población. El procedimiento de cálculo se puede ilustrar de la siguiente manera:

$$V_{per} = PIB/Pob$$

Donde: V<sub>per</sub>=Valor per cápita; PIB=Producto interno bruto; Pob=Población.

## Anexo 2. Cuadros estadísticos

Tabla 1. Principales indicadores económicos globales y del sector primario de México, 2013.

Entidad	Población Total (Miles)	Población Rural (Miles)	PEAO (Miles)	PEAOSI (Miles)	PIB Total (Millones de pesos)	PIBSI (Millones de pesos)
EUM (Total)	118,395	28,425	49,227	6,666	16,121,442	545,902
EUM (Prom.)	3,700	888	1,538	208	503,795	17,059
RCS (Total)	52,838	16,850	21,458	3,947	5,680,184	172,091
RCS (Prom.)	4,403	1,404	1,788	329	473,349	14,341
Campeche	880	201	389	70	741,160	4,401
Quintana Roo	1,485	176	701	47	249,787	2,200
Yucatán	2,064	740	962	109	239,189	9,279
P.de Yucatán (Total)	4,430	1,116	2,053	226	1,230,136	15,880
P. de Yucatán (Prom.)	1,477	372	684	75	410,045	5,293
Tabasco	2,335	743	899	145	510,165	8,043
Veracruz	7,923	2,673	3,110	762	848,562	42,459
Golfo Sur (Total)	10,258	3,416	4,009	908	1,358,727	50,501
Golfo Sur (Prom.)	5,129	1,708	2,005	454	679,363	25,251
México	16,364	3,550	6,727	324	1,508,525	21,823
Morelos	1,874	732	782	83	192,485	6,210
Puebla	6,068	2,454	2,525	660	520,808	24,282
Tlaxcala	1,243	728	509	74	90,641	3,647
Centro (Total)	25,549	7,464	10,543	1,140	2,312,460	55,962
Centro (Prom.)	6,387	1,866	2,636	285	578,115	13,991
Chiapas	5,119	1,392	1,902	747	287,569	22,229
Guerrero	3,524	1,062	1,374	421	235,936	11,826
Oaxaca	3,959	2,399	1,576	504	255,356	15,692

Pacífico Sur (Total)	12,602	4,854	4,853	1,673	778,862	49,748
Pacífico Sur (Prom.)	4,201	1,618	1,618	558	259,621	16,583
Otros (Total)	65,557	11,576	27,770	2,719	10,441,258	373,811
Otros (Prom)	3,278	579	1,388	136	522,063	18,691

Donde: PEAO=Población económicamente activa ocupada; PEAOSI=Población económicamente activa ocupada en el sector primario; PIB=Producto interno bruto; PIBSI=Producto interno bruto del sector primario; EUM=Estados Unidos Mexicanos; RCS=Región Centro Sur.

Fuente: Presidencia de la República (2015); INEGI (2015b) y Banco Mundial (2015a).

Tabla 2. Principales indicadores económicos globales y del sector primario de México (Pesos), 2013.

Entidad	PIBTPEAO Total	PIB Per cápita	PIB PPEAOSI	PIB Per cápita SI
EUM (Total)	327,489.8	136,166.5	81,899.3	19,204.9
EUM (Prom.)	359,879.0	152,418.2	116,431.9	26,194.3
RCS (Total)	264,714.8	107,501.9	43,601.1	10,213.4
RCS (Prom.)	390,642.2	166,078.4	51,934.5	11,407.0
Campeche	1,903,103.5	841,940.1	62,874.5	21,929.4
Quintana Roo	356,170.2	168,206.4	46,528.1	12,523.1
Yucatán	248,641.5	115,875.0	85,163.9	12,538.6
P. de Yucatán (Total)	599,263.9	277,714.3	70,191.5	14,224.4
P. de Yucatán (Prom.)	835,971.7	375,340.5	64,855.5	15,663.7
Tabasco	567,410.4	218,532.8	55,284.3	10,827.4
Veracruz	272,828.9	107,098.4	55,700.3	15,885.4
Golfo Sur (Total)	338,890.0	132,459.2	55,633.7	14,785.5
Golfo Sur (Prom.)	420,119.7	162,815.6	55,492.3	13,356.4
México	224,256.9	92,184.4	67,451.1	6,146.6
Morelos	246,196.3	102,702.6	74,973.1	8,487.1
Puebla	206,278.3	85,834.3	36,800.8	9,895.3
Tlaxcala	177,986.5	72,939.2	49,513.7	5,010.2
Centro (Total)	219,343.2	90,511.8	49,096.0	7,497.6
Centro (Prom.)	213,679.5	88,415.1	57,184.7	7,384.8
Chiapas	151,167.7	56,174.7	29,739.2	15,967.1
Guerrero	171,680.8	66,953.1	28,064.5	11,132.5
Oaxaca	161,985.5	64,500.1	31,120.7	6,541.1
Pacífico Sur (Total)	160,490.6	61,804.1	29,733.7	10,249.8
Pacífico Sur (Prom.)	161,611.4	62,542.7	29,641.5	11,213.6
Otros (Total)	375,996.5	159,269.7	137,501.9	32,292.7
Otros (Prom)	341,421.1	144,222.0	155,130.3	35,066.8

Donde: PIBPPEAO=Producto interno bruto total per cápita de la población económicamente activa ocupada; PIB per cápita=Producto interno bruto per cápita; PIBPPEAOSI=Producto interno bruto per cápita de la población económicamente activa ocupada en el sector primario; PIB per cápita SI=Producto interno bruto per cápita en el sector primario; RCS=Región Centro Sur; EUM=Estados Unidos Mexicanos.

Fuente: Presidencia de la República (2015); INEGI (2015b) y Banco Mundial (2015a).

Tabla 3. Principales indicadores económicos globales y del sector primario de México (Dólares), 2013.

Entidad	PIBPPEAO	PIB Per cápita	PIBPPEAOSI	PIB Per cápita SI
EUM (Total)	25,641.2	10,661.3	6,412.4	1,503.7
EUM (Prom.)	28,177.2	11,933.8	9,116.2	2,050.9
RCS (Total)	20,726.2	8,417.0	3,413.8	799.7
RCS (Prom.)	30,585.8	13,003.3	4,066.3	893.1
Campeche	149,005.9	65,920.8	4,922.8	1,717.0
Quintana Roo	27,886.8	13,169.9	3,643.0	980.5
Yucatán	19,467.7	9,072.6	6,668.0	981.7
P. de Yucatán (Total)	46,920.1	21,744.0	5,495.7	1,113.7
P. de Yucatán (Prom.)	65,453.5	29,387.8	5,077.9	1,226.4
Tabasco	44,426.1	17,110.3	4,328.6	847.7
Veracruz	21,361.5	8,385.4	4,361.1	1,243.8
Golfo Sur (Total)	26,533.8	10,371.1	4,355.9	1,157.6
Golfo Sur (Prom.)	32,893.8	12,747.9	4,344.8	1,045.8
México	17,558.5	7,217.7	5,281.2	481.3
Morelos	19,276.3	8,041.2	5,870.1	664.5
Puebla	16,150.8	6,720.5	2,881.4	774.8
Tlaxcala	13,935.7	5,710.9	3,876.7	392.3
Centro (Total)	17,173.8	7,086.7	3,844.0	587.0
Centro (Prom.)	16,730.3	6,922.6	4,477.3	578.2
Chiapas	11,835.9	4,398.3	2,328.5	1,250.2
Guerrero	13,442.0	5,242.2	2,197.3	871.6
Oaxaca	12,682.9	5,050.1	2,436.6	512.1

Pacífico Sur (Total)	12,565.8	4,839.0	2,328.0	802.5
Pacífico Sur (Prom.)	12,653.6	4,896.9	2,320.8	878.0
Otros (Total)	29,439.1	12,470.2	10,765.9	2,528.4
Otros (Prom)	26,732.0	11,292.0	12,146.1	2,745.6

Donde: PIBTPPEAO=Producto interno bruto total per cápita de la población económicamente activa ocupada; PIB per cápita=Producto interno bruto per cápita; PIBPPEAOSI=Producto interno bruto per cápita de la población económicamente activa ocupada en el sector primario; PIB per cápita=Producto interno bruto per cápita; PIB per cápita SI=Producto interno bruto per cápita en el sector primario; RCS=Región Centro Sur; EUM=Estados Unidos Mexicanos.

Fuente: Presidencia de la República (2015); INEGI (2015b) y Banco Mundial (2015a).

*Capítulo 16**pp 167-179***EL DISEÑO DEL RECORRIDO PARA LA RECOGIDA DE RESIDUOS SOLIDOS. CASO DE ESTUDIO: MUNICIPIO PLAZA DE LA REVOLUCION**

Joanna Campbell Amos & Sira Allende Alonso  
Facultad Matemática Computación de la Universidad de la Habana, Cuba.

**ABSTRACT**

Waste management is an essential part of any municipality's service towards the public. This includes the collection, transport and disposal of solid waste. In recent years waste management has become an area of concern for municipalities worldwide due to population growth (especially in urban areas), environmental concerns and the progressive increase in waste management cost. These factors force municipalities to assess their solid waste management to identify improvement opportunities like the optimal location of dumping sites, location of waste bins, the number of vehicles required and the optimal routing of waste collection vehicles.

The Routing Design for the Waste Collection Problem can be developed by using Graph Theory, in which all arcs or edges of an oriented, not-oriented or mixed network are traversed; subject to vehicle capacity restrictions, and the goal is to minimize the total cost of the routes traversed by the waste collection vehicles.

The objective of this work is to design, for each sector of urban areas, the routes of wasted collection by reducing the distance to travel and the transportation cost of the vehicles, in such way that the total collection of solid wastes is guaranteed. To accomplish this goal, Vedado deal was chosen for designing a routing system. Vedado was divided in two sectors; each sector is assigned to a specialized collection truck. For each sector the collection was modeled as the Chinese Postman Problem on directed graphs, implementing an exact algorithm. Additionally, the geographic data for the graph were obtained from a Geographic Information System.

**KEYWORDS:** CPP, design routing, solid wasted

**RESUMEN.**

La gestión de los residuos sólidos constituye una función esencial de prestación de servicios a la comunidad. Esta incluye la recogida, el transporte y la eliminación de los residuos sólidos. Debido al incremento acelerado, en los últimos tiempos, de la generación de residuos sólidos, su impacto medioambiental y del aumento del costo de la gestión es de vital importancia la elaboración de mecanismos que permitan la disminución de los costes así como la total recogida de desechos sólidos.

El problema del diseño del recorrido para la recogida de residuos sólidos que puede ser modelado a través de la Teoría de Grafos, como un grafo, dirigido, no dirigido o mixto, en dependencia de la red vial, donde los arcos o aristas se corresponden con las calles o carreteras con costes asociados, generalmente distancias de atravesar la calle o carretera, las intersecciones representan el conjunto de vértices, cuyo objetivo es minimizar el coste total recorrido por los vehículos sujetos en general a determinadas restricciones de capacidad o barrenderos.

Este trabajo tiene por objetivo diseñar, para cada sector de zonas urbanas, las rutas de recogida de los residuos mediante la reducción de la distancia a recorrer, costo de transportación de los vehículos, de manera tal que se garantice la total recogida de desechos sólidos. Para lograr este objetivo, se escogió del municipio Plaza de la Revolución el reparto Vedado. El reparto Vedado fue dividido en dos sectores. Cada sector es signado a un camión especializado para la recogida. Para cada sector se modeló la recogida como un Problema del Cartero Chino sobre grafos dirigidos, implementando un algoritmo exacto. Además, los datos geográficos para obtener el grafo se obtuvieron de un Sistema de Información Geográfica.

**1. INTRODUCCIÓN**

Según la Ley General de Equilibrio Ecológico y Protección al medio Ambiente (1993), “los residuos sólidos son cualquier material generado en los procesos de extracción, beneficio, transformación, producción, consumo, utilización o tratamiento cuya finalidad no permite incluirlos nuevamente en el proceso que los generó”.



Los residuos sólidos o comúnmente llamados basura, los podemos clasificar según su composición en residuo orgánico, residuo inorgánico, residuos peligrosos. La gestión de residuos sólidos se puede clasificar en tres categorías: los residuos comerciales, residuos residenciales o domésticos, roll-off-residuos (Toth & Vigo, 2002).

En los últimos años las naciones del mundo industrializado han cuadruplicado su producción de desechos domésticos, incrementándose esta cifra en un 2% o en un 3% por año. El volumen de producción de desechos es directamente proporcional al nivel de desarrollo del país que se trate. El crecimiento poblacional y los estilos de vida de la población, hace que en las ciudades se produzcan un volumen de residuos sólidos muy superior a su capacidad de recogida y eliminación, y ese volumen aumenta a medida que se elevan los ingresos. El servicio de recolección en la mayoría de los países en desarrollo, consume un 30 % a 60% de las rentas municipales disponibles, a pesar de lo cual, queda sin recoger una gran parte de los residuos sólidos, permitiendo el surgimiento de tiraderos clandestinos, que se convierten en potenciales focos de infección en la ciudades.

Tanto la capacidad de los camiones, las distancias a recorrer por estos, el crecimiento acelerados de nuevos centros de población, así como el mal diseño de rutas, contribuyen al mal funcionamiento del servicio de recolección de residuos sólidos, imposibilitándolo de cumplir con sus objetivos.

En la mayoría de los países el diseño de las rutas se hace de manera intuitiva, posibilitando el incremento de los tiempos y gastos de combustible. Las personas también contribuyen al mal funcionamiento de la recogida de basura, al depositar en cualquier lugar sus residuos, obligando a los camiones a tener que detenerse por más tiempo y recorrer distancias más largas no contempladas en la ruta original (Obras Públicas, 2005).

En los últimos años, debido a una serie de aumentos en los costos de la recolección de basura, el impacto que tiene sobre la salud, y el medio ambiente, muchos países se han visto obligados a evaluar su gestión de los residuos sólidos y examinar la relación coste-eficacia y el impacto en el medio ambiente, en términos de diseño de rutas de recolección de residuos sólidos. Para ello se han apoyado en técnicas computacionales que permitan mejorar el diseño de rutas (Nuortioa, Kytöjokib, Niskaa, & Bräysy, 2005). Para lograr un buen diseño de las rutas se han basado en los siguientes aspectos:

1. El diseño de rutas debe aumentar la distancia productiva en relación con la distancia total.
2. Los recorridos no deben fragmentarse ni traslaparse. Cada uno debe consistir en tramos que queden dentro de una misma área de la ciudad o localidad en estudio
3. El inicio de una ruta debe estar cerca del punto de partida y el final cerca del lugar de disposición final de residuos
4. En lugares con pendientes elevadas o desniveles debe procurarse hacer el recorrido de la parte alta a la parte baja. Hay que procurar atenderlas al comienzo del viaje cuando el camión va con poca carga.
5. Tratar de recolectar simultáneamente a ambos lados de la calles. Sin embargo, ello no es aconsejable en avenidas anchas y de mucho tráfico.
6. Se debe respetar el sentido de circulación y la prohibición de ciertos virajes
7. Evitar los giros a la izquierda y las vueltas en U
8. Evitar los recorridos en horas de mucho tránsito
9. Evitar hacer el recorrido cuando hay estacionamiento de vehículos
10. En calles cortas o sin salida, es preferible que los vehículos no entren en ellas sino que esperen en la esquina y el personal vaya a buscar la basura o que los ciudadanos depositen la basura en la esquina
11. Cuando la recolección se hace simultáneamente a ambos lados de la calle, debe hacerse en recorridos largos y rectos, con pocas vueltas

12. Cuando la recolección se hace primero a un lado de la calle y después al otro, generalmente es mejor tener recorridos con muchas vueltas a la derecha alrededor de las manzanas

13. Es preciso conocer las características propias de la ciudad para que las rutas de los camiones de basura no causen muchos problemas (SEDESOL, 1999)

Es preciso conocer las características propias de la ciudad para que las rutas de los camiones de basura no causen muchos problemas (SEDESOL, 1999)

El trabajo se ha organizado en 7 secciones donde en la sección 1 se introduce la motivación del problema. La sección 2 describe como es la gestión de los residuos sólidos en la capital cubana. En la sección 3 se explica el problema de la gestión de la basura. La sección 4 se propone un modelo matemático con su correspondiente algoritmo para lograr un buen diseño del recorrido. En este caso de estudio se trabaja con el callejero de La Habana, escogiendo el municipio Plaza de la Revolución, del cual se extrae una región, Vedado, el cual es dividido en dos sectores, que son transformados en un grafo dirigido para aplicarle el algoritmo correspondiente. Finalizamos con un resumen de los principales resultados obtenidos en la investigación y las líneas de trabajo futuro identificadas.

## **2. LA RECOGIDA DE BASURA EN CIUDAD DE LA HABANA**

En Cuba el manejo de los residuos sólidos urbanos constituye también un problema aún no resuelto que se acrecienta por las limitaciones económicas propias de los países del tercer mundo, y de modo muy especial por ser un país bloqueado económica y tecnológicamente; no podía ser de otra forma. Una ciudad como La Habana, con más de dos millones de habitantes, genera diariamente alrededor de 17 000 metros cúbicos de desechos sólidos; por municipio, como promedio, entre 1 000 y 2 000.

Los medios de transporte que se utilizan en Cuba son muy variados, tales como equipos de tracción animal y vehículos motorizados que consumen gasolina o petróleo. Tomando como hipótesis que estos vehículos posean una capacidad para transportar once toneladas de RSU, con un consumo de siete litros de combustible promedio, tendremos que se necesitarán cien viajes diarios para trasladar los RSU hasta los vertederos establecidos, con un consumo de 700 L de combustible. La combustión de cada litro de petróleo que se consume genera más de 800 g de CO<sub>2</sub>, que si lo multiplicamos por la cantidad de combustible fósil que es necesario consumir diariamente para la transportación de más de cuatro mil toneladas diarias en todo el país constituyen volúmenes considerables.

Para que la basura no llegue al suelo la ciudad necesita 13 000 contenedores más y una mejor situación de los medios de transporte, fundamentalmente los especializados para la recogida. Son 62 los camiones colectores con lo que cuenta la ciudad para el trabajo con los contenedores, y de esos, 48 pertenecen a la Unidad Provincial de Higiene. 15 están paralizados permanentemente por roturas, y los 33 aptos para el trabajo tienen inestabilidad por ponches, alguna afectación transitoria o falta de luces, que imposibilita, en este último caso, el trabajo en horario nocturno. La etapa actual se caracteriza por una recogida no diferenciada de la basura, bajos niveles de aprovechamiento y la disposición final en los vertederos no controlados del territorio. Ausencia de una planificación adecuada de las actividades de barrido, recolección, almacenamiento, transporte y destino final de los residuos sólidos, determinada por limitaciones organizativas, de preparación del personal y económicas; carencia de políticas nacionales y territoriales relacionadas con la minimización de los residuos, reúso, ubicación de infraestructuras, capacitación y tecnologías; falta de un sistema de información básica para el control económico y de costos; insuficiencia en el sistema de control sobre la gestión de los residuos sólidos y sus afectaciones al ambiente, en especial, a nivel local; deficiente incorporación de la dimensión ambiental en los planes de desarrollo relacionados con esta temática;

## **3. EL PROBLEMA DE LA GESTIÓN DE LA BASURA**

El sistema de recolección de residuos sólidos incluye varias decisiones como la determinación de la frecuencia de recogida, la asignación de vehículos a los puntos de recolección, la ubicación de contenedores y el diseño de las rutas. El diseño del recorrido para la recolección de basura es el objetivo de este trabajo. Para el recorrido por vehículos se tomaron las siguientes hipótesis: es dado el sector de recogida con demanda homogénea con su vehículo o barrendero asignado. Se conoce además la ubicación de los contenedores. Además para el caso del recorrido por vehículos existe un solo depósito y un solo relleno sanitario y el vehículo asignado cuenta con suficiente capacidad para la recolección del sector asignado.

### **Consideraciones:**

1. Depósito: Un punto inicial del recorrido, de donde parten los vehículos para comenzar y al que deben regresar una vez concluida la recolección de basura.
2. Relleno Sanitario: Es el lugar de disposición final al cual se lleva la basura.
3. Flota de vehículos: Determinada por el número de vehículos, los tipos de vehículos, la capacidad de cada tipo de vehículo, la densidad de la basura en el vehículo recolector y su peso. Asumiremos que la flota es heterogénea.
4. Demanda: Cada punto de recolección tiene una cierta cantidad y tipos de residuos a ser recogidos que definen su demanda. La demanda de una zona urbana se estima a partir del área total, la densidad de población, la tasa de generación de residuos sólidos por habitante.
5. Sectores de usuarios: Zonas delimitadas geográficamente a las cuales se le asigna un vehículo para la recogida de desechos sólidos, de acuerdo a sus características. La sectorización de la región ha de tomar en cuenta las cantidades y tipos de demandas de los usuarios, desarrollo social, estado de las vías, etc.
6. Contenedores: En ciertas zonas el sistema de recogida es basado en contenedores de basura, diferenciados según tipos de residuos o no. Un problema de decisión importante es la ubicación de los contenedores, dado un conjunto de posibles puntos.
7. Red: Se asume conocida la red vial de la zona en que se realiza la recogida.

A partir de la red vial se construye el grafo asociado a la red  $G = (V, L)$ . Las calles y avenidas son representadas por arcos cuya orientación corresponde con el sentido de la circulación del tráfico o por aristas si es posible el tráfico en ambos sentidos. En consecuencia el grafo puede ser dirigido, no dirigido o mixto. Los vértices de  $G$  pueden representar las intersecciones de las calles o los posibles puntos de ubicación de los contenedores, según se trate de representar un sistema de recogida por acera o por puntos de contenedores. Se asume que la red es tal que para todo par de puntos  $X$  y  $Y$  de la red existe un camino de  $X$  a  $Y$  y un camino de  $Y$  a  $X$ . El grafo asociado a la red ha de ser fuertemente conexo. La selección de una ruta de recolección tomara necesariamente el costo de transporte y operación, que puede representar la distancia física a recorrer, el tiempo necesario ello u otro criterio; en el grafo asociado se da entonces un valor que llamaremos costo del arco/arista.

## **4. MODELACION MATEMATICA. ALGORITMO DE SOLUCION PARA EL DISEÑO DEL RECORRIDO PARA EL CASO DIRIGIDO**

De manera general se ha llegado al consenso que el diseño del recorrido de los vehículos para la recogida de residuos sólidos, se modelan como problemas de rutas por arcos con capacidad (capacity arc routing problem) como el Cpp (por sus siglas en inglés) o problemas de rutas por nodos con capacidad (capacity node routing problem por sus siglas en inglés). El tipo de modelo depende del tipo de servicio que se preste para recoger los residuos en una área geográfica específica, por ejemplo, normalmente si el servicio de recogida es residencial se trata como un problema de rutas por arcos, mientras que si es el caso del servicio de recogida de residuos comercial o industrial o de puntos de contenedores, sería un problema de rutas por nodos.

Antes de explicar el modelo matemático de optimización y algoritmo para resolver el problema del recorrido es necesario establecer algunas definiciones y teoremas para una mayor comprensión.

**Definición 1.** Sea  $G(V, A)$  un grafo dirigido, donde  $V$  es el conjunto de vértices,  $A$  el conjunto de arcos con costo  $c_a$  no negativos asociados. El grado de entrada de un vértice, denotémosle  $d^+(v)$  es el número de arcos que tienen a  $v$  como vértice final y el grado de salida de un vértice, denotémosle  $d^-(v)$ , es el número de arcos que tienen a  $v$  como vértice inicial. Un vértice es simétrico si tiene el mismo grado de entrada y salida, es decir,  $d^+(v) = d^-(v)$ . Un grafo dirigido es simétrico si todos sus vértices son simétricos.

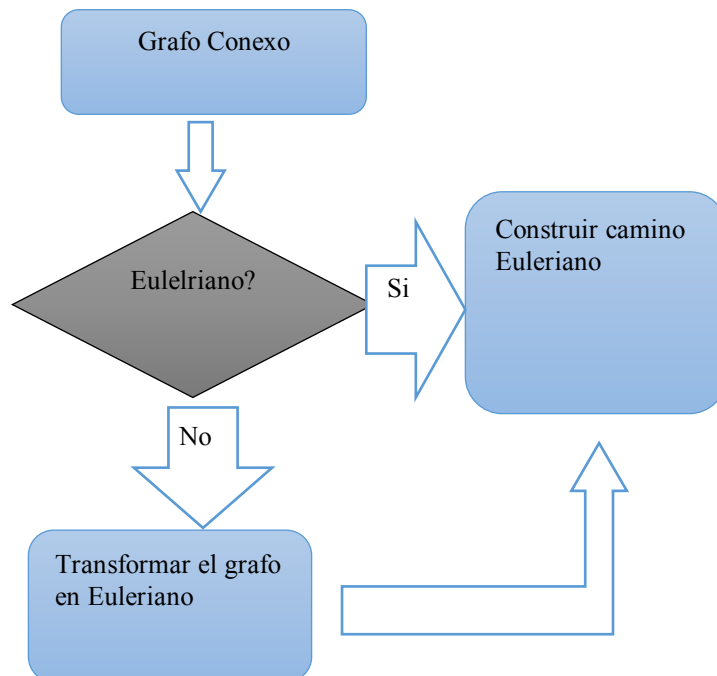
**Definición 2.** Un camino  $P$ , de  $v_0$  (vértice inicial) a  $v_k$  (vértice final) en un grafo dirigido es una sucesión de vértices y arcos  $P = (v_0, l_1, \dots, l_k, v_k)$  (los vértices y o arcos pueden repetirse, donde los arcos  $l_i$  están dirigidos desde  $v_{i-1}$  a  $v_i$ ). Diremos que los arcos son atravesados por el camino  $P$ . Un circuito es un camino cuyo vértice inicial y final coinciden. El costo de un camino, circuito o ciclo, es la suma de los costes de las aristas/arcos que contiene.

**Definición 3.** Un grafo dirigido es fuertemente conexo si cada par de vértices  $u$  y  $v$ , existe un camino de  $u$  a  $v$  y un camino de  $v$  a  $u$ .

**Definición 4.** El camino que contiene a todos los arcos de un grafo, se le denomina camino Euleriano. El camino es un camino cerrado, es decir, comienza y termina en el mismo vértice.

**Definición 5.** Si  $G$  contiene un camino Euleriano se denomina grafo Euleriano.

**Teorema 1.** Un grafo fuertemente conexo y dirigido, contiene un camino Euleriano ssi los grados de entrada y salida son iguales, es decir, el grafo es simétrico.



**Figura. 1:** Esquema de Flujo de la metodología para resolver el problema del cartero chino.

De manera general la metodología a seguir para determinar el recorrido óptimo por arcos del grafo, ya sea dirigido, no dirigido o mixto, es la siguiente:

1. Hacer simétrico el grafo.
2. Construir el camino Euleriano.

La forma de realizar estos dos pasos depende de la estructura del grafo; si el grafo es dirigido se resuelve como un problema del cartero chino dirigido (DCPP), de ser no dirigido se modela como un problema del cartero chino no dirigido (UCPP) y por ultimo si el grafo es mixto se soluciona como un problema del cartero chino mixto (MCPD).

#### 4.1. Formulación matemática para el DCPP

Para el caso direccionado si el grafo es simétrico entonces contiene un camino euleriano. En caso contrario, empleando el Teorema 1, el problema se reduce a determinar el conjunto de arcos de costo mínimo para transformar en grafo en simétrico. Edmonds y Johnson en 1973, Orloff en 1974 y Boldin y Beltrami en 1974 mostraron como un grafo Euleriano de costo mínimo puede ser construido resolviendo un problema de transporte. Sea  $I$  el conjunto de vértices  $v_i$  tal que la diferencia entre el número de arcos que entran en  $v_i$ , y el número de arcos que salen, sean igual a  $s_i$  y  $J$  el conjunto de vértices  $v_j$  tal que la diferencia entre el número de arcos que salen de  $v_j$  y el número que entran sea igual a  $d_j$ . De esta forma,  $s_i$  puede ser interpretado como una oferta en el vértice  $i$ , y  $d_j$  como una demanda en el vértice  $j$ . Además de esto,  $c_{ij}$  denotará el costo del menor camino de  $v_i$  a  $v_j$ . A continuación se presenta la formulación del modelo de PLI para el DCPP:

$$\text{Min} \sum_{v_i \in I} \sum_{v_j \in J} c_{ij} x_{ij} \quad (1)$$

s. a

$$\sum_{j \in J} x_{ij} = s_i \quad i \in I \quad (2)$$

$$\sum_{i \in I} x_{ij} = d_j \quad j \in J \quad (3)$$

$$x_{ij} \geq 0, \text{ entero } (v_i \in I, v_j \in J) \quad (4)$$

En esta formulación a la variable  $x_{ij}$  representa el número de veces extra que cada arco  $(i,j)$  será atravesado, es decir la cantidad de veces que hay que adicionar el arco  $(i,j)$  al grafo  $G$ . De esta forma, la función objetivo en la relación (1), busca aumentar solo aquellos arcos de menor costo. En (2), las restricciones garantizan que los arcos aumentados, partiendo del vértice  $v_i$ , atiendan sus ofertas, mientras que en (3) las restricciones garantizan que los arcos aumentados, entrando en los vértices  $v_j$ , atiendan sus demandas. En la relación (4), las restricciones garantizan la integridad de las variables.

#### 4.2. Solución algorítmica para el DCPP

Como se puede apreciar en la Figura 1 la metodología general para resolver el CPP puede dividirse en dos fases. Una fase es transformar el grafo en Euleriano, si éste no lo es, para después pasar a la fase de determinar el camino Euleriano.

Para esta última fase existen en la literatura algunos algoritmos exactos, que dado un grafo Euleriano permiten encontrar un camino Euleriano (Eiselt et al., 1995). Entre ellos se encuentran:

Algoritmo de Fleury para grafos no dirigidos, Algoritmo End- Pairing para grafos dirigidos y no dirigidos, Algoritmo de Van Aardenne-Ehrenfest y Brujin para grafos dirigidos.

El algoritmo “End- Pairing” puede ser utilizado para el caso dirigido y no dirigido. Además tiene la ventaja de ser flexible para soportar restricciones adicionales como las penalidades y prohibiciones en los giros.

Para cumplir con la fase 1, el modelo matemático para transformar el grafo dirigido en Euleriano puede ser resuelto como un problema de Flujo de Costo Mínimo. A continuación se mostrará en pseudocódigo del algoritmo de solución para resolver el problema DCPP.

1. Obtener los conjuntos I (conjunto de vértices de grado positivo) y J (conjunto de vértices de grado negativo)
2. Calcular los caminos mínimos entre cualquier par de vértices, empleando el algoritmo de Floyd –Warshall
3. Sobre el grafo aplicar el algoritmo Cycle-Canceling (Goldberg & Tarjan, 1989) que soluciona el problema de flujo de costo mínimo, obteniendo un grafo simétrico.

## 5. CASOS DE ESTUDIO

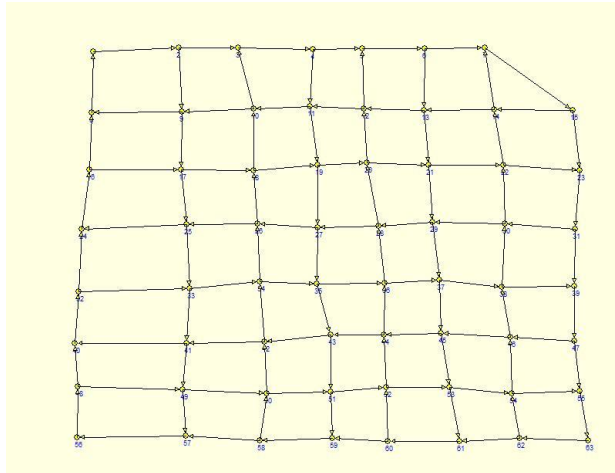
Se escogió el reparto Vedado del municipio Plaza de la Revolución. Las calles del Vedado tienen orientación por lo que el problema del recorrido se asocia con un DCPP. El Vedado fue dividido en dos sectores. A cada sector se le asigna un vehículo especializado con la suficiente capacidad para recoger toda la basura generada por el sector que le corresponde en un viaje.



**Figura 1** Reparto Vedado dividido en 2 sectores

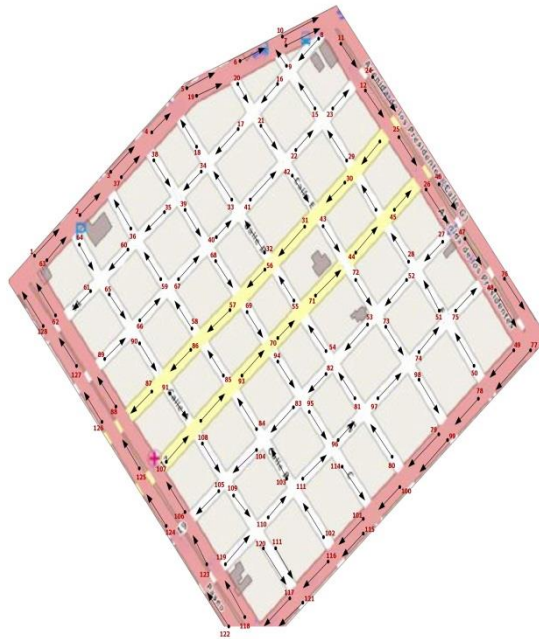
### Sector 1

- Perímetro 10904 m
- Manzanas 49
- Calles orientadas
- Grafo dirigido asociado al sector 1  $G(V=63,A=111)$



**Figura 2** Grafo asociado al sector 1 del Vedado

Para el sector 1 la distancia total a recorrer obtenida por el algoritmo fue de 12431 m. El vehículo tiene que pasar dos veces por 17 calles. Esta solución fue obtenida en 0:0:0.031 s.



**Figura 3** Descripción del recorrido para el sector 1 del Vedado

### Sector 2

- Perímetro 8911 m
- Manzanas 38
- Calles orientadas
- Grafo dirigido asociado al sector 2  $G(V=52,A=91)$

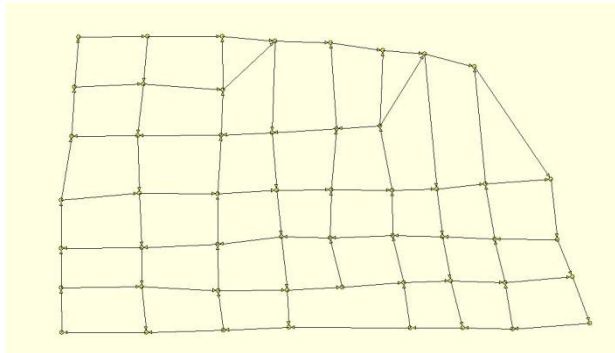


Figura 4 Grafo asociado al sector 2 del Vedado

Para el sector 2 la distancia total a recorrer obtenida por el algoritmo fue de 11788 m. El vehículo tiene que pasar tres veces por 3 calles y dos veces por 21 calles. Esta solución fue obtenida en 0:0:0.063 s.



Figura 5 Descripción del recorrido para el sector 2 del Vedado

### 5.1. CONCLUSIONES Y RECOMENDACIONES

En el trabajo se presentó una propuesta de solución matemático-computacional al problema del diseño del recorrido para la recolecta de basura sobre redes viales con orientación. Se tomaron como casos de estudio dos sectores del reparto Vedado municipio Plaza de la Revolución. Se diseñaron 2 recorridos donde la distancia total a recorrer es la mínima. Por la naturaleza de la red vial del Vedado los grafos asociados fueron dirigidos por lo que se aplicó el algoritmo exacto para solucionar el problema del cartero chino en grafos dirigidos. El tiempo de ejecución fue menor que 1 s. Las dimensiones de las instancias fueron para el sector 1  $V=63$  y  $A=111$ , y para el sector 2  $V=52$  y  $A=91$ .



Con estos casos de estudio se demuestra la viabilidad de la matemático-computacional para dar un a propuesta de recorrido.

Como futuras líneas de investigación sería aconsejable comparar con los recorridos establecidos por comunales en el municipio en cuanto a distancia total a recorrer, tiempo, entre otros.

También sería interesante realizar casos de estudios para los recorridos de los carros de fumigación para controlar el dengue, ya que este problema también se modela con un CPP.

## REFERENCIAS

1. . Análisis Sectorial de Residuos Sólidos en Cuba. (2010, 06 10). Cuba. Retrieved 05 20, 2010
2. Ayala, A., & Gonzáles, E. (2001). Asignación de rutas de vehículos para un sistema de recolección de residuos sólidos en la acera. *Revista de Ingeniería* ISSN 0121-4993 .
3. Belenguer, J., Benavent, E., Lacomme, P., & Prin, C. (2006). Lower and upper bounds for the mixed capacitated arc routing problem. *Computers & Operations Research* .
4. Beltrami, E., & Bodin, L. (1974). Networks and Vehicle Routing for Municipal Waste Collection. *Networks* , vol. 4, pp. 65 - 94.
5. Benavent, E., & Solver, D. (1999). The directed rural postman problem with turn penalties. *Transportation Science*
6. Bodin, L., & Golden, B. (1981). Classification in Vehicle-Routing and Scheduling. *Networks*, 11(2) , pp. 97-108.
7. Bothale, R., Bothale, V., Sharma, J., & Adiga, S. (2002). Solid waste collection transport and disposal site selection through GIS. In *Proceedings of the National Seminar on Solid Waste Management-Current Status and Strategies for Future*, (pp. 262–265.). Bangalore, India.
8. Calita, C. (2008). Optimal routing of waste collection vehicles. *Science Direct* .
9. Cardoso, M. G., C., O., & N. H, M. J. (2009). Experiencia piloto para el manejo integral de los residuos sólidos urbanos en localidades con características urbanas y semi urbanas. caso de estudio en ciudad de La Habana, Cuba. . *Mapping Interativo* .
10. Cerrón, M. (2000). Diseño de rutas optimas de recolección de residuos sólidos domésticos mediante el software MARS. Presentado en XXVII Congreso Interamericano de Engenharia Sanitária e Ambiental.
11. Chang, N., & Lin, Y. (1997). An analysis of recycling impacts on solid waste generation by time series intervention modeling. . *Resour. Conserv. Recy.* , 19: 165–186.
12. Chang, N.-B., & Wei, Y. (2000). Siting recycling drop-off stations in an urban area by genetic algorithm-based fuzzy multi-objective nonlinear integer programming modeling. *Fuzzy Sets Syst.* , 114: 133–149.
13. Chang, N.-B., Lu, H., & Wei, Y. (1997). GIS technology for vehicle routing and scheduling in solid waste collection systems. *J. Envir. Engrg. Div* , 123: 901–910.
14. Chang, Y., & Chang, N.-B. (2001). Optimal shipping strategy of solid waste streams with respect to throughput and energy recovery goals of incineration facilities. *Civ. Eng. Environ. Syst.* , 18: 193–214.
15. Chiplunkar, A., Mehndiratta, S., & Khanna, P. (1981). Optimization of refuse collection systems. *J. Envir. Engrg. Div.* , 107: 1203–1210.
16. Cox-Little-O'Shea. (1996). *Ideals , Variaties and Algorithms*. UTM Springer.
17. Eiselt, H., & Laporte, G. (1995). Arc routing problems, part ii the rural postman problem. *Operations Research* .
18. Eisentein, D., & Iyer, A. (1997). Garbage Collection in Chicago: A Dinamic Scheduling Model. *Mangemnet Science*, Vol 43, No 7 .
19. Evans, J., & Minieka, E. (1992). *Optimization Algorithms for Networks and Graphs*. Marcel Dekker Inc, New York , 278-315.

20. Goldberg, A. V., & Tarjan, R. E. (1989.). Finding Minimum-Cost Circulations by Canceling Negative Cycles. *Journal of the ACM* , 36(4):873–886.
21. Golden, B., & Wong, R. (1981). Capacitated arc routing problems. *Networks* , 305-315.
22. Golden, B., DeArmon, J., & Baker, E. (1983). Computational Experiments with Algorithms for a Class of Routing Problems. *Computers and Operations Research* , 47-59.
23. Kao, J., Lin, H., & Chen, W. (1997). Network geographic information system for landfill siting. *Waste Manag. Res.* , 15: 239–253.
24. Karadimas, N., Doukas, N., Kolokathi, M., & Defteraiou, G. (2008). Routing Optimization Heuristics Algorithms for Urban Solid Waste. *Science Direct* .
25. Komilis, D. (2008). Conceptual modeling to optimize the haul and transfer of municipal solid waste. *Science Direct* .
26. Leao, S., Bishop, I., & Evans, D. (2004). Spatial-temporal model for demand and allocation of waste landfills in growing urban regions. *Comput. Environ. Urban Syst.* , 28: 353–385.
27. Liebman, J., Male, J., & Wathne, M. (1975). Minimum cost in residential refuse vehicle routes. *J. Envir. Engrg. Div* , 101:399–412.
28. Lysgaard, J. (1992). Dynamic transportation networks in vehicle routing and schedule. *Interfaces* , 22: 45–55.
29. Male, J., & Liebman, J. (1978). Districting And Routing For Solid Waste Collection. *J. Envir. Engrg. Div* , 104: 1–14.
30. Massie, K. (1995). Using GIS to Improve Solid Waste Mangement and Recyclig Programs, *Procedures. ESRI User Conf.*, 18 .
31. Mohan, S., & Sarasavani, S. (2002). Optimal routing of solid waste transport using GIS. *GIS India* , 2: 8–13.
32. Montilla, J., Gallardo, A., & Vidal, R. (2000). Diseño de un plan de recogida de residuos sólidos urbanos utilizando tecnología SIG. *Mapping Interactivo* .
33. Mourão, M., Nunes, A., & Prins, C. (2008). Heuristic methods for the sectoring arc routing problem. *Science Direct* , 856-868.
34. Muttiah, R., Engel, B., & Jones, D. (1996). Waste disposal site selection using GIS-based simulated annealing. *Comput. Geosci* , 22: 1013–1017.
35. Nuortioa, T., Kytöjokib, J., Niskaa, H., & Brärys, O. (2005). Improved Route Planning and Scheduling of Waste Collection and Transport.
36. Or, I., & Curi, K. (1993). Improving the efficiency of the solid waste collection. *Waste Management and Research* 11 , 297–311.
37. Osegueda, R., Garcia, A., & Ashur, S. (1999). GIS-Based Network Routing Procedures for Overweight and Oversized Vehicles. *Journal of Transportation Engineering* , 324-331.
38. Perlack, R., & Willis, C. (1985). Multiobjective decision-making in waste disposal planning. *J. Environ. Eng* , 111: 373–385.
39. Racero, J., & Perez, E. (2006). Optimización del sistema de rutas de recolección de residuos sólidos domiciliarios. Presentado en X Congreso de Ingeniería de Organización.
40. Redlands, A. N. (2006). *Network Analyst*. California: Environmental System Research Institute, inc.
41. Rodriguez, A. (2007). Integración de un GIS con un modelo de cálculo y optimización de rutas de vehículos CVRP y software de gestión de flotas. Presentado en XI Congreso de Ingeniería de Organización .
42. Sakurai, K. (1980). Diseño de las rutas de recolección de residuos sólidos.
43. Santos, L., & Coutinho-Rodrigues, J. (n.d.). Implementação em SIG de uma Heurística para o Estudo da Recolha da Resíduos Sólidos Urbanos. *Science Direct* .
44. SEDESOL. (1999). *Manual de Técnicas Administrativas para el Servicio de limpia Municipal*. Mexico: Ingeniería para el Control de Residuos Municipales e Industriales S.A de C.V.D.F Mexico.
45. Shekdar, A., Bhide, A., & Tikekar, V. (1987). Optimization of routes of refuse transportation vehicles. *Indian J. Environ. Health* , 29: 1–15.
46. Toth, P., & Vigo, D. (2002). *The Vehicle Routing Problem*. Philadelphia,PA:SIAM.

47. Vijay, R., Gautam, A., Kalamdhad, A., Apurba, A., & Devotta, S. (2008). GIS-based locational analysis of collection bins in municipal solid waste management systems. *J. Environ. Eng. Sci.*

## Capítulo 17

pp 180-201

# APROXIMACIÓN A LA SITUACIÓN DE POBREZA Y DESIGUALDAD EN MÉXICO. FACTORES CAUSALES Y ALTERNATIVAS DE SOLUCIÓN

José Félix García Rodríguez\*, Ignacio Caamal Cauich\*\*, Rilton Goncalo B. Primo\*\*\*

&

Oscar Priego Hernández\*

\* Universidad Juárez Autónoma de Tabasco, México

\*\* Universidad Autónoma de Chapingo, México

\*\*\*Centro de Estudios por la Amistad de Latinoamérica, Asia y África (CEALA), Brasil.

### SUMMARY

Worldwide, the economic crisis associated with the process of globalization has impacted negatively on the economic and social welfare conditions of all countries, whether poor or rich. In Latin America and particularly Mexico, the crisis has significantly accentuated the problems associated with underdevelopment, particularly poverty, inequality, economic stagnation, unemployment and insecurity. In this context, there are two great scenes in Latin America that concern and worry today: The insertion of local economies to the context of globalization in terms of competitive advantage, and stagnation of economic growth and the drop in quality life of the population, problems associated with the global economic crisis. In both scenarios, the national state is forced to intervene decisively by implementing public policies aimed at social change.

**KEYWORDS:** poverty, inequality, globalization, economic crisis, quality of life

### RESUMEN

En el mundo, la crisis económica asociada al proceso de la globalización ha impactado negativamente en las condiciones económicas y de bienestar social de todos los países, sean pobres o ricos. En América Latina y México particularmente, la crisis ha acentuado notablemente los problemas asociados al subdesarrollo, particularmente la pobreza, desigualdad, estancamiento económico, desempleo e inseguridad. En este contexto, son dos los grandes escenarios en América Latina que ocupan y preocupan en la actualidad: La inserción de las economías locales al contexto de la globalización en términos de ventajas competitivas, y el estancamiento del desarrollo económico y la caída en la calidad de vida de la población, problemas asociados a la crisis económica global. En ambos escenarios, el estado nacional está obligado a intervenir de manera decidida mediante la instrumentación de políticas públicas orientadas al cambio social.

## 1. POBREZA

La pobreza, entendida como una condición socioeconómica que limita el bienestar de las personas y el desarrollo económico de los países, es un problema de naturaleza compleja y multidimensional. De manera tradicional, se consideran como pobres aquellas personas, familias y grupos de personas cuyos recursos monetarios comparados con una línea de bienestar predeterminada son tan limitados que los obligan a estar excluidos de una forma de vida mínimamente aceptable. Como puede apreciarse, el ingreso monetario es la única variable utilizada para la medición de la pobreza, lo que se conoce también como pobreza absoluta. Este criterio de medición constituye uno de los problemas fundamentales en la lucha contra la pobreza, ya que al reducir ésta al propósito único de que los pobres cuenten con un ingreso monetario que los ubique por encima de una línea de pobreza predeterminada, deja de lado múltiples factores determinantes y condicionantes del problema, puesto que la pobreza es de naturaleza multidimensional y compleja. Esto es, la pobreza tiene múltiples dimensiones que no pueden reducirse simplemente al aspecto monetario (Salama, 2011).

Por ello, es necesario identificar las variables determinantes de los flujos de entrada y salida de la pobreza, así como los factores que determinan y condicionan la presencia tanto de la pobreza crónica como la transitoria (López y Beltrán, 2011). De esta manera, la eficacia de las políticas públicas contra la pobreza pasa primero por un conocimiento de las necesidades específicas de los pobres (aspecto microeconómico), pero a la vez, depende también del crecimiento económico y de la reducción de las desigualdades sociales, así como del incremento de los gastos sociales en educación, salud, vivienda e infraestructura básica. Ello sin dejar de reconocer que las políticas públicas de transferencias monetarias condicionadas a los pobres extremos son básicas y deben ser mantenidas y ampliadas. Sin embargo, ésta política asistencialista no debe sustituir a una política de gasto social agresiva enfocada a elevar las capacidades básicas de la población.

La pobreza constituye un problema sociocultural complejo ampliamente estudiado y debatido desde diversas disciplinas del conocimiento. En su estudio destacan los trabajos de Rawls acerca del contrato social (1971), y las investigaciones de Sen (2000), quien bajo el enfoque teórico del bienestar social aporta importantes resultados y evidencias sobre los factores determinantes, y cuya propuesta centrada en las capacidades básicas y libertad de elección, ha contribuido sustancialmente a la instrumentación de políticas públicas que abordan la pobreza a partir de la visión del desarrollo humano (PNUD 2010). En la práctica, son notables los resultados alcanzados en el combate a la pobreza en el mundo mediante la aplicación del enfoque del bienestar social, la empresa social y el microcrédito (Yunus, 2010).

### 1.1 Conceptualización

La pobreza representa una condición socioeconómica que afecta directamente el bienestar de las personas y por lo tanto limita el desarrollo económico de los países (Dieterlen, 2003). Por su trascendencia social, económica, ética y política, su análisis está siempre presente en la agenda y las políticas públicas de todos los países, tanto emergentes y en vías de desarrollo, como aquellos considerados de primer mundo (Nussbaum y Sen, 2004).

El estado del arte acerca de la pobreza, sus nociones y corrientes teóricas puede agruparse en tres enfoques:

1. Pobreza como necesidades básicas insatisfechas
2. Pobreza como insuficiencia de capacidades y libertades
3. Pobreza como problema complejo y dinámico

Los dos primeros corresponden a lo que podría denominarse como enfoque estático de la pobreza, en tanto que el segundo correspondería a un enfoque dinámico del problema. Esta distinción es relevante, toda vez que el diseño de las políticas públicas instrumentadas para enfrentar el problema está condicionado por el enfoque asumido. Cabe decir que en la actualidad, las políticas públicas contra la pobreza en el plano mundial están guiadas por el enfoque estático, y son contadas las experiencias de políticas aplicadas en base al enfoque dinámico.

La primera corriente teórica, misma que justifica la pobreza en razón de que existen necesidades básicas insatisfechas entre la población comprende cinco criterios de clasificación y medición del problema:

- a). Insuficiencia de ingresos
- b). Necesidades básicas insatisfechas
- c). Insuficiencia de ingresos junto con necesidades básicas insatisfechas
- d). Rezago en las condiciones socioeconómicas de vida
- e). Insuficiencia de la política económica y social contra la pobreza

Típicamente, las políticas públicas implementadas en base a estos criterios son las predominantes en el abanico de posibilidades de intervención por parte del Estado, sobresaliendo la clasificación de la pobreza en función del ingreso disponible frente a una línea de bienestar, y la clasificación denominada como multidimensional que considera no solo la insuficiencia de ingresos sino también la presencia de necesidades básicas insatisfechas. Si bien estas políticas públicas permiten establecer alguna compensación a determinados grupos sociales mediante un proceso de focalización, no constituyen una respuesta a la erradicación del problema, y tienden a fomentar y perpetuar la dependencia de los beneficiarios a los recursos monetarios y en especie otorgados a través de dichos programas.

Por su parte, la corriente que asocia la pobreza a la insuficiencia de capacidades y libertades de las familias y las personas comprende cinco enfoques del problema:

- a). Negación de la democracia
- b). No correspondencia entre necesidades y satisfactores
- c). Falta de bienestar y autonomía
- d). Negación de oportunidades de desarrollo
- e). No expansión de las capacidades

En contraste con el enfoque de necesidades básicas insatisfechas, este enfoque avanza en el reconocimiento de que además de ello, la pobreza se corresponde también con una insuficiencia en las capacidades básicas del individuo, lo que no le permite integrarse de manera adecuada a la economía y la sociedad en la que subsiste. Entre dichas capacidades básicas se encuentran el poder vivir una vida de manera prolongada en condiciones de salud adecuadas, acceso pleno a la educación así como a otros satisfactores básicos, todos los cuales permitirían al individuo interactuar con toda libertad en la sociedad (Gallego 2010).

En lo que corresponde al enfoque dinámico, éste parte de la noción de que la pobreza constituye en realidad un problema multidimensional y complejo que no puede ser explicado únicamente en función del ingreso y/o algunos cuantos satisfactores y capacidades del individuo. De esta manera, el problema tiene muchas ramificaciones y bucles, por lo cual no puede ser explicado en base a un enfoque estático. No obstante, el campo de estudio de la pobreza ha estado dominado por la aplicación de enfoques estáticos, mismos que a partir de una línea de bienestar predeterminada bajo el enfoque del ingreso percibido, analizan la evolución de la pobreza en los hogares como una sucesión de cortes transversales. En contraste, los nuevos enfoques dinámicos incorporan al estudio del problema herramientas metodológicas para el análisis de múltiples variables determinantes y condicionantes del problema, así como la dimensión temporal mediante el análisis de datos longitudinales que explican la variación de la pobreza en el tiempo.

## 1.2 Orígenes

El problema de la pobreza en el mundo no se puede descontextualizar del sistema económico en que se genera y reproduce. De esta manera, en el contexto del fenómeno conocido como globalización, en el cual se ha insertado la economía neoliberal, ha tomado impulso el individualismo y el desmantelamiento de la actividad pública en la economía. Si bien se ha generado progreso económico y bienestar en los países, también se ha acentuado la pobreza y desigualdad entre sus habitantes. Particularmente, la globalización impulsa la apertura económica mundial, la libertad de fronteras y el libre tránsito de capitales, mercancías y personas. Además de ello, impone las leyes del mercado y la búsqueda incesante de la productividad y rentabilidad de las actividades económicas, no importando que ello genere estancamiento económico, desempleo, caída de los salarios reales y pérdida de poder adquisitivo, principales factores causales de la pobreza y desigualdad en el mundo.

Aunado a lo anterior, la tendencia mundial a la desincorporación del Estado de sus responsabilidades tradicionales relacionadas con el denominado “estado del bienestar”, ha venido a agravar la situación de pobreza en el plano mundial y local. Y es que una de las razones que explican el nivel de bienestar y prosperidad de los países ricos, es el papel jugado por el Estado como promotor del desarrollo económico y social. Ello ha significado largos períodos de esfuerzos y sacrificios, esfuerzos que han descansado básicamente en el ahorro interno, en la inversión en capital y tecnología, y sobre todo en la inversión pública realizada para la formación de capital físico y humano. Todo ello, se ha traducido en altos niveles de progreso económico y desarrollo humano de sus habitantes. En contraste, los países pobres no cuentan con una base económica que les permita invertir los recursos necesarios para la superación de la pobreza y la erradicación del hambre en el mundo.

### 1.3 Medición

Como ya se ha mencionado con anterioridad, la dimensión del problema de pobreza en el contexto mundial puede ser estimada bajo dos enfoques diferentes: Uno estático que predomina a nivel global, mismo que comprende el análisis de la variable ingreso y algunas necesidades y capacidades básicas insatisfechas, y un enfoque dinámico centrado en una perspectiva compleja y multidimensional del problema. A continuación se establece una distinción más específica:

#### 1.3.1 Enfoque estático

Tradicionalmente, la medición de la pobreza se ha sustentado en indicadores simples construidos en base al criterio de ingreso percibido y línea de bienestar, a partir del cual se obtienen tres dimensiones de pobreza: incidencia, intensidad y desigualdad, siendo su fundamento el siguiente:

##### a. Tasa de incidencia (proporción de pobres)

$H=q/n$  donde:

H= Pobreza como porcentaje de una población (n), a partir de la definición de una línea de pobreza (z).

q= Población total

n= Población con ingreso por debajo de la línea de pobreza

z= Ingreso mínimo necesario para satisfacer las necesidades básicas de la familia, equivalente a 2 USD según el Banco Mundial.

La medición de la pobreza en función de este criterio tiene como principal limitante su definición y cálculo a partir de la base a una línea de pobreza (z) expresada en recursos monetarios, lo cual no reconoce que además del ingreso existen otras variables causales. Por otro lado, no reconoce que aún bajo (z), existen desigualdades de ingreso.

##### b. Intensidad de la pobreza (brecha del ingreso de los pobres o brecha estandarizada del ingreso).

$I= (Z- up)/Z$ , donde:

I= Refleja el déficit porcentual del ingreso medio de los pobres con respecto a la línea de pobreza (Z).

up= Ingreso medio de la población pobre

z-yi mide la «profundidad» de la pobreza de una persona

Al igual que el criterio de medición anterior, el indicador de intensidad de pobreza considera únicamente el ingreso, además de no prestar atención a la desigualdad y regresividad en torno a la distribución del mismo.

### c. Desigualdad de la pobreza

$Ps = H [ I + (1-I) * Gq ]$ , donde:

Ps= Indica que tan desigual es la pobreza

H= Tasa de incidencia de la pobreza

I= Intensidad de la pobreza

Gq= Coeficiente de concentración del ingreso

Este indicador, también conocido como Índice de pobreza de Sen, al incorporar al análisis de la pobreza la distribución del ingreso, supera notablemente la insensibilidad de los indicadores de incidencia e intensidad. No obstante, tiene la limitación de considerar únicamente el ingreso y su distribución, lo que en materia de políticas contra la pobreza lleva a errores al incluir hogares que no son verdaderamente pobres y excluir a otros que sí lo son. Ello es más evidente en los países donde la focalización de las políticas públicas se establece en base a un padrón de pobres, como es el caso de México (García, 2014).

Otro criterio de estimación de pobreza bajo el enfoque estático lo constituye la medición multidimensional, cuya principal ventaja radica en que no sólo toma en cuenta el ingreso de la población, sino que incorpora también el espacio de los derechos sociales y el contexto territorial. La inclusión de estos espacios al análisis posibilita el establecimiento de dos grupos: El primero integra a los **pobres**, entre los cuales se puede ubicar dos subgrupos: a). Población en situación de pobreza multidimensional: Es aquella cuyos ingresos son insuficientes para adquirir los bienes y los servicios que requiere para satisfacer sus necesidades esenciales y que presenta carencia en al menos uno de los siguientes seis indicadores: acceso a la alimentación, acceso a los servicios de salud, acceso a la seguridad social, servicios básicos en la vivienda, rezago educativo, y por último, calidad y espacios de la vivienda; b). Población en pobreza extrema: Es la que tiene tres o más de estas carencias sociales y un ingreso menor a la línea de bienestar mínimo. El segundo grupo, los **no pobres** se puede clasificar en tres subgrupos: **Vulnerables por carencias sociales**, comprende a aquellos que superan la línea de bienestar, pero que tienen más de dos carencias sociales, **Vulnerables por ingresos y No pobre multidimensional y no vulnerable por ingreso**, ambos grupos no presentan carencia social alguna (Coneval, 2010).

Con la aplicación de esta metodología es posible la obtención de tres dimensiones del problema: incidencia, profundidad e intensidad de la pobreza. Las medidas de incidencia permiten conocer el porcentaje o proporción de la población que es pobre multidimensional o que tiene carencias de educación, salud, seguridad social, vivienda y alimentación. Asimismo, las medidas de profundidad e intensidad contabilizan el número promedio, la proporción media y el total de carencias sociales de las personas, así como la proporción de carencias de la población pobre respecto al máximo posible que podría experimentar la población total del país.

Como puede apreciarse, la medición de la pobreza multidimensional tiene entre otras, las siguientes ventajas: a) permite valorar la proporción y el número de personas que disponen de un ingreso insuficiente para satisfacer sus necesidades; b) incorpora al análisis de la pobreza otras dimensiones asociadas con los derechos sociales de la población; c) amplía la riqueza analítica de la medición al identificar tanto a la población pobre como a la vulnerable, ya sea por carencias o por ingreso; c) permite conocer el número de carencias que padece, en promedio, la población pobre y la población vulnerable por carencias, y d) posibilita identificar el número total de carencias sociales que deben ser resueltas, tanto en el país en su conjunto como en cada entidad federativa.

### 1.3.2 Enfoque dinámico



El análisis dinámico de la pobreza requiere de un conjunto de observaciones sobre una cierta cantidad de variables para un mismo grupo de individuos en dos o más tiempos. Esto es, las variaciones inter temporales en la situación de los hogares requiere de información de encuestas panel que aporten datos longitudinales. La principal limitante para su estimación es la dificultad de su diseño y levantamiento, pues suponen el seguimiento en el tiempo de los mismos hogares y personas sujetas de estudio. Por otro lado, en la mayoría de los países subdesarrollados la disponibilidad de este tipo de datos es limitada. Finalmente, los estudios basados en datos de panel dificultan la comparación en variación inter temporal de la pobreza entre países (García, 2014).

El estudio de la pobreza bajo un enfoque dinámico es reciente Lillard y Willis (1978), propusieron una metodología econométrica que toma en cuenta diversos aspectos socioeconómicos y la movilidad en la distribución del ingreso. En base a información de encuesta panel se analiza la frecuencia de entrada y salida de la pobreza así como su duración. De esta manera pudieron concluir que en Estados Unidos los años de escolaridad y de experiencia laboral son las principales variables determinantes de la situación de pobreza, siendo la diferencia racial la variable más explicativa. Por otro lado, Hutchens (1998), identifica diversos eventos que provocan la entrada y/o salida de la pobreza, identificando su duración y su naturaleza crónica y transitoria, concluyendo que los cambios en el ingreso del cabeza de familia es la variable prevaleciente en la pobreza transitoria, y los ingresos permanentemente bajos de la pobreza crónica (Salama, 2011).

Por su parte, Bane y Ellwood (1986), pronostican la probabilidad de escapar de la pobreza y estiman la extensión en que el intervalo comienza y termina, siendo los cambios en el ingreso familiar la principal variable determinante. Así, al menos el 40% del inicio de un intervalo de pobreza obedece a la caída del ingreso familiar; el 60% de la finalización de un intervalo responde al incremento del ingreso del jefe del hogar; la probabilidad de un individuo de salir de la pobreza el primer año es del 45%, el segundo del 28%, el tercero de 24%, y sólo una pequeña fracción permanece en pobreza de manera crónica. Finalmente, únicamente el 19% de las personas que escapan de la pobreza se debe a las transferencias recibidas.

Mcculloch y Baulch (2000), llevaron a cabo un interesante estudio de la dinámica de la pobreza de los hogares rurales en Pakistán. En el mismo captan el efecto de la geografía y los shocks como determinantes del ingreso. El modelo aplicado considera un vector de variables que aproximan el entorno socioeconómico en el cual está inserto el hogar, los activos con los que cuenta y los shocks adversos como determinantes del ingreso familiar:

$$\ln(a_i) = k + H_{ci} + E_{dui} + L_{andi} + A_{ssetsi} + D_{villi} + \epsilon_{it}$$

Donde:

$a_i$  = promedio del ingreso por adulto en el hogar  $i$

$k$  = constante

$H_{ci}$  = vector de acuerdo a la composición del hogar

$L_{andi}$  = posesión de tierras en el hogar

$A_{ssetsi}$  = valor del ganado, vehículos y otras pertenencias

$D_{villi}$  = variable dicotómica que indica si el hogar se encuentra o no en la región de estudio.

En lo que corresponde a América Latina, Santillán y Laplante (2009), estudiaron la dinámica de la pobreza en Argentina. Su análisis considera diversas variables sociodemográficas, concluyendo que el riesgo de entrar y salir de la pobreza es determinada principalmente por el nivel educativo del jefe del hogar, el empleo formal y las transferencias recibidas. Asimismo, La interacción entre el perfil sociodemográfico y el contexto macroeconómico son determinantes en la dinámica de la pobreza.

Particularmente, el riesgo de caída en pobreza aumenta en los hogares con mayor número de dependientes y bajo nivel educativo del jefe de familia.

Por su parte, Maurizio, Perrot y Villafañe (2008), estudian la reducción de la pobreza suscitada en Argentina (del 54% al 27%) en el contexto del auge macroeconómico de 2003-2006, buscando determinar la influencia del mercado laboral, los eventos demográficos y las políticas públicas en la superación de la pobreza. Encontraron que sólo el 30% de los hogares pobres lograron salir de la pobreza; los eventos relacionados con el mercado de trabajo incidieron en un 45%; el incremento en ingresos no laborales sólo influyó en un 11%. Por su parte, los eventos laborales adversos determinaron la caída en pobreza de los hogares en un 41%.

En Chile, Contreras, Cooper y Neilson (2005), estudian los determinantes de la pobreza en los hogares; las causas que explican la entrada y salida de la misma, y analizan la movilidad posicional de los hogares en la distribución de ingresos. Para ello parten de un modelo con un vector de variables que aproxima la relación entre las características del hogar, la geografía y los eventos inesperados:

$$Y = a + \beta_1 X \text{ geography} + \beta_2 \text{ activos hogar} + \beta_3 X \text{ composición hogar} + \beta_4 X \text{ shocks} + \epsilon_i$$

Entre sus principales hallazgos se encontró una alta vulnerabilidad a la pobreza transitoria en los primeros siete deciles de ingreso y una baja movilidad en el decil más rico (coeficiente de inmovilidad de 55.6%). Los hogares con jefe de familia adulto, empleo fijo y mayor educación, son menos vulnerables a la pobreza. Por el contrario, hogares jóvenes y con un mayor número de niños o dependientes económicos son más vulnerables y tienen menor probabilidad de salir de la pobreza. Por otro lado, la educación media y universitaria entre los integrantes de la familia reducen la probabilidad de caer en pobreza, pero no tiene peso en la probabilidad de salir de la misma. La educación técnica es más efectiva en este sentido. Asimismo, los problemas de salud de la familia reducen la probabilidad de salir de la pobreza (Salama, 2011).

En México, Urzúa y Brambila (2009), analizan la dinámica de la pobreza y concluyen que ésta se correlaciona directamente con el crecimiento económico y la distribución del ingreso. En esta línea de investigación, López y Perry (2008), comprueban la hipótesis de que la educación se asocia con una menor desigualdad, mientras que la apertura comercial y la profundidad financiera inciden en un mayor crecimiento con desigualdad. Calderón y Chong (2004), estudian el impacto del gasto en infraestructura en el crecimiento económico con disminución en la brecha de ingresos. Urzúa (2008), analiza la influencia de los monopolios y/o oligopolios en la dinámica de la pobreza, así como el impacto de la migración y las remesas recibidas. Urzúa, Macías y Sandoval (2008), analizan la evolución de los salarios mínimos reales, en tanto que Cortés y Escobar (2005), estudian el impacto las transferencias vía programas oficiales (Progres- Oportunidades). Por su parte, Meléndez (2000), plantea un modelo para analizar la pobreza crónica y transitoria así como identificar los factores que la producen, encontrando que la educación del jefe de familia y las transferencias de ingreso explican las transiciones de pobreza y que la probabilidad de entrar en pobreza aumenta con el desempleo del jefe de familia.

#### 1.4 Situación de pobreza en México

México vive una preocupante situación de pobreza y rezago social. El Estado mexicano ha enfrentado la pobreza mediante una política social de naturaleza asistencialista, integrada por una gran cantidad de programas públicos desarticulados entre sí, lo que ha implicado un enorme gasto de recursos fiscales sin resultados significativos. La pobreza en México obedece a causas estructurales y al impacto negativo de la crisis económica, pero sobre todo, a la falta de efectividad de las políticas públicas implementadas, lo que constituye una llamada de atención a una acción más decidida por parte del Estado y la sociedad mexicana. En México, la realidad de la vida cotidiana revela la severidad y

profundidad de la pobreza y las cifras oficiales no hacen más que confirmarla. El problema de la pobreza es creciente y persistente con el tiempo, de ahí la necesidad de investigar en torno al diseño y ejecución de nuevos enfoques de política pública contra la pobreza. En este sentido, toma relevancia una propuesta de política de Estado contra la pobreza.

En México, la pobreza es uno de los problemas estructurales más apremiantes social y económicamente hablando. La medición de la pobreza en México ha tenido una amplia difusión a partir de la creación del Comité Técnico de Medición de la Pobreza (CTMP), y ahora con el Consejo Nacional de Evaluación de la Política Social (Coneval). Por definición, una persona se encuentra en situación de pobreza por ingresos cuando su ingreso está por debajo del monto mínimo necesario que le permite satisfacer sus necesidades esenciales (línea de bienestar) (Coneval, 2013).

#### **1.4.1 Pobreza por ingresos**

El umbral de ingreso mínimo necesario para satisfacer las necesidades esenciales de la población se denomina línea de pobreza y se encuentra expresado a partir del valor monetario de una canasta de bienes y servicios básicos predeterminada. De acuerdo con la Metodología para la Medición de la Pobreza elaborada por el Comité Técnico para la Medición de la Pobreza del Coneval, el ingreso empleado para esta medición es el Ingreso Neto Corriente Total Per Cápita (INTPC) y los umbrales de pobreza se definen en tres niveles:

- 1) Pobreza alimentaria: Incapacidad para obtener una canasta básica alimentaria, aún si se hiciera uso de todo el ingreso disponible en el hogar para comprar sólo los bienes de dicha canasta.
- 2) Pobreza de capacidades: Insuficiencia del ingreso disponible para adquirir el valor de la canasta alimentaria y efectuar los gastos necesarios en salud y en educación, aun dedicando el ingreso total de los hogares nada más para estos fines.
- 3) Pobreza de patrimonio: Insuficiencia del ingreso disponible para adquirir la canasta alimentaria, así como para realizar los gastos necesarios en salud, educación, vestido, vivienda y transporte, aunque la totalidad del ingreso del hogar sea utilizado exclusivamente para la adquisición de estos bienes y servicios.

De acuerdo con el informe de medición de la pobreza 2010 del Coneval, en ese año, 58.3 millones de mexicanos eran pobres de patrimonio, es decir, no contaban con un ingreso suficiente para satisfacer sus necesidades de alimentación, de salud, de educación, de vivienda, de vestido y de transporte público, aún si dedicaran la totalidad de sus recursos económicos a este propósito. Asimismo, 21.2 millones eran pobres alimentarios, es decir, quienes tienen ingresos insuficientes para adquirir una canasta básica de alimentos, incluso si los destinaran exclusivamente para ese fin.

Entre el período 2008-2010, la pobreza por ingresos aumentó de manera sustancial. Específicamente, la pobreza de patrimonio entre los mexicanos aumentó del 47.7% a 51.3%, en tanto que la pobreza alimentaria pasó del 18.4% al 18.8%. En términos absolutos, de una población cercana a los 112 millones de habitantes, la pobreza de patrimonio pasó de 52.3 a 57.7 millones de mexicanos, es decir un incremento de 5.4 millones de habitantes, en tanto que la pobreza alimentaria pasó de 20.2 a 21.2 millones de habitantes, un incremento de 1 millón de personas.

Visto el problema en un horizonte de tiempo más largo, entre 2006 y 2010, la incidencia de la pobreza de patrimonio y de la pobreza alimentaria aumentó 5.0 y 8.6 puntos porcentuales respectivamente, lo cual se traduce en un aumento de 6.5 y 1.3 millones de personas pobres de patrimonio y alimentarios, respectivamente. Para el periodo 1992-2010, la pobreza de patrimonio

disminuyó de 53.1% a 51.3% y la alimentaria de 21.4% a 18.8%. A pesar de ello, se incrementó en 11.6 millones el número de personas pobres de patrimonio y en 2.6 millones el número de pobres alimentarios.

#### 1.4.2 Pobreza multidimensional

A partir del 2008, México incorpora al estudio de la pobreza el criterio de medición bajo el enfoque multidimensional. Por definición, se considera en pobreza multidimensional a aquella población con ingreso inferior al valor de la línea de bienestar y que observa al menos una carencia social (Coneval, 2013). Bajo esta metodología de cálculo, el problema de pobreza en México arroja nuevas dimensiones. Así, según el informe de pobreza multidimensional 2008 del Coneval, el 44.2% de la población mexicana era pobre multidimensional. Esto es, había 47.2 millones de personas que tenían al menos una carencia social en alguno de los siguientes rubros: educación, salud, seguridad social, calidad de la vivienda, servicios básicos en la vivienda y alimentación, así como un ingreso inferior a la línea de bienestar económico.

En lo que corresponde al año 2012 la situación de pobreza multidimensional en México se incrementó al 45.5% según el informe de Coneval para ese año. De esta manera, entre 2008-2012 la incidencia de pobreza multidimensional se incrementó casi en un 3%, y la población mexicana en situación de pobreza pasó de 47.2 millones a 53.3 millones de habitantes, esto es, un incremento cercano al 3%, lo cual es reflejo de la crisis económica del 2008. Por su parte, entre 2010-2012 el porcentaje de pobreza multidimensional disminuyó 0.6% al pasar de 46.1% a 45.5%. Considerando que las personas se encuentran en condición de pobreza extrema cuando tienen tres o más carencias sociales y perciben un ingreso inferior a la línea de bienestar mínimo (equivalente al costo de la canasta alimentaria básica), entre 2010 y 2012 ésta se redujo del 11.3% a 9.8%. Esto es, una disminución de 13.0 a 11.5 millones de mexicanos.

Para el mismo período, la población vulnerable por carencia social (aquella que enfrenta al menos una carencia y cuenta con un ingreso superior a la línea de bienestar) pasó de 32.1 a 33.5 millones de mexicanos. Esto es, el porcentaje de vulnerabilidad por carencia social creció del 28.1% al 28.6%. Por su parte, la población vulnerable por ingreso (la que no enfrenta carencias sociales pero su ingreso es inferior a la línea de bienestar) pasó de 6.7 a 7.2 millones de personas. Un incremento del 5.9% al 6.2% del total poblacional. Finalmente, la población no pobre ni vulnerable (la que no enfrenta ninguna carencia social y tiene ingresos superiores a la línea de bienestar, creció de 22.8 a 23.2 millones de mexicanos.

En la actualidad la situación de pobreza en México sigue siendo creciente y persistente. De esta manera, el informe Coneval 2014 registra una incidencia del 46.6% entre la población. Comparativamente con la pobreza observada en 2012 (45.5%), ésta repuntó en un 1.5%. Asimismo, entre 2012-2014 la población detectada en situación de pobreza se incrementó en 2 millones de habitantes al pasar de 53.5 a 55.3 millones. Un dato alentador es que en lo que corresponde a pobreza extrema, ésta disminuyó un 3% durante el período de análisis, pasando de 9.8% a 9.5%, reduciéndose la población en dicha situación de 11.5 millones a 11.4 millones de habitantes.

Como puede apreciarse de las cifras anteriores, en México la medición de la pobreza continúa siendo estática y reduccionista, pues su dimensión únicamente considera la variable ingreso y ciertos indicadores subjetivos de carencia social. Ello no ayuda a explicar de manera objetiva el problema, ya que únicamente cuantifica su incidencia en función del número de hogares y personas pobres. Por ello, surge la inquietud de investigadores y funcionarios públicos acerca de la necesidad de incorporar nuevas metodologías al estudio de la pobreza, con el propósito de hallar explicaciones y soluciones dinámicas al problema. Al respecto, en el ámbito mundial se vienen estudiando nuevos enfoques sustentados en

análisis dinámicos de la pobreza, aplicando en ellos las herramientas de la investigación operacional. Ello ha permitido superar los resultados obtenidos mediante los enfoques estáticos tradicionales, mismos que se centran en la medición y clasificación de la pobreza.

### **1.5 Políticas públicas aplicadas contra la pobreza**

El estudio de la pobreza en México abarca varios años, y ha formado parte importante de la agenda y las políticas públicas de los gobiernos en turno. Las primeras aproximaciones a la pobreza en México abarcan desde la segunda mitad del siglo pasado. No obstante, la intervención directa del Estado en el combate a la pobreza arranca apenas a partir de 1997 con la puesta en marcha del programa institucional conocido como Progresá. A partir del 2001, éste adopta el nombre de Oportunidades, denominación que el gobierno federal cambió recientemente a Prospera. Por otra parte, a partir del 2013 la nueva administración federal puso en marcha la Cruzada Nacional contra el Hambre, programa oficial cuya prioridad es la erradicación de la pobreza extrema en poco más de 400 municipios del país.

Oportunidades, hoy Prospera, constituye un programa de transferencias monetarias condicionadas ampliamente reconocido a nivel internacional por lo organismos multinacionales (Banco Mundial, Fondo Monetario Internacional). Atiende a poco más de cinco millones 800 mil hogares, lo que significa que uno de cada cuatro mexicanos forman parte de su padrón de beneficiarios. Sin embargo, a pesar de su carácter de largo plazo y la enorme cantidad de recursos públicos ejercidos, sus resultados en términos de reducción de hogares en pobreza extrema no son satisfactorios, pues según las más recientes cifras oficiales, en México el 53.3% de la población se encuentra en situación de pobreza (Coneval, 2013), la cual ha aumentado significativamente no obstante la gran cantidad de programas oficiales implementados y la enorme cantidad de recursos públicos ejercidos en el combate a la pobreza.

Visto el problema de la pobreza en México, tanto del lado del ingreso como desde la perspectiva multidimensional, los magros resultados obtenidos demuestran el impacto negativo de la crisis económica mundial, pero hacen evidentes también los pobres resultados de las políticas públicas instrumentadas contra la pobreza, lo que constituye una fuerte llamada de atención a una acción más decidida por parte del Estado y la sociedad mexicana. Ello debido a que dichos programas únicamente se han enfocado a combatir la pobreza coyuntural, no la estructural, la cual tiene amplias y profundas raíces en la economía mexicana (Pick y Sirkin, 2011). En este sentido, en México es necesario pasar de la teoría a la práctica, por lo que se requiere de la intervención del estado nacional con el propósito de enfrentar de manera decidida el problema, para lo cual es necesario el establecimiento de diversas políticas públicas integradas en una política de Estado contra la pobreza.

## **2. DESIGUALDAD**

La desigualdad constituye un tema de relevancia social y económica en el mundo, y se encuentra siempre en el centro del debate político (Piketty, 2015). Desde el punto de vista económico, la desigualdad alude a la forma en que el ingreso de un país o una región se encuentra distribuido entre la población. Para ello, se emplea un indicador conocido como *coeficiente de Gini*, un parámetro que mide el nivel de desigualdad en la distribución del ingreso en una escala que va de 0 a 1, dependiendo del grado de concentración. De esta manera, un coeficiente de 0 sería indicativo de la inexistencia de desigualdad, en tanto que un parámetro de 1 significaría una total concentración de la riqueza. Históricamente, el coeficiente de Gini en todos los países del mundo ha observado una tendencia creciente, independientemente de su grado de desarrollo y orientación ideológica. Así por ejemplo, en los últimos 30 años el Gini de China pasó de 0,27 a 0,48, en Brasil se mantiene en 0,50, en Suecia de 0,20 a 0,25, y en Estados Unidos de 0,30 a 0,38. Incluso, el coeficiente a escala mundial (0,70) demuestra una alarmante concentración de la riqueza. Analizado el indicador a nivel de deciles de población, significa que el 46 por ciento de la riqueza del mundo está en manos del 1 por ciento de sus

habitantes, dicho de otra manera: 70 millones de personas acumulan la misma riqueza que los otros 7.000 millones de habitantes.

## 2.1 Conceptualización

En cierto modo, pobreza y desigualdad van de la mano, lo que implica fuertes riesgos de inestabilidad económica, social y política en los países. Últimamente, los principales economista teóricos en el mundo reconocen tal riesgo, puesto que la creciente desigualdad en el ingreso de la población puede generar efectos colaterales dañinos. Teóricamente, desigualdad y prosperidad tienen una relación ambigua (Caparrós, 2014).

Por otro lado, la desigualdad puede contribuir al crecimiento económico, pues quienes normalmente concentran la riqueza son aquellos que más invierten y trabajan duro en función de los incentivos económicos. Sin embargo, grandes disparidades en la distribución del ingreso entre la población puede tornarse en ineficiencia económica, ya que cierran el acceso a la educación, la salud y otros satisfactores básicos, pilares del capital humano y la productividad; además de ello, una población con limitado poder adquisitivo no contribuye a la formación de un adecuado mercado interno.

De acuerdo con Piketty (2014), ya es tiempo de que los economistas y los hacedores de políticas públicas reubiquen el tema de la desigualdad en el centro del análisis económico, tal como lo hacían los economistas del siglo XIX. Y es que el asunto de la desigualdad en la distribución del ingreso pasó a segundo término en las prioridades económicas a medida que la visión optimista de la economía impuesta por las conclusiones de Kuznets apuntaban a un autoequilibrio de la economía y a una disminución de las desigualdades en el largo plazo.

En su importante investigación acerca de la desigualdad en la distribución de ingreso, divulgada en su libro "*El capital en el siglo XXI*", Piketty llega a importantes conclusiones:

1. En materia de desigualdad en la distribución del ingreso hay que desconfiar de todo determinismo económico: la historia de la distribución de la riqueza es siempre profundamente política y no podría resumirse en mecanismos puramente económicos.
2. La dinámica de la distribución de la riqueza pone en juego poderosos mecanismos que empujan alternativamente en el sentido de la convergencia y la divergencia. Y no existe ningún proceso natural y espontáneo que permita evitar que las tendencias desestabilizadoras y no igualitarias prevalezcan permanentemente (2014, pág. 36).

Para Piketty, el principal mecanismo de convergencia hacia la reducción de las desigualdades en una sociedad, es la difusión de los conocimientos derivados de la investigación, la ciencia y la tecnología, así como la inversión en la formación de capital humano, es decir el impulso de la capacitación y formación de habilidades. Estos elementos inciden directamente en el aumento de la productividad y la disminución de las desigualdades en la distribución del ingreso al nivel de las familias y los países que adoptan dichas medidas de política económica. La experiencia histórica de países pobre y emergentes así lo indica. Tal es el caso de China y otros países, quienes han adoptado los modos de producción y formación de capital humano de los países ricos, con lo que han logrado incrementar su productividad. El proceso de apertura comercial ayuda en este sentido. Por otra parte, si bien el libre juego de oferta y demanda, así como la movilidad del capital y del trabajo constituyen también elementos que pueden conducir a la reducción de desigualdades, sin embargo, sus resultados suelen ser ambiguos y contradictorios.

Entre las principales fuerzas de divergencia, es decir aquellos elementos que no solo impiden o limitan la disminución de las desigualdades sociales sino que tienden a aumentarla, se encuentran principalmente la falta de inversión en la formación de capital humano y de capacidades y habilidades

que impulsen la productividad de los trabajadores. Otro poderoso elemento divergente de la desigualdad es la tendencia hacia una excesiva acumulación y concentración de la riqueza en unas cuantas manos, una economía mundial que no crece y por lo tanto, no genera los empleos necesarios que demanda una mano de obra en crecimiento, y un elevado rendimiento del capital frente al trabajo, factor que constituye el principal obstáculo a la igualdad en la distribución de ingresos en el largo plazo.

## 2.2 Distribución del ingreso

En su renombrado libro *“El capital en el siglo XXI”*, Piketty plantea de entrada que la distribución de la riqueza constituye en la actualidad uno de los temas más debatidos y controversiales dentro de la economía política contemporánea, y que los debates acerca del tema están más cargados de especulaciones teóricas y prejuicios políticos y sociales que de información y hechos objetivos. El análisis de Piketty sobre la riqueza parte de algunas interrogantes entre las que se encuentran: “¿Acaso la dinámica de la acumulación del capital privado conduce inevitablemente a una concentración cada vez mayor de la riqueza y del poder en unas cuantas manos, como lo creyó Marx en el siglo XIX? ¿Acaso las fuerzas equilibradoras del crecimiento, la competencia y el progreso técnico conducen espontáneamente a una reducción y a una armoniosa estabilización de las desigualdades en las fases avanzadas del desarrollo, como lo pensó Kuznets en el siglo XX?” (2014).

Si bien es cierto, que las predicciones de Marx acerca de la concentración del ingreso en manos de la clase social dueña del capital no sucedieron tal como lo planteaba, no es menos cierto que históricamente el sistema capitalista imperante en el mundo conduce inevitablemente a profundas desigualdades derivadas de la concentración del ingreso, dado que históricamente la tasa de rendimiento del capital ha superado con creces la tasa de crecimiento de la producción y del ingreso.

Cabe mencionar, que el tema de la distribución del ingreso ha sido el centro de análisis desde el momento mismo del surgimiento de la ciencia de la economía política clásica, a fines del siglo XVIII, teniendo como marco de referencia la revolución industrial, el éxodo del campo a la ciudad y un descomunal crecimiento demográfico. Dichos sucesos históricos significaron importantes interrogantes y análisis respecto a sus implicaciones en el reparto de la riqueza, la estructura social y el equilibrio de las sociedades europeas.

Para Thomas Malthus, importante economista clásico famoso por sus catastróficas previsiones demográficas contenidas en su *“Ensayo sobre el principio de población”* publicado en 1798, el problema económico central era la sobrepoblación derivada de una alta tasa de natalidad, misma que frente a los escasos de la tierra y los medios de producción había provocado el estancamiento de los salarios agrícolas y el incremento de la renta de la tierra en beneficio de sus propietarios, lo que derivaba en la concentración de la renta nacional en sus manos. Además, podría generar catástrofes, hambrunas y por supuesto pobreza extrema; como solución extrema, propuso la limitación de la ayuda pública a los pobres y el control estricto de la tasa de natalidad.

Por su parte, David Ricardo, otro de los padres de la economía clásica dedicó también buena parte de sus investigaciones al tema de la distribución de la riqueza. Éste al igual que Marx preveía una tendencia creciente a la concentración de producción y la renta en una clase social. Ricardo, en su famosa obra *“Principio de economía política y tributación”* publicado en 1817, sustentaba que a largo plazo el precio de la tierra y su renta observaba una tendencia creciente en beneficio de sus propietarios. A partir del postulado de Malthus acerca del crecimiento de la población y la producción, frente a una limitada disponibilidad de tierras cultivables, introduce el principio económico básico de *escasez*.

Dadas las leyes de oferta y demanda, el proceso debería conducir al incremento en el precio de la tierra y su arrendamiento y con el tiempo a la concentración del grueso del ingreso nacional en manos

de los terratenientes. Por esta razón, conduciría lógicamente al desequilibrio social, proponiendo como solución una medida fiscal novedosa: el establecimiento de un impuesto más gravoso sobre la renta del suelo. Si bien, el mecanismo del precio y el libre juego de oferta y demanda como mecanismos de asignación de la riqueza y fundamentos básicos de la economía capitalista suponen la corrección de distorsiones económicas, la realidad histórica demuestra que no bastan para impedir la concentración de la riqueza en unas cuantas manos.

Medio siglo después, en un contexto social y económico distinto al que le tocó vivir a Ricardo, Marx publica el primer tomo de su obra cumbre el *Capital*, referencia obligada de todo economista. La preocupación de los teóricos como Marx ya no era si la agricultura sería capaz de alimentar a una población creciente y la desigualdad provocada por la concentración del ingreso en manos de los terratenientes, sino más bien comprender la incidencia de la revolución industrial y el sistema capitalista en la sobre explotación de la mano de obra, así como en la generación de pobreza, desigualdad y depauperización de grandes masas urbanas emigradas del campo a las ciudades, provocando el estancamiento de los salarios por un período prolongado, surgiendo el concepto básico en su tesis del proletariado industrial. Correspondiente al estancamiento y caída de los salarios se generó un intenso proceso de concentración de la riqueza nacional en manos de los dueños del capital, ya sea a través de beneficios industriales, renta del suelo o rentas urbanas. Como consecuencia de la situación anterior se gestaron los primeros movimientos comunistas y socialistas.

En el contexto socioeconómico anterior, Marx desarrolla su investigación a partir de dos interrogantes centrales: ¿Para qué sirvió el desarrollo de la industria, para qué sirvieron todas esas innovaciones técnicas...si al cabo de medio siglo de desarrollo industrial la situación de las masas siguió siguiendo igual de miserable? ¿Qué se puede decir de la evolución que tendría semejante sistema a largo plazo? En su célebre *Manifiesto comunista* publicado en 1848 expone las contradicciones internas del capitalismo a partir del cual inicia un largo período de estudio para demostrar su tesis, producto del cual en 1867 se publica el primer tomo del *Capital*, obra que queda inconclusa a su muerte en 1883, pero que es concluida por Engels con la publicación de dos tomos más a partir de sus notas.

Marx, al igual que Ricardo centró su investigación en el análisis de las contradicciones internas del capitalismo y sus consecuencias, marcando distancia de los economistas burgueses para quien el sistema capitalista se autorregulaba gracias al sistema de precios, las leyes de oferta y demanda y el concepto de mano invisible, postulados en su momento por Adam Smith y Say. De la misma manera fue más allá de los socialistas utópicos (proudhonianos), a quienes criticó por contentarse únicamente con criticar el estado de las cosas y no plantear un análisis científico de la situación de pobreza imperante y medidas de solución, que en su caso fueron radicales.

De esta manera, partiendo de los principios de escases y aumento del precio del capital propuestos por Ricardo, Marx analizó a profundidad la dinámica del capital, que en un mundo de carácter industrial observaría una tendencia natural a la acumulación absoluta, proponiendo su famosa obra "*principio de acumulación infinita*", que en pocas palabras conduciría irremediabilmente a la concentración del capital como proporción del ingreso nacional, y a una situación de equilibrio social y económico inestable, que tarde o temprano llevaría a la rebelión y toma del poder por parte del proletariado. Si bien dicha previsión marxista nunca tuvo lugar debido al potencial del progreso técnico, la innovación tecnológica y el incremento de la productividad, elementos que tienden a equilibrar el proceso de concentración del ingreso, hasta la fecha no han sido capaces de evitar las graves desigualdades sociales y la pobreza imperante en el mundo.

Es imposible referirse a la distribución del ingreso y desigualdad sin mencionar los trabajos de Simon Kuznets, quien a diferencia de los presagios apocalípticos de Malthus, Ricardo y Marx, en 1955 inaugura una etapa de optimismo económico. Según Kuznets, a largo plazo y en etapas avanzadas de



desarrollo del capitalismo la desigualdad en el ingreso observa una tendencia a la disminución, independientemente de las políticas públicas adoptadas por los países y las condiciones específicas de los países. De esta manera, bastaba con esperar y dejar que las leyes de la economía capitalista operen libremente a través de los mecanismos de la mano invisible, el libre mercado y el sistema de precios, popularizándose la expresión: “El crecimiento es una marea ascendente que levanta todos los barcos” (Piketty, 2014, p. 25). El problema es que en la realidad actual ni existe crecimiento económico sostenible ni operan las leyes de mercado según lo establece la teoría. Congruente con la visión económica optimista de Kuznets, Robert Solow postula en 1956 su análisis centrado en una situación de crecimiento económico en la cual todas las variables macroeconómicas (producción, ingresos, beneficios, sueldos.) se alinean en este sentido, situación sistematizada en la conocida expresión: *Un sendero de crecimiento equilibrado*; así, todos los grupos sociales se ven beneficiados.

La teoría optimista de Kuznets tuvo gran influencia en la economía política mundial cuando menos hasta los años 90's. Su investigación estaba ampliamente respaldada en su profundo análisis estadístico y fue pionera en el estudio de los ingresos y su distribución en Estados Unidos durante los 35 años que abarca el período 1913-1948. Su análisis se sustentó en las declaraciones de ingresos tomadas del impuesto federal sobre el ingreso creado en 1913 y en las estimaciones del ingreso nacional calculadas por Kuznets. De esta manera, constituyó el primer análisis objetivo acerca de la desigualdad en la distribución de ingresos de una sociedad. El principal hallazgo fue una considerable reducción en la desigualdad de ingresos para el período 1913-1948. Dicha conclusión tuvo una gran influencia en la política económica mundial a partir de la posguerra, influencia que abarca hasta nuestros días y que hace que el estado nacional poco intervenga en la actividad económica como mecanismo de compensación social.

Del mismo modo, hasta los años setentas, la teoría económica ortodoxa planteaba que una economía en crecimiento impulsaría el desarrollo y el bienestar social, tal como una marea alta hace subir los barcos. De hecho, renombrados economistas teóricos como Robert Lucas, premio nobel de economía, planteaban que una economía vigoriza tenía mejores efectos redistributivos que una política social con dicho propósito. Por ello, Lucas (2003) citado por Caparrós expresó: “de todas las tendencias nocivas para una economía eficaz, la más seductora y venenosa consiste en hacer foco en cuestiones de distribución” (2014, p. 350).

No obstante, la tendencia a la concentración del ingreso en todos los lugares del mundo, junto al incremento en el precio de los alimentos y otros factores de orden socioeconómico y político ha tenido como consecuencia pobreza, desigualdad e inestabilidad económica y política en muchos países. En la actualidad, los organismos multinacionales (FMI, BM), se han manifestado al respecto considerando que la desigualdad económica es causa del retraso en el crecimiento económico, de crisis financieras y debilidad en la demanda y formación del mercado interno. Más allá de los simples efectos económicos nocivos, se considera que la desigualdad también tiene serio impacto en las condiciones de salud y educación de la población, limitando de esta manera la formación de capital humano en los países. Asimismo, es fuente de inestabilidad política, por ejemplo, el pasado reciente de América Latina, la región más desigual del mundo es muestra fehaciente de ello.

A pesar de tendencia a la concentración del ingreso en los últimos 30 años y los nocivos efectos económicos y sociales de la desigualdad, muchos países ricos siguen siendo escépticos en este sentido, manteniendo una postura ortodoxa al respecto. No obstante, a partir de la crisis financiera global de 2008, misma que impactó de manera directa la economía de todos los países, tanto a pobres como ricos, hubo un reconocimiento unánime de la necesidad de establecer acciones urgentes de recate y control financiero, y aunque en menor medida, invertir para reducir la pobreza y desigualdad asociada a la crisis.

Para Caparrós (2014), si bien la excesiva concentración del ingreso preocupa a todos los países, frente a la desigualdad lo que se busca no es la igualdad sino evitar los extremos en la distribución, la medida...”Como dice una declaración contra la desigualdad de Oxfam. La desigualdad ha sido vinculada a diferentes problemas sociales, incluyendo la violencia, la enfermedad mental, el crimen y la obesidad. Es más: se ha demostrado que la desigualdad no solo es mala para los pobres sino también para los ricos. Las personas más ricas viven más saludables y más felices si viven en sociedades más igualitarias” (p. 355).

### 2.3 Situación de desigualdad en México

La Encuesta Nacional de Ingresos y Gastos de los Hogares (Enigh), levantada de manera bianual por el Instituto Nacional de Estadística y Geografía (Inegi), constituye el instrumento estadístico idóneo para el análisis de la situación de desigualdad en el país. Como punto de partida es importante mencionar que de acuerdo a las proyecciones de población del Consejo Nacional de Población (Conapo, 2013), México contaba a mediados del 2013 con una población de 117,418,216 habitantes, quienes cohabitaban en 31,559,379 hogares integrados por 3.7 personas en promedio. De acuerdo al reporte de la Enigh correspondiente al 2012, el ingreso promedio trimestral a precios constantes considerando transferencias era de \$38,125. Asimismo, entre el período 2010-2012, la desigualdad económica medida a través del coeficiente de concentración del ingreso, conocido también como coeficiente de Gini, observó un repunte del 1.15%, al pasar de un coeficiente del 0.487 en el 2010 a 0.496 en 2012. Vista la desigualdad en un período de mediano plazo, entre los años 2006 al 2012, ésta observa una discreta disminución del 1%, pasando el coeficiente de Gini de 0.50 a 0.496 del 2006 al 2012. (Tabla 1).

Tabla 1. INGRESO CORRIENTE TOTAL PROMEDIO TRIMESTRAL EN MÉXICO POR DECILES DE INGRESO (Precios Constantes 2012)

DECILES DE HOGARES <sup>1</sup>	AÑO DE LEVANTAMIENTO			
	2006	2008	2010	2012
<b>INGRESO CORRIENTE TOTAL</b>	<b>43 698</b>	<b>42 865</b>	<b>37 574</b>	<b>38 125</b>
I	7 796	7 136	6 633	6 997
II	13 506	12 460	11 673	11 794
III	17 780	16 792	15 611	15 734
IV	22 161	20 986	19 650	19 513
V	27 072	25 628	23 973	23 914
VI	32 611	31 501	29 059	28 862
VII	40 357	39 381	35 605	35 570
VIII	50 788	50 084	45 089	44 849
IX	69 194	69 159	61 133	61 014
X	155 715	155 525	127 313	133 003
<b>COEFICIENTE DE GINI<sup>2</sup></b>	<b>0.445</b>	<b>0.457</b>	<b>0.435</b>	<b>0.440</b>

<sup>1</sup> Los hogares están ordenados en los deciles de acuerdo con su ingreso corriente total trimestral.

<sup>2</sup> El coeficiente de Gini es una medida de concentración del ingreso: toma valores entre cero y uno. Cuando el valor se acerca a uno, indica que hay mayor concentración del ingreso; en cambio cuando el valor del Gini se acerca a cero la concentración del ingreso es menor.

NOTA: Cifras expandidas conforme a las proyecciones de población del CONAPO actualizadas en abril de 2013.

FUENTE: INEGI. Encuesta Nacional de Ingresos y Gastos de los Hogares. Año correspondiente.

Fuente: Tomada de Inegi-Enigh 2012

Un reflejo de la desigualdad en México es la concentración del ingreso en los deciles más altos de la población. Así, la Enigh 2012 muestra también que el ingreso promedio trimestral captado por el

decil X (\$133,003), es equivalente a 19 veces el que obtiene el decil I (\$6,997). Asimismo, el ingreso concentrado en los deciles IX y X (\$194,017), representa casi el 51% del ingreso promedio trimestral total (\$381,250). (Tabla 1). De todo lo anterior se puede concluir que el problema de desigualdad en México está presente en el tiempo, con tendencia a incrementarse. Los efectos de la crisis económica del 2008, el ritmo insuficiente del crecimiento económico, el desempleo y la precaridad del trabajo así como la caída del poder adquisitivo de los salarios explican en gran medida la persistencia de la desigualdad. (Tabla 1).

Por otra parte, descontando el efecto de los ingresos por transferencias efectuadas del gobierno e instituciones privadas a las familias, el panorama de la desigualdad en México cambia sustancialmente. Así, para el año 2012 el ingreso promedio trimestral sin transferencias, comparado con el ingreso que sí las incluye se reduce en un 24.5 % al pasar de \$38,125 pesos a \$30,608. Derivado de ello, el coeficiente de Gini se eleva en un 12.7%, al pasar de 0.440 con transferencias a 0.496 sin éstas. De la misma manera, el decil I que engloba los hogares de más bajos ingresos se ve fuertemente impactado al momento de descontar las transferencias, ya que observan una disminución en el ingreso del 2.8%, al d de \$6,997 pesos a \$2,493. Asimismo, sin considerar transferencias, el ingreso promedio trimestral del decil X equivale a 47.3 veces del ingreso captado por el decil I, y los deciles IX y X concentran el 82.3 del ingreso promedio total (Tabla 2). Todo ello denota el fuerte impacto de las transferencias en la economía familiar, principalmente las derivadas de la política pública asistencialista y de las remesas familiares, sobre todo en los deciles de hogares de más bajos ingresos.

Tabla 2. INGRESO CORRIENTE TOTAL PROMEDIO TRIMESTRAL EN MÉXICO  
SIN TRANSFERENCIAS POR DECILES DE INGRESO  
(Precios Constantes 2012)

DECILES DE HOGARES <sup>2</sup>	AÑO DE LEVANTAMIENTO			
	2006	2008	2010	2012
<b>INGRESO CORRIENTE TOTAL</b>	<b>35 913</b>	<b>35 804</b>	<b>31 004</b>	<b>30 608</b>
I	2 401	2 345	2 296	2 439
II	7 478	6 929	6 487	6 250
III	12 074	11 548	10 621	9 983
IV	16 291	15 914	14 470	13 807
V	20 758	20 381	18 657	18 107
VI	26 285	25 872	23 631	22 762
VII	33 090	33 163	29 611	28 756
VIII	43 281	43 086	38 278	37 483
IX	60 166	60 258	52 726	51 167
X	137 308	138 544	113 266	115 322
<b>COEFICIENTE DE GINI<sup>3</sup></b>	<b>0.500</b>	<b>0.507</b>	<b>0.487</b>	<b>0.496</b>

<sup>1</sup> No se consideran los ingresos por transferencias del gobierno (becas, oportunidades, etc.), transferencias de instituciones privadas (beneficencia) y transferencias de otros hogares, (regalos, remesas).

<sup>2</sup> Los hogares están ordenados en los deciles de acuerdo con su ingreso corriente total trimestral antes de transferencias.

<sup>3</sup> El coeficiente de Gini es una medida de concentración del ingreso: toma valores entre cero y uno. Cuando el valor se acerca a uno, indica que hay mayor concentración del ingreso; en cambio cuando el valor del Gini se acerca a cero la concentración del ingreso es menor.

NOTA: Cifras expandidas conforme a las proyecciones de población del CONAPO actualizadas en abril de 2013.

FUENTE: INEGI. Encuesta Nacional de Ingresos y Gastos de los Hogares. Año correspondiente.

Fuente: Tomada de Inegi-Enigh 2012

Del análisis de desigualdad derivado del ingreso per cápita promedio trimestral (el ingreso del hogar dividido entre su número de integrantes), se desprenden también hechos relevantes. Así, los datos de la Enigh 2012 muestran un Gini de 0.470, el cual frente al 0.460 prevaleciente en 2010 representa una disminución del 2.2% en la situación de desigualdad. Entre 2006 y 2012 se observó una disminución marginal del 1.7%, lo cual se explica entre otros factores por la crisis económica iniciada en 2008, la cual generó una caída del 5% en el Gini reportado entre ese año y el 2010 al pasar de 0.483 a 0.460. Asimismo, en el 2012 el ingreso promedio trimestral per cápita del decil X (39,534 pesos) representó 23 veces el obtenido por el decil I (1,715 pesos). (Tabla 3).

Tabla 3 INGRESO CORRIENTE TOTAL PROMEDIO TRIMESTRAL PER CÁPITA POR DECILES DE PERSONAS (Precios Constantes 2012)

DECILES DE PERSONAS <sup>2</sup>	AÑO DE LEVANTAMIENTO			
	2006	2008	2010	2012
<b>INGRESO CORRIENTE TOTAL</b>	<b>11 046</b>	<b>10 705</b>	<b>9 694</b>	<b>10 259</b>
I	1 754	1 602	1 551	1 716
II	3 099	2 826	2 827	2 948
III	4 134	3 867	3 820	3 932
IV	5 180	4 923	4 768	4 930
V	6 282	6 099	5 811	5 985
VI	7 700	7 528	7 084	7 233
VII	9 475	9 361	8 726	8 893
VIII	12 170	12 038	11 123	11 417
IX	17 302	16 918	15 415	16 002
X	43 361	41 893	35 817	39 534
<b>COEFICIENTE DE GINI<sup>3</sup></b>	<b>0.478</b>	<b>0.483</b>	<b>0.460</b>	<b>0.470</b>

<sup>1</sup> El ingreso per cápita se calcula dividiendo el ingreso del hogar por el número de sus integrantes.

<sup>2</sup> Las personas están ordenadas en los deciles de acuerdo con su ingreso corriente total trimestral per cápita.

<sup>3</sup> El coeficiente de Gini es una medida de concentración del ingreso; toma valores entre cero y uno. Cuando el valor se acerca a uno, indica que hay mayor concentración del ingreso; en cambio cuando el valor del Gini se acerca a cero la concentración del ingreso es menor.

NOTA: Cifras expandidas conforme a las proyecciones de población del CONAPO actualizadas en abril de 2013.

FUENTE: INEGI. Encuesta Nacional de Ingresos y Gastos de los Hogares. Año correspondiente.

Fuente: Tomada de Inegi-Enigh 2012

Como se mencionó con anterioridad, en México es relevante el impacto que tienen las transferencias públicas y privadas en las condiciones de bienestar de la población. De esta manera, del análisis del ingreso promedio trimestral per cápita sin transferencias reportado en la Enigh 2012 se observa un fuerte repunte en las condiciones de desigualdad, pues el Gini sin transferencias (0.470) se eleva a 0.503 una vez consideradas éstas, lo que representa un incremento del 5.5% en la concentración del ingreso. Asimismo, el ingreso promedio trimestral sin transferencias del decil X es 41 veces más alto respecto al obtenido por el decil I. (Tabla 4).

En síntesis, México ha experimentado durante las últimas décadas una persistente desigualdad producto del estancamiento económico, la minusvalía del poder adquisitivo de los salarios, el desempleo y la prevalencia de la pobreza en más de la mitad de la población. Ello no obstante ser la decimocuarta

economía del mundo. En palabras de Esquivel (2015), la desigualdad en la distribución de los ingresos provoca que en un país rico sigan persistiendo millones de pobres.

Tabla 4. INGRESO CORRIENTE TOTAL PROMEDIO TRIMESTRAL PER CÁPITA SIN TRANSFERENCIAS EN DECILES DE PERSONAS (Precios Constantes 2012)

DECILES DE PERSONAS <sup>3</sup>	AÑO DE LEVANTAMIENTO			
	2006	2008	2010	2012
<b>INGRESO CORRIENTE TOTAL</b>	<b>9 078</b>	<b>8 942</b>	<b>7 999</b>	<b>8 236</b>
I	814	760	737	808
II	2 023	1 887	1 875	1 884
III	3 027	2 865	2 826	2 801
IV	3 979	3 851	3 688	3 679
V	5 021	4 945	4 633	4 667
VI	6 218	6 225	5 767	5 764
VII	7 806	7 867	7 205	7 166
VIII	10 155	10 195	9 328	9 266
IX	14 525	14 454	13 051	13 073
X	37 210	36 371	30 881	33 251
<b>COEFICIENTE DE GINI<sup>4</sup></b>	<b>0.510</b>	<b>0.513</b>	<b>0.492</b>	<b>0.503</b>

<sup>1</sup> El ingreso per cápita se calcula dividiendo el ingreso del hogar por el número de sus integrantes.

<sup>2</sup> No se consideran los ingresos por transferencias del gobierno (becas, oportunidades, etc.), transferencias de instituciones privadas (beneficencia) y transferencias de otros hogares, (regalos, remesas).

<sup>3</sup> Las personas están ordenadas en los deciles de acuerdo con su ingreso corriente total trimestral per cápita antes de transferencias.

<sup>4</sup> El coeficiente de Gini es una medida de concentración del ingreso: toma valores entre cero y uno. Cuando el valor se acerca a uno, indica que hay mayor concentración del ingreso; en cambio cuando el valor del Gini se acerca a cero la concentración del ingreso es menor.

NOTA: Cifras expandidas conforme a las proyecciones de población del CONAPO actualizadas en abril de 2013.

FUENTE: INEGI. Encuesta Nacional de Ingresos y Gastos de los Hogares. Año correspondiente.

Fuente: Tomada de Inegi-Enigh 2012

### 3. POLÍTICA DE ESTADO COMO ALTERNATIVA DE SOLUCIÓN A LA POBREZA Y DESIGUALDAD EN MÉXICO.

De acuerdo a cifras oficiales del Coneval y el Inegi, la pobreza y desigualdad en México aumentó significativamente a partir de la crisis económica global iniciada en 2008. Ante ello, es necesario investigar las causas más profundas que explican dichos problema y actuar en consecuencia, tanto por parte del Estado como de la sociedad mexicana mediante la instrumentación y puesta en marcha de políticas públicas efectivas.

Visto el problema de la pobreza en México, tanto del lado del ingreso como desde la perspectiva multidimensional, así como la situación de desigualdad imperante, es evidente el impacto negativo de la crisis económica mundial, pero también los magros resultados de las políticas públicas instrumentadas, lo que constituye una fuerte llamada de atención a una acción más decidida por parte del Estado y la sociedad mexicana. Ello en virtud de que tales programas únicamente se enfocan en combatir la pobreza coyuntural, no la estructural (Levy, 2010). En este sentido, en México es necesario pasar de la teoría a la

práctica, por lo que se requiere de la intervención del estado nacional con el propósito de enfrentar de manera decidida el problema de pobreza y desigualdad, para lo cual es necesario el establecimiento de una política de Estado.

A continuación se presentan de manera general algunos lineamientos propuestos para el diseño e instrumentación de una política de Estado contra la pobreza y desigualdad en México, partiendo de los aspectos conceptuales a los contenidos programáticos establecidos a través de políticas públicas específicas que en conjunto definen e integran a la política de Estado en su conjunto.

### **3.1 Política de Estado. Aspectos conceptuales**

En cualquier economía del mundo queda claro que el Estado tiene la importante responsabilidad de hacerse cargo de la provisión de ciertos bienes públicos, en cuyo mercado se presentan externalidades, monopolios naturales y asimetrías de información que afectan en gran medida a la población más pobre y marginada. Dichos bienes son conocidos como *bienes meritorios*, toda vez que dicha población merece el acceso a los mismos en razón de su dignidad humana y/o de su ciudadanía nacional (Aguilar, 2007).

Entre los bienes meritorios sobresalen el acceso a la alimentación, la salud y la educación. Dicha responsabilidad del Estado, de profundo carácter ético y de justicia social, debe ser llevada a cabo mediante la instrumentación de políticas públicas racionales. Hablar de políticas públicas, y de gobernar por políticas públicas supone decisiones de gobierno que incorporan la opinión, la participación y la corresponsabilidad de los gobernados, en su calidad de ciudadanos electores y contribuyentes.

En esta perspectiva, el establecimiento de una política de Estado frente a la pobreza en México supone formular y desarrollar un conjunto de políticas públicas articuladas en torno a este objetivo primordial, sujetas a un fundamento legal respaldado constitucionalmente, con todo el apoyo político necesario para su implementación, y viables desde el punto de vista administrativo y de racionalidad económica. Teóricamente, una política de Estado es entendida como un conjunto de acciones integrales de política pública, mismas que son incorporadas de manera permanente a la gestión gubernamental, cuyo propósito deberá constituirse en una prioridad nacional. De esta manera, dichas actividades deberán trascender en beneficio de la sociedad en su conjunto, y dada su importancia económica, social y política, deben evitarse los vaivenes político-administrativos.

De manera específica, una política de Estado contra la pobreza y desigualdad en México deberá reunir las siguientes características:

1. Debe quedar inscrita en el texto constitucional.
2. Debe trascender los límites sexenales; es decir el largo plazo como condición fundamental.
3. Debe aislarse de apasionamientos político partidistas o de grupos.
4. La flexibilidad como característica natural de lo público estatal.
5. La factibilidad de consolidarse para beneficio de la sociedad mexicana.
6. Debe constituirse como elemento fundamental para lograr la gobernabilidad del país.
7. Debe lograr consensos entre los diversos grupos sociales y políticos.
8. Debe contribuir a abatir rezagos, generar riqueza, empleo y sobre todo, contribuir a lograr la estabilidad económica, social y política del país.

#### **3.1.1 Aspectos instrumentales de una política de Estado contra la pobreza y desigualdad en México.**

Las acciones contra la pobreza y desigualdad implementadas en la mayor parte de los países de América Latina, presuponen el crecimiento económico y la estabilidad macroeconómica como prerrequisitos de acción. De esta manera, la política económica, el gasto social y el combate a la pobreza y desigualdad se supedita a una política ortodoxa de equilibrio fiscal sustentada por el Banco Mundial y el Fondo Monetario Internacional. Sin embargo, muchos países, han podido mejorar la calidad de vida de sus habitantes y disminuir la brecha de la desigualdad social a partir de la aplicación del enfoque de bienestar social y justicia distributiva que sustenta el desarrollo humano de sus habitantes, lo que se ha traducido a su vez en exitosas políticas públicas, como es el caso de Brasil con el programa Bolsa Familia.

Ambos enfoques de la pobreza y desigualdad tienen implicaciones relevantes al momento de diseñar e implementar las políticas públicas. En principio, habrá que definir cómo conceptualizamos a quienes habrán de recibir los bienes y/o servicios. Este aspecto es relevante, pues la consideración de los pobres como **agentes pasivos** conducirá a políticas públicas paternalistas; por el contrario, si se asumen como **agentes activos**, se diseñarán políticas públicas que los involucran en el proceso de cambio. Podría considerarse en términos generales, que las políticas contra la pobreza y desigualdad impulsadas por los organismos multinacionales antes mencionados se identifican con el primer enfoque, en tanto que las segundas se enfocan al desarrollo humano a través de las capacidades básicas de los individuos.

Un segundo aspecto que habrá que definirse, es el tipo de bienes a distribuir entre los pobres. Si éstos son alimentos y productos que satisfacen sus necesidades básicas, o constituyen bienes que les permiten expresar sus preferencias e impulsan sus capacidades básicas. Un tercer elemento a considerar son los criterios aplicados para la distribución de los bienes y servicios entre los pobres, Este último es un aspecto relevante de la justicia distributiva, ya que aplicar un criterio u otro en la distribución significará dejar fuera a estratos sociales que también necesitan ayuda. En este sentido, la escasez de recursos obliga a tomar no la alternativa mejor, sino la menos mala.

Finalmente, hay que considerar los criterios utilizados para la medición de la pobreza. En el campo de las políticas públicas persiste un fuerte debate acerca del medio más adecuado para su medición. Así, por una parte podemos encontrar metodologías consideradas como tradicionales tales como el Índice de Calidad de Vida, Necesidades Básicas Insatisfechas, y Porcentaje de Pobres, las cuales se conceptualizan como **métodos directos**. Por otro lado, están la medición de la Línea de Pobreza y la estimación del Ingreso Per cápita, considerados ambos como métodos **indirectos**, ya que su cálculo parte del ingreso monetario proveniente de la fuerza de trabajo disponible para la satisfacción de sus necesidades fundamentales. A ello habría que sumarle el criterio de medición multidimensional de la pobreza calculado a partir de las carencias sociales.

Como puede apreciarse de las cifras de pobreza y desigualdad analizadas, en México la medición de la pobreza continúa siendo estática y reduccionista, pues su dimensión únicamente considera la variable ingreso y ciertos indicadores subjetivos de carencia social. Ello no ayuda a explicar de manera objetiva el problema, ya que únicamente cuantifica su incidencia en función del número de hogares y personas pobres. Por ello, surge la inquietud de investigadores y funcionarios públicos acerca de la necesidad de aplicar nuevas metodologías al estudio del problema, con el propósito de hallar explicaciones y soluciones dinámicas al objeto de estudio. Al respecto, en el ámbito mundial se vienen aplicando nuevos enfoques sustentados en análisis dinámicos de la pobreza. Ello ha permitido superar los resultados obtenidos mediante los enfoques estáticos tradicionales, mismos que se centran en la medición y clasificación de la pobreza. El reto es lograr aplicar en México estos criterios dinámicos de medición.

### 3.1.2 Contenidos programáticos de una política de Estado contra la pobreza y desigualdad en México.

- a). Gasto público orientado a la creación de infraestructura social
- b). Inversión social en educación y salud, componentes fundamentales del capital humano y elementos básicos para la disminución de la desigualdad y el rezago social.
- c). Enfoque central hacia el desarrollo humano del país en su conjunto
- d). Impulso al crecimiento económico mediante la reactivación del mercado interno nacional.
- e). Impulso a la generación de empleos en la economía formal y la redistribución de la riqueza mediante una política fiscal efectiva.
- f). Reforma salarial progresiva que impulse el poder adquisitivo del salario mínimo
- g). Impulso decidido a la investigación, ciencia, tecnología e innovación
- h). Impulso a la creación de micro, pequeñas y medianas empresas, así como del microcrédito.
- i). Aplicación de un programa eficiente de aplicación de las remesas internacionales en beneficio de la sociedad mexicana.
- j). Diseño de estrategias de orientación a la población relacionada con el fenómeno de la migración, acerca del uso más adecuado de las remesas recibidas, con propósitos de superación de la pobreza y desigualdad.

## REFERENCIAS

1. Aguilar, L. (2007). *El estudio de las políticas públicas. Primera antología*. México, Miguel Ángel Porrúa.
2. Bane, M. and Ellwood, D. (1986). Slipping into and out of Poverty: The Dynamics of Spells, *Journal of Human Resources*, Vol. 21, No.1.
3. Calderón, C. and Chong A. (2004). Volume and Quality of Infrastructure and the Distribution of Income: An Empirical Investigation, *Review of Income and Wealth*, 50, 87-106.
4. Caparrós, M. (2014). *El hambre*. Argentina, Ed. Planeta
5. Coneval (2010). Informe de pobreza multidimensional en México, 2008, México, Coneval.
6. Coneval (2013). Informe de pobreza multidimensional en México, 2012, México, Coneval.
7. Coneval (2014). Informe de pobreza multidimensional en México, 2014, México, Coneval.
8. Contreras, D. Et. Al. (2005). Dinámica de la Pobreza y Movilidad Relativa de los ingresos: Chile 1996-2001, *Serie documentos de trabajo, No. 232*, Departamento de Economía, Universidad de Chile, Santiago de Chile.
9. Cortés, F. and Escobar, A. (2005). Movilidad social intergeneracional en el México urbano, *Revista de la CEPAL*, número 85, abril.
10. Dieterlen, P. (2003). *La pobreza: un estudio filosófico*, México, Universidad Nacional Autónoma de México-Fondo de Cultura Económica.
11. Esquivel, G. (2015). Desigualdad extrema en México, México, Oxfam
12. Gallego, L. (2010). *Acercamiento al problema social de la pobreza. De las nociones de pobreza a los mecanismos causales*, España, Universidad Autónoma de Barcelona.
13. García, J. (2014). La dinámica de la pobreza en México. Una aproximación multidimensional, *Rev. Investigación Operacional*, Vol. 35, No. 2, pp. 141-147.
14. Hutchens, R. (1981). Entry and Exit Transitions in a Government Transfer Program: The Case of Aid to Families with Dependent Children, *The Journal of Human Resources*, Vol. 16, No. 2 (Spring): pp. 217-237.
15. Inegi (2014). Enigh 2012, México, Autor
16. Krugman, P. (2009). *De vuelta a la economía de la Gran Depresión y la crisis del 2008*. México, Grupo Editorial Norma.
17. Levi S. (2010). *Buenas intenciones, malos resultados. Política social, informalidad y crecimiento económico en México*, México, Ed. Oceano.
18. Lillard, L. and WILLIS, R. (1978). Dynamic aspects of earnings mobility, *Econometrica*, Vol. 46, No. 5, pp. 985-1012.



19. López, A. Beltrán, R. (2011). *El análisis dinámico de la pobreza. Enfoques, metodología y hallazgos*, México, UAM-Xochimilco.
20. López, J. and Perry, G. (2008). Inequality in Latin America: Determinants and Consequences. *Policy Research*, Working Paper 4504, Washington. D.C., World Bank.
21. Maurizio, R., Perrot, B. and Villafañe, S. (2008). *Dinámica de la pobreza y mercado de trabajo en Argentina Post-convertibilidad*, Buenos Aires, PNUD.
22. McCulloch, N. and Baulch, B. (2000). Simulating the impact of policy upon chronic and transitory poverty in rural Pakistan. *Journal of Development Studies*, Vol. 36, número 6. pp. 100-130.
23. Meléndez, A. (2000). *Educación y empleo como determinantes de la dinámica de la pobreza en México*, Tesis de Licenciatura, Universidad de las Américas, Puebla.
24. Micklethwait, J y Wooldridge, A. (2003). *Un Futuro perfecto. El desafío y la promesa secreta de la globalización*. México, Editorial Océano.
25. Nussbaum, M. and Sen, A. (2004). *La calidad de vida*, México, Fondo de Cultura Económica.
26. Pick, S., Sirkin, J. (2010). *Pobreza. Cómo romper el ciclo a partir del desarrollo humano*, México, Ed. Limusa.
27. Piketty, T. (2014). *El capital en el siglo XXI*. México, Fondo de Cultura Económica
28. Piketty, T. (2015). *La economía de las desigualdades. Cómo implementar una redistribución justa y eficaz de la riqueza*, México, Siglo veintiuno editores.
29. Programa de las Naciones Unidas para el Desarrollo (2010): *Informe sobre Desarrollo Humano 2010*, Washington, D.C., PNUD.
30. Rawls, J. (2002). *Teoría de la Justicia*, México, Fondo de Cultura Económica.
31. Salama, P. (2011). Luchas contra la pobreza en América Latina. El caso de la pobreza rural en Brasil. *Problemas del Desarrollo. Rev Latinoamericana de Economía*. Vol. 42/No. 165, abril-junio; págs. 7-34.
32. Santillán, M. and Laplante, B. (2009). La dinámica de la pobreza y las variables de población en la Argentina, un análisis longitudinal a partir de la encuesta permanente de hogares (1995-2003), *Notas de población*, no. 89, CEPAL.
33. Sen, A. (2000). *Desarrollo y Libertad*, México, Editorial Planeta
34. Soros, S. (1999). *La crisis del capitalismo global*. México, Plaza Janés
35. Stiglitz, J. (2007). *El malestar en la globalización*. México, Taurus
36. Urzúa, C. (2008). Evaluación de los efectos distributivos y espaciales de las empresas con poder de mercado en México, *Reporte técnico para la Comisión Federal de Competencia, EGAP*, Tecnológico de Monterrey, México.
37. Urzúa, C. and Brambila, C. (2010). Determinantes de la Pobreza Estatal, En: Aparicio, R. Villarespe, V. and Urzúa, C. (compiladores). *Pobreza en México: Magnitud y Perfiles*, Instituto de Investigaciones Económicas, UNAM-Instituto Tecnológico y de Estudios Superiores de Monterrey-Consejo Nacional de Evaluación de la Política Social, México.
38. Urzúa, C., Macías, A. and Sandoval, H. (2008). Tips for the Analysis of Poverty in Mexico, 1992-2005, *Revista de Administración, Finanzas y Economía*, Vol. 2, pp. 74-91.
39. Yunnus, M. (2010). *Empresas para todos. Hacia un nuevo modelo de capitalismo que atiende las necesidades más urgentes de la humanidad*. Colombia, Grupo Editorial Norma.

## Capítulo 18

pp. 202-214.

# FACTORES QUE IMPACTAN EL CRECIMIENTO ECONÓMICO: UN ESTUDIO COMPARATIVO ENTRE PAÍSES INDUSTRIALIZADOS Y ECONOMÍAS EMERGENTES USANDO MODELACIÓN MULTINIVEL

Ignacio Ortiz Betancourt, Mario Miguel Ojeda Ramírez & Patricia Tapia Blásquez  
Universidad Veracruzana

### RESUMEN

El gasto público es concebido como un instrumento de cambio social, a partir del cual el Estado puede impulsar el bienestar de los ciudadanos, y de esta manera los recursos públicos pueden ser empleados de una manera eficiente. No obstante, existe una profunda discusión acerca de qué áreas del gasto público deben ser favorecidas para generar un mayor impacto sobre la tasa de crecimiento del producto interno bruto (PIB). En este trabajo se estudió el efecto del gasto ejercido en salud pública, el gasto en inversión y la tasa de crecimiento poblacional como factores explicativos del patrón de comportamiento en la tasa de crecimiento del PIB para países industrializados y economías emergentes, durante el periodo 1995-2010. Se aplicó un modelo multinivel para analizar la variabilidad que hay en este periodo de tiempo y entre los países. Se encontró que sólo el gasto en inversión es significativo para explicar el ritmo de crecimiento del PIB.

### ABSTRACT

Public expenditure is conceived as an instrument for social change, from which the State can foster welfare of citizens, and thus public resources will be used efficiently. However, there is a deep discussion about which areas of public spending should be favored to a greater extent, as well as their impact on the growth rate of gross domestic product (GDP). This work studied the effect of public health expenditure, investment spending and population rate growth as factors for explaining the pattern of behavior in the growth rate of GDP of industrialized countries and emerging economies during the period 1995-2010. A multilevel model was applied to analyze the variability in this period of time and across countries, and found that only investment spending is significant to explain the growth rate of GDP.

**KEYWORDS:** public finances; economic development; multilevel modelling

## 1. INTRODUCCIÓN

El gasto público se define como el conjunto de erogaciones que llevan a cabo los entes económicos dirigidos por el Estado para adquirir los medios necesarios en la realización de sus actividades de producción de bienes o servicios, ya sean públicos o privados, así como también, para garantizar el bienestar de los ciudadanos, a través del gasto social en educación, salud, vivienda, entre otros.

Desde mediados del siglo XX, se ha discutido sobre el modelo de crecimiento adecuado para combatir los problemas sociales, como son el desempleo, la pobreza, la marginación, por mencionar algunos. Esta preocupación se refleja en el trabajo de Keynes (1936), quien fue el principal promotor del gasto público como mecanismo para acabar con los periodos de estancamiento económico y las consecuencias sociales derivadas de éste. En este sentido, la escuela de pensamiento keynesiano constituye una teoría económica básica para entender la estrecha relación entre el crecimiento económico y el desarrollo social.

Draibe y Riesco (2006) señalan que durante los últimos años ha cobrado fuerza, dentro del campo de los estudios del desarrollo económico, una postura integradora entre crecimiento económico y desarrollo social, dejando de lado aquellas que solo privilegian la primera sobre la segunda. Por parte,

Güendell y Barahona (2005) abordan el vínculo existente entre el acceso a la riqueza, a la libertad y al poder como premisa clave para alcanzar el desarrollo económico, lo cual se relaciona de manera directa con la teoría del desarrollo humano, comúnmente representada por Sen (2000) y por el trabajo desarrollado por la Organización de las Naciones Unidas (ONU,1990)<sup>11</sup>.

Ahora bien, de acuerdo con el Programa de las Naciones Unidas para el Desarrollo (PNUD, 2002), dentro de la teoría del desarrollo humano se destacan una serie de factores que inciden de manera directa en la capacidad de desarrollo del individuo, los cuales deben ser operados por el Estado a través de políticas públicas. Estos son: los derechos humanos, la equidad de etnia y religión, el empleo, la salud y la educación. Dentro de esta serie de elementos decisivos, surge la interrogante acerca de cuál de estos factores es el más importante. Al respecto, Lang (2012) afirma que la salud, la educación y el estándar de vida son los elementos más representativos para poder bosquejar el avance de un país en términos de desarrollo humano. No obstante, Picazzo et al. (2011) señalan al gasto en salud como un elemento trascendental para lograr impulsar el desarrollo humano, acompañado de los mecanismos institucionales que permiten el acceso a la misma. De manera similar, Rivera y Currais (2005) y Bakare (2011) mencionan el papel de la salud pública como un derecho fundamental del ser humano, así como también, el efecto multiplicador que el gasto en salud genera en el empleo y en el ritmo de crecimiento económico. Sepúlveda (2006) y Boyacioglu (2012) por su parte, afirman que el gasto en salud es fundamental para explicar el patrón de crecimiento y desarrollo económico de

En el plano institucional, en México, la Secretaría de Salud, por medio de la Comisión Nacional de Protección Social en Salud,<sup>12</sup> señala que al impulsar el gasto en salud se podrá contar con una población más sana como condición para incrementar el desarrollo económico nacional, toda vez que la salud representa la principal de las necesidades humanas. En esta línea, Wang (2011), Safdari, Mehrizi y Elahi (2001) y Bakare (op. cit. 2011) encuentran una fuerte relación positiva entre el gasto público y la tasa de crecimiento económico de un país. De manera contraria, Elmi y Sadeghi (2012) señalan que ésta relación sólo se observa en el corto plazo, más en el largo plazo no resulta significativa. En todo caso, el gasto en salud representa una variable importante para estudiar el ritmo de crecimiento económico de un país, y si éste ha sido capaz de acelerar su camino al desarrollo, el cual involucra aspectos económicos y sociales.

Otra variable importante es el factor demográfico. De acuerdo con un estudio realizado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2012,1), si bien el gasto en salud es, y será, uno de los elementos más importantes en el crecimiento económico global de los próximos cincuenta años, la caída en las tasas de natalidad y el proceso de envejecimiento de la población representan un factor negativo para el crecimiento económico; toda vez que esto se traduce en una reducción de la fuerza laboral necesaria para incrementar la producción nacional. En esta misma perspectiva Weil (2007) y Ashraf, Lester y Weil (2008) mencionan que la relación entre dichas variables no es significativa.

Con base en lo anterior, se espera que las bajas tasas de crecimiento poblacional reduzcan la tasa de crecimiento del PIB, salvo en aquellos países donde se genere el llamado “bono demográfico”, el cual es definido por Bloom et al. (2007) como la transición demográfica de un país de un escenario de equilibrio con altas tasas de natalidad y mortalidad, a otro con bajas tasas de fertilidad y mortalidad. La importancia de este fenómeno radica en que puede generar las condiciones para entrar en un círculo virtuoso de crecimiento económico. Resultados similares son presentados por Quang (2012), quien

<sup>11</sup> En dicho año se emite el Primer Informe Sobre Desarrollo Humano, el cual señala la necesidad de medir el bienestar no sólo en términos de ingreso. En este sentido, se plantea un enfoque del desarrollo de las personas, por las personas y para las personas.

<sup>12</sup> En este caso, se hace mención a las bondades de contar con un esquema de cobertura en salud, como es el Seguro Popular. Disponible en [http://www.seguro-popular.salud.gob.mx/index.php?option=com\\_content&view=article&id=332&Itemid=326](http://www.seguro-popular.salud.gob.mx/index.php?option=com_content&view=article&id=332&Itemid=326) [Revisado el 18 de marzo de 2013]

menciona la existencia de una relación positiva entre la tasa de crecimiento poblacional y el ritmo de crecimiento del PIB. Por lo dicho hasta el momento, se puede afirmar que el gasto en salud pública y la tasa de crecimiento poblacional son factores que inciden en la tasa de crecimiento del PIB.

No obstante, existen numerosos trabajos que pretenden explicar el comportamiento del PIB a partir de su relación con el gasto en inversión productiva,<sup>13</sup> principalmente en infraestructura. Al respecto, Canning y Fay (1993), Easterly y Revelo (1993) y Torrisi (2010) encuentran una relación positiva entre el gasto en infraestructura y la tasa de crecimiento económico. Aunque, trabajos como el de Barro (1991), Devarajan, Swaroop y Zou (1996) y Bukhari, Ali y Saddagat (2007) afirman que el gasto en inversión no impacta de manera significativa en la tasa de crecimiento económico.

Otro factor importante a considerar radica en el grado de desarrollo con el cual una nación es catalogada dentro del panorama económico global. En este sentido, el Banco Mundial (2002) señala que tradicionalmente a los países se les clasifica en dos grandes grupos, con base en su nivel de ingreso: desarrollados (industrializados) o en desarrollo. De manera similar, el Fondo Monetario Internacional (FMI)<sup>14</sup> clasifica a los países con base en su nivel de ingreso per cápita, a partir de lo cual bien podríamos formar nuestra propia idea acerca de lo que representa un país industrializado. No obstante, dicha clasificación supone una serie de dificultades, toda vez que no existe un consenso acerca de los requisitos que debe cumplir un país para considerarse industrializado, o poseer una economía desarrollada.

Dada la dificultad de hallar una definición precisa sobre éste término, para este trabajo de investigación se toman como países industrializados a aquellos que pertenecen al Grupo de los 7 (G-7)<sup>15</sup>, el cual agrupa a las siete economías más industrializadas del mundo: Estados Unidos de América (EUA), Gran Bretaña, Italia, Francia, Alemania, Japón y Canadá.

Por otra parte, durante los últimos años, el concepto de *economía emergente* se ha posicionado con fuerza en la esfera económica internacional, opacando al de *economía en vías de desarrollo*<sup>16</sup>. Tal como lo señala García-Herrero (2011), una economía emergente es aquella que presenta un gran potencial de crecimiento debido al ritmo ascendente que ha experimentado para disminuir la brecha con las economías desarrolladas, principalmente en términos del ingreso per cápita.<sup>17</sup>

No obstante, dentro del grupo de las economías emergentes, se halla un conjunto de países que ha destacado por sus patrones de crecimiento económico, llamado el bloque de los BRICS. Este término fue acuñado en 2001 por el economista Británico Jim O'Neill<sup>18</sup> para referirse a las cuatro economías emergentes más importantes del mundo: Brasil, Rusia, China, India y Sudáfrica<sup>19</sup>. Tal como lo menciona Santos (2005), este acrónimo se refiere a un grupo de países que destacan por la cuantía de sus recursos naturales y energéticos, el volumen y ritmo de crecimiento de su PIB, el flujo de sus exportaciones y su tasa de crecimiento poblacional.

<sup>13</sup> De acuerdo con el Ministerio de Finanzas de Dinamarca (2001), el gasto público en inversión abarca la construcción de infraestructura: nuevos edificios, caminos y carreteras, puertos, entre otros. Disponible en <http://www.fm.dk/publikationer/2001/investing-in-denmarks-future/2-what-is-public-investment/> [Revisado el 30 de abril de 2013]

<sup>14</sup> Con base en <http://datos.bancomundial.org/quienes-somos/clasificacion-paises> [Revisado el 10 de mayo de 2013]

<sup>15</sup> Con base en información del Departamento Académico de Estudios Internacionales del Instituto Tecnológico Autónomo de México (ITAM). Disponible en <http://g8mexico.itam.mx/preguntas.html> [Revisado el 14 de abril de 2013]

<sup>16</sup> De acuerdo con Frenkel y Rapetti (2009), el alza del concepto se debe a que las economías emergentes son aquellas economías en desarrollo que han experimentado elevadas tasas de crecimiento en su PIB. Dado su mayor peso en la economía mundial, éstas economías han sido el foco de atención, dejando de lado a aquellas que aún siguen en desarrollo y no han logrado detonar el ritmo de crecimiento de su PIB.

<sup>17</sup> De acuerdo con la OCDE (2012,2), México se constituye como una de las economías emergentes más prometedoras, toda vez que durante los últimos diez años ha demostrado solidez macroeconómica y una tasa de crecimiento constante.

<sup>18</sup> Presidente del banco de inversiones Goldman Sachs (EUA).

<sup>19</sup> Sudáfrica fue añadido a finales de 2010 no por el propio O'Neill, sino por medio de una invitación del presidente Chino Hu Jintao.

Sin embargo, tal como lo señala García-Herrero (2012), el grupo de los BRICS se caracteriza por ser restrictivo y emplear un horizonte de tiempo muy amplio para realizar sus proyecciones de crecimiento (más de 25 años). Es por esto que en los últimos años han surgido nuevos criterios para agrupar a las economías emergentes, tal como la presentada a partir del 2010 por el BBVA Research,<sup>20</sup> quien ha propuesto una nueva clasificación: economías *águilas* (EAGLES<sup>21</sup>) y economías *nido* (NEST). En este sentido, las *economías águilas*<sup>22</sup> se caracterizan por ser economías emergentes con alto potencial de crecimiento, las cuales se espera que en la próxima década superen el promedio de las tasas de crecimiento de los países industrializados.<sup>23</sup> Mientras que las *economías nido* se refiere a economías emergentes de menor crecimiento, pero que en los próximos diez años podrán superar a la economía de menor desempeño dentro del grupo de países industrializados. Es decir, si bien no tienen el empuje de las águilas, si se espera que jueguen un papel importante dentro de la economía global.

Es importante analizar el patrón de crecimiento del PIB entre estos grupos de países, ya que, de acuerdo con el Programa de las Naciones Unidas para el desarrollo (2013), en el año 2020, la producción combinada de Brasil, China e India superará a la producción total de Canadá, Francia, Alemania, Reino Unido y EUA. Aunado a lo anterior, un estudio publicado por la firma consultora PWC (2013), señala que para el año 2017 China habrá superado a EUA como la mayor economía del mundo (en cuanto al tamaño de su PIB); además, se espera que en el 2050 India y Brasil se consoliden como la tercera y cuarta economía más grande del mundo, respectivamente. Además, dicho informe apunta que para mediados de este siglo, Rusia podría convertirse en la economía más grande de Europa, y México puede superar a las economías de Gran Bretaña y Francia.

Bajo esta tesis, el documento tiene por objetivo realizar un estudio comparativo entre países industrializados y economías emergentes, sobre el impacto del gasto público en inversión, el gasto en salud pública y la tasa de crecimiento poblacional sobre la tasa de crecimiento del PIB durante el periodo 1995-2010. Esto con la meta última de identificar los factores donde los Estados deberían enfocarse para potencializar el crecimiento y desarrollo económico. Se espera que el gasto en salud pública, la tasa de crecimiento poblacional y el gasto en inversión tengan un impacto positivo y significativo sobre la tasa de crecimiento del PIB, sobre todo en el caso de las economías emergentes economías, las cuales se prevé destaquen sobre los países industrializados.

## 2. METODOLOGÍA

El presente es un estudio con las siguientes características: observacional, longitudinal, y correlacional (Ojeda et al., 2011). Dentro de las unidades de estudio se consideraron 26 países para el periodo 1995-2010 (resultado en un total de 416 observaciones), los cuales se clasificaron en tres grupos. En primer lugar, se consideraron 5 economías emergentes<sup>24</sup>: Brasil, Rusia, India, China y Sudáfrica; el segundo grupo se conformó por 14 economías nido<sup>25</sup>: México, Turquía, Chile, Indonesia, Corea del Sur, Malasia, Tailandia, Colombia, Perú, Argentina, Egipto, Nigeria, Polonia y Bangladesh. Como parte del último grupo se consideraron a las economías industrializadas que integran al G-7. Para el desarrollo del análisis se utilizaron datos obtenidos a través del Fondo Monetario Internacional (World Economic

<sup>20</sup> Disponible en <http://www.bbvarsearch.com/KETD/ketd/esp/nav/eagles.jsp> [Revisado el 8 de mayo de 2013]

<sup>21</sup> El término es un acrónimo en inglés, que significa "economías emergentes y líderes del crecimiento".

<sup>22</sup> Incluye al grupo BRIC, además de Indonesia, México, Corea del Sur, Turquía y Taiwán.

<sup>23</sup> Cabe mencionar que la categoría EAGLE es dinámica, toda vez que el desempeño del grupo es revisado anualmente, y un país puede perder dicho status. Egipto, por ejemplo, perdió en 2013 su sitio.

<sup>24</sup> Se considera al grupo de los BRICS.

<sup>25</sup> Para efectos de este trabajo se ha decidido modificar los países que integran esta categoría, manteniendo por separado al grupo de los BRICS.

Outlook Databases<sup>26</sup>) y del Banco Mundial (Data Catalog<sup>27</sup>). De éstos, se obtuvo la siguiente información: tasa de crecimiento del PIB; gasto en salud pública (% del PIB); gasto en inversión (% del PIB); y, tasa de crecimiento poblacional. Además, se empleó como variable contextual la característica del país, con el fin de comparar su ritmo de crecimiento económico. La TABLA 1 describe cada una de dichas variables de estudio:

**TABLA 1:** Descripción de las variables de estudio

Variable	Tipo	Unidad de medida	Descripción
Tasa de crecimiento del PIB	Dependiente	Porcentaje	Tasa anual de crecimiento del producto interno bruto a precios de mercado.
Gasto en salud pública	Independiente	% del PIB	Total de recursos canalizados para la provisión de servicios de salud a cargo del Estado.
Gasto en inversión		% del PIB	Total de recursos empleados la construcción de infraestructura productiva.
Tasa de crecimiento poblacional		Porcentaje	Tasa de crecimiento exponencial de la poblacional, del año t-1 a t.
Tipo de país		Industrializado Emergente Nido	Clasificación de las economías con base a su patrón de crecimiento económico.

Fuente: elaboración propia con información del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

Se empleó la técnica de modelación multinivel (Goldstein, 2003). De acuerdo con Murillo (2008), los modelos multinivel pueden definirse como una extensión de los modelos de regresión simple, ya que implican varios modelos lineales en cada uno de los niveles. De esta manera, aquellos modelos ubicados en el primer nivel se vinculan con los de segundo nivel, en que los coeficientes de regresión del primer nivel se regresan a un segundo nivel de variables explicatorias, y así con cada uno de los diferentes niveles empleados. Así, se colocaron en el primer nivel los años y en el segundo nivel los países para estudiar la variabilidad en la tasa de crecimiento de cada país en el periodo de tiempo y también compararla con los otros países. La finalidad de emplear un modelo de este tipo consistió en poder identificar cuáles variables son o no estadísticamente significativas para explicar el patrón de crecimiento económico, así como también hacer un análisis de la varianza dentro de cada país y entre países, y así determinar qué tanto la clasificación de los países se ajusta a su patrón de crecimiento. Así, se plantea el siguiente modelo:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4j} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

<sup>26</sup> Disponible en <http://www.imf.org/external/ns/cs.aspx?id=28> [Revisado el 20 de marzo de 2013]

<sup>27</sup> Disponible en <http://datacatalog.worldbank.org/> [Revisado el 20 de marzo de 2013]

Donde:

$Y_{ij}$ = tasa de crecimiento del PIB para el año  $i$  en el país  $j$ .

$B_{0j}$ = intercepto.

$X_1$ = gasto en salud pública (% PIB).

$X_2$ = tasa de crecimiento poblacional.

$X_3$ = gasto en inversión (% del PIB).

$X_4$ = tipo de país (BRIC, nido o industrializado). Se crearon dos variables dummy para las categorías: economías nido y países industrializados.

$u_{0j}$ = varianza entre países.

$e_{ij}$ = residuos de primer nivel o varianza en los años.

### 3. RESULTADOS

En una primera fase exploratoria de los datos se observó que la media del gasto en salud pública como porcentaje del PIB fue de 6.68%. Sin embargo, el rango fue de 16.12% y la varianza de 8.164 (Véase TABLA 2). Esta diferencia puede ser explicada por el comportamiento de la variable en los casos de Francia, Alemania y EUA, destacando éste último, toda vez que dichos países destinaron al menos el 10% de su PIB en dicho rubro durante el periodo. Por otra parte, los niveles más bajos se encontraron en Indonesia durante los años 1995-1997. Como se puede observar en la GRÁFICA 1, EUA representa un caso atípico para el gasto en salud pública como porcentaje del PIB, toda vez que ningún otro país alcanzó cifras tan elevadas, incluso aquellos que forman parte del G-7.

**TABLA 2:** Estadísticas descriptivas. Países industrializados, economías emergentes y economías nido.

	Rango	Mínimo	Máximo	Media	Varianza
Salud pública (% PIB)	16.12	1.77	17.89	6.6816	8.144
Crecimiento poblacional	3.05	-0.54	2.52	1.0462	0.494
Gasto en inversión (% PIB)	37.46	10.78	48.24	22.9079	37.461
Tasa de crecimiento del PIB	29.63	-13.13	16.50	3.8932	15.009
N válido					

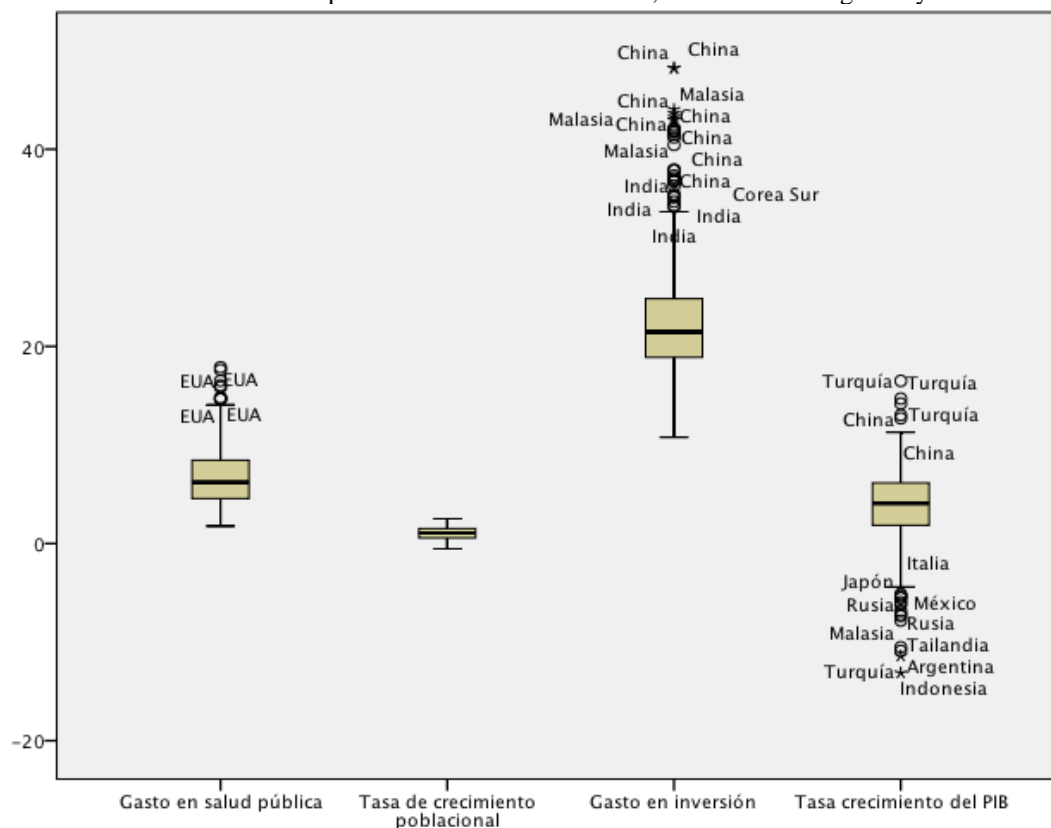
Fuente: elaboración propia con información del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

Por otra parte, se encuentra que la variable crecimiento poblacional es la que presenta una menor varianza, ya que ésta encuentra su nivel máximo en 2.52% para el caso de Nigeria durante el 2010, y su mínimo en -0.54% para el caso de Polonia durante los años 2000 y 2001. Esta variable presenta mínima dispersión entre los grupos de países, incluso tomando en cuenta las elevadas tasas registradas por Nigeria, Sudáfrica y Egipto durante gran parte del periodo, siendo éstos países africanos.

En lo que se refiere al gasto en inversión como porcentaje del PIB, se tiene que esta variable presenta la mayor variabilidad entre el grupo de países durante el periodo de estudio (véase TABLA 2). Destaca el hecho de que todos los casos sean de países asiáticos, destacando China y Malasia. En el caso de China en 2009 toda vez que dicha nación destinó casi el 50% de su PIB a inversión. (Véase el gráfico 1). Además, cabe mencionar que China durante los años 1995-1996 y 2003-2010 destinó por lo menos el 40% de su PIB a dicho rubro, contrastando con el caso de Argentina en 2002, donde el país Sudamericano destinó el 10.78%, siendo esta la cifra más baja dentro de este análisis. En relación a la tasa de crecimiento del PIB se puede señalar que dicha variable también presenta una varianza elevada,

toda vez que se ubica en un rango que va del -13.13% al 16.50%, siendo éstos los casos de Indonesia en 1998 y el de Turquía en 1999. Este último dato es relevante toda vez que superó a la máxima tasa de crecimiento reportada por China durante el periodo, la cual fue de 14.20% en 2007 (véase ver GRÁFICO 1). También se destacan los casos atípicos de China y Turquía, países que obtuvieron las tasas más elevadas en el periodo. Por otra parte, las tasas más bajas son representadas por los casos extremos de Argentina e Indonesia.

**GRÁFICA 1:** Estadísticas descriptivas. Países industrializados, economías emergentes y economías nido



Fuente: elaboración propia con información del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

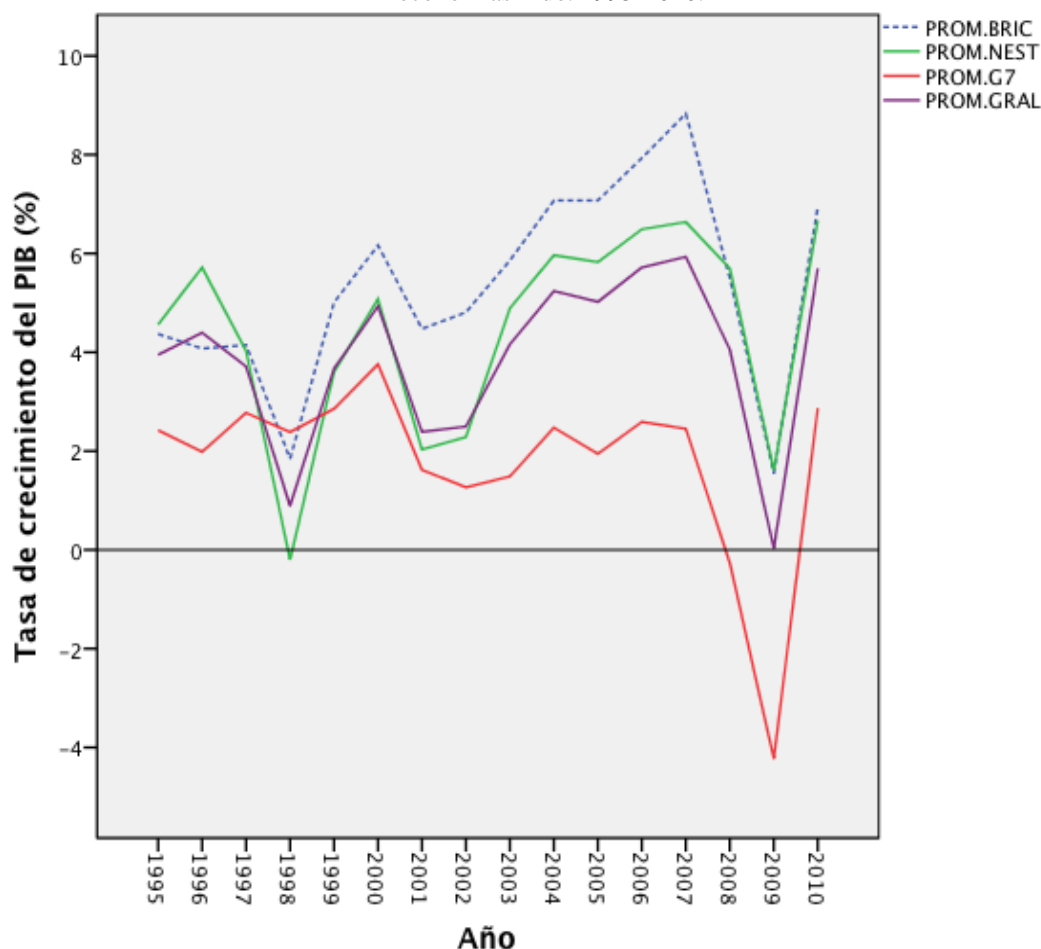
En la gráfica 2 se observa, que el grupo de países industrializados se ubica por debajo del promedio general durante la mayor parte del periodo, a diferencia de las economías emergentes, ya sean BRICS o nido. Además, se destacan las variaciones pronunciadas durante los años de 1998 y 2009, donde los tres grupos de países sufrieron caídas significativas. No obstante, en 2010 se observa una recuperación generalizada.

El análisis de correlación mostró una relación negativa entre la tasa de crecimiento del PIB y el gasto en salud pública. De manera contraria, la tasa de crecimiento poblacional y el gasto en inversión presentaron una relación positiva con la tasa de crecimiento del PIB, siendo esta última la más significativa de todas.

Para comprobar el modelo multinivel anteriormente descrito,<sup>28</sup> a continuación se presenta la TABLA 3 con el resultado de las estimaciones.

<sup>28</sup> Se empleó el software MLWin 2.27



**GRÁFICA 2:** tasa de crecimiento del PIB.Promedio de países industrializados, economías emergentes y economías nido. 1995-2010.

Fuente: elaboración propia con datos de del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

Como se puede observar en la TABLA 3, el promedio de la tasa de crecimiento del PIB (o el valor que toma la variable respuesta cuando las variables explicativas son cero) para el conjunto de países es de 3.893%. No obstante, al añadir al modelo la variable inversión esta cambia a -1.312%. Además, cuando se agrega al modelo la variable categórica y se selecciona como referencia al grupo de los BRICS, se observa que éstos tuvieron una tasa de crecimiento negativa durante el periodo, la cual fue de -1.237%. A pesar de esto, su ritmo de crecimiento estuvo por encima del resto de los países, toda vez que las economías nido tuvieron un crecimiento promedio de 0.480% menor que el de los BRICS y los países industrializados de 2.178% menos. Respecto a la variable salud pública, se tiene que ésta es significativa y presenta una relación negativa con la tasa de crecimiento del PIB, es decir, que al aumentar el gasto en salud, la tasa de crecimiento del PIB disminuye (véase la TABLA 3). Sin embargo, desde el momento en que se introduce al modelo la variable inversión se observa que su error estándar aumenta y al compararlo en tablas con una distribución normal estándar, resulta no ser significativa.

TABLA 3: Estimaciones del modelo multinivel

Elementos fijos						
Coefficientes	$\beta_{0j}$	$\beta_1$ Salud	$\beta_2$ Población	$\beta_3$ Inversión	$\beta_4$ Nido	$\beta_5$ Industrializado
Estimaciones	3.893 (0.381)					
	6.818 (0.728)	-0.438 (0.099)= -4.42				
	7.035 (0.981)	-0.449 (0.105)= -4.27	-0.137 (0.412)= <b>-0.33*</b>			
	-1.312 (1.403)	-0.176 (0.098)= <b>-1.79*</b>	0.146 (0.362)= <b>0.40*</b>	0.272 (0.037)= 7.35		
	-1.237 (1.452)	-0.009 (0.114)= <b>-0.078*</b>	0.056 (0.349)= <b>0.16*</b>	0.261 (0.035)= 7.45	-0.480 (0.628)	-2.178 (0.813)
Elementos aleatorios						
$S_{u0}^2$	3.030 (1.049)	1.634 (0.660)	1.712 (0.680)	1.110 (0.493)	0.668 (0.372)	
$S_e^2$	11.943 (0.855)	11.817 (0.846)	11.790 (0.844)	10.597 (0.759)	10.625 (0.761)	

Fuente: elaboración propia con datos de del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

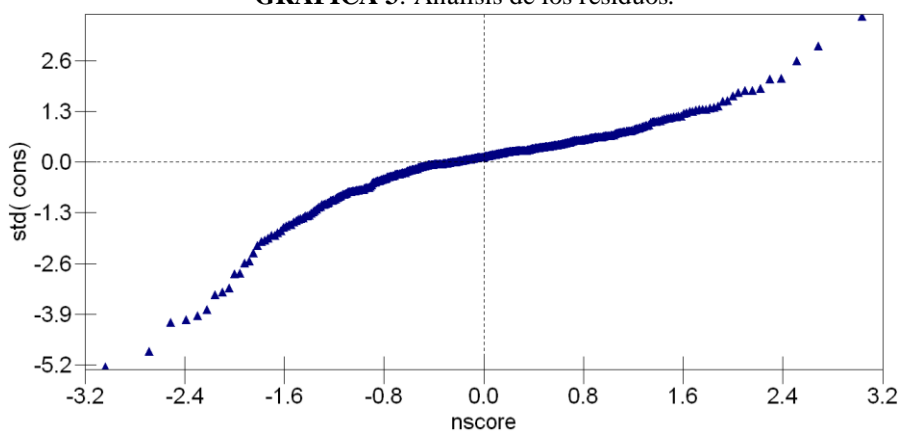
\*Variable no significativa.

En cuanto a la tasa de crecimiento poblacional, se tiene que ésta no es significativa para explicar la tasa de crecimiento del PIB en los casos seleccionados durante el periodo. Toda vez que su primer valor es de -0.33, el cual se reduce hasta 0.16 en el modelo final. A partir de este resultado, se puede decir que el ritmo de crecimiento poblacional no impacta el patrón de crecimiento económico. Por el contrario, el gasto en inversión es significativo y mantiene una relación positiva con la tasa de crecimiento del PIB. En primera instancia, un incremento de 1% en el gasto en inversión provoca un incremento de 0.272% en el PIB; posteriormente, dicho el impacto se reduce a 0.262% aunque sigue siendo significativo. Esto reduce la varianza entre los años y entre los países.

En lo que se refiere a la parte aleatoria del modelo, se tiene que la varianza entre países se reduce de manera significativa, ya que esta pasa de 3.03 a 0.668. Por otra parte, la varianza entre años, si bien se reduce de 11.94 a 10.62, es más elevada que la observada entre países. Esto puede ser producto de la gran variabilidad en algunos países en sus tasas de crecimiento, tales como Argentina, Turquía y México. En cuanto al porcentaje de variación del modelo por países, se tiene que éste explica un 6% su variabilidad. De esta manera, tanto las tasas de crecimiento de un país en el periodo, se parecen a las tasas de crecimiento de otro país en el mismo periodo.

Finalmente, el análisis de los residuos comprueba el supuesto de normalidad, con lo cual puede decirse que las estimaciones obtenidas son eficientes (Goldstein, op. cit., 2003). Destacan los casos atípicos de Turquía y México en el extremo inferior, y, en el extremo superior, el grupo de países representado por Malasia, China, India y Turquía, siendo este último el más volátil dentro de las unidades de estudio.

GRÁFICA 3: Análisis de los residuos.



Fuente: elaboración propia con datos de del Fondo Monetario Internacional (World Economic Outlook Databases) y del Banco Mundial (Data Catalog).

#### 4. DISCUSIÓN

Con base en estas estimaciones, se observa que, a diferencia de lo señalado por Boyacioglu (op. cit., 2012), Picazzo et al. (op. cit., 2011), Lang (op. cit., 2012), Bakare (op. cit., 2011) y Wang (op. cit., 2011), el gasto en salud pública no representa una variable significativa para explicar el patrón de crecimiento económico de un país. En este sentido, los resultados de este trabajo concuerdan con los presentados por Elmi y Sadeghi (op. cit., 2012), en cuanto a la ausencia de relación entre las variables en el largo plazo, además de los obtenidos por Acemoglu y Johnson (op. cit., 2007), quienes señalan la existencia de una relación negativa.

Sin embargo, no debe dejarse de lado el hecho de que el gasto en salud pública es un mecanismo importante para reducir las desigualdades sociales y combatir la injusticia social, por lo cual la Comisión Europea (2011) y la Comisión para Reducir las Desigualdades en Salud en España (2011) coinciden en la necesidad de seguir impulsando el gasto en este rubro. Más allá de los resultados del modelo es preciso recordar el papel de la salud como un derecho fundamental de la población, así como los efectos que se derivan de contar con una población saludable, capaz de poder desempeñar sus tareas y así contribuir con las actividades que impactan en el crecimiento y desarrollo económico, tal como lo mencionan Rivera y Currais (op. cit., 2005).

Por otra parte, el hecho de que no se haya presentado una relación significativa entre la tasa de crecimiento poblacional y la tasa de crecimiento del PIB, contrasta con los hallazgos presentados por Quang (op. cit., 2012), quien menciona que dichas variables se relacionan de manera directa, e incluso, afirma la existencia de una tasa óptima de crecimiento poblacional para impactar sobre el crecimiento económico. Con los datos disponibles, no tenemos evidencia suficiente para resaltar la importancia del bono demográfico, tal como lo señalan Bloom et al. (op.cit., 2007).

En cuanto a la relevancia del gasto en inversión sobre la tasa de crecimiento del PIB, el modelo indica que sólo esta variable es significativa para explicar el ritmo de crecimiento de una nación. Esto concuerda con los hallazgos presentados por Canning y Fay (op. cit., 1993), Easterly y Revelo (op. cit., 1993) y Torrissi (op. cit., 2010) en cuanto a la fuerte relación entre dichas variables. Esto puede explicar las altas tasas de crecimiento de las economías emergentes asiáticas, destacando los casos de China, India y Corea del Sur.

Finalmente, se observó que la clasificación de países entre grupos sí refleja las características de las economías y el impacto de éstas en sus tasas de crecimiento. En este sentido, se puede señalar que los países industrializados destinan una mayor cantidad de recursos al gasto en salud pública, y presentan las menores tasas de crecimiento poblacional y del PIB; el grupo de los BRICS registran las tasas más altas de gasto en inversión y de crecimiento del PIB. En el caso de las economías nido, éstas superaron el promedio de las tasas de crecimiento del PIB de los países industrializados, así como también del gasto en inversión, lo cual, con base en la relación observada entre las variables, respalda su alto potencial de crecimiento de cara al futuro.

## 5. CONCLUSIÓN

A lo largo de este trabajo se indagó acerca de la relación existente entre el gasto en salud pública, el gasto en inversión y la tasa de crecimiento poblacional sobre la tasa de crecimiento del PIB de un conjunto de economías emergentes, economías nido y países industrializados durante el periodo 1995-2010. Dicho análisis permitió identificar que sólo el gasto en inversión es significativo para explicar el patrón de crecimiento de la tasa de crecimiento del conjunto de países.

De esta manera, se hace evidente la necesidad de impulsar el gasto en inversión como mecanismo para potencializar la tasa de crecimiento del PIB, tal como lo mencionan Canning y Fay (op. cit., 1993), Easterly y Revelo (op. cit., 1993) y Torrissi (op. cit., 2010); hecho que se refleja en el desempeño de gran parte de las economías emergentes y economías nido mencionadas en este análisis. De acuerdo a las estimaciones del documento, el gasto en inversión podría sentar las condiciones para alcanzar mejores escenarios de desarrollo económico y social.

Finalmente, se confirmó la importancia de clasificar a los países en categorías, toda vez que éstos presentaron patrones de comportamiento muy similares entre grupos. Esto, corrobora la pertinencia de catalogar a las economías con base en su patrón de crecimiento económico, tal como en su momento lo hizo O'Neill (op. cit., 2001). No obstante, el constante cambio del escenario económico global ha permitido el desarrollo de nuevas clasificaciones, como la presentada por García-Herrero (op. cit., 2012). De esta manera, se pueden identificar a las economías emergentes por su ritmo de crecimiento económico sostenido y altas tasas de crecimiento poblacional; a las economías nido como aquellas que poseen un mayor potencial de crecimiento; y, a los países industrializados como aquellos con las menores tasas de crecimiento económico, pero con mayor nivel de gasto en salud pública. Sin embargo, de continuarse con esta tendencia a la baja en el ritmo de crecimiento de las economías industrializadas, esta clasificación podría considerarse obsoleta.

## REFERENCIAS

1. Acemoglu, D. y Johnson, S., (2007), "Disease and development: the effect of life expectancy on economic growth", *Journal of Political Economy*, Univeristy of Chicago Press. Vol. 115. December.
2. Ashraf, Q., Lester, A. y Weil, D., (2009), "When does improving health raise GDP?", NBER Macroeconomics Annual.
3. Bakare, A., (2011). "Health care expenditure and economic growth in Nigeria: an empirical study. *Journal of Emerging Trends in Economics and Management Sciences*", Vol. 2. Scholarlink Research Institute Journals.
4. Banco Mundial, (2002). *Más allá del crecimiento económico*, Grupo Banco Mundial, EUA.
5. Barro, R., (1991). "Economic growth in a cross-section of countries", *Quarterly Journal of Economics*, No. 106, 407-443.
6. Bloom, D. et. al. (2007), "Realizing the demographic dividend: is Africa any different?", Program on the global demography of aging, Harvard University.

7. Boyacioglu, E., (2012), "The importance of health expenditures on sustainable development", *International Journal of Social Sciences and Humanity Studies*, Vol. 4, No. 2.
8. Bukhari, S., Ali, L. y Saddaqt, M., (2007), "Public investment and economic growth in the Three Little Dragons: evidence from heterogeneous dynamic panel data", *International Journal of Business and Information*, Vol. 2, No.1, June, 57-79.
9. Canning, D. y Fay, M., (1993), "The effects of transportation on economic growth. Department of Economics", Columbia University, EUA.
10. Comisión Europea, (2011), "Reducir las desigualdades en materia de salud en la Unión Europea". Dirección General de Empleo, Asuntos Sociales e Inclusión, Luxemburgo.
11. Comisión para Reducir las Desigualdades Sociales en Salud en España, (2011), "Propuesta de políticas e intervenciones para reducir las desigualdades sociales en salud en España", Gac. Sanit.
12. Devajaran. S., Swaroop, V. y Zou, H., (1996), "The composition of public expenditure and economic growth", *Journal of Monetary Economics*, No. 37, 313-344.
13. Draibe, S. y Riesco, M., (2006), "Estado de bienestar, desarrollo económico y ciudadanía: algunas lecciones de la literatura contemporánea", Sede subregional de la CEPAL en México, Serie estudios y perspectivas, No. 55. Unidad de Desarrollo Social, México.
14. Easterly, W. y Rebelo, S., (1993), "Fiscal policy and economic growth", *Journal of Monetary Economics*, No. 32, 417-458.
15. Elmi, Z. y Sadeghi, S., (2012), "Health care expenditures and economic growth in developing countries: panel co-integration and causality", *Midle East Journal of Scientific Research*, 12 (1): 88-91.
16. García-Herrero, A., (2011), "Economías emergentes", Servicio de Estudios Económicos BBVA, México.
17. García-Herrero, A., -, (2012), "Economías emergentes líder: BBVA EAGLES", BBVA Research, México.
18. Goldstein, H., (2003), "Multilevel Statistical Models", Arnold, New York.
19. Güendell, L. y Barahona, M., (2005), "Programación social con enfoque de derechos: hitos y derroteros en la configuración de un nuevo paradigma", Costa Rica.
20. Keynes, J., (1936), *General theroy of employment, interest and money*. MacMillan, Cambridge University Press, Royal Economic Society.
21. Lang, J., (2012), "The most influential factors in determining the happiness of nations. Major Themes in Economics", Spring.
22. Murillo, F., (2008), "Los modelos multinivel como herramienta para la investigación educative", *Magis. Revista Internacional de Investigación en Educación*, No. 1, 45-62.
23. Ojeda, M. et al., (2011), *Metodología Estadística Aplicada a las Finanzas Públicas*, México.
24. Organización para la Cooperación y el Desarrollo Económicos, OCDE, (2012), "Looking to 2060: long-term global growth prospects. A going for growth report", OECD Economic Policy Papers.
25. - Organización para la Cooperación y el Desarrollo Económicos, OCDE (2012), "Perspectivas OCDE: México, reformas para el cambio", México.
26. Organización de las Naciones Unidas, ONU, (1990), *Primer Informe sobre Desarrollo Humano*, EUA.
27. Picazzo, E. et al., (2011), "La teoría del desarrollo humano y sustentable: hacia el reforzamiento de la salud como un derecho y libertad universal", *Estudios Sociales*, Vol. 19. Núm. 37, enero-junio.
28. Programa de las Naciones Unidas para el Desarrollo, PNUD, (2002) "Importancia del desarrollo humano: una base conceptual", Oficina del Programa de las Naciones Unidas para el Desarrollo, Panamá.

29. Programa de las Naciones Unidas para el Desarrollo, PNUD, (2013), “Informe sobre desarrollo humano 2013. Programa de las Naciones Unidas para el Desarrollo. El ascenso del sur: progreso humano en un mundo diverso”, EUA.
30. PWC, (2013), “World in 2050: The BRICS and beyond: prospects, challenges and opportunities”, EUA.
31. Quang, M., (2012), “Population and economic growth in developing countries”, *International Journal of Academic Research in Business and Social Sciences*, Vol. 2, Núm. 1, January.
32. Rivera, B. y Currais, L., (2005), “La inversión en salud como gasto público productivo: un análisis de su contribución al crecimiento económico”, *Presupuesto y gasto público*, Instituto de Estudios Fiscales, No. 39, España.
33. Safdari, M., Mehrizi, M. y Elahi, M., (2011), “The effect of Health Expenditure on Economic Growth in Iran”, *American Journal of Scientific Research*, EuroJournals Publishing Inc., Issue 37.
34. Santos, G., (2005), “Grupo BRIC: Brasil, Rusia, India y China”, LXI Legislatura. Cámara de Diputados, Centro de investigación y análisis, Subdirección de Política Exterior, México.
35. Sepúlveda, J., et al, (2006), “Macroeconomía y Salud. Invertir en salud para el desarrollo económico”, *Aspectos Macroeconómicos, Regionales e Individuales de la Inversión en Salud en relación con el crecimiento económico*, Secretaría de Salud, México.
36. Sen, A., (2000), *Desarrollo y libertad*, Editorial Planeta, Barcelona.
37. Torrisi, G., (2010) “Infrastructures and Economic Performance: A critical Comparison Across Four Approaches” *Theoretical and Practical Research in Economics Fields. Association for Sustainable Education, Research and Science*, Vol. 1, June, 86-96.
38. Wang, K., (2011), “Health care expenditure and economic growth: Quantile panel-type analysis”, *Economic Modelling*, No. 28.
39. Weil, D., (2007), “Accounting for the effect of health on economic growth”, *The Quarterly Journal of Economics*, Vol. 122.

