



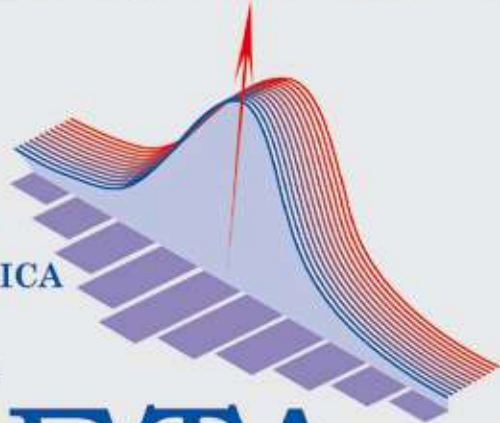
Universidad Mayor
de San Andrés

Varianza

Revista del Instituto de Estadística Teórica y Aplicada



UMSA
FCPN
CARRERA
ESTADÍSTICA



IETA
Instituto de Estadística
Teórica y Aplicada



Varianza

Revista de la Carrera de Estadística

Publicación del Instituto de Estadística Teórica y Aplicada

UMSA
FCPN
CARRERA
ESTADÍSTICA



IETA

Instituto de Estadística
Teórica y Aplicada

Número 15

Noviembre, 2018

ISSN 9876-6789

DEP. LEGAL: 4-3-156-18 P.O.

REVISTA VARIANZA

Nº 15 – Noviembre, 2018

**DIRECTOR CARRERA DE
ESTADÍSTICA**

Rivero Suguiura, Fernando Oday

DIRECTOR a.i. I.E.T.A.

Valdez Blanco, Dindo

AUTORES DE ARTÍCULOS

Chirino Gutierrez, Alvaro

Almeida, María Eugenia

Mancilla Flores, Emma Martha

Miranda Gonzales, Yuri

Paredes Alarcón, Marisol

Rivero Suguiura, Fernando Oday

Ruiz Aranibar, Gustavo

Valdez Blanco, Dindo

Vargas Salazar, Fernando Rodrigo

REVISIÓN DE TEXTO

Pinto Ajhuacho, Jaime Tito

DISEÑO Y DIAGRAMACIÓN

Encinas Choque, Luis Carlos

*Los artículos escritos son entera
responsabilidad de los autores.*

PRESENTACIÓN

La ciencia de la estadística es un componente indispensable en investigaciones que utiliza la recopilación de información, tratamiento de datos, diseño de metodologías, predicción y análisis entre otras; ya sea en proyectos académicos, empresariales o gubernamentales. Conforme el incremento de la información estadística, se ha vuelto mucho más importante la necesidad de un creciente número de empresas e instituciones que requieren profesionales idóneos para comprender el lenguaje estadístico y sus aplicaciones.

La estadística tiene amplia participación en la investigación que se realiza en ámbitos tan diversos como: la industria, educación, medicina, salud, psicología, ingeniería, historia, política, demografía, economía, mercadotecnia, finanzas, entre muchas otras. Es así que la revista Varianza de esta gestión, presenta y difunde un conjunto de artículos científicos que emanan del trabajo investigativo de profesionales vinculados con la estadística, como también del proceso de enseñanza – aprendizaje y labor aplicativa que despliegan docentes y estudiantes de nuestra carrera.

Finalmente, agradezco al director del Instituto de Estadística Teórica y Aplicada, Lic. Dindo Valdez Blanco y a todo el equipo por esta nueva edición de la revista Varianza, así como también a los profesionales, docentes y estudiantes que aportan con sus artículos.

M.Sc. Fernando O. Rivero Suguiura

DIRECTOR CARRERA DE ESTADÍSTICA

Carrera de Estadística
Instituto de Estadística Teórica y Aplicada (I.E.T.A.)
Facultad de Ciencias Puras y Naturales
Universidad Mayor de San Andrés

La Paz - Bolivia
Edificio Antiguo - Planta Baja
Telefax: 2442100 -2612844
Correos:estadistica@umsa.bo - ieta@umsa.bo

*Dedicado a los estudiantes de la
Carrera de Estadística, futuros
investigadores estadísticos.*

Contenido

El efecto del racionamiento de agua en el municipio de La Paz sobre los casos de diarrea en niños menores a 5 años.....	1
<i>Lic. Chirino Gutierrez, Alvaro & Almeida, María Eugenia</i>	
Aplicación del REDATAM.....	8
<i>Lic. Mancilla Flores, Emma Martha</i>	
Estimación de la economía informal en Bolivia.....	15
<i>Lic. Miranda Gonzales, Yuri</i>	
Una mirada a la educación en Bolivia desde las proyecciones de población e indicadores educativos censales.....	30
<i>Lic. Paredes Alarcón, Marisol</i>	
Regresión lineal simple con términos de perturbación no normales	40
<i>M.Sc. Rivero Suguiura, Fernando Oday</i>	
Estadística multivariable análisis de componentes principales	44
<i>Dr. Cs. Ruiz Aranibar, Gustavo</i>	
Descomposición del test de Friedman aplicado a pruebas de degustación.....	62
<i>Lic. Valdez Blanco, Dindo</i>	
El soporte de la economía del conocimiento: “Investigación + Desarrollo + Innovación (I+D+I)”	67
<i>PhD. Vargas Salazar, Fernando Rodrigo</i>	

EL EFECTO DEL RACIONAMIENTO DE AGUA EN EL MUNICIPIO DE LA PAZ SOBRE LOS CASOS DE DIARREA EN NIÑOS MENORES A 5 AÑOS

Una aplicación del Método de Diferencia en Diferencia

Lic. Chirino Gutierrez, Alvaro & Almeida, María Eugenia

✉ alvaro_rqsbo@hotmail.com

RESUMEN

En noviembre de 2016 se declaraba emergencia nacional por la presencia de sequía y falta de recursos hídrico en diferentes regiones del país, específicamente la ciudad de La Paz atravesaba un régimen de racionamiento de agua en distintos puntos de la zona Sur y la ladera Este de la ciudad, dejando así a varios hogares sin acceso a agua salubre y a los niños vulnerables a infecciones o enfermedades diarreicas potencialmente mortales.

Para determinar el impacto del racionamiento de agua en La Paz sobre el promedio de niños que asistieron a los establecimientos de salud por enfermedad diarreica aguda se utilizó el método Diferencia en diferencias. El método consiste en aplicar una doble diferencia para comparar los cambios en el tiempo en la variable de interés entre el grupo de tratamiento y el grupo control, para el caso de la ciudad de La Paz se compararon los establecimientos de salud cercanos a una zona con racionamiento de agua y como controles a los establecimientos fuera de las zonas de racionamiento.

Los resultados demostraron que el racionamiento de agua que sufrió la ciudad de La Paz en el mes de noviembre de 2016 tuvo un impacto negativo sobre la salud de los niños. El promedio de niños con enfermedad diarreica aguda en el mes noviembre de 2016 aumento significativamente respecto al año 2015, en los centros de salud cercanos a las zonas de racionamiento, teniendo en cuenta que la EDA representa la segunda enfermedad más mortal en la infancia.

PALABRAS CLAVE

Diferencia en diferencias, crisis del agua, racionamiento, enfermedades diarreicas.

1. ANTECEDENTES

El 8 de noviembre de 2016 la Empresa Pública Social de Agua y Saneamiento (EPSAS), comunicó a la población de La Paz el **plan de racionamiento del suministro de agua en 94 barrios de la Zona Sur y la ladera Este de la ciudad de La Paz**, el motivo fue el descenso de los niveles de agua en la represa de Hampaturi, debido a la escasez de lluvia en los meses previos.

La primera etapa de la restricción estableció el racionamiento día por medio en horas

de la noche, entre las 20:00 y las 08:00; posteriormente los barrios afectados se dividieron en tres zonas que recibirían el suministro de agua en diferentes horarios por periodos de 12 horas; las restricciones se ampliaron hasta llegar a las 72 horas de corte y solo tres de abastecimiento, además se incorporaban otros barrios de La Paz y de la ciudad de El Alto.

Finalmente la mañana del 21 de noviembre de 2016 el Jefe de Estado en una conferencia de prensa declaró emergencia nacional, contenida en el Decreto Supremo N° 2987

que establece en su primer artículo: “El presente Decreto Supremo tiene por objeto declarar situación de emergencia nacional, debido a la presencia de sequía y déficit hídrico en diferentes regiones del territorio nacional, provocadas por fenómenos climáticos adversos”.

2. MOTIVACIÓN

Según datos de la Organización Mundial de la Salud (OMS) las enfermedades diarreicas representan la **segunda mayor causa de muerte de niños menores de cinco años**, ocasionando la muerte de 525.000 niños cada año.

La prevención de las enfermedades diarreicas en particular se concentra en el **acceso al agua potable**, el acceso a buenos sistemas de saneamiento y el lavado de las manos con jabón que permiten reducir el riesgo.

En base a la relación entre las enfermedades diarreicas y el acceso al agua potable, es razonable querer conocer el efecto que existió del racionamiento de agua sobre la salud de una población vulnerable como son los niños menores a cinco años.

3. OBJETIVO

El objetivo del estudio es determinar el impacto de la “crisis del agua” que vivió la ciudad de La Paz en la salud de los niños menores de cinco años.

La investigación se centra en la variable de enfermedad diarreica aguda del “Formulario de notificación para la vigilancia epidemiológica”, del Sistema Nacional de Información de Salud (SNIS).

La pregunta de investigación del estudio es:

“¿Cuál fue el efecto del racionamiento de agua que existió en la ciudad de La Paz el último trimestre de 2016 sobre la salud de los niños menores a cinco años?”.

4. METODOLOGÍA

Dado que se desea conocer el efecto del racionamiento del agua en la salud de los niños menores a 5 años, se plantea un problema de causalidad entre el racionamiento de agua y la salud de los niños. Es inevitable preguntarse en este tipo de problemas qué hubiera sucedido con la salud de los niños (resultado) si no hubiera existido el racionamiento del agua (tratamiento), es decir, se necesita construir un contra factual para responder a los objetivos de la investigación. Debido a que el fenómeno ya ha ocurrido y que además éste es basado en un experimento natural, las alternativas para responder la pregunta de investigación planteada caen dentro de los llamados diseños cuasi-experimentales, estos diseños normalmente tratan de recrear un contexto experimental, con la idea central de estimar el efecto del tratamiento de la mejor forma posible.

La unidad de análisis son los niños y niñas menores de cinco años de la ciudad de La Paz, debido a que no se cuenta con un registro se define como una unidad agregada que alberga a las unidades de interés a los centros de salud. Estas unidades son útiles debido a que existen registros de los reportes de casos de diarrea por centro de salud de gestiones pasadas y a niveles mensuales.

Se elige el método de Diferencia en Diferencia como modelo que permitirá aproximar el impacto debido a que se puede hacer seguimiento en el tiempo a los centros de salud y en base a la cercanía del centro de salud con las zonas afectadas es posible

identificar centros de salud en zonas de tratamiento (afectadas por el racionamiento).

4.1. MÉTODO DE DIFERENCIA EN DIFERENCIAS

El método de diferencias en diferencias se concentra en controlar dos tipos de diferencia: la diferencia en el tiempo y la diferencia entre los grupos tratados y de control.

Para calcular el estimador de diferencia en diferencias, primero se calculan los cambios en el tiempo del grupo tratado y los cambios del grupo control (no tratados) para luego restar estos dos resultados como la diferencia

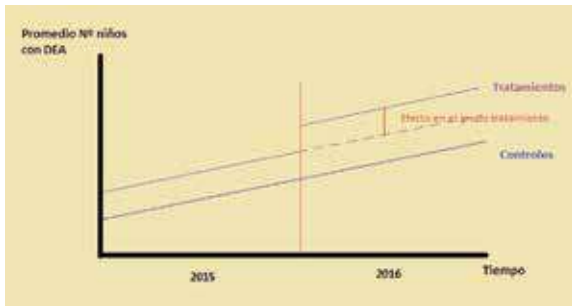
entre ambos grupos.

- Población: Centros de salud de la ciudad de La Paz.
- Tratamiento: Centros de salud cercanos a una zona afectada por la crisis del agua en La Paz (zona de racionamiento).
- Controles: centros de salud fuera de las zonas de racionamiento de agua.
- Variable: (Y_{it}) – Número de niños menores de cinco años que reportaron enfermedad diarreica en el centro de salud i en el año t .

E.D.A	Tratamientos	Controles
2016	$Y_1/T = 1$	$Y_1/T = 0$
2015	$Y_0/T = 1$	$Y_0/T = 0$
DIF.	$(Y_1/T = 1) - (Y_0/T = 1)$	$(Y_1/T = 0) - (Y_0/T = 0)$

$$DD = \left[(Y_1/T = 1) - (Y_0/T = 1) \right] - \left[(Y_1/T = 0) - (Y_0/T = 0) \right]$$

Gráficamente:



Puede representarse el método como una regresión múltiple, el coeficiente α de la interacción entre el grupo tratado y el periodo post tratamiento representa la estimación de diferencia en diferencias.

$$Y_{it} = \beta_0 + \beta_1 * T_i + \beta_2 * P_t + \alpha * T_i * P_t + \epsilon_{it}$$

Donde:

- Y_{it} = Número de niños menores de cinco años que reportaron enfermedad diarreica aguda en el centro de salud i en el año t .
- T_i = Es una variable dicotómica que toma valor 1 cuando el centro de salud i está cercano a una zona de racionamiento.
- P_t = Es una variable dicotómica que toma el valor 1 cuando la información pertenece al año 2016 (año de la crisis del agua).
- α = Representa el estimador de diferencia en diferencias.

$$\text{cov}(\epsilon_{it}, T_i) = 0$$

$$\text{cov}(\epsilon_{it}, t) = 0$$

Demostración:

$$E(Y_{i1}/T_i = 1) = \beta_0 + \beta_1 * 1 + \beta_2 * 1 + \alpha * 1 * 1 + E(\epsilon_{i1}/T_i = 1) = \beta_0 + \beta_1 + \beta_2 + \alpha$$

$$E(Y_{i0}/T_i = 1) = \beta_0 + \beta_1 * 0 + \beta_2 * 1 + \alpha * 0 * 1 + E(\epsilon_{i0}/T_i = 1) = \beta_0 + \beta_2$$

$$E(Y_{i1}/T_i = 0) = \beta_0 + \beta_1 * 1 + \beta_2 * 0 + \alpha * 1 * 0 + E(\epsilon_{i1}/T_i = 0) = \beta_0 + \beta_1$$

$$E(Y_{i0}/T_i = 0) = \beta_0 + \beta_1 * 0 + \beta_2 * 0 + \alpha * 0 * 0 + E(\epsilon_{i0}/T_i = 0) = \beta_0$$

$$DD = [E(Y_{i1}/T_i = 1) - E(Y_{i0}/T_i = 1)] - [E(Y_{i1}/T_i = 0) - E(Y_{i0}/T_i = 0)]$$

$$DD = [(\beta_0 + \beta_1 + \beta_2 + \alpha) - (\beta_0 + \beta_2)] - [(\beta_0 + \beta_1) - \beta_0]$$

$$DD = [(\beta_1 + \alpha) - \beta_1] = \alpha$$

4.2. SUPUESTOS DEL MODELO

un grupo muy afectado respecto al otro.

- El sesgo por selección es separable y no varía en el tiempo, por esta razón es controlado al aplicar la diferencia a través del tiempo.
- Las observaciones dentro de la unidad tiempo/grupo están correlacionadas.
- La expectativa que genera la futura implementación del tratamiento no debe afectar los resultados del periodo cero.

5. DATOS

La información para el estudio proviene de dos fuentes de información:

- Los errores no deben estar correlacionados.

- La base de datos del “Formulario de notificación para la vigilancia epidemiológica” que todos los centros de salud reportan mensualmente de los años 2015 y 2016, contiene la variable de estudio (Y_{it}) – Número de niños menores de cinco años que reportaron enfermedad diarreica en el centro de salud i en el año t . Esta base de datos también contiene las características y ubicación (latitud y longitud) de 101 centros de salud en la ciudad de La Paz. Fuente: Sistema Nacional de Información de Salud (SNIS).

4.3. VENTAJAS Y DESVENTAJAS DEL MÉTODO

- El método de diferencia en diferencias es útil si el tratamiento ya fue implementado en las observaciones.

El método diferencia en diferencias presenta problemas si:

- El tamaño de la respuesta depende del tamaño de la intervención y comparamos

El Efecto del Racionamiento de Agua en el Municipio de La Paz sobre los casos de diarrea en niños menores a 5 años

- Base de datos de zonas afectadas por el racionamiento de agua en la ciudad de La Paz (con latitud y longitud). Fuente: EPSAS.

Para la construcción de la variable tratamiento T_i se empleó la fórmula Haversine para calcular la distancia entre los centros de salud y las zonas afectadas y se definieron dos niveles del tratamiento; el primero, centros de salud hasta una distancia lineal de 0.5 kilómetros a zonas de racionamiento, y la segunda hasta 1

6. RESULTADOS

6.1. Considerando a los centros de salud cercanos a las zonas afectadas con una distancia menor a 0.5 km.

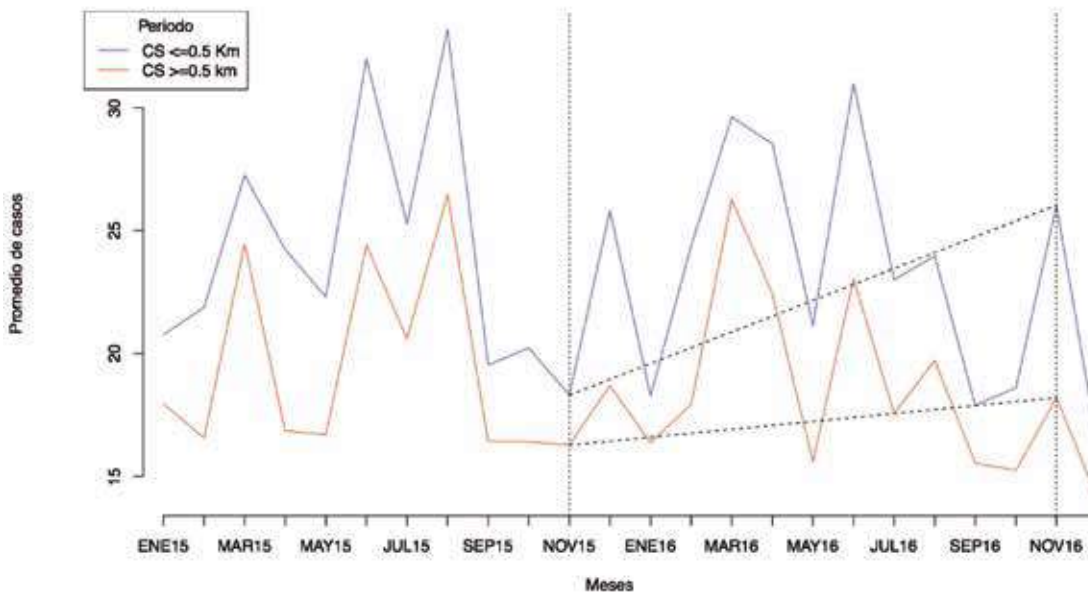
- Tratamientos= 29
- Controles= 72

Regresión:

Tabla N° 1
Modelo de diferencia en diferencia para distancia de hasta 0.5 km.

Dif	Coefficiente	Std. Err	t	p
T	6.2073	2.7502	2.26	0.026**
const	2.2727	1.4415	1.58	0.118
** p<0.05				

Figura N° 1
Casos reportados de diarrea en los centros de salud por grupo de tratamiento y control (0.5 km)



6.2. Considerando a los centros de salud cercanos a las zonas afectadas con una distancia menor a 1 Km.

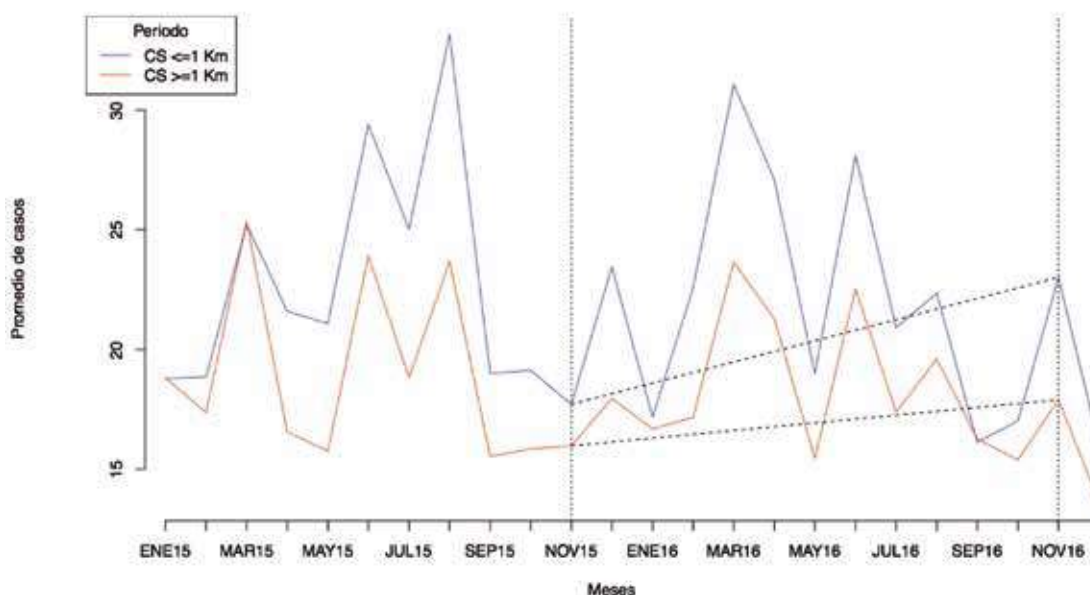
- Tratamientos= 51
- Controles= 50

Regresión:

Tabla N° 2: Modelo de diferencia en diferencia para distancia de hasta 1 km.

Dif	Coefficiente	Std. Err	t	p
T _i	3.3868	2.5	1.35	0.179
const	2.3404	1.74	1.35	0.182

Figura N° 2
Casos reportados de diarrea en los centros de salud por grupo de tratamiento y control (1 km)



Las tablas 1 y 2 presentan los resultados del modelo de diferencia en diferencia considerando los niveles de tratamiento de 0.5 km. y 1 km. respectivamente.

En el primer caso se puede afirmar estadísticamente que existió un efecto del racionamiento a un nivel de significancia del 5%, mientras que en el segundo caso no existe evidencia estadística que permita verificar el efecto del racionamiento en centros de salud a una distancia de hasta 1 km de distancia a las zonas de racionamiento. Este efecto en el caso significativo es positivo, lo que implica que el racionamiento tuvo un efecto viéndose incrementado los casos de diarrea

en las zonas cercanas (hasta 0.5 km) al área de racionamiento, este efecto se lo puede apreciar de forma visual en la figura 1, donde se percibe la pendiente existente respecto los datos reportados en 2016, el valor del efecto es de 6.2, que se traduce en que se incrementaron 6 casos aproximadamente dentro los centros cercanos hasta 0.5 km del área de racionamiento de agua.

7. CONCLUSIONES Y RECOMENDACIONES

- Los centros de salud que se encontraban a 0.5 km o menos de las zonas afectadas por el racionamiento de agua tuvieron

*El Efecto del Racionamiento de Agua en el Municipio de La Paz
sobre los casos de diarrea en niños menores a 5 años*

- un incremento significativo en el reporte de los niños menores de cinco años con enfermedad diarreica aguda (EDA).
- Se puede llegar a la conclusión de que la “crisis del agua” que sufrió la ciudad de La Paz en el mes de noviembre de 2016 tuvo un impacto negativo sobre la salud de los niños que se encontraban cerca de las zonas de racionamiento, teniendo en cuenta que la EDA representa la segunda mayor causa de muerte de niños menores de cinco años es importante tomar medidas de prevención.
 - En cuanto al modelo propuesto, se recomienda explorar otras alternativas que permitan validar o refutar los resultados de este documento y de esta forma se pueda contribuir a una mejora en la evidencia existente de fenómenos similares al estudiado. A partir de los hallazgos de este documento los autores profundizarán un estudio que permita añadir otras variables que pudieron ser afectadas y buscarán plantear otros modelos que sirvan para hacer un análisis de sensibilidad y permita tener un panorama más amplio del fenómeno de estudio.

BIBLIOGRAFÍA

- Gertler, Martínez, Premand, Rawling, Vermeersch, C. (2011). Evaluación de impacto en la práctica. pp.235.
- THE WORLD BANK. Lima (2009). Sesión Técnica III Diferencia en diferencias (Dif-en - dif) y Datos panel. pp 1-25.
- Cansino, Sánchez A. (2009). Cálculo del estimador de diferencias en diferencias aplicado a la evaluación de programas públicos de formación: métodos alternativos para su obtención a partir de datos simulados. pp 1-26.
- Card, Krueger, A. (1999). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*. vol: 84 (4). pp: 772-793.
- SEI, Buxton, Escobar, Purkey, D. (2013). El Cambio Climático y la Crisis del Agua en La Paz y El Alto.
- UDAPE, Ministerio de Planificación del desarrollo, y Comité Interinstitucional Metas de Desarrollo del Milenio MDM. (2013). *Objetivos de Desarrollo del Milenio en Bolivia. Séptimo informe de progreso 2013.*
- Montero, B. (21 de noviembre 2016). Gobierno declara emergencia nacional por sequía. *La Razón Digital*.

APLICACIÓN DEL REDATAM

Lic. Mancilla Flores, Emma Martha ¹

✉emmartha3@gmail.com

RESUMEN

El presente artículo es una descripción breve del software libre REDATAM , en base a la experiencia del Taller: “Uso y procesamiento de datos censales para la generación de Indicadores Sociodemográficos utilizando en REDATAM”²

PALABRAS CLAVE

REDATAM, mapa temático.

1. INTRODUCCIÓN

La Agenda 2030 y Objetivos de Desarrollo Sostenible (ODS)- Una oportunidad para América Latina y el Caribe; es un plan de acción mundial cuyo objetivo fundamental es asegurar el progreso social y económico sostenible en todo el mundo y el fortalecimiento de la paz universal³.

Son un llamado universal a la adopción de medidas para poner fin a la pobreza, proteger el planeta y garantizar que todas las personas gocen de paz y prosperidad.

Mediante la aprobación de la Agenda 2030 para el Desarrollo Sostenible, los Estados Miembros de las Naciones Unidas trazaron el camino hacia una mayor dignidad, prosperidad y sostenibilidad para las personas y el planeta, y se comprometieron a no dejar a nadie atrás.

La agenda está integrada por 17 objetivos, desagregadas en 169, que abarcan 5 esferas

de acción: personas, planeta, prosperidad, paz y alianzas; cuyas metas globales son el crecimiento sustentable, prosperidad energética, cuidado del medio ambiente, seguridad, paz e igualdad de género, las cuales se medirán a través de 230 indicadores globales.

Ante este nuevo reto surge la necesidad emergente de tener datos que coadyuven a una óptima toma de decisiones de forma rápida, responsable y de manera precisa generados a través de censos de población y vivienda, censos agrícolas, encuestas laborales, encuestas demográficas de salud, encuestas de migración, encuesta integrada de hogares entre otras.

Y anterior a este reto, debido a las restricciones tecnológicas y búsqueda de soluciones a cuestionamientos de esta naturaleza, es que la CELADE⁴ a partir de la década de los 80's empieza la búsqueda de respuestas y es así que en 1986 aparece la primera versión de REDATAM cuya

1 Docente de la carrera de Estadística-UMSA.

2 Desarrollado en la UMSA auspiciado por IDIS-carrera de Sociología y el Fondo de Población de las Naciones Unidas (UNFPA).

3 www.cepal.org/celade

4 Centro Latinoamericano y Caribeño de Demografía (CELADE), que desde 1997 es la División de Población de la Comisión Económica para América Latina y el Caribe (CEPAL)

principal característica era el procesamiento de un gran volumen de datos que se podían manejar en un computador personal, aunque la rapidez no era su principal característica.

A partir de la primera generación se fue mejorando, a más de dos décadas se cuenta con la cuarta generación REDATAM+SP⁵ cuyas características son mucho mejores que la inicial y posteriormente dando respuesta a muchas necesidades de los usuarios se incorpora la plataforma REDATAM WebServer, que permite el procesamiento de información en línea yendo a la par del desarrollo tecnológico mundial.

2. ¿QUÉ ES REDATAM?

REDATAM es un acrónimo de *Recuperación de datos* para áreas pequeñas por *microcomputador*.



Es un software de distribución y uso libre, constituye una herramienta para administrar bases de datos de gran volumen, se usa para el procesamiento, análisis, mapeo y generación de indicadores de la población de información de censos, encuestas, estadísticas vitales y otras fuentes de datos, a partir de una estructura jerárquica, con desagregación geográfica como desee el usuario, permitiendo el análisis nacional, regional y local.

Permite obtener estadísticas básicas y descriptivas: Media, mediana, moda, frecuencias, rango, otros; así como tablas, gráficos, indicadores, mapas temáticos.

2.1 CARACTERÍSTICAS

- Es de fácil uso, su interfaz de usuario está acorde a las tecnologías actuales. La pueden emplear especialistas o no.
- El almacenamiento de los datos sigue un formato encriptado lo que no permite la identificación de registros de individuos.
- Base de datos altamente comprimida, se compone de tres tipos de archivos:

» rdf, archivo de variables

» ptr, archivo de punteros

» dicX, el diccionario

Estos son almacenados en forma codificada y comprimida para proteger la confidencialidad de los mismos.

- Permite una interface con Sistemas de Información Geográficas como Arc-Gis, QGis, ArcView, otros.

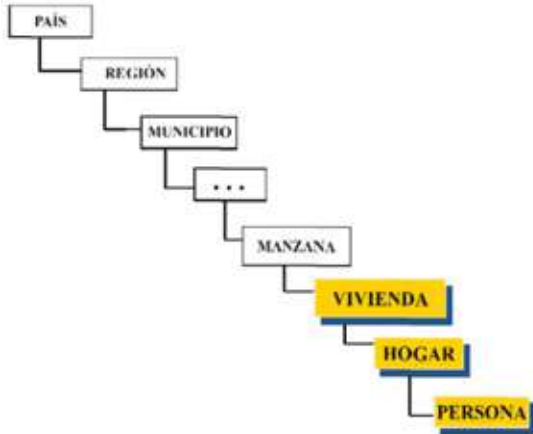
2.2 BENEFICIOS

Al usar esta herramienta, se puede mencionar que:

- Ayuda a manejar grandes volúmenes de datos organizados de manera jerárquica, facilitando el procesamiento de la información para unidades geográficas según vea necesario el usuario.

5 Originalmente llamada REDATAM+G4.

Figura N° 1
Estructura jerárquica



- Obtiene indicadores estadísticos de manera rápida y confiable mediante frecuencias, cruces de variables, promedio, mediana, máximo, mínimo.
- Posibilita la selección de subconjuntos de datos dentro de una zona geográfica de interés, así como la definición de filtros y universos.
- Elabora mapas temáticos de zonas de interés mediante listas de áreas y conteos.
- Presenta la información en forma gráfica, incorporando los tipos de gráficos más utilizados: barras, torta, pirámide de edades y otros.
- Provee de un lenguaje de programación para recodificación y construcción de nuevas variables y generación de indicadores.
- Permite el intercambio de información con otras fuentes de datos, como: SPSS, DBase, Excel, otros.
- Además cuenta con un servidor en línea WebServer (Figura N° 2), mediante el cual se puede realizar desde el procesamiento pasando por el análisis de información hasta conseguir resultados mediante tablas, gráficos y mapas.

Figura N° 2
Servidor en línea



- Inicia desde la validación de los datos, verificación de la consistencia y la creación y generación una base de datos
- Para el proceso: Análisis de datos, definición y cálculo de indicadores además de nuevas variables, genera

tablas, gráficos y mapas.

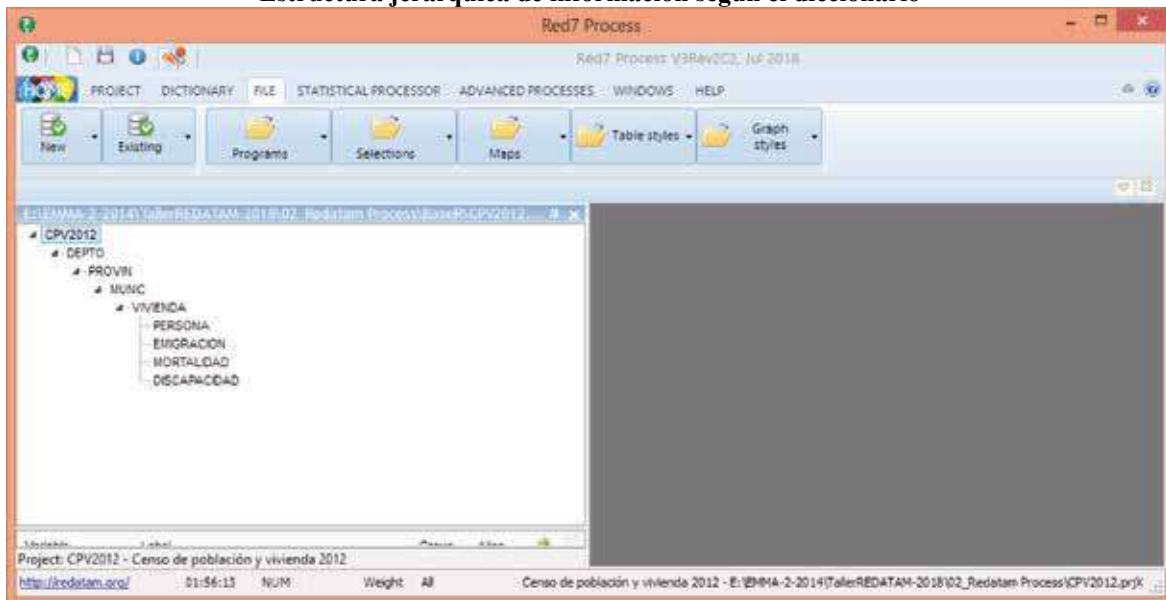
REDATAM R+SP.

3. EJEMPLO DE APLICACIÓN

En el taller “Uso y procesamiento de datos censales para la generación de indicadores sociodemográficos utilizando REDATAM”, se nos facilitó para el trabajo la base de datos del Censo Nacional de Población y Vivienda 2012 (CPV2012) y se empleó el software

Cuya base se encuentra distribuida de manera jerárquica como se observa en la Figura N°3, cuya desagregación permite analizar desde lo general a nivel nación como por departamentos, provincias, municipios, por vivienda, persona con características sociodemográficas como son emigración, mortalidad, discapacidad.

Figura N°3
Estructura jerárquica de información según el diccionario



En base a lo cual de acuerdo a la necesidad que se tenga se pueden generar tablas, gráficos como Tabla N° 1, Figuras N° 1 y N° 2 a nivel nacional la descripción de la

variable Edad en grupos generales y por grupos quinquenales, observándose que aún Bolivia tiene una población joven.

Tabla N° 1

CEPAL/CELADE Redatam7 03/10/2018

Base de datos

C:\tallerRedatam\02_Redatam Process\BaseR\CPV2012Comunidad.dicX

Área Geográfica

Toda la Base de Datos

Frecuencia

de PERSONA.EDAD3GR (Edad en grandes grupos)

PERSONA.EDAD3GR (Edad en grandes grupos)	Casos	%	Acumulado %
1. niños-jóvenes	6 063 051	60,27%	60,27%
2. adultos	3 398 660	33,78%	94,05%
3. adultos mayores	598 145	5,95%	100,00%
Total	10 059 856	100,00%	100,00%

Gráfico N°1
Bolivia: Distribución de la población según grandes grupos de edades
 CPV -2012

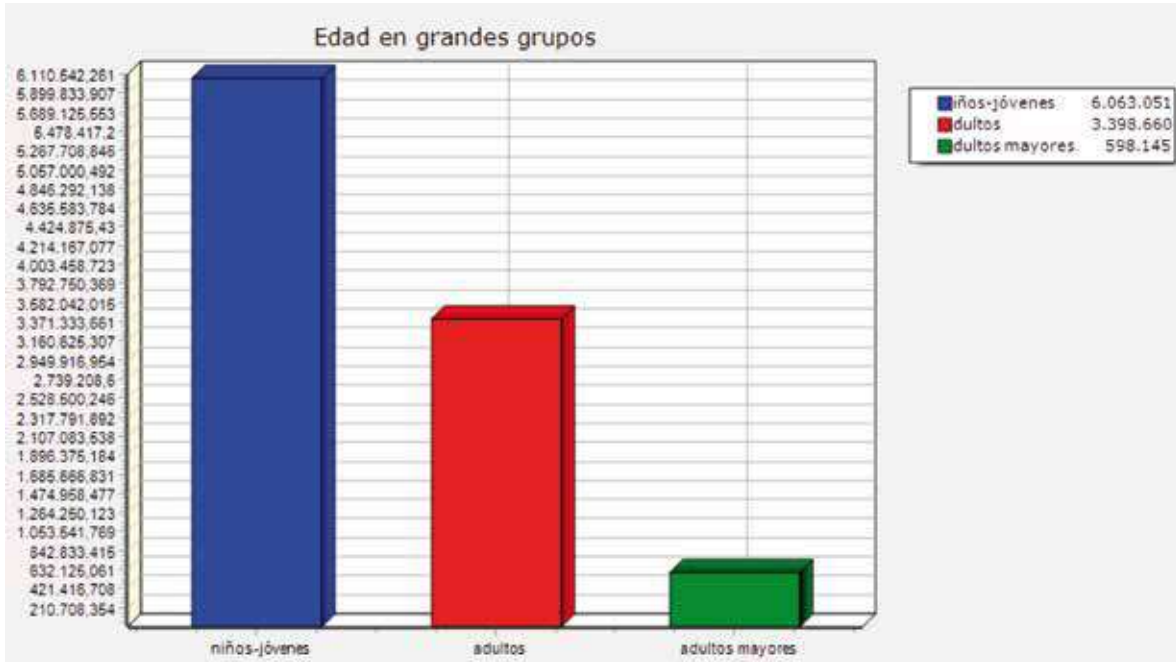
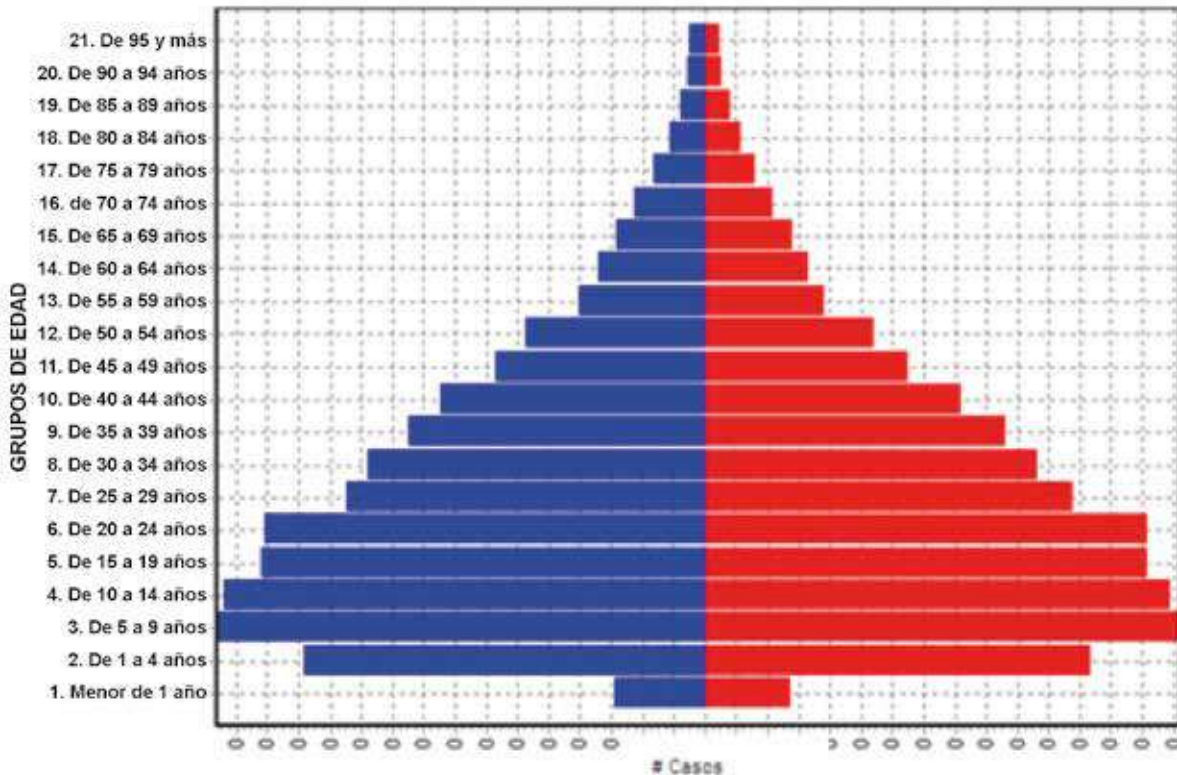


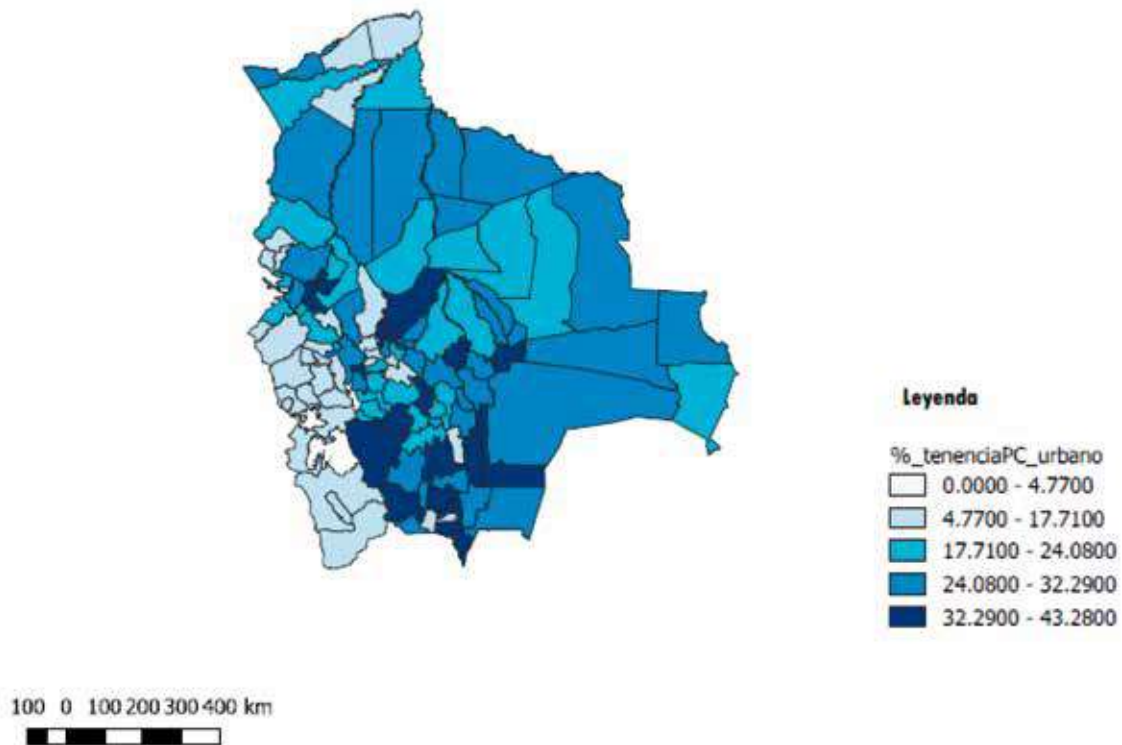
Gráfico N°2
Bolivia: Distribución de grupos de edad quinquenales por sexo
 CPV -2012



Como se mencionó líneas arriba, también se pueden calcular y generar indicadores ya establecidos así como otro tipo, en este caso se calculó porcentaje de tenencia de teléfono

celular en el área urbana y para observar los resultados Gráfico N°3 mediante el mapa temático que se pudo generar empleando el QGis⁶ para este objetivo.

Gráfico N°3
Bolivia: Porcentaje de tenencia de teléfono en área urbana



4. CONCLUSIONES

- El REDATAM constituye una herramienta útil para el trabajo estadístico con sus ventajas y desventajas combinando las capacidades del SPSS⁷ y sobre todo proporciona un apoyo para el análisis de aplicación de políticas públicas nacionales o locales.

- Es un software empleado mundialmente para el análisis principalmente de diferentes censos y encuestas en varios países como: Argentina, Chile, Bolivia, Costa Rica, Brasil, países de África, Asia y muchos otros.
- Se aplicó el REDATAM en el análisis del Censo de Población y Vivienda 2012 del Estado Plurinacional de Bolivia.

6 QGIS (Quantum GIS) es un Sistema de Información Geográfica (SIG) de código libre para plataformas GNU/Linux, Unix, Mac OS, Microsoft Windows y Android.
7 Paquete estadístico para las Ciencias Sociales.

BIBLIOGRAFÍA

Material del taller, responsable Especialista Redatam Informa , Vol.17 Diciembre 2011
Alejandra Silva

<http://www.cepal.org/celade>

Redatam Informa , Vol.14 Diciembre 2008

ESTIMACIÓN DE LA ECONOMÍA INFORMAL EN BOLIVIA

Lic. Miranda Gonzales, Yuri¹

✉ yymiranda@gmail.com

RESUMEN

La investigación desarrolla una aproximación de la magnitud del sector informal en Bolivia respecto al empleo y al producto, el periodo de análisis es de 1999 al 2013 –debido a la disponibilidad de la Matriz Insumo Producto – es así que, con relación al empleo los resultados evidencian una participación de la informalidad en torno al 69,7% (con un ligero descenso a partir del 2008) y alrededor del 26,7% respecto al producto, los resultados señalan que la magnitud de la informalidad no presenta mejoras significativas con relación a otros indicadores como pobreza, desigualdad y desempleo.

El documento está dividido en cuatro secciones, la primera está referida a la síntesis de aspectos conceptuales que le dan un marco teórico referencial al documento; la segunda sección contiene un análisis sobre la estimación de la dimensión de la informalidad en Bolivia, a través de dos puntos de vista respecto al empleo y al producto; la tercera sección contiene la descripción de las características de la informalidad. En la cuarta sección se desarrolla un resumen de propuestas de políticas y finalmente las principales conclusiones.

PALABRAS CLAVE

Matriz Insumo Producto, Valor agregado, Ingreso bruto, Productividad, Conciliación temporal

1. INTRODUCCIÓN

La informalidad, como fenómeno de estudio, no es algo nuevo y durante las últimas décadas han surgido numerosas investigaciones para tratar de caracterizarla y encontrar sus causas, así como su incidencia en la economía con relación a la generación de empleo y su participación en el valor agregado. La necesidad de comprender mejor la informalidad, se basa en su tamaño y en la constatación empírica de que en este ámbito se observa la falta de cumplimiento de normativas por parte de empresas y trabajadores, la baja productividad del trabajo y demás factores de producción,

la marginalización de las empresas que se desenvuelven en este marco, los más bajos niveles de formación de empresarios y trabajadores y los menores niveles de ingresos que perciben en relación al sector formal.

El presente trabajo busca ayudar a una mejor comprensión del sector informal, mediante el análisis de las encuestas a hogares, realizadas por el Instituto Nacional de Estadísticas (INE) de Bolivia como principal fuente de información. Por otra parte, se pretende responder a preguntas que son recurrentemente estudiadas, como, por ejemplo: ¿cuál es la participación de la informalidad en el empleo?, ¿cuál es su

1 Docente Investigador, Carrera de Estadística – UMSA.

2 Centro de Estudios para el Desarrollo Laboral y Agrario (CEDLA) publicó varios estudios sobre el sector informal entre los que destacan: “Los trabajadores por cuenta propia en La Paz: funcionamiento de las unidades económicas, situación laboral e ingresos” 1988. “Informalidad e ilegalidad una falsa identidad 1989 Larrazabal H. Adicionalmente están los trabajos de: Fernández M. “Sector informal y movimiento obrero, La Paz Centro de Estudios Laborales (ILDIS), 1987. Mercado, A. “Principales problemas del universo laboral, ILDIS, 1994.

contribución a la actividad económica?, ¿cuáles son las causas que llevan a las empresas a pertenecer o no al sector informal?

En Bolivia se realizaron varias investigaciones especialmente a partir de los años 80², existiendo diferentes documentos de investigación que sirven como antecedentes, entre los que se pueden mencionar: el Centro de Apoyo al Desarrollo Laboral (LABOR) 2002, en el cual toman como base las zonas urbanas del país y determinan la existencia de una gran segmentación productiva en Bolivia. Un segmento mayoritario (62% de los ocupados urbanos), está integrado por personas calificadas como pobres, que desarrollan actividades informales con escaso capital, bajos niveles educativos, baja productividad y bajos ingresos. Otro, minoritario (38% de los ocupados urbanos), realizan actividades en el sector moderno de la economía, con mayor acceso al capital, alta productividad (seis veces la de los informales) y mayores ingresos, estando la mayoría por encima de la línea de la pobreza.

Julio Humérez (2005), realiza una investigación para determinar el tamaño de la economía informal en el Producto Interno Bruto (PIB). A partir de información sobre los agregados monetarios y utilizando el supuesto de que el circulante es el medio de pago más usado por quienes pertenecen a este sector, concluye que el tamaño de la economía informal se sitúa alrededor del 47% del total.

Landa y Yañez (2007), tratan de determinar cuál es la proporción de trabajadores bolivianos que pertenecen al sector informal. En base a las Encuestas de Hogares, representativas a nivel urbano, llega a la conclusión de que en la economía informal se encontrarían alrededor del 60% de los

trabajadores, con tendencia decreciente a lo largo del tiempo.

Martínez y Chumacero, en su investigación “El Sector informal urbano en Bolivia 1995 – 2005“, donde realizan un análisis sobre el empleo, ingreso, productividad y contribución al Producto Interno Bruto urbano, afirman que la participación de la informalidad en el PIB urbano alcanza al 23,5% para 2005 y que el empleo en el sector informal ocupa al 62,4% de los trabajadores.

Es así que la informalidad es un fenómeno socioeconómico con distintas vertientes de análisis. La definición es amplia y el factor común es la escasez de información para cuantificarla adecuadamente, por otro lado, la manera como se intenta reducir, varía según los países y regiones, debido a las diferentes causas que la generan. En lo que hay consenso, es en la importancia de conocer las características de la informalidad, como afecta el crecimiento de una economía, cuáles son los impactos sobre la sociedad y de qué forma se supera un fenómeno que es perjudicial para el desarrollo humano y económico.

Para una mejor comprensión del lector, el presente documento está dividido en cuatro secciones. La primera describe una síntesis de aspectos conceptuales que dan un marco teórico referencial al documento. La segunda sección contiene la estimación de la informalidad en Bolivia durante los últimos 14 años, a partir de la información de las Encuestas a Hogares (EH) realizadas por el INE. El estudio se enfoca en dos componentes: la incidencia en el empleo y su participación en el Producto Interno Bruto. Se obtiene una estimación del tamaño del sector informal como generador de empleo para el período que va desde 1999

al 2013. El segundo elemento se enfoca en la estimación del tamaño de la economía informal con relación al Producto Interno Bruto (PIB) urbano. La tercera sección contiene la descripción de las características generales del mercado informal. Finaliza con la sección de propuestas de líneas de acción y las principales conclusiones.

2. ASPECTOS CONCEPTUALES

El concepto de Sector Informal surgió tras el conocido “Informe sobre Kenia” en 1972, cuando una misión técnica de la Oficina Internacional de Trabajo (OIT) constata que el desempleo en ese país era bajo pero que había un enorme contingente de trabajadores pobres que habían conseguido sobrevivir desempeñando oficios con muy poco o ningún capital y sin someterse a las normas legales vigentes, muy costosas para ellos. Los integrantes de la misión de la OIT denominaron a este contingente de trabajadores pobres como “sector no institucionalizado”.

La OIT definió la informalidad en dos momentos diferentes que merecen destacarse. El primero (1993) utilizando la denominación de “sector informal” y, el segundo (2003) llega al concepto de “economía informal”. También en 2002 propuso reconocer la diversidad del trabajo informal y pidió a los Estados miembros de la OIT tomar medidas relativas al déficit de trabajo decente en la economía informal.

En el Sistema de Cuentas Nacionales de las Naciones Unidas (SCN-1993) ya se habló de la producción informal, explicando que la misma comprende aquellas actividades que son productivas desde un punto de vista económico legal, pero que son escondidas del Estado con el objetivo de evadir pago de

impuestos sobre la renta, el valor agregado y otros impuestos a la producción. Luego la versión del SCN – 2008, dedica un capítulo a la economía informal y su articulación con las cuentas nacionales y la necesidad de cuantificarla.

El Fondo Monetario Internacional (FMI) realizó un estudio en el que analiza algunos determinantes de la economía informal, como la rigidez del mercado laboral, la importancia de la agricultura, la inflación, la fortaleza del sistema en aplicación de las normas y la presión tributaria.

Hernando De Soto planteó un enfoque diferente de lo que es el Sector Informal Urbano, cómo nace y cómo se reproduce. La informalidad es resultado de la imposibilidad que tienen muchas personas, en su mayor parte migrantes del campo, de acceder a la legalidad, dado el alto costo en dinero y tiempo que esa legalidad implica. Si bien el enfoque de De Soto tiene la debilidad que desconoce el problema de la heterogeneidad estructural que está detrás del fenómeno de la informalidad, tiene la virtud que incorpora un aspecto: los altos costos de transacción, un tanto ausentes en los enfoques estructuralistas.

El Banco Mundial (Perry et al. 2007) realizó una investigación sobre la informalidad en América Latina y el Caribe bajo los enfoques de la exclusión y del escape. La exclusión significa que los trabajadores se encuentran en esta situación debido a razones estructurales, más allá de su voluntad. El concepto del escape habla de los trabajadores por cuenta propia que están en la línea de vulnerabilidad porque encuentran pocos beneficios y elevados costos en la formalidad, lo que les lleva a tomar una decisión propia de ir hacia la informalidad. En este documento destaca la definición que proponen sobre informalidad, como “una manifestación de las relaciones

entre los agentes económicos y el Estado, el cual desempeña una función importante en cuanto a mitigar las fallas del mercado, asegurar el suministro de bienes públicos y mantener condiciones para la igualdad de oportunidades”.

Por lo tanto, es importante evitar confusiones terminológicas, como recomienda la OIT en su publicación “Medición de la Economía Informal – 2012”, por ello a continuación se describen algunas definiciones extraídas del mismo documento y de la Decimoquinta Conferencia Internacional de Estadísticos del Trabajo (CIET):

- a) **Economía Informal:** Todas las actividades económicas de trabajadores o unidades económicas que, en la legislación o en la práctica, no recaen en el ámbito de mecanismos formales o estos son insuficientes (basado en Conferencia Internacional del Trabajo 2002).
- b) **Sector Informal:** Grupo de unidades de producción (empresas no constituidas de propiedad de jefes de hogares), incluidas las “empresas informales por cuenta propia” y las “empresas de trabajadores informales” (basado en la Decimoquinta CIET)
- c) **Empresa del sector informal:** empresas no registradas o empresas privadas pequeñas no constituidas que realizan actividades no agrícolas y cuya producción de bienes o servicios es para la venta o el trueque, al menos en parte (basado en la decimoquinta CIET)
- d) **Empleo en el sector informal:** Todos los trabajadores en empresas del sector informal (inciso c) o todas las personas

que estuvieron empleadas en por lo menos una empresa del sector informal, independientemente de su situación laboral y de si se trató de su trabajo principal o de un trabajo secundario.

- e) **Empleo asalariado informal:** Todos los empleos asalariados caracterizados por una relación de trabajo que no está sujeta a la legislación laboral nacional, al impuesto sobre la renta, a la protección social o a determinadas prestaciones relacionadas con el empleo.
- f) **Empleo Informal:** El número de trabajadores informales, tanto en empresas del sector formal, empresas del sector informal u hogares, incluidos los empleados que tienen empleos informales (e); los empleadores y trabajadores por cuenta propia que trabajan en sus propias empresas del sector informal, los miembros de cooperativas de productores informales; los trabajadores familiares auxiliares en empresas del sector formal o informal; y los trabajadores por cuenta propia que producen bienes para el consumo final por su hogar (basado en la decimoquinta CIET)



Fuente: Elaboración propia en base a OIT

En base a los aspectos conceptuales y los antecedentes revisados, se puede establecer que la **informalidad** – en su amplio conjunto de categorías de definiciones – tiene “al menos” las siguientes características:

3. MAGNITUD DE LA INFORMALIDAD EN BOLIVIA

En Bolivia y en muchos países no existe una estimación oficial del tamaño del sector informal. Es por eso que existen diferentes aproximaciones metodológicas, entre las cuales destacan los métodos directos (o en otras ocasiones son denominados métodos microeconómicos), las cuales se basan básicamente en el uso de Encuestas a Hogares y Encuestas de Empleo. Dichos métodos están condicionados a los alcances de las encuestas, tanto en cobertura geográfica como en diferencias metodológicas, pero no por ello dejan de ser una importante herramienta de análisis.

Por otro lado, los métodos indirectos se caracterizan por el uso de información de agregados macroeconómicos³. Los métodos indirectos que destacan – por su uso frecuente – son la estimación monetaria, el análisis de consumo de energía eléctrica y los modelos con múltiples causas y múltiples indicadores (MIMIC).

4. PARTICIPACIÓN DEL EMPLEO INFORMAL

Debido al alcance de las encuestas a hogares y las características de la economía boliviana, se estimará la magnitud de la economía informal urbana en la población

económicamente activa.

Es importante aclarar, antes de realizar dicha estimación que se realizó un riguroso trabajo de homogenización de la información, en cuanto a las variables usadas para cada encuesta a hogares. Como el objetivo es obtener una serie de tiempo que muestre la evolución y la magnitud de la informalidad en el empleo, es necesario realizar una conciliación de las variables a ser usadas. Con el objetivo de no realizar supuestos fuertes que pudieran originar discrepancias en los resultados, se decidió no usar la Encuesta continua de Hogares realizada entre el 2003 y 2004 y tampoco incluir en la serie la Encuesta Trimestral de Empleo (ETE) del año 2010, debido a sus características significativamente diferentes en metodología de muestreo y a su alcance a nivel nacional. La no inclusión de estos resultados en el periodo de observación, no hacen perder generalidad a la serie analizada y los resultados muestran con solidez el comportamiento de la informalidad en el periodo mencionado.

La clasificación de los informales se basó en el documento “La Medición de la Informalidad: Manual Estadístico Sobre el Sector Informal y el Empleo Informal – OIT primera edición 2013 p.42”. Por ello se consideró como trabajadores del sector informal urbano a:

- a. Ocupados en establecimientos económicos donde hay máximo 5 trabajadores
- b. Trabajadores por cuenta propia

3 Camargo, 2001. “estos métodos se caracterizan en buscar residuos no explicados de acuerdo con el comportamiento de una variable macroeconómica, a partir de la cual puede inferirse la magnitud de la actividad económica informal. Cabe destacar que estos métodos, contrariamente a los métodos microeconómicos, permiten trabajar con series estadísticas de tiempo que hacen posible observar la tendencia de las actividades informales”

4 No se toman en cuenta las trabajadoras y los trabajadores del hogar. Puesto que esta categoría se la considera de manera separada, existen varios trabajos que siguen esta línea debido a sus características individuales y de pertenencia a una unidad familiar, para mayor información se sugiere revisar las publicaciones de: Martínez, INE, UDAPE, J. Evia y N. Pacheco

- c. Patrones socios o empleadores que no reciben remuneración
- d. Trabajadores aprendices familiares sin remuneración⁴

económica)⁵. La estimación del empleo informal resulta:

$$SI = \frac{\sum_{i=1}^4 X_i}{PEA} * 100\%$$

Durante la elaboración de la investigación (2016) y en la recopilación de información histórica, la última encuesta a hogares que se dispone corresponde a la gestión 2013, para fines descriptivos la mayoría de las series se muestran hasta ese año.

SI= Empleo informal

X₁= Cantidad de ocupados en microempresas con menos de 5 trabajadores.

X₂= Cantidad de trabajadores por cuenta propia.

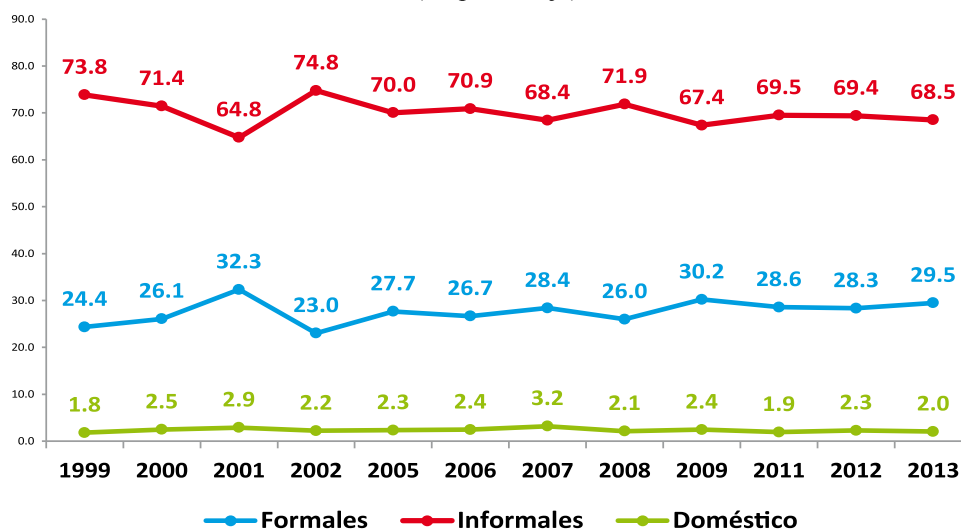
X₃= Cantidad de patrones socios empleadores sin remuneración

X₄= Cantidad de trabajadores aprendiz, familiar sin remuneración

Según el Instituto Nacional de Estadística (INE) se considera Población Económicamente Activa (PEA) “personas que en determinado periodo de tiempo trabajaron o buscaron trabajo activamente” y que pertenecen a la Población en Edad de Trabajar (PET, “Habitantes de 10 años y más de edad y personas capacitadas física e intelectualmente para ejercer una actividad

PEA= Población económicamente activa

Gráfico N° 1
Bolivia: Participación del Sector Informal en la PEA.
(En porcentaje)



Fuente: Elaboración propia en base a la encuesta a hogares INE

Los resultados son concordantes con estudios anteriores⁶ y se evidencia que la población

económicamente activa encuentra como principal fuente de ingreso al sector informal,

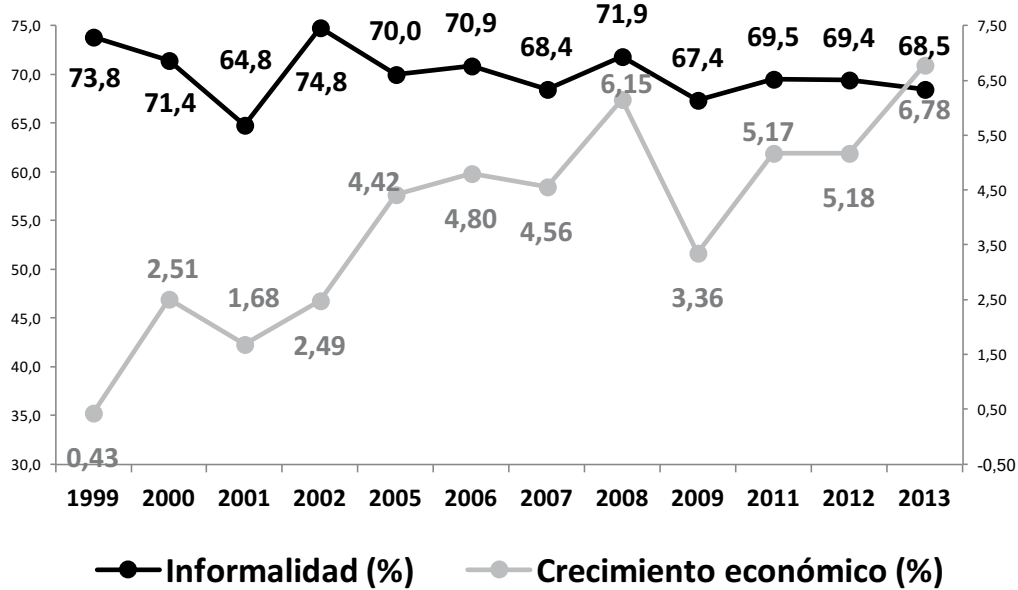
5 www.ine.gob.bo

6 Martinez, INE, UDAPE, J. Evia N. Pacheco

con un 68,5% en el 2013. La evolución de este es prácticamente estable entre 1999 y el 2013, en torno al 69,7%, con un comportamiento ligeramente decreciente a partir del año 2008 y continúa decreciendo aunque de manera muy leve, pese a que el crecimiento

económico presentó un comportamiento favorable durante el periodo de análisis. Es decir, que mientras la economía del país presentó crecimientos históricamente elevados, la informalidad prácticamente se mantuvo constante (Gráfico N° 2).

Gráfico N° 2
Bolivia: Informalidad y Crecimiento Económico, 1999 – 2013
(En porcentaje)



Fuente: Elaboración propia en base a la encuesta a hogares INE

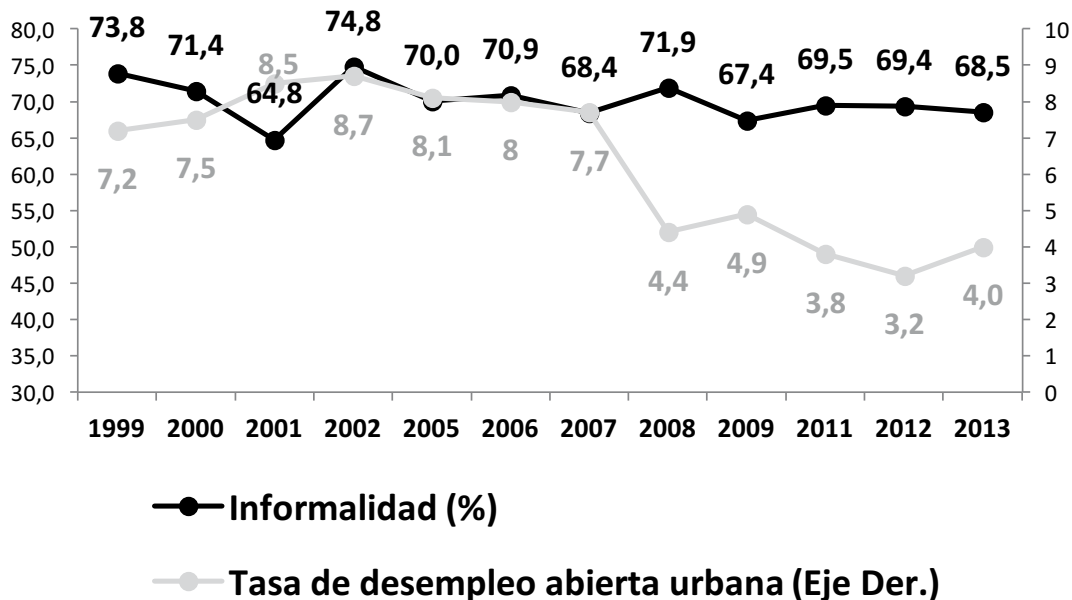
Si bien en los últimos años el crecimiento del Producto Interno Bruto fue explicado principalmente por la demanda interna, se puede observar que la informalidad se mantuvo constante. Este hecho puede ser explicado debido a que al interior de la demanda interna el principal componente – en cuanto a valores – es el consumo de los hogares. Analizando la matriz insumo

producto⁷, se tiene que en el vector de consumo de los hogares, muestra que las actividades económicas más importantes son: carnes frescas y elaboradas, transporte y almacenamiento, restaurantes y productos agrícolas no industriales, actividades donde la informalidad históricamente tiene una alta incidencia.

7

La Matriz Insumo Producto disponible en la página oficial del INE es de la gestión 2012

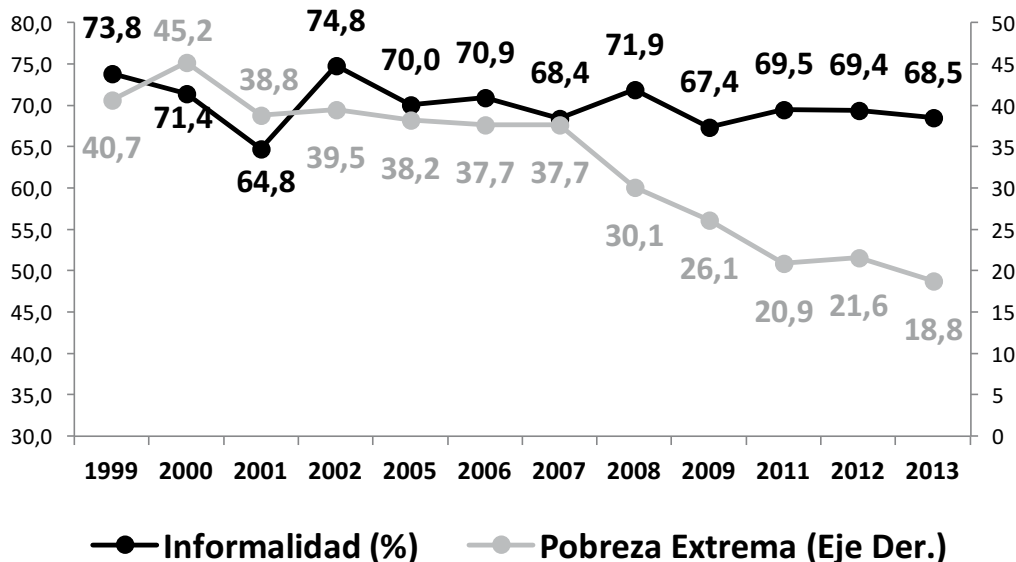
Gráfico N° 3
Bolivia: Informalidad y Tasa de desempleo Abierta Urbana, 1999 – 2013
 (En porcentaje)



Fuente: Elaboración propia en base a la encuesta a hogares INE

La Tasa de Desempleo de 8,7% en el 2002 reflejando que la Población Económicamente descendió a menos de la mitad para el 2013 Activa encontró empleo principalmente en el (4%). A pesar de ello, la informalidad no sector informal. presentó una disminución al mismo ritmo,

Gráfico N° 4
Bolivia: Informalidad y Pobreza, 1999 – 2013
 (En porcentaje)



Fuente: Elaboración propia en base a la encuesta a hogares INE

El crecimiento económico, pobreza extrema y el empleo mostraron mejoras en los últimos años y, al mismo tiempo, la disminución de la informalidad en el empleo no acompañó la variación de estos indicadores. Esto puede ser interpretado de distintas formas, pero la más inmediata surge del hecho de que la población económicamente activa encuentra trabajo en actividades del sector informal, reflejándose como un problema estructural, en el sentido de que la informalidad tiene alta incidencia tanto en ritmos de crecimientos significativos como en etapas de bajo crecimiento.

4.1. CONTRIBUCIÓN DE LA INFORMALIDAD AL PIB

La estimación de la contribución del sector informal al Producto Interno Bruto urbano, sigue la línea de los métodos directos y el enfoque usado se basa en el planteamiento de M.A. Castigulia⁸.

“Un método simple para aproximarlos proviene de considerar que en su enorme mayoría el ingreso de los informales se declara en las encuestas de hogares, porque ninguno de ellos tiene clara la separación entre la retribución a su trabajo y el pago por los servicios del capital de su propiedad. Así, el producto del sector informal urbano puede aproximarse mediante la sumatoria de los resultados – convenientemente expandidos – de las encuestas de hogares”.

“Este *producto informal* no debe compararse directamente con las estimaciones del producto que reflejan las cuentas nacionales, por un lado, porque hay diferencias metodológicas importantes y, también, debido a que una parte del sector informal

urbano ya está computado en el Producto (lamentablemente, no es posible estimar en qué proporción ocurre tal fenómeno). Pero si puede obtener de la misma fuente, el total de los ingresos laborales del sector moderno y asumir que es correcta y adecuada para la realidad del empleo moderno, la estimación de las cuentas nacionales sobre la participación de los salarios en el producto. Así puede construirse una estimación del producto del sector moderno comparable con la del sector informal urbano y entonces estimar, con los datos fácilmente accesibles, la participación del sector informal urbano en el producto interno bruto”.⁹

Siguiendo esa línea se tiene:

$$PSIUPBI = \frac{\sum y_i}{\sum y_i + (\sum y_f \times psp)} * 100\%$$

Donde:

PSIUPBI = Es el porcentaje del Producto Interno Bruto total que se genera en el sector informal.

Y_i = Es el ingreso de los trabajadores del sector informal urbano según la encuesta a hogares.

Y_f = Es el ingreso laboral de los asalariados del sector moderno según la encuesta a hogares.

PSP = Es la participación de los salarios del sector moderno en el Producto Interno Bruto, según las cuentas nacionales.

Las estimaciones encontradas se describen en el siguiente cuadro:

8 La Microempresa en América Central, La experiencia de PROMICRO – OIT, “Sector Informal Urbano: Una Aproximación a su Aporte al Producto” 1994.
9 La Microempresa en América Central, La experiencia de PROMICRO – OIT, “Sector Informal Urbano: Una Aproximación a su Aporte al Producto” 1994. P.47

Cuadro N° 1
Bolivia: Participación del Sector Informal en el PIB urbano, 1999 – 2013
(En porcentaje)

	1999	2000	2001	2002	2005	2006	2007	2008	2009	2011	2012	2013
Participación	28.1%	27.4%	24.8%	28.5%	26.8%	27.1%	26.0%	27.8%	25.8%	26.3%	26.2%	25.9%

Fuente: Elaboración propia en base a la encuesta a hogares INE

Los resultados son comparables con los expuestos por D. Martinez¹⁰ y J.L, Evia y M.A. Pacheco. Al respecto es importante mencionar que los dos enfoques (magnitud en el empleo y en el Producto Interno Bruto) no tienen por qué resultar coincidentes. Por el contrario, como señala D. Martinez, no existe una evidencia teórica que logre sustentar una magnitud significativa en los dos enfoques.

“Es frecuente leer, escuchar o decir que el Sector Informal Urbano abarca el 50% de los ocupados en el área urbana y que generan el 60 o 70% de la actividad económica, o que el tamaño del sector es menor a esa cifra, pero su actividad económica mayor, [...]. Muchas de estas afirmaciones son totalmente erróneas. Así, si el Sector Informal Urbano soporta al 60% del empleo y sustenta el 70% u 80% de la actividad económica, entonces la productividad media del sector es mayor que uno, y la del Sector Formal (40% restante del empleo y 20% o 30% restante de la actividad económica) será menor de uno, es decir, que la productividad del Sector Informal Urbano sería mayor que la del Sector Formal, y, por tanto, el ingreso del Sector Informal Urbano también sería superior. Esto es, obviamente, un absurdo teórico que no se sustenta y no puede sustentarse en la menor evidencia empírica.”

5. CARACTERÍSTICAS DEL MERCADO INFORMAL

La creciente aglutinación de empresas denominadas microempresas y familiares, donde la informalidad se visualiza de manera directa y sin necesidad de establecer ciertas hipótesis, deriva en su característica de generación de empleo con baja inversión, sujetos al comercio y algunos servicios, a raíz de su manera de subsistencia en la realidad económica de Bolivia.

Según el Instituto Nacional de Estadística (INE), mediante la Encuesta a las Micro y Pequeñas Empresas¹¹, el comercio es la actividad más predominante dentro de las Mypes con un 51%, seguido de los servicios con un 39% debido a que en estas actividades económicas no se necesitan fuertes capitales de inversión.

Según el análisis del presente estudio, del 68,5% de la participación de la informalidad en el empleo para el año 2013, queda en evidencia que las mujeres son más propensas a encontrar refugio en la informalidad (Cuadro N° 2).

10 El sector informal urbano en Bolivia 1995- 2005.

11 El INE concretó la ejecución de la “Encuesta Trimestral a las Micro y Pequeñas Empresas MYPES 2010” dirigida a las empresas “tipo local visible” y no a los puestos móviles debido al tipo de muestreo. La cobertura de esta fue: Sucre, La Paz, Cochabamba, Oruro y Santa Cruz y la ciudad de El Alto.

Cuadro N° 2
Bolivia: Participación de la Informalidad en el Empleo Según Género, 1999 – 2013
 (En porcentaje)

Gestión	Formal		Informal		Empleo doméstico	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
1999	29.3%	18.4%	70.6%	77.8%	.2%	3.8%
2000	29.8%	21.4%	70.0%	73.2%	0.1%	5.4%
2001	39.4%	23.9%	60.4%	70.0%	0.2%	6.1%
2002	26.0%	18.0%	73.8%	77.3%	0.1%	4.7%
2005	32.2%	22.1%	67.7%	72.9%	0.1%	5.0%
2006	30.7%	21.8%	69.0%	73.2%	0.3%	5.0%
2007	33.2%	22.6%	66.6%	70.7%	0.3%	6.7%
2008	29.5%	21.9%	70.0%	74.0%	0.5%	4.1%
2009	34.9%	24.5%	64.7%	70.5%	0.4%	4.9%
2011	32.9%	23.2%	66.9%	72.7%	0.2%	4.1%
2012	32.3%	23.4%	67.6%	71.6%	0.1%	5.0%
2013	33.6%	24.4%	66.3%	71.2%	0.1%	4.4%

Fuente: Elaboración propia en base a la encuesta a hogares INE

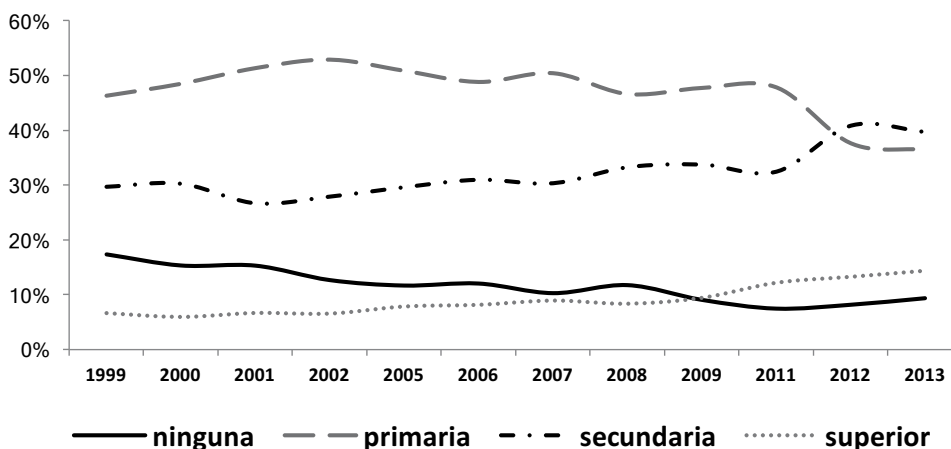
De la Población Económicamente Activa femenina, el 77,8% pertenecían al sector informal en el año 1999, pasando esta cifra a un 71,2 % para el 2013. Estos porcentajes son superiores con relación a la PEA masculina y esta característica se mantiene en todo el periodo de análisis – las mujeres mayoritariamente encuentran empleo en el sector informal con relación a los hombres – si bien la participación de la mujer es superior en comparación con la de los hombres, es importante destacar que existe un incremento en el sector formal de la participación de la mujer en los últimos años de 18,4% en el 1999 al 24,4% en el 2013.

Con relación al nivel de instrucción, el sector informal se concentra en los grados primario y secundario (superando así más del 50% en todos los años). Es importante destacar que la evolución del nivel de instrucción presenta importantes características, entre las cuales destaca:

- En 1999 la mayor participación del nivel de instrucción en el empleo informal correspondía a *primaria* con 46%. Posteriormente esta proporción paso a ser la segunda en el 2013, donde las personas con un nivel de instrucción *aprobado de secundaria* que pertenecen al sector informal alcanzan el 40%.
- El nivel superior (donde se incluye a profesionales o grado superior) en 1999 representaba solo el 7% de todos los informales. Luego, su participación se duplico a un 14% en el año 2013. Si bien pueden surgir varias interpretaciones de esta evolución, una de las conclusiones a las que se puede arribar es que existen carencias para la generación de empleos y oportunidades para mano de obra altamente calificada en la economía formal.

Cuadro N° 3
Bolivia: Participación del Nivel de Instrucción en la Informalidad, 1999 – 2013
 (En porcentaje)

Categoría	1999	2000	2001	2002	2005	2006	2007	2008	2009	2011	2012	2013
ninguna	17%	15%	15%	13%	12%	12%	10%	12%	9%	7%	8%	9%
primaria	46%	48%	51%	53%	51%	49%	50%	47%	48%	48%	38%	37%
secundaria	30%	30%	27%	28%	30%	31%	30%	33%	34%	32%	41%	40%
superior	7%	6%	7%	7%	8%	8%	9%	8%	9%	12%	13%	14%



Fuente: Elaboración propia en base a la encuesta a hogares INE

Por otro lado, la pobreza es una característica de la informalidad analizada en diferentes estudios que generalmente la relacionan con la falta de oportunidades de empleo y la falta de niveles de instrucción alcanzados. Mediante las estimaciones que realiza el INE en cada Encuesta a Hogares¹², se marca una

tendencia en cuanto al comportamiento de la pobreza, bajo el marco del análisis se tiene que la PEA catalogada como pobre tiene a un 81% de informales en el 1999 pasando este porcentaje a 86% en el 2013, decir hay más informales dentro de las personas catalogadas como pobres en el 2013.

Cuadro N° 4
Bolivia: Participación de la Informalidad en el Empleo Según Pobreza, 2000 – 2013
 (En porcentaje)

Gestión	Formal		Informal		Empleo doméstico	
	No Pobre	Pobre	No Pobre	Pobre	No Pobre	Pobre
2000	40%	18%	56%	81%	4%	1%
2001	47%	23%	49%	75%	4%	2%
2002	45%	20%	52%	79%	3%	2%
2005	43%	17%	55%	82%	2%	1%
2006	43%	15%	55%	83%	2%	2%
2007	45%	16%	53%	81%	2%	3%
2008	38%	16%	59%	83%	2%	1%
2009	42%	18%	56%	80%	2%	2%
2011	38%	16%	60%	83%	2%	2%
2012	37%	15%	61%	83%	2%	2%
2013	40%	12%	58%	86%	2%	1%

Fuente: Elaboración propia en base a la encuesta a hogares INE

12 El INE realiza una estimación de la pobreza mediante el enfoque indirecto o enfoque monetarista de la Línea de Pobreza, clasificando a las personas en “pobres” y “no pobres”.

Del total de personas clasificadas como “no pobres” en el año 2000, el 56% tenía como fuente de ingreso algún empleo en el sector informal, pasando esta cifra al 58% para el 2013 y prácticamente manteniendo ese comportamiento a lo largo del periodo de análisis.

6. PROPUESTAS DE LÍNEAS DE ACCIÓN

Se pueden establecer algunas propuestas de líneas de acción para atenuar o enfrentar el problema de la informalidad. Estas propuestas se enmarcan bajo el tradicional enfoque de:

- Incentivos a la formalización.
- Cumplimiento de legislaciones laborales.
- Integración de políticas públicas y privadas.

En esa línea, se considera importante realizar una breve descripción de algunas acciones que ya se realizaron en Bolivia hasta el año 2016, para enfrentar este problema con el objetivo de realizar propuestas de líneas de acción más específicas, que ayuden a fortalecer y/o aumentar las alternativas para enfrentar este problema.

- A partir del 2009 se creó el Ministerio de Producción y Economía Plural, dentro del marco de la nueva Constitución Política del Estado, donde uno de los objetivos de dicho ministerio es “diseñar políticas para estimular la productividad y la coordinación entre pequeñas medianas y grandes empresas”, en ese entorno se intentó incorporar al sector informal dentro del plan nacional de

desarrollo, con el enfoque de economía plural.

- Se establecieron incentivos para la formalización, uno de los más destacables es la implementación de un banco de segundo piso (Banco de Desarrollo Productivo – BDP) para mejorar el acceso al crédito por parte de los pequeños y micro empresarios y también productores rurales.
- La inscripción de garantías no convencionales para el acceso al crédito.
- El ofrecimiento por parte del gobierno de subsidios para estimular la formalización y certificación de calidad.
- En la nueva Constitución Política del Estado (CPE) se establece que “el Estado reconocerá priorizará y fomentará el desarrollo de las micro, pequeñas y medianas empresas productoras urbano y rurales”. Así también señala que “El Estado protegerá y fomentará el trabajo por cuenta propia, el comercio minorista y las micro y pequeñas empresas”.¹³

Las líneas de acción se enfocan en los siguientes puntos:

a) **Productividad:**

- a. Reducir los requisitos del acceso a financiamiento de nuevas tecnologías.
- b. Mejoramiento de capacidades y estructuras de apoyo públicas y privadas, con el objetivo de incrementar la inversión

13 Plan Nacional de Desarrollo “Bolivia, Soberana, Productiva y Democrática Para Vivir Bien” 2006 - 2011

en innovación, investigación, gestión y desarrollo empresarial, especialmente orientadas hacia las micro, pequeñas y medianas empresas.

- c. Capacitación técnica a trabajadores ocupados y desocupados y el otorgamiento de certificaciones, con el objetivo de fortalecer el capital humano en las primeras etapas de crecimiento de una empresa.

b) Incentivos a la formalización:

- a. Alivianar la carga impositiva en términos generales, pero especialmente durante los primeros años de vida de una nueva empresa formal.
- b. Disminuir los trámites administrativos y el tiempo que insume el registro de las nuevas empresas y de las empresas informales que buscan pasar a la formalidad.
- c. Promover la asociatividad empresarial entre las microempresas, teniendo especialmente en cuenta a los grupos más vulnerables de la sociedad, ofreciéndoles programas de capacitación empresarial y creando mecanismos que faciliten su interacción comercial con empresas de mayor tamaño.
- d. Promoción de Ferias a la Inversa exclusivamente para informales asociados, con el objetivo de fomentar la asociatividad y de esa

manera obtener beneficios para acceso al crédito y nuevos mercados.

- e. Buscar nuevos mercados externos para incentivar a formalizarse y beneficiarse de estos.

c) Incentivos a cumplir con la legislación

- a. Mejorar la legislación laboral con el objetivo de incorporar formas más flexibles de contratación de manera de emplearse pro jornadas de trabajo parciales y flexibles que se pueda realizar dentro de un marco de formalidad.
- b. Promover ventajas de la formalidad laboral y empresarial¹⁴.
- c. Cambio cultural en la población. Mostrar a la informalidad como un alejamiento de los principios básicos de convivencia ¹⁵

7. CONCLUSIONES

- Se estimó la magnitud de la informalidad en el empleo y en el Producto bajo el concepto de *empleo informal*. La evolución de los resultados señala que el problema es estructural, debido a que no se cuenta con elementos para sustentar que niveles de alto crecimiento garanticen mejores empleos y así reflejar una disminución de la incidencia en la informalidad.
- El bajo nivel educativo de los empresarios y trabajadores informales es una característica que resalta de los resultados,

14 Villamil R. “La economía informal causas, consecuencias y ejes de solución” en “23° Congreso de Gestión de Personas, OIT.

- en línea con diferentes investigaciones. La evolución de las incidencias de las categorías de nivel de instrucción, señala que cada vez existe más gente con grado de instrucción superior en el sector informal.
- Las mujeres son más vulnerables ante el mercado de trabajo, provocando que se refugien en empleos del sector informal. Pese a ello en los últimos años se percibe mejoras, como se refleja en el Cuadro N° 2.
 - La participación del sector informal urbano en el PIB urbano y la magnitud en el empleo, confirman que los empleos en este sector son de baja productividad.

BIBLIOGRAFÍA

- Banco Mundial (2008). “Informalidad: Escape y Exclusión”
- Castiglia M.A. y D. Martinez (1994). “Sector Informal Urbano: Una aproximación a su arte al producto”, La Micro Empresa en América Latina – La expresión de PROMICO, OIT 1991-1998.
- INE. “Glosario de Terminología Estadística”
- Evia, J y M.N. Pacheco (2010). “Sector Informal en América Latina”, SOPLA.
- OIT (2013) “La Medición de la Informalidad: Manual Estadístico Sobre el Sector Informal y el Empleo Informal” p.42
- Sanchez, D (2009). “El Sector Informal Urbano en Bolivia 1995-2005. Empleo, ingreso, productividad y contribución al Producto Interno Bruto”, Centro de Apoyo al Desarrollo Laboral.
- Séruzier, M (2011). “Economía Informal y Contabilidad Nacional”, Convergencia y Preguntas Entre Definiciones OIT y Punto de Vista del SCN 2008, Caracas Venezuela.
- Vargas M. (2011). “¿Ser o no ser informal?: Una simulación estructural para Bolivia”, Fundación ARU.
- Villamil R. “La economía informal causas, consecuencias y ejes de solución” en “23º Congreso de Gestión de Personas, OIT.

UNA MIRADA A LA EDUCACIÓN EN BOLIVIA DESDE LAS PROYECCIONES DE POBLACIÓN E INDICADORES EDUCATIVOS CENSALES

Lic. Paredes Alarcón, Marisol¹

✉ marycorreo7@yahoo.es

RESUMEN

Este artículo identifica de forma descriptiva, una serie de información estadística mostrada en dos ámbitos, el primero referido a proyecciones demográficas elaboradas por la Comisión Económica para América Latina y el Caribe (CEPAL), desde la perspectiva de una mirada sobre el largo plazo, acerca del futuro comportamiento de la población en edad escolar en Bolivia, y en el segundo ámbito, algunos rasgos importantes sobre indicadores educativos en Bolivia, provenientes de los censos 1992, 2001 y 2012. Ambos brindan un punto de vista sobre la educación en Bolivia, a fin de planificar adaptaciones en el nivel primario y una reasignación de recursos a mediano y largo plazo hacia el nivel de educación secundaria y en la educación superior.

PALABRAS CLAVE

Educación; Indicadores Educativos; Proyecciones de población.

1. INTRODUCCIÓN

Una de las funciones principales, entre otras que le concierne cumplir al Estado es la educación de la población. El Estado Plurinacional de Bolivia asume este principio y lo instituye como mandato constitucional, cuando considera a la educación como su función suprema y primera responsabilidad financiera, y va más allá, incluso, cuando a su vez la define como un derecho fundamental de toda persona, es decir, de todos sus habitantes. (Capítulo I - Ley de la Educación “Avelino Siñani –Elizardo Pérez” N° 070). Y la sociedad debe ejercer tuición de manera conjunta con el Estado sobre esta función, el sistema educativo al que ella da lugar, en todos sus alcances y procesos, sobre la base de criterios de armonía y coordinación. (Art.1, numeral 3 de la Ley).

En este sentido, para una formulación

adecuada de las políticas educativas, resulta importante conocer la estructura de la población y la situación de los cambios demográficos que atravesará en el futuro. Este conocimiento permitirá prever la demanda de servicios educativos, que dependerá de las tendencias esperadas en la población en edad escolar.

2. OBJETIVO

El objetivo de este trabajo es ilustrar una serie de datos e indicadores estadísticos oficiales, desde la perspectiva de una mirada sobre el largo plazo acerca del futuro comportamiento de la población en edad escolar en Bolivia, y algunos rasgos importantes sobre indicadores censales del periodo 1992-2001.

3. METODOLOGÍA

El presente trabajo es de tipo descriptivo.

¹ Docente de la Carrera de Estadística, Universidad Mayor de San Andrés.

La metodología empleada incluyó la revisión documental y recopilación de datos e indicadores estadísticos oficiales, provenientes por una parte de las proyecciones demográficas elaboradas por la Comisión Económica para América Latina y el Caribe (CEPAL) para el Estado Plurinacional de Bolivia, revisión 2017², y por otra de los datos comparativos de los últimos tres censos de Población y Vivienda realizados en 1992, 2001 y 2012 por el Instituto Nacional de Estadística (INE). A fin de comparar aspectos relevantes respecto de indicadores seleccionados provenientes de los censos de población, se muestra la situación de género y de ámbito geográfico de cada indicador escogido, comparando las diferencias o brechas presentes en el sistema educativo.

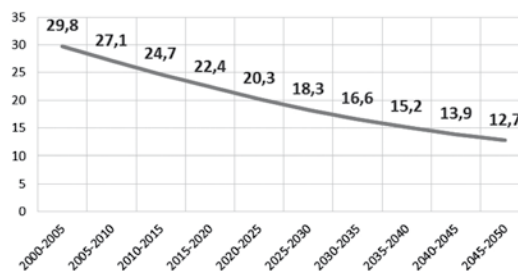
4. POBLACIÓN Y POBLACIÓN ESCOLARIZADA

Resulta fundamental, dentro del planteamiento de las políticas públicas y referido en este caso específico al ámbito educativo, efectuar una mirada en el largo plazo sobre aspectos relevantes en las tendencias y cambios demográficos de Bolivia durante los próximos 32 años, tomando en consideración las proyecciones demográficas elaboradas por la Comisión Económica para América Latina y el Caribe (CEPAL) para el Estado Plurinacional de Bolivia.

La población de Bolivia durante los últimos 25 años, se ha incrementado en alrededor de 59% hasta 2018; aunque este crecimiento

demográfico ha sufrido una desaceleración en las últimas décadas. Esto se refleja en el comportamiento de la Tasa Media de Crecimiento Anual de la población que disminuyó de 2,7% en el período 1992-2001, a 1,7% en el período 2001-2012.

Figura N° 1
Bolivia: Tasa Bruta de Natalidad (por 1000)
Período 2000-2050



Fuente: Elaboración propia con base en datos de la CEPAL Revisión 2017

De acuerdo a las proyecciones de la CEPAL, se prevé un descenso en la Tasa Bruta de Natalidad³; ésta se reduciría de 29,8 por cada 1.000 habitantes según las previsiones para el período 2000-2005 a 22,4 por cada 1.000 habitantes durante el período 2015-2020; y luego a sólo 12,7 por cada 1000 habitantes para el período 2045-2050.

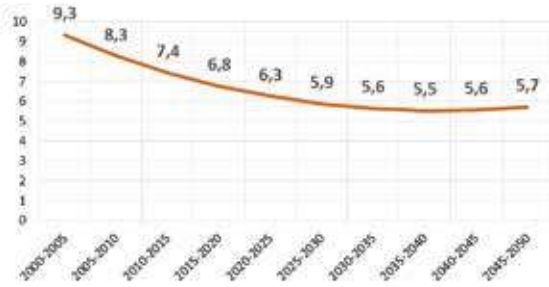
En tanto que la Tasa Bruta de Mortalidad⁴ descendería de 9,3 personas por cada 1.000 habitantes entre 2000-2005 a 6,8 por cada 1.000 habitantes durante el período 2015-2020; y disminuiría a 5,7 personas por cada 1000 habitantes en el período 2045-2050, aunque con un leve crecimiento al final de este período.

2 Estado Plurinacional de Bolivia, Estimaciones y proyecciones de población a largo plazo, 1950-2100. CELADE – División de Población de la CEPAL. Revisión 2017.

3 **Tasa Bruta de Natalidad:** Mide la frecuencia de los nacimientos ocurridos en un período en relación a la población total. Es el cociente entre el número medio anual de nacimientos ocurridos durante un período determinado y la población media del período.

4 **Tasa Bruta de Mortalidad:** Mide la frecuencia de las defunciones ocurridas en un período en relación a la población total. Es el cociente entre el número medio anual de defunciones ocurridas durante un período determinado y la población media de ese período.

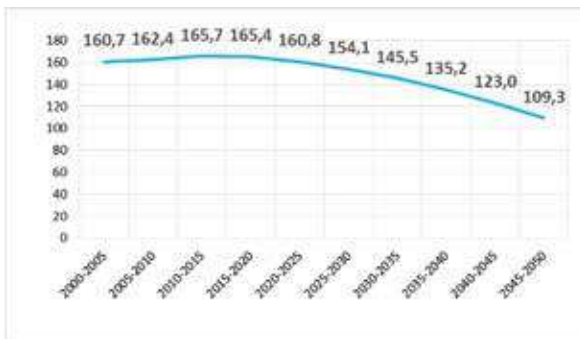
Figura N° 2
Bolivia: Tasa Bruta de Mortalidad (por 1000)
Período 2000-2050



Fuente: Elaboración propia con base en datos de la CEPAL
Revisión 2017

En suma, de acuerdo a las proyecciones de la CEPAL, el crecimiento anual de la población del país muestra una tendencia ligeramente ascendente hasta el período 2010-2015, y después empieza a decrecer de manera sostenida en el tiempo llegando a acentuarse hasta el final del periodo 2045-2050. La Figura N°3 da cuenta de esta dinámica poblacional.

Figura N° 3
Bolivia: Crecimiento anual de la población (en miles) Período 2000-2050



Fuente: Elaboración propia con base en datos de la CEPAL
Revisión 2017

Como resultado de estas tendencias, el crecimiento anual de la población⁵ en Bolivia tendrá importantes transformaciones demográficas en los siguientes 32 años, lo que tendrá implicaciones en los ámbitos

económico y social, y por ende en el sector educativo.

Tomando en consideración las proyecciones de la CEPAL para la población en edad escolar, es decir, para los niveles de educación inicial (de 4 a 5 años), primaria (6 a 12 años) y secundaria (de 13 a 17 años), aquella aumentaría ligeramente en total de 3,22 millones de estudiantes en 2015 a 3,26 millones para el año 2020, a partir de donde comienza a decrecer y de forma mucho más acentuada a partir de 2030, hasta alcanzar a 2,87 millones para el año 2050.

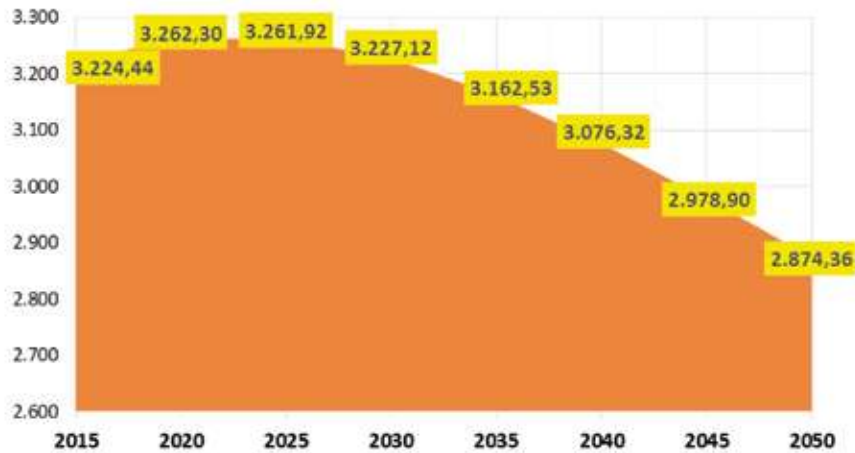
Analizando estas variaciones para los niveles inicial y primario, esto significa que entre los años 2015 y 2030, el número de niños y niñas en edad escolar, para los niveles inicial y primario, disminuirá en alrededor de 39.000 estudiantes, mientras que los estudiantes en el nivel secundario aumentarán en cerca de 42.000 estudiantes.

Este es un aspecto fundamental para la formulación de estrategias para cada uno de los niveles educativos de Educación Regular en la planificación sectorial. Ello requerirá probablemente de adaptaciones en los dos niveles mencionados y una reasignación de recursos hacia el nivel de educación secundaria. No obstante, su uso debe ser apropiadamente planificado y gestionado frente a las proyecciones posteriores a 2030.

Entre 2030 y 2050 la tendencia a la caída de la población en edad escolar se acentuaría, cayendo la población en edad escolar para los niveles inicial y primario en 250 mil niños y niñas, y a la misma vez, será también decreciente para los estudiantes en edad escolar correspondiente al nivel secundario en alrededor de 102 mil estudiantes.

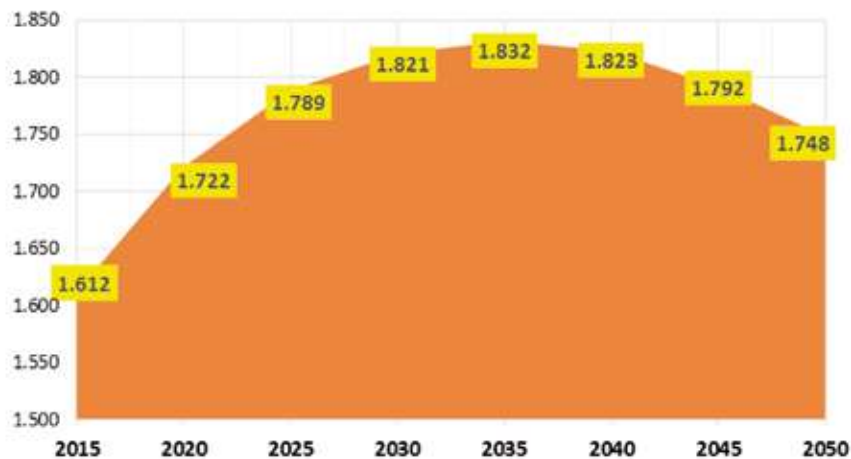
⁵ **Crecimiento Anual:** Es el incremento medio anual total de una población, vale decir el número de nacimientos menos el de defunciones, más el de inmigrantes y menos el de emigrantes, durante un determinado período.

Figura N° 4
Bolivia: Proyecciones de la población de 4 a 17 años. Período 2015 - 2050 (en miles)



Fuente: Elaboración propia con base en datos de la CEPAL Revisión 2017

Figura N° 5
Bolivia: Proyecciones de la población de 18 a 25 años. Período 2015 - 2050 (en miles)



Fuente: Elaboración propia en bases a datos de la CEPAL Revisión 2017

En tanto que, tomando en cuenta el grupo poblacional de 18 a 25 años, como una aproximación a la población en educación superior, ésta se incrementaría de manera importante de 1,61 millones de jóvenes en 2015 a 1,79 millones hasta el año 2025, situación que iría revirtiéndose hasta el año 2050, cuando alcanzaría a 1,75 millones de jóvenes. Esto implicaría un importante crecimiento potencial de la demanda en

educación superior, ya que entre los años 2015 y 2030, la población joven entre 18 y 25 años llegaría a crecer en alrededor de 208.000 personas. Este crecimiento podría relacionarse también con el incremento de la matrícula en el nivel secundario, lo que a su vez podría ejercer presión sobre las universidades públicas principalmente y en algún grado sobre los institutos técnicos y tecnológicos. Ello puede ser visto como una

oportunidad, e implicaría ciertamente costos de oportunidad ya que puede derivar en una exigencia de adaptación y establecimiento de condiciones favorables a corto y mediano plazo en la educación superior.

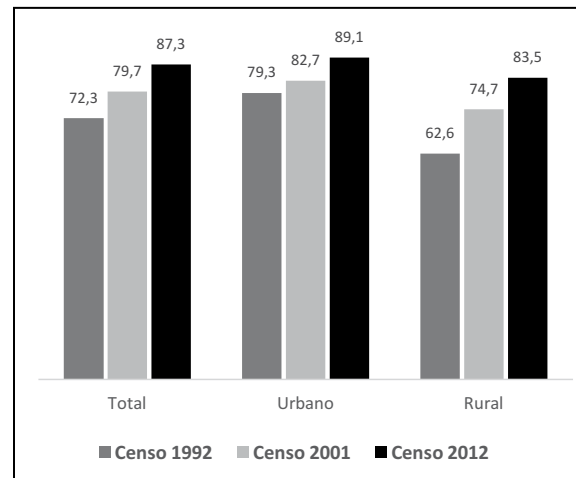
La situación presentada indica que se iniciaría un proceso de transición demográfica en Bolivia –es decir, la disminución del ritmo de crecimiento poblacional y el envejecimiento relativo de la población-; por lo que el país podría beneficiarse del denominado “bono demográfico”, concepto que hace referencia a aquella fase en la que el equilibrio entre edades representa una oportunidad para el proceso de desarrollo en diferentes ámbitos. Lo que suele ocurrir cuando cambia, en este caso, de manera favorable, la relación de dependencia entre la población en edad potencialmente productiva (jóvenes y adultos) y aquella en edad dependiente (niños y personas mayores), con un mayor peso relativo de la primera con respecto a la segunda. Este “bono demográfico” abre, a su vez, una importante oportunidad para la expansión de los niveles de educación secundaria y superior.

5. TASAS DE ASISTENCIA ESCOLAR DE LA POBLACIÓN DE 4 A 19 AÑOS

Los resultados del último Censo Nacional de Población y Vivienda (2012), proporcionan un punto de partida a partir del cual se pueden medir tanto los avances en los períodos intercensales, como los puntos de atención y futura intervención. Los esfuerzos realizados en el sector educativo durante los últimos años han mejorado en aspectos que se ven reflejados en los resultados censales a través de indicadores educativos. Al efecto, se muestra una comparación de los resultados del Censo 2012 con aquellos de los Censos de 2001 y 1992, reflejando las tasas totales y las tasas según área geográfica.

Uno de estos indicadores, es la tasa de asistencia escolar, que representa el cociente entre el número de personas entre 6 y 19 años (que considera sólo los niveles de primaria y secundaria), que asisten al Subsistema de Educación Regular y el total de la población en el mismo rango de edad.

Figura N° 6
Bolivia: Tasa de asistencia escolar de la población de 6 a 19 años por censo y área, Censos 1992, 2001 y 2012 (en porcentaje)



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican la pregunta de asistencia escolar

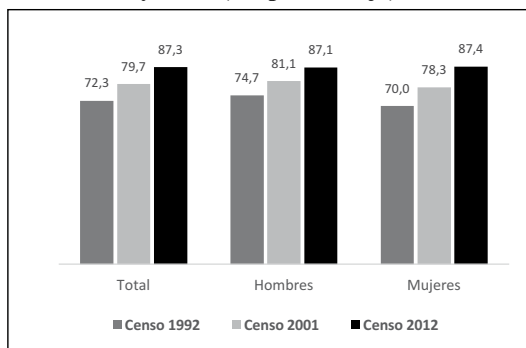
Comparativamente, la Tasa Total de Asistencia Escolar reflejada según el Censo 2012 ha mejorado en alrededor de 8 puntos porcentuales en relación a su similar del Censo 2001, y en 15 puntos porcentuales en relación a la tasa calculada según el Censo 1992. El desglose de la tasa de asistencia escolar para el grupo etario de 6 a 11 años indica que en 2001 este grupo registró una tasa de asistencia de 92,9 %, subiendo a 96,5 % en 2012. En tanto que para el grupo de edad de 12 a 19 años la tasa de asistencia escolar en 2001 llegó a 68,4%, incrementándose hasta 81% en 2012, lo que implica una mejoría de alrededor de 13 puntos porcentuales en relación a los datos del Censo 2001. Esta mejoría ha sido más acentuada en el área rural, pues la brecha que existía entre las

tasas de asistencia de lo urbano y rural se ha acortado de 16,7% del censo 1992 a 5,6% en el censo 2012.

En el período 2001-2012, en el área urbana, la Tasa de Asistencia Escolar para el grupo de edad de 6 a 11 años aumenta de 95,4% a 97,3% y para las edades de 12 a 19 años de 72,9% a 83,8%. En tanto que, en el área rural en el mismo periodo, esta tasa se incrementa de 89,2% a 95,0% para el grupo de 6 a 11 años de edad, y de 60,0% a 75,1% para el grupo de 12 a 19 años.

De acuerdo a los datos de los tres últimos censos 1992, 2001 y 2012, se advierte que los niveles de asistencia escolar entre hombres y mujeres de 6 a 19 años de edad en el último censo presentan diferencias favorables para las mujeres, en relación a los censos anteriores. La Tasa de Asistencia Escolar para los hombres en el período 1992-2012 se incrementó de 74,7% a 87,1%, esto es alrededor de 12 puntos porcentuales, pero este crecimiento durante el mismo periodo en el caso de la tasa para las mujeres sube de 70% a 87,4%, alcanzando a 17 puntos porcentuales, con el aditamento de que según los datos del Censo 2012 las tasas para ambos géneros se equiparan.

Figura N° 7
Bolivia: Tasa de asistencia escolar de la población de 6 a 19 años por censo y sexo, Censos 1992, 2001 y 2012 (En porcentaje)

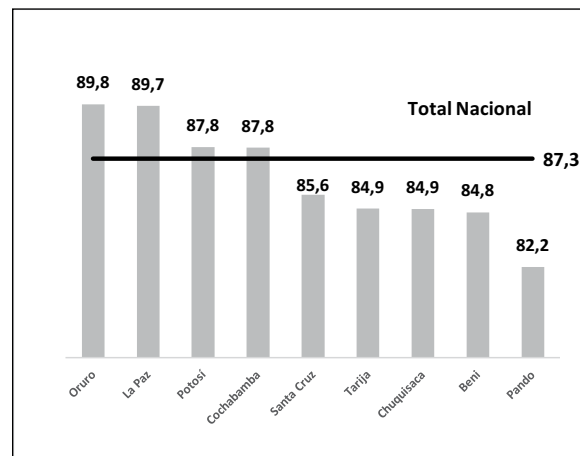


Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta.

A nivel departamental, según el último censo, Oruro es el departamento que presenta la mayor tasa de asistencia escolar, seguido por La Paz. En tanto que el departamento con la más baja tasa de asistencia escolar es Pando. Otros departamentos que están por debajo del total nacional son Santa Cruz, Tarija, Chuquisaca, Beni y Pando.

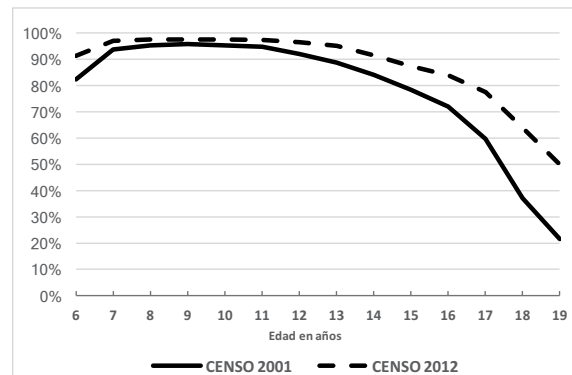
Figura N° 8
Bolivia: Tasa de asistencia escolar de la población de 6 a 19 años por departamento, Censo 2012 (En porcentaje)



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta.

Figura N° 9
Bolivia: Tasa de asistencia escolar por edades simples de la población de 6 a 19 años, Censo 2001 y 2012



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta.

De acuerdo a resultados del Censo 2012, las tasas de asistencia específicas por edad para el tramo de 6 a 19 años muestran que entre los 6 y 14 años más del 90% se encuentran dentro del sistema educativo. A partir de los 13 años ya comienza un declive gradual hasta alcanzar a cerca de 78% a los 17 años, terminando en alrededor de 50% para la población con 19 años. Sin embargo, en relación al Censo 2001, se advierte una importante mejoría en la asistencia escolar, sobre todo a partir de los 12 años, aunque queda aún mucho por hacer para alcanzar la universalización en la Educación Regular.

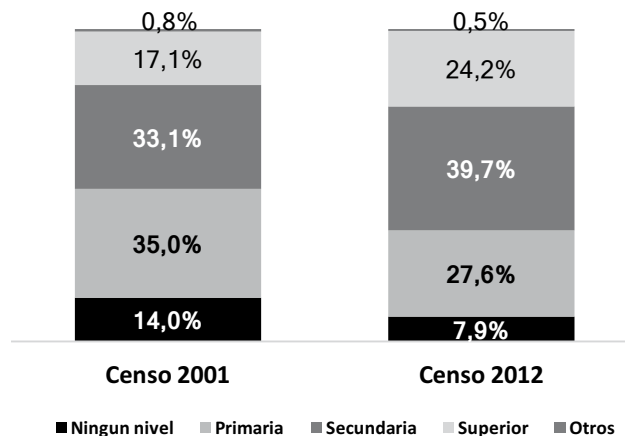
6. NIVEL DE INSTRUCCIÓN MÁS ALTO ALCANZADO

En Bolivia, la educación obligatoria consta de 14 años de estudio, hasta el bachillerato. Una

vez concluida, los bachilleres pueden acceder al mundo laboral o continuar con carreras técnicas de nivel superior u optar por el ingreso a las carreras universitarias. En cambio, aquellas personas que solamente cursaron el nivel primario o las que no alcanzaron nivel educativo algunos se encuentran menos preparadas y en desventaja para ingresar al mundo laboral, y tampoco pueden acceder a niveles educativos superiores.

De acuerdo al último Censo 2012, en Bolivia, la mayor parte de la población de 19 años o más que tiene nivel de instrucción hasta secundaria alcanza a 39,7%; le siguen segmentos que tienen nivel primario, superior y ningún nivel. Comparativamente, en el Censo 2001 la mayor parte de la población de 19 años o más contaba con instrucción hasta el nivel primario en un 35,0%; le seguían el nivel secundario, superior y ningún nivel.

Figura N° 10
Bolivia: Nivel de instrucción más alto alcanzado de la población de 19 años o más de edad, Censos 2001 y 2012 (en porcentaje)



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta.

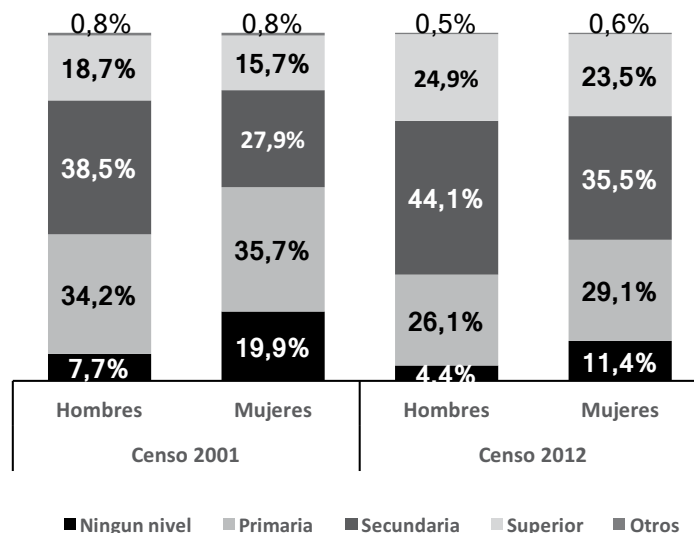
Se advierte un importante crecimiento del nivel de instrucción más alto alcanzado por la población de 19 o más años de edad en el

nivel superior, que pasa de 17,1% en 2001 a 24,2% en 2012; es decir, se incrementó en siete puntos porcentuales. En el lado opuesto,

el segmento de personas que no tienen ningún nivel de instrucción ha disminuido de 14% en 2001 a 7,9% en 2012; es decir, decreció en seis puntos porcentuales entre ambos censos. En el nivel de instrucción más alto alcanzado,

correspondiente a secundaria, se registró un aumento de 33,1% a 39,7%, en tanto que el nivel primario disminuyó de 35% a 27,6% entre ambos censos, que refleja un giro importante.

Figura N° 11
Bolivia: Distribución porcentual de la población de 19 años o más de edad por nivel de instrucción más alto alcanzado y sexo, censos 2001 y 2012



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta

Los esfuerzos por disminuir la población que no tiene ningún nivel educativo han suscitado avances interesantes ya que en el Censo 2001, 14 de cada 100 personas no tenían instrucción educativa alguna, en tanto que en 2012 sólo 8 de cada 100 personas se sitúan en esta categoría. En 2001, de la población masculina de 19 años o más, 38,5% tenía nivel de educación secundaria, 34,2% alcanzó el nivel primario, el 18,7% declaró haber aprobado algún grado del nivel superior, 7,7% no contaba con nivel de instrucción. En la población femenina los porcentajes eran de 35,7% para el nivel primario, le sigue el 27,9% con educación en el nivel secundario, 19,9% no contaba con ningún nivel de instrucción, y el 15,7% habían alcanzado el nivel superior.

En los resultados del Censo 2012, la situación se ha modificado tanto para hombres como para mujeres, empero muestra términos más favorables para el género femenino. La población masculina de 19 años o más se ha incrementado a 44,1% en el nivel secundario en relación a 2001; le sigue en magnitud el segmento correspondiente al nivel primario con 26,1%, luego el de nivel superior con 24,9% que aumentó en relación al anterior censo, y aquellas personas sin nivel de instrucción disminuyeron a 4,4%. La población femenina en relación al Censo 2001, aumentó su participación en ocho puntos porcentuales en el nivel secundario, llegando a 35,5%, habiendo cursado al menos algún grado de este nivel; le sigue aquella que logró el nivel primario con 29,1%, y después el 23,5% con algún grado aprobado del nivel

superior, en el que también se incrementó su participación en alrededor de ocho puntos porcentuales con relación a 2001, en tanto que la población femenina sin ningún nivel de instrucción alcanza a 11,4%, aunque ha disminuido en 8 puntos porcentuales con respecto al Censo 2001.

Las diferencias por género señalan características concretas en cuanto al nivel de educación alcanzado, que en términos generales determinan una situación que presentaría menores grados de marginación y segregación de las mujeres durante el último período intercensal.

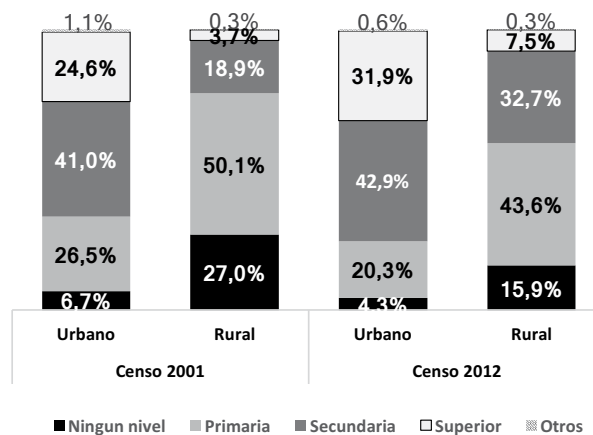
Analizando este indicador según área geográfica, se advierte que en el área rural también se ha logrado cambios favorables en las categorías del nivel de instrucción más alto alcanzado en el periodo 2001-2012.

En el año 2001, la desagregación por área muestra que la mayoría de la población de 19 o más años de edad del área urbana tiene nivel de educación secundaria en un 41%; le sigue aquella en el nivel primario con 26,5%; en el

nivel de educación superior llega a participar en un 24,6% y un reducido 6,7% de personas se muestra sin ningún nivel de instrucción. En tanto que en el área rural la mitad de la población tiene nivel de instrucción más alto alcanzado en primaria, le sigue un 27% con ningún nivel; 18,9% con nivel de educación secundaria y tan sólo el 3,7% cursó algún grado de nivel superior.

Según el Censo 2012, la población del área urbana tiene un 42,9% con nivel secundario, 31,9% cuenta con educación de nivel superior, mayor en siete puntos porcentuales respecto a la del Censo 2001; en tanto que el nivel primario alcanzó a 20,3%, y tan sólo un 4,3% se muestra sin nivel de instrucción. La población del área rural presenta un 43,6% con educación en el nivel primario; muestra un incremento de 14 puntos porcentuales para el nivel secundario, llegando a 32,7%; registra una importante disminución de la población rural sin nivel de instrucción de 11 puntos porcentuales inferior al Censo 2001, llegando a 15,9%, y finalmente un leve ascenso en la población con un nivel de educación superior que alcanza a 7,5%.

Figura N° 12
Bolivia: Distribución porcentual de la población de 19 años o más de edad por nivel de instrucción más alto alcanzado y área, censos 2001 y 2012



Fuente: Instituto Nacional de Estadística

No incluye personas que residen habitualmente en el exterior y que no especifican respuesta.

Una visión complementaria de este indicador e incluso tal vez, sustitutiva a futuro, por ser más fidedigno del nivel más alto de educación alcanzado, podría corresponder a un análisis de los años promedio cursados en cada nivel de educación, como suele realizarse en muchos países de la región y el mundo. Su ventaja resulta evidente por su mayor verosimilitud respecto del análisis practicado aquí que sólo considera los datos globales en cada nivel. En la práctica, cuando se efectúa el relevamiento de la información algunas personas pueden señalar que han cursado un nivel educativo más alto por el hecho de haber cursado únicamente un grado del mismo o sin haber concluido estudios en dicho nivel y ello puede llevar a una apreciación errónea de lo que acontece en la realidad. Esta información fue recolectada para los censos 1992 y 2001, pero la boleta censal para el censo 2012 no brinda exhaustividad en este dato.

7. CONCLUSIONES

- De los resultados mostrados, en primera instancia, sobre el comportamiento futuro en la población en edad escolar, estas proyecciones deben ser tomadas en cuenta para la planificación del sector

educativo. Esto implica que se requerirá de adaptaciones en el nivel primario y una reasignación de recursos hacia el nivel de educación secundaria, además de una exigencia por el crecimiento que obliga a planificar el establecimiento de condiciones favorables a mediano y largo plazo en la educación superior.

- Por otra parte, los indicadores del sector educativo, en la Educación Regular, durante los últimos tres censos, muestran una mejora significativa. Las diferencias por género señalan que, en términos generales, se da una situación que presentaría menores grados de marginación y segregación de las mujeres durante el último período intercensal. También las brechas por área geográfica han disminuido, ya que se advierte que en el área rural se han logrado cambios favorables en la Educación Regular en el periodo 1992-2012.
- Finalmente, se espera promover la reflexión respecto de los cambios poblacionales y avances en los indicadores educativos censales en nuestro país, para la planificación educativa.

BIBLIOGRAFÍA

CELADE- División de Población de la CEPAL, *Estado Plurinacional de Bolivia: Estimaciones y proyecciones de población a largo plazo 1950-2100. Revisión 2017.* Disponible en <http://www.cepal.org>.

Ministerio de Educación, Bolivia, *Ley N° 070 de 20 de diciembre de 2010. Ley de*

la Educación “Avelino Siñani – Elizardo Pérez”. Disponible en: <http://www.minedu.gob.bo>.

INE, Bolivia (2001). *Censo de Población y Vivienda 2001. Características de la Población.* La Paz-Bolivia.

REGRESIÓN LINEAL SIMPLE CON TÉRMINOS DE PERTURBACIÓN NO NORMALES

M.Sc. Rivero Sugiura, Fernando Oday

✉ fors2004@yahoo.com.ar

RESUMEN

Existen muchas funciones de distribución de probabilidad discreta o continua que no son consideradas en el análisis del comportamiento aleatorio de un conjunto de datos y se asume normalidad en la mayoría de los casos, aunque este conjunto de datos a veces no presente cercanía a una distribución normal. El presente artículo describe el método de regresión lineal simple considerando valores de perturbación basados en otro tipo de distribución como es el de la función secante hiperbólica generalizada (SHG), que complica su tratamiento matemático al momento en que se aplica el método de estimación por máxima verosimilitud.

PALABRAS CLAVE

Distribución secante hiperbólica generalizada (SHG), Verosimilitud modificada.

1. DISTRIBUCIÓN SECANTE HIPERBÓLICA GENERALIZADA (SHG)

La distribución secante hiperbólica debida a Vaughan en el año 2002, es una distribución de probabilidad continua, cuya función de densidad y función característica son proporcionales a la función secante

hiperbólica. Es simétrica de media cero y varianza, enfocada en este caso a la variable aleatoria no observable de perturbación. Es una distribución leptocúrtica simétrica, con pico agudo y coeficiente de curtosis entre 3 y 9, muy similar a la normal estándar y t-student.

La función de densidad, se define como:

$$f(u; t) = \frac{1}{\sigma_u} \left(\frac{ae^{bu/\sigma_u}}{e^{2bu/\sigma_u} + 2ce^{bu/\sigma_u} + 1} \right) \quad t: \text{parámetro y } -\infty < u < \infty$$

Si: $-\pi < t < 0$

$$a = \frac{\text{sen } t}{t} b; \quad b = \sqrt{\frac{\pi^2 - t^2}{3}}; \quad c = \cos t$$

Si: $t = 0$

$$a = b = \frac{\pi}{\sqrt{3}}; \quad c = 1$$

Si: $t > 0$

$$a = \frac{\text{senh } t}{t} b; \quad b = \sqrt{\frac{\pi^2 + t^2}{3}}; \quad c = \cosh t$$

2. MODELO DE REGRESIÓN LINEAL SIMPLE

la forma:

$$Y = \beta_0 + \beta_1 X + u$$

Un modelo de regresión lineal simple es de

A menudo se asume que los valores de perturbación u_i , $i = 1, \dots, N$ son independientes e idénticamente distribuidos con distribución normal, media cero y varianza σ_u^2 . Sin embargo, no siempre puede ocurrir esto, tal el ejemplo de que las respuestas sean dicotómicas, es decir, 0 o 1, entonces tiene distribución Bernoulli, y en otras oportunidades, puede darse una distribución continua sesgada.

En este artículo se considera que los valores de perturbación del modelo siguen distribución Secante Hiperbólica Generalizada (SHG), es decir, $u \sim SHG(0, \sigma_u^2; t)$.

Se toma una muestra aleatoria de tamaño n de una población de tamaño N , considerándose

el modelo de estimación

$$Y = \hat{Y} + e = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

luego

$$e = Y - \hat{Y} = Y - \hat{\beta}_0 - \hat{\beta}_1 X$$

con

$$E(e) = 0 \text{ y } V(e) = \sigma_u^2$$

3. FUNCIÓN DE VEROSIMILITUD

Sea

$$z_i = \frac{u_i}{\sigma_u} = \frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma_u} \quad i = 1, \dots, n$$

la función de verosimilitud, es

$$L = \frac{a^n}{\sigma_u^n} \prod_{i=1}^n \left(\frac{ae^{bz_i}}{e^{2bz_i} + 2ce^{bz_i} + 1} \right) = \frac{a^n}{\sigma_u^n} \prod_{i=1}^n \left[\frac{ae^{b\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma_u}\right)}}{e^{2b\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma_u}\right)} + 2ce^{b\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma_u}\right)} + 1} \right]$$

Tomando logaritmo de L , derivando respecto a β_0 , β_1 y σ_u e igualando a cero, se tiene:

$$\frac{\partial \ln L}{\partial \beta_0} = -\frac{bn}{\sigma_u} + \frac{2b}{\sigma_u} \sum_{i=1}^n g(z_i) = 0 \quad (1)$$

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{b}{\sigma_u} \sum_{i=1}^n X_i + \frac{2b}{\sigma_u} \sum_{i=1}^n X_i g(z_i) = 0 \quad (2)$$

$$\frac{\partial \ln L}{\partial \sigma_u} = -\frac{n}{\sigma_u} - \frac{b}{\sigma_u} \sum_{i=1}^n z_i + \frac{2b}{\sigma_u} \sum_{i=1}^n z_i g(z_i) = 0 \quad (3)$$

donde

$$g(z_i) = \frac{e^{2bz_i} + ce^{bz_i}}{e^{2bz_i} + 2ce^{bz_i} + 1} \quad i = 1, \dots, n \quad (4)$$

Las expresiones de (1), (2) y (3) no admiten soluciones explícitas dado a los términos relacionados con la función no lineal de (4). Para establecer soluciones en las ecuaciones anteriores, se emplean

expresiones de verosimilitud modificadas mediante el uso de los dos primeros términos de la serie de Taylor alrededor de $t_{(i)}$ dado por Tiku y Suresh.

$$g[z_{(i)}] \cong g[t_{(i)}] + g'[t_{(i)}][z_{(i)} - t_{(i)}] = \theta_i + \alpha_i z_{(i)} \quad 1 \leq i \leq n \quad (5)$$

Donde $\theta_i = g [t_{(i)}] - \alpha_i t_{(i)}$ y $\alpha_i = g' [t_{(i)}]$ obtienen las ecuaciones de verosimilitud modificada L^* , es decir:

$$\frac{\partial \ln L}{\partial \beta_0} \cong \frac{\partial \ln L^*}{\partial \beta_0} = -\frac{bn}{\sigma_u} + \frac{2b}{\sigma_u} \sum_{i=1}^n [\theta_i + \alpha_i z_{(i)}] = 0 \quad (6)$$

$$\frac{\partial \ln L}{\partial \beta_1} \cong \frac{\partial \ln L^*}{\partial \beta_1} = -\frac{b}{\sigma_u} \sum_{i=1}^n X_{[i]} + \frac{2b}{\sigma_u} \sum_{i=1}^n X_{[i]} [\theta_i + \alpha_i z_{(i)}] = 0 \quad (7)$$

$$\frac{\partial \ln L}{\partial \sigma_u} \cong \frac{\partial \ln L^*}{\partial \sigma_u} = -\frac{n}{\sigma_u} - \frac{b}{\sigma_u} \sum_{i=1}^n z_{(i)} + \frac{2b}{\sigma_u} \sum_{i=1}^n z_{(i)} [\theta_i + \alpha_i z_{(i)}] = 0 \quad (8)$$

4. SOLUCIONES DE ESTIMACIÓN con

De (6), (7) y (8), se logran las soluciones de estimación de los parámetros

$$\hat{\sigma}_u = \left(-C + \sqrt{C^2 + \frac{4nD}{b}} \right) / \frac{2n}{b}$$

$$\hat{\beta}_0 = \bar{Y}_{[.]} - \hat{\beta}_1 \bar{X}_{[.]} \quad \hat{\beta}_1 = A - \hat{\sigma}_u B \quad y$$

$$\bar{X}_{[.]} = \frac{\sum_{i=1}^n \alpha_i X_{[i]}}{\sum_{i=1}^n \alpha_i} \quad \bar{Y}_{[.]} = \frac{\sum_{i=1}^n \alpha_i Y_{[i]}}{\sum_{i=1}^n \alpha_i}$$

$$A = \frac{\sum_{i=1}^n \alpha_i (X_{[i]} - \bar{X}_{[.]}) Y_{[i]}}{\sum_{i=1}^n \alpha_i (X_{[i]} - \bar{X}_{[.]})^2} \quad B = \frac{1/2 \sum_{i=1}^n X_{[i]} - \sum_{i=1}^n \theta_i X_{[i]}}{\sum_{i=1}^n \alpha_i (X_{[i]} - \bar{X}_{[.]})^2}$$

$$C = \sum_{i=1}^n Y_{[i]} - A \sum_{i=1}^n X_{[i]} - 2 \sum_{i=1}^n \theta_i Y_{[i]} + 2A \sum_{i=1}^n \theta_i X_{[i]}$$

$$D = 2 \left[\sum_{i=1}^n \alpha_i (Y_{[i]} - \bar{Y}_{[.]})^2 - A \sum_{i=1}^n \alpha_i (X_{[i]} - \bar{X}_{[.]}) Y_{[i]} \right]$$

BIBLIOGRAFÍA

- Bent Jorgensen (1993). The Theory of Linear Models. New York, London.
- S. R. Searle (1971). Linear Models. New York.
- G. A. F. Seber (1977). Linear Regression Analysis. New York.
- David W. Hosmer. Applied Logistic Regression. New York.

ESTADÍSTICA MULTIVARIABLE ANÁLISIS DE COMPONENTES PRINCIPALES

Dr. Cs. Ruiz Aranibar, Gustavo¹

✉ ruizaranibargustavo@gmail.com.bo

RESUMEN

El análisis de componentes principales es probablemente la transformación más comúnmente utilizada para realizar varias tareas en muchas aplicaciones, se utiliza generalmente para reducir la dimensionalidad de los datos multivariados de modo que los modelos de rasgos univariados puedan ajustarse a los componentes principales individuales. Convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables sin correlación lineal llamadas componentes principales, cuyos resultados se los debe interpretar.

PALABRAS CLAVE

Ecuación característica, matriz de correlación, valor propio, vector propio, ejes principales, mutivariante, análisis de componentes principales.

1. INTRODUCCIÓN.

En el Análisis de los Componente Principales (ACP), se dispone de datos estadísticos que se deberían representar en un espacio de varias dimensiones, lo que significa que si se tiene un número bastante grande de variables y observaciones; por ejemplo, 10 características económicas (variables) medidas en 50 regiones (observaciones), u 80 caracteres (variables) medidas en 2.000 individuos (observaciones); esta información no se la podría representar. Por lo tanto, se tiene una matriz rectangular de valores numéricos cuyas dimensiones son tales que son un obstáculo para la rápida asimilación por el estadístico de la información contenida en este conjunto de números, lo que se pretende en la presente investigación, es efectuar un estudio de datos estadísticos representándolos en un espacio pequeño.

El objetivo principal que persigue el ACP es la representación de las medidas numéricas

de varias variables en un espacio de pocas dimensiones donde nuestros sentidos puedan percibir relaciones, porque de otra manera permanecerían ocultas en dimensiones superiores. Dicha representación debe ser tal que al desechar dimensiones superiores (de cuarta dimensión en adelante) se realiza con la mínima pérdida de información.

El ACP consiste en una transformación lineal de m variables originales en m nuevas variables, donde cada nueva variable es una combinación lineal de los valores originales. El proceso se realiza de una manera que requiere que cada nueva variable explique, sucesivamente, la mayor cantidad posible de las varianzas totales. Cuando se han calculado las nuevas variables de las diferentes muestras, se tendrá en cuenta toda la variante original.

El ACP es una técnica estadística que fue propuesta por Karl Pearson (1901) como parte del análisis de factores y Hotelling

¹ Se agradece a la UAGRM por la beca otorgada con fondos del IDH, para cursar y culminar exitosamente el Doctorado en Ciencias en Educación Superior. Especializado en Estadística e Informática.

(1933) quienes introdujeron el concepto de ACP. Sin embargo, la complejidad de los cálculos retrasó su desarrollo y aplicación hasta la aparición de los computadores.

2. DESCRIPCIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES.

El ACP es una técnica estadística descriptiva, que tiene como punto de partida una matriz de datos con una serie de individuos, a los que se les ha medido varias variables. Por eso se la clasifica como una técnica multivariante. Dada una matriz de n filas (muestras) y m columnas (variables; $n \geq m$), se puede efectuar un gráfico considerando dos variables para todas las muestras, según sus valores en el eje de abscisas y de ordenadas, también se puede hacer una representación de tres de esas variables en un gráfico tridimensional, pero si el número de variables es mayor a 3, ya no se puede visualizar una representación a m dimensiones, siendo imposible ver en un gráfico una representación respecto a todas las variables al mismo tiempo.

El ACP tiene como objetivo básico inicial suplir este déficit, realizando la representación de una nube de puntos multidimensional (de más de tres dimensiones), en dos o tres dimensiones. En definitiva, se trata de visualizar lo que no se ve. El ACP trata de hacer una representación de las n muestras en dos o tres dimensiones, pero contemplando todas las variables, sin prescindir de ninguna de ellas en el análisis.

Como consecuencia de su propio procedimiento, se consigue crear expresiones matemático-estadísticos (los componentes) en realidad, también podrían considerarse objetivos de la técnica porque en muchas ocasiones permiten establecer relaciones

entre las variables, ver cómo se asocian, relacionan, distancian, etc.

3. CONCEPTOS PRELIMINARES.

Se entenderá la esencia de la técnica del ACP de la manera siguiente: si se representa en forma bidimensional dos variables x_1 y x_2 , en un sistema de coordenadas rectangulares X_1 y X_2 , luego sin mover los datos, se traza desde el origen un eje Y_1 que se ajuste mejor a los datos donde se proyecta cada punto a este eje, luego un segundo eje Y_2 desde el origen y perpendicular a Y_1 , lo que se realizó es una rotación con un ángulo θ , del nuevo sistema de coordenadas, la relación existente entre las variables originales y las nuevas variables son expresadas como una combinación lineal de las variables originales, a esta técnica se la denomina como componentes, observando la expresión de esta combinación de esta forma:

$$y_1 = a_{11}x_1 + a_{12}x_2$$
$$y_2 = a_{21}x_1 + a_{22}x_2$$

Estos coeficientes que multiplican a las variables originales son los vectores propios de la matriz de correlaciones, es la expresión de la transformación lineal realizada. Ahora si en lugar de estar trabajando con dos variables originales, se trabaja con m variables originales la expresión de las m componentes sería:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m$$

.....

$$y_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mm}x_m$$

Para mayor comprensión, se tienen tres puntos A, B y C (muestras) y dos variables X_1 y X_2 , los cuales pueden representarse en un sistema de coordenadas rectangulares (variables), obteniendo luego los promedios \bar{X}_1 y \bar{X}_2 , para determinar:

$$x_1 = X_1 - \bar{X}_1$$

$$x_2 = X_2 - \bar{X}_2$$

Se rotan con un ángulo Θ , los ejes con coordenadas \bar{X}_1 y \bar{X}_2 , como nuevo origen, proyectándose los valores de los tres puntos sobre Y_1 , de tal forma que la variación a lo largo del nuevo eje Y_1 es grande, pero la variación a lo largo del eje Y_2 es pequeña, porque a lo largo del eje Y_2 es igual a cero. En este caso, el fenómeno indicado por las dos variables (variables explicativas) X_1 y X_2 , puede ser explicado por la nueva variable Y_1 , cuando se emplea el nuevo sistema de coordenadas Y_1 y Y_2 .

Además, si se puede estadísticamente hacer independientes a las variables Y_1 y Y_2 , entonces los efectos de Y_1 y Y_2 pueden ser diferenciados pudiendo utilizarse esas variables para ulteriores investigaciones.

Conociendo las coordenadas de tres puntos A, B y C, en el nuevo sistema de coordenadas, considerando el ángulo θ , se las determina efectuando operaciones geométricas, algebraicas y trigonométricas, de manera que se encuentra para estos puntos:

Primeras coordenadas en Y_1 :

$$y_{11} = x_{11} \cos\theta + x_{12} \text{sen}\theta$$

$$y_{21} = x_{21} \cos\theta + x_{22} \text{sen}\theta$$

$$y_{31} = x_{31} \cos\theta + x_{32} \text{sen}\theta$$

Segundas coordenadas en Y_2 :

$$y_{12} = -x_{11} \text{sen}\theta + x_{12} \cos\theta$$

$$y_{22} = -x_{21} \text{sen}\theta + x_{22} \cos\theta$$

$$y_{32} = -x_{31} \text{sen}\theta + x_{32} \cos\theta$$

Una vez determinada estas nuevas coordenadas de los tres puntos en X_1 y X_2 , se tendrá en el nuevo sistema de coordenadas Y_1 y Y_2 ; trasladando los resultados a notación matricial:

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = [X_1, X_2]$$

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} \cos\theta \\ \text{sen}\theta \end{bmatrix} = Y_1$$

$$\begin{bmatrix} y_{12} \\ y_{22} \\ y_{32} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} -\text{sen}\theta \\ \cos\theta \end{bmatrix} = Y_2$$

$$Y = [Y_1, Y_2] = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{bmatrix}$$

Este cambio de coordenadas expresada en forma matricial para los tres puntos, permite

generalizar el análisis para m variables, el cual es utilizado en el programa computacional.

4. FUNDAMENTOS DEL ACP.

La técnica del ACP construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos, en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para esta transformación lineal debe construirse primero la matriz de varianza-covarianza o la matriz de coeficientes de correlación. Debido a la simetría de estas matrices, existe una base completa de vectores propios de las mismas. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensionalidad de datos. Además, las coordenadas en la nueva base dan la composición en factores subyacentes de los datos iniciales.

Una de las ventajas del ACP para reducir la dimensionalidad de un grupo de datos, es que retiene aquellas características del conjunto de datos que contribuyen más a su varianza, manteniendo un orden de bajo nivel de los componentes principales e ignorando los de alto nivel. El objetivo es que esos componentes de bajo orden a veces contienen el aspecto más importante de esa información.

5. CÁLCULOS Y FÓRMULAS MATEMÁTICO – ESTADÍSTICOS EN EL ACP.

Por definición, un conjunto de datos constituye una muestra aleatoria multivariada, si cada individuo ha sido extraído al azar de una población de n individuos y en él se han medido u observado m características se tiene:

Matriz de datos multivariados:

$$X = [x_{ij}] = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}$$

$$i = 1, 2, 3, \dots, n \quad j = 1, 2, 3, \dots, m \quad n \geq m \quad (\text{Si } n < m \text{ se trabaja con la } X'_{ij})$$

A partir de esta matriz, que contiene toda la información estadística de la muestra, es posible calcular las funciones que permitan extraer conclusiones de los datos.

Matriz de suma de cuadrados y productos.

$$C = [c_{ij}] = \sum_{i=1}^n x_{i1} * x_{ij}$$

Suma de columnas.

$$X = x_j = \sum_{i=1}^n x_{ij}$$

$$A = [a_{ii}] = SCE_i = \sum_{\alpha=1}^n (x_{i\alpha} - \bar{x}_i)^2 = \sum_{\alpha=1}^n x_{i\alpha}^2 - \frac{1}{n} \left(\sum_{\alpha=1}^n x_{i\alpha} \right)^2$$

$$A = [a_{ij}] = SPE_{ij} = \sum_{\alpha=1}^n (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = \sum_{\alpha=1}^n x_{i\alpha} x_{j\alpha} - \frac{1}{n} \left(\sum_{\alpha=1}^n x_{i\alpha} \right) \left(\sum_{\alpha=1}^n x_{j\alpha} \right)$$

Matriz de suma de cuadrados y suma de productos de desviaciones.

Matriz de varianza – covarianza o matriz de covarianza.

$$S = [s_{ij}] = \begin{bmatrix} s_{11} & \dots & s_{1m} \\ \dots & \dots & \dots \\ s_{m1} & \dots & s_{mm} \end{bmatrix} = \begin{bmatrix} a_{11}/n & \dots & a_{1m}/n \\ \dots & \dots & \dots \\ a_{m1}/n & \dots & s_{mm}/n \end{bmatrix}$$

Vector de promedios.

$$\bar{X} = [\bar{x}_j] = \left(\sum_{i=1}^n x_{ij} / n \right) = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_m \end{bmatrix}$$

Vector de desviaciones estándar.

$$D = [d_j] = \left[\sqrt{a_{jj}} \right]$$

$$r = \frac{\text{cov}(x,y)}{s_x s_y} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} * \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad -1 \leq r \leq +1$$

6. VECTORES Y VALORES PROPIOS.

El ACP se basa en una transformación lineal de las observaciones originales, esta transformación lineal debe satisfacer las exigencias del análisis que se tiene en el campo de álgebra vectorial como generación de vectores y valores propios.

Se procede a encontrar de la matriz de correlación, las raíces características λ (escalar) donde I es una matriz unitaria:

$$R - \lambda I = \begin{bmatrix} 1 - \lambda & r_{12} & \dots & r_{1m} \\ r_{21} & 1 - \lambda & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 - \lambda \end{bmatrix}$$

Matriz estandarizada.

$$X = x_{ij} = \frac{\sum_{j=1}^n (x_{ji} - \bar{x}_j)}{d_j}$$

Matriz de coeficientes de correlación.

$$R = [r_{ij}] = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}$$

Para dos columnas cualesquiera (identificadas por los subíndices), el coeficiente de correlación se determina por:

El determinante de esta matriz es:

$$|R - \lambda I| = \begin{vmatrix} r_{11} - \lambda & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} - \lambda & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mm} - \lambda \end{vmatrix} = 0$$

Igualando a cero este determinante, y desarrollando permite encontrar un polinomio de grado m en λ , que se designa por $f(\lambda)$, denominado polinomio característico de la matriz; considerando el teorema fundamental del algebra lineal, que indica que todo polinomio de orden m posee m raíces. El polinomio característico de la matriz R de orden m puede escribirse de una manera general como:

$$f(\lambda) = c_m \lambda^m + c_{m-1} \lambda^{m-1} + \dots + c_1 \lambda + c_0$$

$$c_m = (-1)^m \quad c_0 = |A| \quad c_{m-1} = (-1)^{m-1} \text{tr}(A)$$

Donde los coeficientes c_j dependen de los elementos de la matriz R.

La ecuación: $|R - \lambda I| = 0$ como origina un polinomio con respecto a la incógnita λ , el cual puede ser resuelto con respecto a λ , y obtener, entonces los vectores característicos.

Considerando los casos a), b) y c):

a) Teniéndose una matriz R de orden 2, se obtendrá:

$$(r_{11} - \lambda)(r_{22} - \lambda) - r_{12}r_{21} = 0$$

Es decir:

$$\lambda^2 - (r_{11} + r_{22})\lambda + (r_{11}r_{22} - r_{12}r_{21}) = 0$$

Cuyas raíces son:

$$\lambda_1 = \frac{1}{2} \left[(r_{11} + r_{22}) + \sqrt{(r_{11} + r_{22})^2 - 4(r_{11}r_{22} - r_{12}r_{21})} \right]$$

$$\lambda_2 = \frac{1}{2} \left[(r_{11} + r_{22}) - \sqrt{(r_{11} + r_{22})^2 - 4(r_{11}r_{22} - r_{12}r_{21})} \right]$$

En el caso particular de una matriz simétrica, $r_{12} = r_{21}$, las raíces valen:

$$f(\lambda) = \lambda^3 - (r_{11} + r_{22} + r_{33})\lambda^2 + \left(\begin{vmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{vmatrix} + \begin{vmatrix} r_{11} & r_{13} \\ r_{31} & r_{33} \end{vmatrix} + \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} \right) \lambda - \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$$

$$f(\lambda) = \lambda^3 - (\text{tr } R)\lambda^2 + (R_{11} + R_{22} + R_{33})\lambda - |R|$$

$$\lambda = \frac{1}{2} \left[(r_{11} + r_{22}) \pm \sqrt{(r_{11} + r_{22})^2 - 4r_{12}^2} \right]$$

Y dado que el radicando es la suma de dos cuadrados, las raíces λ_1 y λ_2 son necesariamente reales para una matriz simétrica.

O también si:

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

Entonces se tiene:

$$f(\lambda) = \lambda^2 - (r_{11} + r_{22})\lambda + \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} = \lambda^2 - (\text{tr } R)\lambda + |R|$$

Donde $\text{tr } R$ denota la traza de R, esto es, la suma de sus elementos diagonales.

b) Teniéndose una matriz R de orden 3, se obtendrá:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Entonces:

Aquí R_{11} , R_{22} , R_{33} denotan, los cofactores de los elementos diagonales r_{11} , r_{22} , r_{33} .

Considerando de nuevo la matriz cuadrada R de orden 3, $R=[r_{ij}]$, como se indicó anteriormente,

$$S_1 = tr R \quad S_2 = R_{11} + R_{22} + R_{33} \quad S_3 = |R|$$

Son los coeficientes de su polinomio característico, con signos alternantes. Por otra parte, cada S_k es la suma de los menores principales de orden k de R .

c) En el caso general, si R es una matriz cuadrada de orden m , su polinomio característico es:

$$f(\lambda) = \lambda^m - S_1 \lambda^{m-1} + S_2 \lambda^{m-2} - \dots + (-1)^m S_m$$

Donde S_k es la suma de los menores principales de orden k .

Sabiendo que la traza de una matriz cuadrada es igual a la suma de sus valores característicos y el determinante de una matriz es igual al producto de sus valores característicos. Para calcular el determinante de una matriz, se emplea la expansión del determinante por los cofactores.

De una manera general, para la matriz R de orden m , se tendrá:

$$|R| = \sum_{j=1}^m r_{ij} R_{ij} \quad |R| = \sum_{i=1}^m r_{ij} R_{ij}$$

La primera expresión corresponde a la expansión del determinante por los cofactores de los elementos de la fila i ; la segunda corresponde a la expansión del determinante por los cofactores de los elementos de la columna j .

De manera que:

$f(\lambda)=0$ Ecuación característica de la matriz, que admite m soluciones:

$\lambda_1, \lambda_2, \dots, \lambda_m$ no todas necesariamente diferentes.

Obteniendo las m soluciones que constituyen los valores característicos o valores propios, cada valor propio podrá generar un vector propio con una dimensión igual a la de la matriz original, y en función de estos valores característicos, ordenados de mayor a menor, se encuentran los m vectores característicos, los cuales multiplicados por la matriz original proporcionan los componentes principales.

La técnica del ACP, transforma a las variables originales en nuevas variables (componentes), las cuales tiene desigualdad en cuanto a la información explicada, lo que significa que se tiene componentes muy informativos y otros no. Por eso se tiene unos componentes principales, que son los que se usan para hacer la representación. Esta desigualdad generada al crear los componentes permite elegir, entre ellos, los principales y eliminar los poco importantes, lo que no sucede con los originales porque todos ellos eran principales, todos eran importantes, no prescindiendo de ninguno de ellos.

Lo importante es que se tiene la nube de puntos desde unos ejes, donde uno es mucho más importante que el otro, ahora los ejes son Y_1 y Y_2 . Si x_1 y x_2 eran dos variables que tenían la misma cantidad de información, ahora y_1 y y_2 no tienen la misma cantidad de información, y_1 tiene mucha más información que y_2 . (en estadística información es equivalente a dispersión). Una variable que no varía no tiene información, una variable

que varía mucho, es muy informativa.

El objetivo de la técnica ACP es conseguir girar los ejes de tal forma que exista la mayor desigualdad posible entre la varianza de la nube de puntos originales en las proyecciones en cada uno de los respectivos nuevos ejes, además, estos ejes, de estas nuevas variables, sean independientes entre sí, o sea, que tengan correlación cero.

La búsqueda de estos nuevos ejes se hace mediante el cálculo de los llamados valores propios y vectores propios de la matriz de correlaciones entre todas las variables del estudio. Puede hacerse también a partir de la matriz de varianzas-covarianzas, pero ésta tiene el problema de que cuando las variables tienen unidades de escala muy diferentes introduce un exceso de influencia por parte de las variables con mayor varianza. Por esto, suele trabajarse con la matriz de correlaciones; de esta forma se unifica el peso de las variables originales. Suele hablarse de variables estandarizadas cuando se trabaja con la matriz de correlaciones. Una variable es estandarizada cuando la muestra se transforma de manera que tiene un promedio cero y la varianza o la desviación estándar tiene por valor la unidad. Esto se consigue restando a cada valor de los valores de la columna considerada de la muestra el promedio y dividiendo por la desviación estándar de dicha columna. De esta forma todas las variables del estudio tienen la misma media y la misma desviación estándar y ninguna pesa más que otra.

7. METODOS DEL ACP.

Existen dos formas básicas de aplicar el ACP:

1. Método basado en la matriz de correlación, cuando la información no

es homogénea o las medidas empleadas para las diferentes variables no son las mismas, o la magnitud de las variables aleatorias es diferente y se las estandariza previamente,

2. Método basado en la matriz de covarianzas, que se usa cuando los datos son dimensionalmente homogéneos y presentan valores medios similares, este método parte de la determinación de la matriz de varianzas-covarianza sin estandarizar la información.

Posteriormente en cada uno de estos métodos se procede a efectuar los diferentes cálculos en forma similar.

Las conclusiones que se sacará con estos cálculos efectuados, es el de interpretar los resultados de la matriz de los coeficientes de correlación, de la matriz de los ejes principales, observando los valores propios generados y habrá que ser muy cuidadoso al interpretar los vectores propios, ya que el método no es independiente de la escala de medición de las variables originales; para mayor confiabilidad de la interpretación, será el profesional entendido en los problemas de su especialidad, donde se tomará en cuenta las variables consideradas en los resultados de las matrices quien interpretará y sacará conclusiones en forma correcta.

8. TRABAJO COMPUTACIONAL.

Siendo el número de cálculos a efectuar por lo general numerosos, especialmente cuando el número de muestras n y el número de variables m , son grandes este se vuelve muy complicado, razón por la cual, se ha desarrollado un programa computacional

(12), para los dos métodos mencionados, que facilita enormemente este trabajo, cuyos resultados se ilustrará para su comprensión con aplicaciones en algunas áreas como ser:

Las variables X_j , designan las asignaturas, identificadas como:

X_1 = Cálculo I

X_2 = Inglés

X_3 = Metodología de la Investigación

X_4 = Física

X_5 = Álgebra lineal

X_6 = Química

X_7 = Geología

X_8 = Geometría descriptiva.

Aplicación 1: Análisis de calificaciones por asignatura en una universidad. En las universidades, el progreso en las asignaturas realizado por los estudiantes se evalúa con una calificación de 0 a 100, y a partir de una calificación mayor o igual a 51 el estudiante se considera aprobado y por debajo de este valor es reprobado. Se presenta en esta aplicación del ACP, las notas finales de 8 asignaturas de 20 estudiantes de un determinado curso de acuerdo al pensum de una carrera y se busca la forma de sintetizar la información contenida en estas asignaturas a fin de obtener relaciones entre materias, como una medida sintética del aprovechamiento estudiantil que pueda ser utilizada con fines descriptivos.

La información disponible de acuerdo a lo especificado, es la siguiente:

N° est.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	50,00	50,00	50,00	50,00	50,00	50,00	50,00	50,00
2	73,00	39,00	32,00	83,00	54,00	65,00	33,00	82,00
3	83,00	49,00	42,00	93,00	64,00	75,00	43,00	92,00
4	49,00	79,00	75,00	64,00	49,00	65,00	69,00	55,00
5	68,00	24,00	45,00	84,00	72,00	67,00	28,00	64,00
6	85,00	89,00	95,00	79,00	83,00	75,00	92,00	42,00
7	38,00	92,00	84,00	43,00	32,00	43,00	74,00	52,00
8	59,00	39,00	42,00	64,00	22,00	51,00	32,00	76,00
9	45,00	75,00	76,00	32,00	31,00	22,00	79,00	29,00
10	52,00	51,00	46,00	49,00	62,00	51,00	42,00	12
11	72,00	44,00	55,00	76,00	82,00	81,00	44,00	65,00
12	76,00	83,00	78,00	72,00	69,00	63,00	75,00	91,00
13	86,00	93,00	88,00	82,00	79,00	73,00	85,00	96,00
14	66,00	73,00	68,00	62,00	59,00	53,00	65,00	76,00
15	43,00	31,00	33,00	42,00	35,00	19,00	12,00	39,00
16	76,00	42,00	44,00	75,00	82,00	71,00	45,00	55,00
17	28,00	48,00	49,00	21,00	30,00	29,00	49,00	53,00
18	47,00	59,00	58,00	49,00	52,00	51,00	63,00	62,00
19	60,00	82,00	84,00	60,00	60,00	60,00	70,00	65,00
20	62,00	73,00	67,00	70,00	65,00	68,00	76,00	77,00

Estadística multivariante
Análisis de componentes principales

Cálculos. Cada uno de los cálculos efectuados por el programa de ACP son los siguientes:

Totales	1218,00	1215,00	1216,00	1250,00	1132,00	1132,00	1126,00	1253,00
----------------	---------	---------	---------	---------	---------	---------	---------	---------

Matriz de suma de cuadrados y productos.

	1	2	3	4	5	6	7	8
1	79396	74686	75038	81664	73726	73505	69671	79485
2	74686	82737	81589	75601	69708	69631	76758	76881
3	75038	81589	81462	76043	70451	70015	76066	76297
4	81664	75601	76043	85060	75773	76454	70349	82392
5	73726	69708	70451	75773	70884	69298	65521	72455
6	73505	69631	70015	76454	69298	69960	65189	74200
7	69671	76758	76066	70349	65521	65189	72178	71301
8	79485	76881	76297	82392	72455	74200	71301	87409

Matriz de suma de cuadrados y productos de las desviaciones.

	1	2	3	4	5	6	7	8
1	5219,797	692,500	983,602	5539,000	4787,203	4566,203	1097,602	3177,297
2	692,500	8925,750	7717,000	-336,500	939,000	862,000	8353,500	761,250
3	983,602	7717,000	7529,203	43,000	1625,399	1189,399	7605,203	114,602
4	5539,000	-336,500	43,000	6935,000	5023,000	5504,000	-26,000	4079,500
5	4787,203	939,000	1625,399	5023,000	6812,801	5226,801	1789,399	1535,203
6	4566,203	862,000	1189,399	5704,000	5226,801	5888,801	1457,399	3280,203
7	1097,602	8353,500	7605,203	-26,000	1789,399	1457,399	8784,199	757,102
8	3177,297	761,250	114,602	4079,500	1535,203	3280,203	757,102	8908,547

Matriz de varianza-covarianza.

	1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000
1	260,990	34,625	49,180	276,950	239,360	228,310	54,880	158,865
2	34,625	446,283	385,850	-16,825	46,950	43,100	417,675	38,063
3	49,180	385,850	376,460	2,150	81,270	59,470	380,260	5,730
4	276,950	-16,825	2,150	346,750	251,150	285,200	-1,300	203,975
5	239,360	46,950	81,270	251,150	340,640	261,340	89,470	76,760
6	228,310	43,100	59,470	285,200	261,340	294,440	72,870	164,010
7	54,880	417,675	380,260	-1,300	89,470	72,870	439,210	37,855
8	158,865	38,063	5,730	203,975	76,760	164,010	37,855	445,427

Promedios.

	60,9	60,75	60,8	62,5	56,6	56,6	56,3	62,65
--	------	-------	------	------	------	------	------	-------

Desviaciones estándar.

	16,16	21,13	19,4	18,62	18,46	17,16	20,96	21,11
--	-------	-------	------	-------	-------	-------	-------	-------

Datos estandarizados.

	1	2	3	4	5	6	7	8
1	-0,6576	-0,4960	-0,5425	-0,6543	-0,3485	-0,3749	-0,2930	-0,5842
2	0,7300	-1,0035	-1,4468	1,0730	-0,1373	0,4771	-1,0836	0,8936
3	1,3333	-0,5421	-0,9444	1,5964	0,3908	1,0452	-0,6186	1,3554
4	-0,7180	0,8420	0,7133	0,0785	-0,4014	0,4771	0,5907	-0,3533
5	0,4284	-1,6956	-0,7937	1,1254	0,8133	0,5907	-1,3162	0,0624
6	1,4540	1,3034	1,9692	0,8637	1,3942	1,0452	1,6603	-0,9537
7	-1,3816	1,4418	1,1654	-1,0207	-1,2991	-0,7725	0,8232	-0,4918
8	-0,1146	-1,0035	-0,9444	0,0785	-1,8272	0,3181	-1,1301	0,6165
9	-0,9593	0,6575	0,7636	-1,5964	-1,3519	-1,9657	1,0557	-1,5540
10	-0,5370	-0,4498	-0,7435	-0,7435	-0,7066	0,2852	-0,6651	-2,3391
11	0,6697	-0,7728	-0,2914	0,7066	1,3414	1,3860	-0,5721	0,1085
12	0,9110	1,0266	0,8640	0,4973	0,6548	0,3635	0,8697	1,3093
13	1,5143	1,4879	1,3664	1,0207	1,1829	0,9316	1,3345	1,5402
14	0,3077	0,5652	0,3617	-0,0262	0,1267	-0,2045	0,4046	0,6165
15	-1,0800	-1,3726	-1,3965	-1,0730	-1,1407	-2,1358	-2,0603	-1,0922
16	0,9110	-0,8651	-0,8439	0,6543	1,3414	0,8180	-0,5255	-0,3533
17	-1,9849	-0,5883	-0,5928	-2,1722	-1,4047	-1,5677	-0,3395	0,4780
18	-0,8386	-0,0807	-0,1407	-0,7066	-0,2429	-0,3181	0,3116	-0,0300
19	-0,0664	0,9804	1,1654	0,13086	0,1796	0,1931	0,6372	0,1085
20	0,0664	0,5652	0,3115	0,3926	0,4436	0,6475	0,9162	0,6627

Matriz de coeficientes de correlación.

	1	2	3	4	5	6	7	8
1	1,000	0,101	0,157	0,921	0,803	0,824	0,162	0,466
2	0,101	1,000	0,941	-0,043	0,120	0,119	0,943	0,085
3	0,157	0,941	1,000	0,006	0,227	0,179	0,935	0,014
4	0,921	-0,043	0,006	1,000	0,731	0,893	-0,003	0,519
5	0,803	0,120	0,227	0,731	1,000	0,825	0,231	0,197
6	0,824	0,119	0,179	0,893	0,825	1,000	0,203	0,453
7	0,162	0,943	0,935	-0,003	0,231	0,203	1,000	0,086
8	0,466	0,085	0,014	0,519	0,197	0,453	0,086	1,000

Valores propios y porcentajes de contribución.

	Valores Propios	% valores Propios	% acumulados valores propios
	1	2	3
1	3,92797	49,09965	49,09965
2	2,73479	34,18494	83,28458
3	0,84449	10,55614	93,84074
4	0,20224	2,52805	96,36879
5	0,16320	2,03995	98,40874
6	0,06228	0,77849	99,18772
7	0,04355	0,54434	99,73157
8	0,02147	0,26843	100,00000

Estadística multivariable
Análisis de componentes principales

Ejes principales o componentes principales.

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7	Vector 8
1	0,45492	-0,17748	-0,06237	-0,52625	-0,45226	0,22511	0,10870	-0,46225
2	0,20206	0,54049	0,10107	-0,14360	0,06183	0,00310	-0,78849	-0,10323
3	0,22742	0,52566	-0,04902	-0,10187	0,01887	-0,67339	0,44251	-0,09776
4	0,43221	-0,27286	0,01655	-0,43593	0,28339	-0,13126	-0,06062	0,66874
5	0,42389	-0,11097	-0,43495	0,56042	-0,45145	-0,16135	-0,18740	0,19949
6	0,45811	-0,15840	-0,09232	0,30228	0,68519	0,10718	0,05185	-0,42600
7	0,23529	0,52168	0,02670	0,12451	-0,02348	0,65547	0,35847	0,31239
8	0,26225	-0,12669	0,88589	0,28474	-0,19268	-0,10712	0,02582	0,00547

Resultados de los componentes principales.

Columnas = Vectores propios

Filas = Observaciones

	1	2	3	4	5	6	7	8
1	134,808	37,071	19,578	3,215	-3,530	-4,113	-2,458	4,842
2	166,178	-7,165	43,230	-6,091	-2,894	-4,788	-6,489	8,447
3	193,139	0,249	47,145	-5,447	-3,600	-5,611	-6,981	9,415
4	164,179	69,258	25,564	-1,310	12,474	-9,233	-7,314	8,602
5	166,910	-10,466	17,312	1,879	-4,202	-18,885	4,960	12,783
6	215,742	85,610	-3,262	-9,537	-4,999	-7,847	0,924	0,152
7	137,877	97,057	33,676	-5,559	5,533	-10,989	-9,744	4,948
8	132,084	11,768	53,167	-15,457	4,275	-8,530	2,323	0,470
9	116,153	94,382	13,863	-14,026	-11,576	1,016	1,845	7,076
10	128,275	34,577	-19,456	-1,926	-1,954	-3,837	-10,784	2,798
11	186,267	11,966	14,132	11,484	-2,343	-13,339	-0,852	3,576
12	199,825	62,688	47,808	1,716	-14,643	-9,578	-7,461	7,678
13	225,475	70,736	47,294	0,935	-14,385	-9,864	-8,082	8,619
14	171,552	55,907	39,463	-0,351	-12,973	-8,219	-7,099	6,683
15	88,075	9,436	17,422	-10,794	-15,584	-17,882	-7,977	4,634
16	177,780	8,038	6,294	5,477	-9,715	-4,251	-4,065	6,858
17	99,332	49,392	51,303	16,692	-11,709	-6,735	0,938	0,692
18	144,161	51,827	30,281	9,581	-4,584	-2,996	-1,633	7,949
19	175,337	73,575	29,237	0,923	-3,621	-15,009	185,962	4,345
20	185,228	56,479	37,079	7,920	-1,798	-1,754	-4,839	12,228

Los valores propios de los componentes indican la cantidad de varianza, la cantidad de información que tiene cada componente. Como se puede observar en la tabla de valores propios y porcentaje de contribución del primer componente tiene un 49.099 %

de información y el segundo un 34.184 %. Los dos juntos tienen un 83.284 %. Por lo tanto, haciendo una representación en dos dimensiones con esas dos primeras componentes se pierde un 16.716% de información únicamente.

En este problema se considera lo siguiente:

- 1) Observar en los resultados de los ejes principales o componentes, el valor absoluto de los coeficientes distinguiendo los que tienen un valor grande y un valor pequeño. En este caso en el primer componente se observa que en las asignaturas X_1 , X_4 , X_5 y X_6 , tienen coeficientes con valor absoluto grande, cercano en todos los casos a 0.5, los demás son más pequeños, pesan mucho menos en este componente. En el segundo componente el peso principal se lo tiene en las asignaturas X_2 , X_3 , y X_7 , con coeficientes cercanos a 0.54. Las demás asignaturas pesan poco.

- 2) Observar entre los coeficientes con valor absoluto grande el juego de signos que hay. En este problema el signo es el mismo, por lo tanto, las variables que pesan en un componente y el otro todos van en la misma dirección. Pero en otro caso se podría encontrar con valores de signos contrarios; por lo tanto, se tiene que interpretar el juego de fuerzas de los signos.

La interpretación es más clara, si se analiza de esta manera, en el primer componente se tiene reunidas las asignaturas de ciencias exactas, donde existe mayor correlación entre ellas, en el segundo componente está relacionado con las asignaturas de humanidades, la asignatura de geometría descriptiva no pesa ni en uno ni en el otro, porque no tiene ninguna relación ni con las materias de ciencias exactas ni con las de letras. Así parece que existe un conjunto de habilidades cognitivas relacionadas con las ciencias exactas y un segundo relacionado con las de humanidades, estos dos conjuntos de habilidades

son estadísticamente independientes. Interpretación realizada por un docente universitario.

Aplicación 2: Análisis comercial. Venta de calzados por una empresa. En la venta de calzados se utilizó 8 muestras y se determinó 5 factores, que afectan las ventas de zapatos en una ciudad, excluyendo el precio de venta (variable dependiente), siendo estos:

Nº	X_1	X_2	X_3	X_4	X_5	Y
	1	2	3	4	5	
1	4,00	5,00	1,00	8,00	105,00	20,00
2	5,00	5,00	2,00	10,00	104,00	24,00
3	5,00	7,00	2,00	11,00	101,00	24,00
4	10,00	7,00	2,00	12,00	95,00	24,00
5	15,00	8,00	5,00	14,00	100,00	32,00
6	16,00	9,00	6,00	15,00	103,00	38,00
7	18,00	9,00	7,00	17,00	110,00	45,00
8	24,00	12,00	7,00	19,00	117,00	55,00

Donde:

X_1 = Número de salidas o ventas de pares del producto

X_2 = Número de vendedores

X_3 = Gastos de propaganda en unidades monetarias (en miles de \$us, Bs, etc.)

X_4 = Ingresos en u. m. (en miles de \$us, Bs, etc.)

X_5 = Índice de las condiciones generales del negocio

Y = Ventas de calzados en u. m. (en cientos de miles de \$us, Bs, etc.)

Cálculos. La determinación de los diferentes cálculos efectuados por el programa computacional de ACP son:

Estadística multivariable
Análisis de componentes principales

Suma de las columnas:

	97	62	32	106	835
--	----	----	----	-----	-----

Matriz de suma de cuadrados y productos:

	1	2	3	4	5
1	1547,00	864,00	509,00	1469,00	10331,00
2	864,00	518,00	284,00	879,00	6538,00
3	509,00	284,00	172,00	406,00	3412,00
4	1469,00	879,00	486,00	1500,00	11169,00
5	10331,00	6538,00	3412,00	11169,00	87465,00

Matriz de suma de cuadrados y productos de las desviaciones:

	1	2	3	4	5
1	370,875	112,250	121,000	183,750	206,625
2	112,250	37,500	36,000	57,500	66,750
3	121,000	36,000	44,000	62,000	72,000
4	183,750	57,500	62,000	95,500	105,250
5	206,625	66,750	72,000	105,250	311,875

Matriz de varianza – covarianza:

	1	2	3	4	5
1	46,359	14,031	15,125	22,969	25,828
2	14,031	4,688	4,500	7,188	8,344
3	15,125	4,500	5,500	7,750	9,000
4	22,969	7,188	7,750	11,938	13,156
5	25,828	8,344	9,000	13,156	38,984

Promedios

	12,125	7750,000	4,000	13,250	104,375
--	--------	----------	-------	--------	---------

Desviaciones estándar

	6,809	2,165	2,345	3,455	6,244
--	-------	-------	-------	-------	-------

Datos estandarizados

	1	2	3	4	5
1	-1,11624	-1,18814	-1,19659	-1,42137	0,09364
2	-0,97886	-1,18814	-0,79772	-0,87989	-0,05618
3	-0,97886	-0,32404	-0,79772	-0,60916	-0,50563
4	-0,29194	-0,32404	-0,79772	-0,33842	-1,40453
5	0,39498	0,10801	0,39886	0,20305	-0,65545
6	0,53236	0,54006	0,79772	0,47379	-0,20600
7	0,80713	0,54006	1,19659	1,01526	0,84272
8	1,63143	1,83621	1,19659	1,55674	1,89143

Matriz de suma de cuadrados y productos:

	1	2	3	4	5
1	7,000	6,663	6,630	6,835	4,253
2	6,663	7,000	6,204	6,726	4,321
3	6,630	6,204	7,000	6,695	4,302
4	6,835	6,726	6,695	7,000	4,269
5	4,253	4,321	4,302	4,269	7,000

Matriz de suma de cuadrados y productos de las desviaciones

	1	2	3	4	5
1	7,000	6,663	6,630	6,835	4,253
2	6,663	7,000	6,204	6,726	4,321
3	6,630	6,204	7,000	6,695	4,302
4	6,835	6,726	6,695	7,000	4,269
5	4,253	4,321	4,302	4,269	7,000

Matriz de varianza – covarianza

	1	2	3	4	5
1	0,875	0,833	0,829	0,854	0,532
2	0,833	0,875	0,775	0,841	0,540
3	0,829	0,775	0,875	0,837	0,538
4	0,854	0,841	0,837	0,875	0,534
5	0,532	0,540	0,538	0,534	0,875

Matriz de coeficientes de correlación:

	1	2	3	4	5
1	1,00000	0,95182	0,94721	0,97636	0,60755
2	0,95182	1,00000	0,88626	0,96084	0,61723
3	0,94721	0,88626	1,00000	0,95645	0,61463
4	0,97636	0,96084	0,95645	1,00000	0,60986
5	0,60755	0,61723	0,61463	0,60986	1,00000

Valores propios y porcentajes de contribución.

	Valores Propios	% valores Propios	% acumulados valores propios
	1	2	3
1	4,29505	85,90107	85,90107
2	0,54577	10,91538	96,81644
3	0,11385	2,27697	99,09334
4	0,02927	0,58536	99,67878
5	0,01606	0,32122	100,00000

Matriz de los ejes principales

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5
1	0,47222	-0,19470	-0,01947	-0,85434	-0,09396
2	0,46463	-0,15160	-0,68869	0,35150	-0,40410
3	0,46340	-0,15316	0,72466	0,31527	-0,37050
4	0,47439	-0,19587	-0,01401	0,21503	0,83055
5	0,34834	0,93664	0,00005	-0,02401	0,02817

Resultado de los componentes principales:

Columnas = Vectores propios = Componentes
Filas = Observaciones = Variables

	1	2	3	4	5
1	45,04660	95,08986	-2,90340	-2,13899	6,83499
2	46,58266	93,41362	-2,22628	-2,22239	8,00346
3	46,94128	90,10464	-3,61781	-1,23153	7,94130
4	47,68672	83,31547	-3,72946	-5,14334	8,13306
5	54,59313	86,02236	-2,36925	-7,80611	7,94957
6	57,51281	88,13694	-2,36660	-7,84987	7,99605
7	62,30783	93,75909	-1,70854	-8,97969	9,29589
8	69,92223	98,30081	-3,91906	-12,78763	9,37809

Como se puede observar en la tabla de valores propios y porcentaje de contribución del primer componente tiene un 85.901 % de información y el segundo un 10.915 %. Los dos juntos tienen un 96.816 %. Por lo tanto, haciendo una representación en dos dimensiones con esos dos primeros componentes se pierde un 3.184 % de información únicamente.

En la interpretación de estos componentes se observa que:

- 1) En los resultados de los ejes principales o componentes, el valor absoluto de los coeficientes distinguiendo los que tienen un valor grande y un valor pequeño. En este caso en el primer componente

$$Z_1 = 0.47222Y_1 + 0.46463Y_2 + 0.46340Y_3 + 0.47439Y_4 + 0.34834Y_5$$

se observa que en las variables X_1 , X_2 , X_3 y X_4 , tienen coeficientes con valor absoluto más grande, cercano en todos los casos a 0.465, el restante es más pequeño, pesan mucho menos en este componente. En el segundo componente el peso principal se lo tiene en la variable X_5 , con coeficiente den 93.967.

- 2) Entre los coeficientes con valor absoluto grande el juego de signos que se obtuvo, en este caso el signo es el mismo, por lo tanto, las variables que pesan en un componente y el otro todos van en la misma dirección. En otros casos se podría encontrar con valores de signos contrarios; por lo tanto, se tiene que interpretar el juego de fuerzas de los signos.

La interpretación es más clara, si se analiza de esta manera, en el primer componente se tiene reunidas las variables de salidas, vendedores, gastos de propaganda e ingresos, donde existe mayor correlación entre ellas, el segundo componente está relacionado con el índice de las condiciones generales del negocio que no tiene mucha correlación con las otras variables.

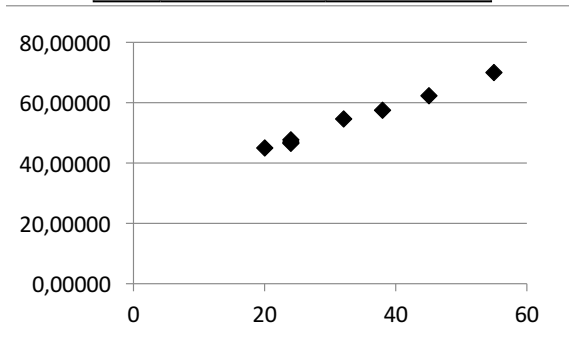
- 3) La función del primer componente principal que contribuye con un 85.901 % de la variación es:

$$Z_i = \sum_{j=1}^n V_j Y_j \quad i = 1, \dots, m$$

Donde V = vector i de la matriz de ejes principales.

Sustituyendo los valores de Y_i se encuentra:

	Y	Z
1	20	45,04660
2	24	46,58266
3	24	46,94128
4	24	47,68672
5	32	54,59313
6	38	57,51281
7	45	62,30783
8	55	69,92223



Efectuando un gráfico con los valores de Y y Z, se observa la relación que existe entre Z_i y Y. Como puede observarse esta relación tiende a ser lineal en consecuencia, se puede ajustar una línea de regresión por el método de los mínimos cuadrados, de acuerdo a su diagrama de dispersión se ajustará a una recta, a una parábola, cúbica, etc., y determinar su coeficiente de correlación, en el caso de ajustarse a una recta, su ecuación sería:

$$Y = 30.1538 + 0.7228Z$$

Los coeficientes de correlación y de determinación son:

$$r = 0.997 \quad r^2 = 0.994$$

Que muestran que existe una buena correlación entre estas variables. Para una mejor interpretación de estos resultados se aconseja a un profesional del área de administración de empresas, economía o ingeniero comercial.

CONCLUSIONES.

- Mediante el ACP se facilita la comprensión de la información multivariante, reduciendo el número de variables en todos los casos considerados, manteniendo la máxima cantidad de información, presentada en una tabla de datos de variables cuantitativas. Además, la interpretación de las fórmulas, resultados y aplicaciones se facilita, cuando se tiene un conocimiento acerca de: estadística, álgebra lineal y álgebra matricial.
- Cabe señalar que en un gráfico de dos dimensiones se observa una buena fotografía de las posiciones relativas de los puntos, en cambio en la representación de ocho dimensiones originales no se puede ver. Se denomina fotografía porque la metáfora es apropiada, puesto que cuando se observa una fotografía en realidad se está viendo una representación bidimensional de una realidad que es tridimensional. En el ACP se está realizando algo similar, se observa una fotografía bidimensional, para que se pueda observar una realidad constituida por varias dimensiones pero que no son visualizadas; por lo tanto, se está observando una fotografía bidimensional de una realidad de ocho dimensiones.
- El objetivo principal que persigue el ACP es la representación de las medidas numéricas de varias variables en un espacio de pocas dimensiones donde nuestros sentidos puedan percibir relaciones que de otra manera permanecerían ocultas en dimensiones superiores. Dicha representación debe ser tal que al desechar dimensiones superiores (generalmente de la tercera o cuarta en adelante) la pérdida de información sea mínima.

- La mejor interpretación de los resultados, la dará el profesional especializado en cada uno de los temas abordados en los diferentes problemas.

COLABORACIÓN

- Prof. M. Nilda Avilés de Ruiz
Lic. en Idiomas. Universidad Autónoma Gabriel René Moreno
Santa Cruz – Bolivia (agosto, 2005)
- M. Cs. Astrid Keitel Ruiz Avilés

Lic. en Adm. Emp. Universidad Católica Boliviana, Santa Cruz – Bolivia (octubre, 1999)

M. Cs. Academia Diplomática Boliviana "Rafael Bustillo", La Paz – Bolivia (abril, 2004)

- Dr. Osman Miranda Lira
Universidad Mayor de San Andrés, La Paz – Bolivia (Julio, 1991).
Especialidad: Radiología e Imagenología. (agosto, 1993)

BIBLIOGRAFÍA

BALESTRA Pietro, Calcul Matricial pour Économistes. Éditions Castella, Lausanne, Suisse, 1972, pp. 241 – X.

BERTIER P., BOUROCHE J. M., Analyse des Données Multidimensionnelles. Presses Universitaires de France, Vendôme, France, 1975, pp. 270 – XI.

DAGNELIE Pierre, Analyse Statistique à Plusieurs Variables. Les Presses Agronomiques de Gembloux, Gembloux, Bélgica, 1975 (2da. Edición), pp. 362 – XIV.

DAVIS C. John, Statistics and Data Analysis in Geology. John Wiley & Sons, New York, Estados Unidos, 1973, pp. 550 – VII.

GANTMACHER F. R., Théorie des Matrices (Tomo I y II), Dunod, Paris, France, 1966, pp. 638 – XV.

JOHNSON A. Richard, Probabilidad y Estadística para Ingenieros. Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (5ta. edición), pp. 630 – XV.

JOHNSTON J., Métodos de Econometría. Editorial Vicens- Vives, Barcelona, España (3ra. edición), pp. 464 – XIII.

JORESKOG J.K., KLOVANJ. E. REYMENT R. A., Geological Factor Analysis. Elsevier Scientific Publishing Company. Amsterdam, Holanda, 1976, pp 178 – VII.

KRUMBEIN W. C., GRAYBILL Franklin A., An Introduction to Statistical Models in Geology. Editorial McGraw-Hill, New York, Estados Unidos, 1965, pp. 475 – XV.

LIPSCHUTZ Seymour, Álgebra Lineal. Editorial McGraw-Hill, Madrid, España, 1998, pp. 553 – XV.

RUIZ Aranibar Gustavo. Análisis de las Componentes Principales. U.T.O. – F.C.E.F. Revista del I. I. E., N° 26-27. Oruro – Bolivia. 1988, pp. 261 – XVI.

RUIZ Aranibar Gustavo. Librería Científica de Programas Informáticos, La Paz -Bolivia.

13. YAMANE Taro, Estadística. Ed. Harla, México, D, F, México, 1979 (3ra. Edición), pp. 771 – XXIV.

WAYNE W. Daniel, Bioestadística. Ed. Limusa Wiley, México, D, F, México, 2004 (4ta. Edición), pp. 755 – XIV.

PENSAMIENTO.



No es lo que comes, sino lo que digieres, lo que te hace fuerte y saludable.

No es lo que ganas, sino lo que ahorras, lo que te hace rico y solvente.

No es lo que aprendes, sino lo que recuerdas lo que te hace sabio y juicioso.

No es lo que investigas, sino lo que creas, desarrollas y das a conocer a tus semejantes, lo que te hace científico.

Gustavo Ruiz Aranibar

DESCOMPOSICIÓN DEL TEST DE FRIEDMAN APLICADO A PRUEBAS DE DEGUSTACIÓN

Lic. Valdez Blanco, Dindo

✉ dindoalvarez@hotmail.com

RESUMEN

Anderson (1959) analizó las tablas de contingencia de doble entrada con los totales marginales constantes desarrollando la descomposición del test χ^2 de Pearson en componentes ortogonales y Rayner (1989) aplicó estas tablas y la prueba χ^2 a las pruebas de degustación de sabor de distintas variedades de un determinado producto. La descomposición del test χ^2 se puede aplicar como una prueba de homogeneidad entre filas (variedades), tal descomposición hace del test más informativo y exhaustivo en su aplicación a este tipo de tablas donde el primer componente de tal descomposición coincide con el estadístico Q_F del test de Friedman, y los subsecuentes componentes son extensiones a esta prueba.

PALABRAS CLAVE

Test Chi cuadrado, Test de Friedman, Escalas de Likert, Descomposición ortogonal

1. DESCOMPOSICIÓN ORTOGONAL DEL TEST CHI CUADRADO DE PEARSON

Primero, consideremos una tabla de contingencia de doble entrada de observaciones N_{ij} ; $i = 1, 2, \dots, r$, $j = 1, 2, \dots, c$, donde los totales de las filas $n_{i\cdot}$ y los totales de las columnas $n_{\cdot j}$ son valores conocidos (constantes), la clasificación de las columnas está ordenada y no el de las filas. Para docimar la hipótesis nula de homogeneidad entre las filas el estadístico χ^2 se define como,

$$\chi^2 = \sum_i \sum_j \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

con $E_{ij} = n_{i\cdot} \cdot n_{\cdot j} / n$, siendo n el total de datos.

Sin embargo, es posible hacer más potente a esta prueba descomponiéndola en $c-1$ componentes ortogonales para analizar efectos de escala, asimetría, curtosis, y

momentos de mayor orden entre las distintas variedades (filas) de la tabla. Para definir tales componentes primero se otorga una puntuación a las categorías de las columnas, luego se definen polinomios ortogonales denotados por $h_r(j)$, los cuales son ortogonales respecto de p_j (la probabilidad marginal).

Tal que:

$$\sum h_r(j) \cdot h_s(j) \cdot p_j = \begin{cases} 0, & \text{si } r = s \\ 1, & \text{si } r \neq s \end{cases}$$

Emerson¹ (1968) definió estos polinomios como:

$$h_r(j) = (jA_r + B_r)h_{r-1}(j) - C_r h_{r-2}(j), \quad (2)$$

$$j, r = 1, 2, \dots, c < n$$

Donde A_r, B_r, C_r son constantes para un r dado tal que $h_r(j)$ sea un polinomio de grado j , y los polinomios iniciales son: $h_{-1}(j)=0$ y $h_0(j)=1$. Tomando las probabilidades estimadas

¹ P.L. Emerson (1968) ha realizado varios trabajos comparando su método con el de Gram Schmidt, llegando a la conclusión que su método es más rápido y preciso.

$p_j = n_{.j}/n$, proporcionales a las columnas marginales, se define al s-avo componente de la partición como:

$$V_{si} = \frac{1}{\sqrt{n_{i\bullet}}} \sum_{j=1}^c N_{ij} \cdot h_s(j), \text{ para } s = 1, \dots, c-1 \text{ e } i = 1, \dots, r$$

De esta manera el estadístico χ^2 resulta particionado en c-1 componentes llamados Q_s tal que:

$$\chi^2 = \sum_{s=1}^{c-1} \sum_{i=1}^r V_{si}^2 = \sum_{s=1}^{c-1} Q_s$$

Los componentes V_{si} satisfacen la condición lineal de ortogonalidad,

$$V_{s1} \sqrt{n_{1\bullet}} + V_{s2} \sqrt{n_{2\bullet}} + \dots + V_{sr} \sqrt{n_{r\bullet}} = 0, \quad s = 1, \dots, c-1$$

Bajo la hipótesis nula de homogeneidad entre las filas, cada componente V_{si} tiene distribución asintótica normal estándar, en consecuencia los componentes $Q_s = \sum_i V_{si}^2$ de la partición tienen distribución asintótica χ^2_{r-1} .

Tales componentes son llamados estadísticos score, y tienen similares propiedades a los estadísticos obtenidos por el método de la razón de verosimilitud (Rayner, 1989, sección 3.4).

2. APLICACIÓN AL TEST DE FRIEDMAN

Supongamos que $r > 2$ tratamientos son rankeados (clasificados) cada uno por n personas (jueces), entonces es posible realizar un cuadro de contingencia de doble entrada con r filas (tratamientos) y r columnas (rangos) como se muestra en la tabla

Cuadro N° 1
Cuadro de contingencia de calificación de tratamientos según rango

Tratamiento	Rango				$n_{i\bullet}$
	1	2	...	r	
T1	N_{11}	N_{12}	...	N_{1r}	n
T2	N_{21}	N_{22}	...	N_{2r}	n
⋮	⋮	⋮	⋮	⋮	⋮
Tr	N_{r1}	N_{r2}	...	N_{rr}	n
$n_{\bullet j}$	n	n	...	n	nXR

Donde:

N_{ij} es el número de veces que el tratamiento i recibe el rango j

r es el número de muestras aleatorias

n es el tamaño de las muestras

Lo anterior queda claro ya que cada tratamiento y cada rango es asignado n veces.

Aplicando la definición, el estadístico χ^2 de Pearson para este tipo de tablas resulta:

$$\chi^2 = \frac{r}{n} \sum_{i=1}^r \sum_{j=1}^r (N_{ij} - \frac{n}{r})^2 = \sum_{s=1}^{r-1} \sum_{i=1}^r V_{si}^2 = \sum_{s=1}^{r-1} Q_s$$

Donde los componentes están dados por

$$V_{si} = \frac{\sqrt{r-1}}{r} \left(\sum_{j=1}^r h_s(j) \frac{N_{ij} - n/r}{\sqrt{n/r}} \right),$$

$$s = 1, \dots, r-1; i = 1, \dots, r. \quad (3)$$

El primer componente de esta descomposición resulta ser el estadístico del test de Friedman, tal que

$$Q_1 = \sum_{i=1}^r V_{1i}^2 = \frac{12}{rn(r+1)} \sum_{i=1}^r \left(R_i - \frac{n(r+1)}{2} \right)^2 = Q_F$$

Donde R_i es la suma de los rangos para el

i-ésimo tratamiento $R_i = \sum_{j=1}^r N_{ij}$.

3. INTERPRETACIÓN DE LOS COMPONENTES

Ya que los componentes V_{si} envuelven momentos de orden s , los componentes Q_s detectan los efectos del s -avo momento bajo la hipótesis nula y por esta razón se puede considerar al test de Friedman como un test de posición ya que envuelve polinomios de primer grado en su cálculo, de esta manera los subsecuentes componentes Q_s para $s = 2, \dots, r-1$, se consideran las extensiones al test de Friedman, puesto que estos miden adicionalmente los efectos en la escala, asimetría, curtosis, y momentos de mayor orden respectivamente.

4. APLICACIÓN A UNA PRUEBA DE DEGUSTACIÓN

Para mostrar la utilidad de tal descomposición se realizó un estudio en un grupo de personas universitarias entre los 18 y 25 años de edad con el objeto de medir la preferencia sobre cinco sabores de yogurt bebible (coco, frutilla, uva, limón y chicle) donde cada persona clasificó a cada sabor del 1° al 5° en el orden de preferencia llegando a realizar 95 observaciones (secuencias de preferencia) de manera que cada sabor recibió un total de 95 valoraciones (entre 1s, 2s, etc.) Es así que con el conjunto de observaciones se elaboró una tabla de contingencia 5X5, donde las filas corresponden a los sabores (tratamientos) y las columnas a los rangos (bloques).

La hipótesis a docimar es que cada variedad de yogurt tiene el mismo nivel de aceptación (homogeneidad), al aplicar la prueba χ^2 de homogeneidad entre filas se obtuvo un valor de 87,47 con 16 grados de libertad

y una significancia de 7.2E-12; por tanto podemos concluir que la distribución en las preferencias de los 5 sabores de yogurt no es homogénea al 5% de significación.

Cuadro N° 2
Frecuencias observadas de escalas de preferencia por sabor de yogurt

Tratamiento (sabor)	Escalas de Preferencia					Total
	1	2	3	4	5	
Coco	12	18	20	19	26	95
Frutilla	8	17	24	22	24	95
Uva	13	16	18	21	27	95
Limón	15	31	21	16	13	95
Chicle	47	13	12	17	6	95
Total	95	95	95	95	95	475

Donde la escala 1 es la menos gustada y 5 es la mayor en preferencia

En cambio si utilizamos los datos brutos y aplicamos el test de Friedman para probar la hipótesis que los 5 sabores tienen similares resultados se obtiene $Q_F=41.364$ con una significancia igual a 0. Consecuentemente rechazamos la hipótesis nula y concluimos que los cinco sabores no produjeron idénticos resultados respecto a sus rangos asignados.

Puesto que existen diferencias entre los tratamientos, es importante analizar la influencia de los componentes ortogonales del test χ^2 para aclarar donde radican estas diferencias exactamente, para lo cual calculamos los cuatro componentes. Aplicando la ecuación 2, los polinomios ortogonales de Emerson resultan:

Cuadro N° 3
Polinomios ortogonales de Emerson basados en los resultados del Cuadro N° 2

J	1	2	3	4	5
$h_1(j)$	1,00	1,00	1,00	1,00	1,00
$h_2(j)$	-1,41	-0,70	0,00	0,70	1,41
$h_3(j)$	1,19	-0,59	-1,19	-0,59	1,19
$h_4(j)$	-0,70	1,41	0,00	-1,41	0,70
$h_5(j)$	0,26	-1,06	1,60	-1,06	0,26

El lector puede comprobar que tales polinomios son orto normales con $p_j=0.2$.

El valor esperado de la variable N_{ij} es 19 en todos los casos.

Efectos de posición y escala. Basados en los anteriores resultados, la media o efecto lineal para el i -ésimo sabor está relacionado con los elementos V_{1i} quedando definida como:

$$V_{1i} = \frac{2}{5} \sum_{j=1}^5 h_1(j) \frac{N_{ij} - 19}{\sqrt{19}}, \quad i = 1, \dots, 5$$

De igual manera la variación llamada también efecto de dispersión o escala queda definido por:

$$V_{2i} = \frac{2}{5} \sum_{j=1}^5 h_2(j) \frac{N_{ij} - 19}{\sqrt{19}}, \quad i = 1, \dots, 5$$

Los subsecuentes componentes son calculados aplicando la ecuación 3 y se presentan en el cuadro N° 4.

Cuadro N° 4
Componentes de la partición del estadístico χ^2 basados en los resultados del Cuadro N° 3

	Lineal	Cuadrático	Asimetría	Curtosis
Variedad	V_{1i}	V_{2i}	V_{3i}	V_{4i}
Coco	1,88	-0,05	0,77	0,24
Frutilla	2,40	-1,26	0,38	0,49
Uva	2,14	0,38	0,25	0,00
Limón	-1,36	-1,91	1,75	-0,85
Chicle	-5,06	2,85	-3,17	0,12

El cuadro N° 5 tiene un estilo parecido al clásico análisis de varianza, basado en los componentes ortogonales del test χ^2 de Pearson. La tabla muestra que la suma de cuadrados de los efectos de ubicación es el estadístico Q_F del test de Friedman para datos ordenados por rangos.

Cuadro N° 5
Análisis de significación de los primeros componentes de la descomposición del Test χ^2

Fuente	Grados de Libertad	Suma de Cuadrados	Significancia
Posición Q^1	4,00	41,36	0,00
Escala Q^2	4,00	13,55	0,01
Residuo Q^3+Q^4	8,00	15,05	0,06
Total χ^2	16,00	69,96	

5. CONCLUSIONES

- El componente Q_2 se considera como un nuevo estadístico, el cuál mide los efectos de escala.
- La prueba extendida de Friedman, confirma una que existen diferencias en cuánto a la posición entre las distintas variedades de yogurt también muestra algo revelador, la existencia de diferencias significativas en la escala, tal diferencia no podría haber sido revelado si no se utilizaba la descomposición del test chi cuadrado. El residuo mide los efectos de asimetría y curtosis simultáneamente y este no es significativo al 5%. En estos estudios de comparación, tales datos usualmente son analizados mediante el tradicional análisis de varianza, pero de una manera distinta al análisis de rangos, ya que esta prueba puede verse afectado por valores atípicos que llevarían a tener una heterogeneidad en la varianza y por la asunción de normalidad de los datos.
- El análisis realizado es más robusto que el ANOVA, en el sentido que ninguno de los dos supuestos: homogeneidad en la varianza y normalidad de los datos, han sido asumidos. También este estudio ha demostrado ser más exhaustivo que la simple aplicación del test de Friedman.

BIBLIOGRAFÍA

- P. Anderson, Revista Electrónica *Biometrics*, septiembre de 1968.
- J.C. Rayner, *Smooth Tests of Goodness of Fit*, Oxford University Press, New Cork, 1989
- Dindo Valdez, *Pruebas Suavizadas de Bondad de Ajuste con Aplicación en Pruebas No Paramétricas*, trabajo de tesis, UMSA, La Paz, 2001
- Emma Mancilla, *Teoría Asintótica con Aplicación a Cuadros de Contingencia*, trabajo de tesis, UMSA, La Paz, 1998
- Alan Agresti, *Categorical Data Análisis*, Wiley Press, New Cork, 1990

EL SOPORTE DE LA ECONOMÍA DEL CONOCIMIENTO: “INVESTIGACIÓN + DESARROLLO + INNOVACIÓN (I+D+I)”

PhD¹ Vargas Salazar, Fernando Rodrigo

✉ cipacohc@gmail.com

RESUMEN

Los datos estadísticos que se proponen en este artículo de las diversas fuentes examinadas dejan ver que la clave para salir de la pobreza es incrementar la economía de un país y elevar el nivel de calidad de vida de su gente, que los políticos brinden la mejor tecnología, ciencia y educación a su nación, no solo se trata de incrementar las exportaciones o atraer inversiones extranjeras o crear un mercado más grande o abrir nuevos mercados. Se trata -como dice el Premio Nobel de economía Joseph Stiglitz, PhD- de insertar a los profesionales cualificados a la economía global y a la economía del conocimiento.

PALABRAS CLAVE:

Economía del conocimiento, PIB educativo, I+D+I

1. INTRODUCCIÓN

Según las investigaciones emprendidas en el año 2008 por el Banco Mundial (BM) los bajos niveles de Investigación + Desarrollo + Innovación (I+D+I) en Latinoamérica se deben a diversas causales en toda la región (*excepto Chile que jurídicamente es el más estable*), entre ellas tenemos que no se ofrece suficiente seguridad jurídica, ni estabilidad económica, hay demasiada burocracia en la apertura de empresas que desmotiva el espíritu emprendedor, no existe capital de riesgo para financiar proyectos innovadores, la forma como funciona el sistema legal es poco confiable, la carga impositiva o tributaria es otra razón, luego **es evidente la falta de preparación educativa de elevada calidad en Latinoamérica.**

En razón a ello, se propone enfatizar cuatro

aspectos económicos relacionados con los bajos niveles de I+D+I.

2. PIB EDUCATIVO

El Producto Interno Bruto (PIB) es la suma de todos los bienes y servicios de una nación y sirve para medir el progreso económico de un país o para indicar si se tuvo un mal año o no, es calculado a través de una fórmula; no obstante ¿se hace lo mismo para medir con una fórmula matemática el progreso educativo de un país?, ¿qué tiene que ver el PIB con la educación?.

Pues bien, la literatura sostiene que hay economías latinas que poseen un PIB elevado al igual que su densidad poblacional, empero, sus universidades se hallan en los últimos puestos en el ranking mundial de universidades comparativamente con sus

1 *Docente de preuniversitario, pregrado y postgrado; investigador independiente; alma mater Universidad Mayor de San Andrés en La Paz, Bolivia; Doctor en Educación Superior por la UMSA/Universidad de Bremen ALEMANIA.*
2 México es considerado la 16^{ma} economía más grande del mundo y la 2^{da} en Latinoamérica.

pares asiáticos, norteamericanos, de Oceanía y europeos. Por ejemplo, datos a 2016 refieren que **México**² posee 127 millones de habitantes con un PIB de 1.047 billones de USD\$ y su mejor universidad está por debajo de la mejor universidad de **Singapur**³ el cual es un país pequeñísimo de solo cinco millones de habitantes, pero cuenta con un sorprendente PIB de 296 mil millones de USD\$, **Chile** posee 17 millones de habitantes con un PIB de 247 mil millones de USD\$ y su mejor universidad está por debajo de la mejor universidad de **Israel** el cual es un país de solo ocho millones de habitantes y de igual modo cuenta con un extraordinario PIB de 317 mil millones de USD\$, finalmente **Bolivia** posee casi 11 millones de habitantes con un funesto PIB de 33 mil millones de USD\$ y ninguna de sus universidades se hallan en el ranking⁴ de las 500 mejores del planeta.

Esto lleva a inferir que “la preparación educativa desde las guarderías y consecuentemente los sistemas educativos libre de adoctrinamiento político” es substancialmente más valioso para una nación si lo que se desea es generar riqueza a largo plazo, que solo generar un PIB elevado en corto o mediano plazo. Si el PIB fuese más relevante que la educación se diría que mientras más recursos económicos se tenga mejor educación se tendrá, y esto sería una imprecisión lógica tal como se mostró líneas anteriores. Por ende, la forma de hacer política debe cambiar.

3. PORCENTAJE DE INVERSIÓN MUNDIAL EN INVESTIGACIÓN + DESARROLLO (I+D) DE PATENTES INDUSTRIALES REGISTRADAS

Las patentes industriales al igual que las publicaciones científicas son vestigios tangibles y medibles, la **primera** indica el nivel de desarrollo tecnológico, el nivel de desarrollo de ciencia y el nivel de innovación en un país; en tanto que la **segunda** indica el tipo de investigación que se lleva a cabo en un país. En general, los países con mayor cantidad de patentes en el mundo son: Estados Unidos, Japón, Alemania, China, Corea del Sur, Israel y Francia.

Al analizar deliberadamente a Corea del Sur se advierte por ejemplo que solo en el año 2008 registró 80.000 patentes en EEUU en comparación con Brasil 582, México 325, Argentina 79, Colombia 12, Costa Rica 9, Perú 7, Cuba 6, Ecuador 2, Bolivia 0. Esta pasmosa y progresiva diferencia puede ser consecuencia de que cuenta con profesionales *-especialmente ingenieros-* de altísimo nivel educativo, lo que conduce a más sofisticación tecnológica en diversas áreas del saber humano, lo que conduce a mayor venta de sus productos, lo que conduce a su vez a una mayor apertura de su mercado al mercado internacional. Desde el punto de vista económico y en correspondencia con el sector productivo de un país esto es muy revelador.

3 Por incomprensible e increíble que parezca son los países que no poseen recursos naturales los más productivos del planeta, ejemplo de ello son: Indonesia, Malasia, Filipinas y Tailandia (*llamados los tigres o dragones menores*); Hong Kong, Taiwán, Singapur y Corea del sur (*llamados los tigres o dragones mayores*). Otros casos dignos de mencionar son Israel, Japón, Australia, Finlandia, Irlanda, etc. Finalmente están aquellas naciones que están escalando a pasos gigantes la montaña de la productividad tales como Brunéi, Vietnam, Laos, Birmania, Camboya y Chile.

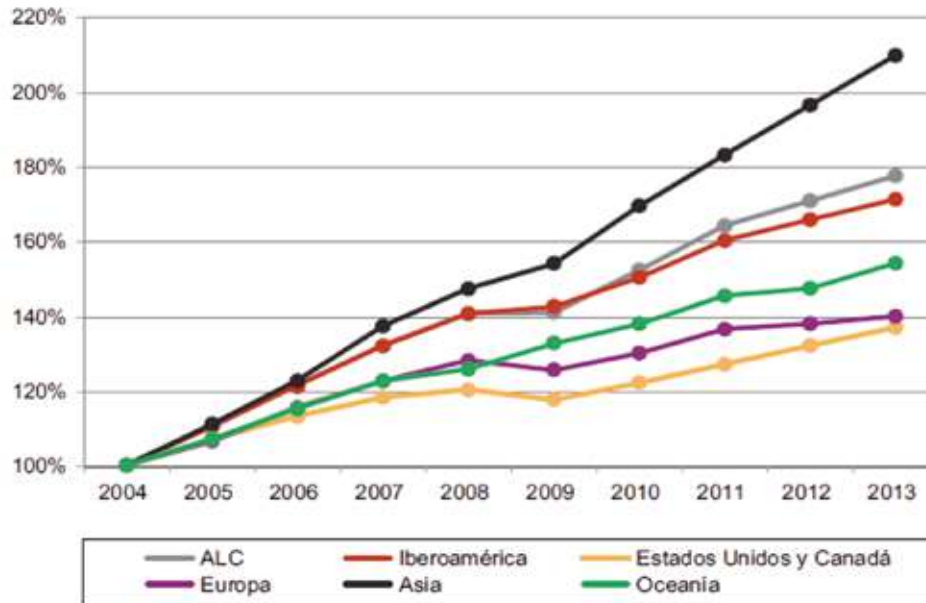
4 Obtenido en fecha 24 de abril del 2018 por acceso a la página https://www.timeshighereducation.com/world-university-rankings/2018/world-ranking#!/page/0/length/-1/sort_by/rank/sort_order/asc/cols/stats; además de ser obtenido en fecha 8 de septiembre del 2017 por acceso a la página <http://www.shanghairanking.com/ARWU2017.html>

Según los datos que refiere el Estado de la ciencia, 2015, pág. 51, a nivel latinoamericano los países que aún colaboran en temas de ciencia y tecnología con patentes en tres tecnologías, a saber, nanotecnología, biotecnología y TIC en comparación con el año 2008 son Chile, Brasil, México, Colombia y Argentina respectivamente. Un aspecto poco entusiasta es que la participación de países latinos en ciencia y tecnología es inversamente proporcional al tiempo.

4. PORCENTAJE DE PIB DESTINADO A INVESTIGACIÓN + DESARROLLO (I+D)

En el Gráfico N° 1, se distinguen tres aspectos; primero, en 10 años seis bloques de países tuvieron un incremento en su PIB lo cual es loable; **segundo**, en el año 2009 la crisis económica mundial afectó en particular a países de primer mundo; **tercero**, tanto a Asia como a Latinoamérica y el Caribe (ALC) se les atribuye un incremento trascendente en su economía incluso después de dicha crisis.

Gráfico N° 1
Evolución del PIB



Fuente: Extraído de Estado de la ciencia: principales indicadores de ciencia y tecnología iberoamericanos e interamericanos, 2015; p 15.

Entonces surge la interrogante ¿dado que ALC supone una estabilidad económica desde el año 2009 hasta el año 2015 no debería destinarse una significativa suma de dinero a I+D?, la respuesta es clara, se trata solo de un anhelo, pues parece ser que las propagandas populistas por diversos medios de comunicación (*con énfasis en la TV*) para los gobernantes del tercer mundo son de mayor prioridad para perpetuarse en el poder. Del mismo modo, resulta contradictorio pero digno de enfatizar que Estados Unidos,

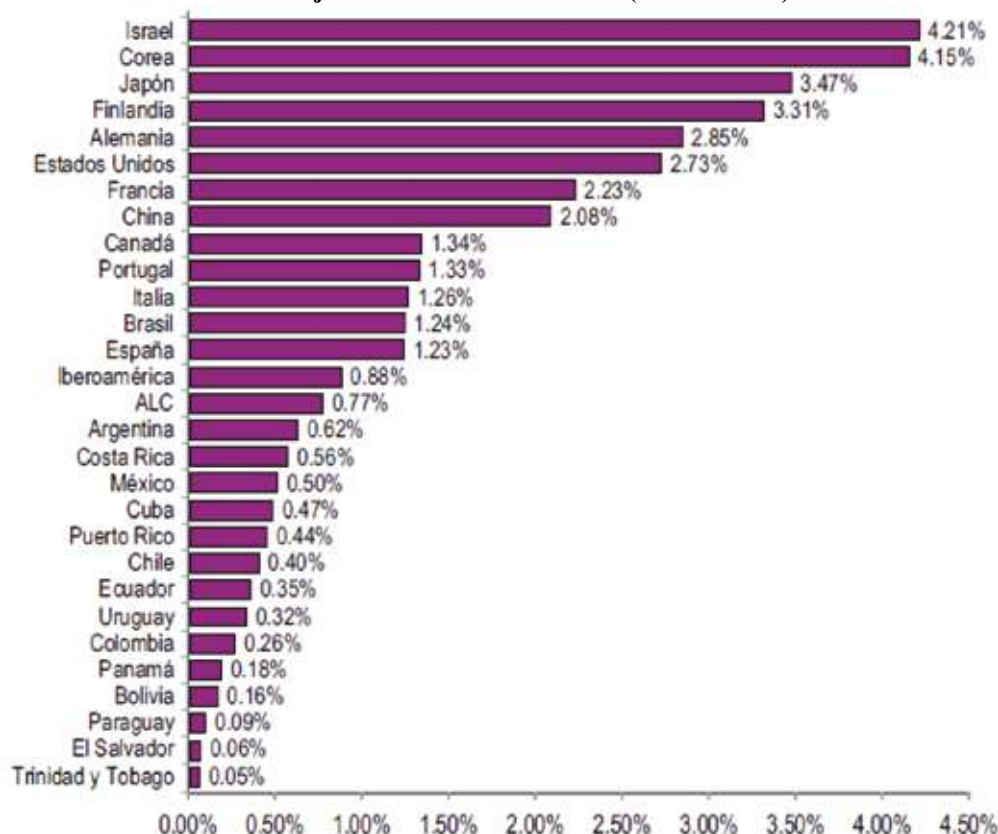
Canadá, Oceanía y Europa hasta el presente supieron sobrellevar dicho aprieto económico con la decidida participación en I+D de forma insistente.

En el Gráfico N° 2, se nota que hasta el año 2015 el total de países de Latinoamérica y el Caribe realizaron una inversión irracional que representó tan solo el 0.77% del PIB destinado a I+D en comparación con cualquier país industrializado, por ejemplo, Israel (4.21%). A su vez, el país del cono sur

que más invirtió hasta esa misma fecha y en esta actividad es Brasil (1.24%) superando a Chile (0.40%) quien fuera antiguo líder en la región, correlativamente entre los que menos invierten esta Bolivia (0.16%). Por tanto, la inversión de PIB destinado a I+D que efectúan los países industrializados continúa

siendo altísimo en relación a los países de segundo y tercer mundo, esto a pesar de la crisis del petróleo que afectó tanto a países de primer mundo (*países industrializados*) como a países exportadores de esa materia prima (*Venezuela, México, Colombia, Brasil, Perú, Ecuador*).

Gráfico N° 2
Porcentaje de PIB destinado a I + D (hasta el 2015)



Fuente: Extraído de Estado de la ciencia: principales indicadores de ciencia y tecnología iberoamericanos e interamericanos, 2015; p 20.

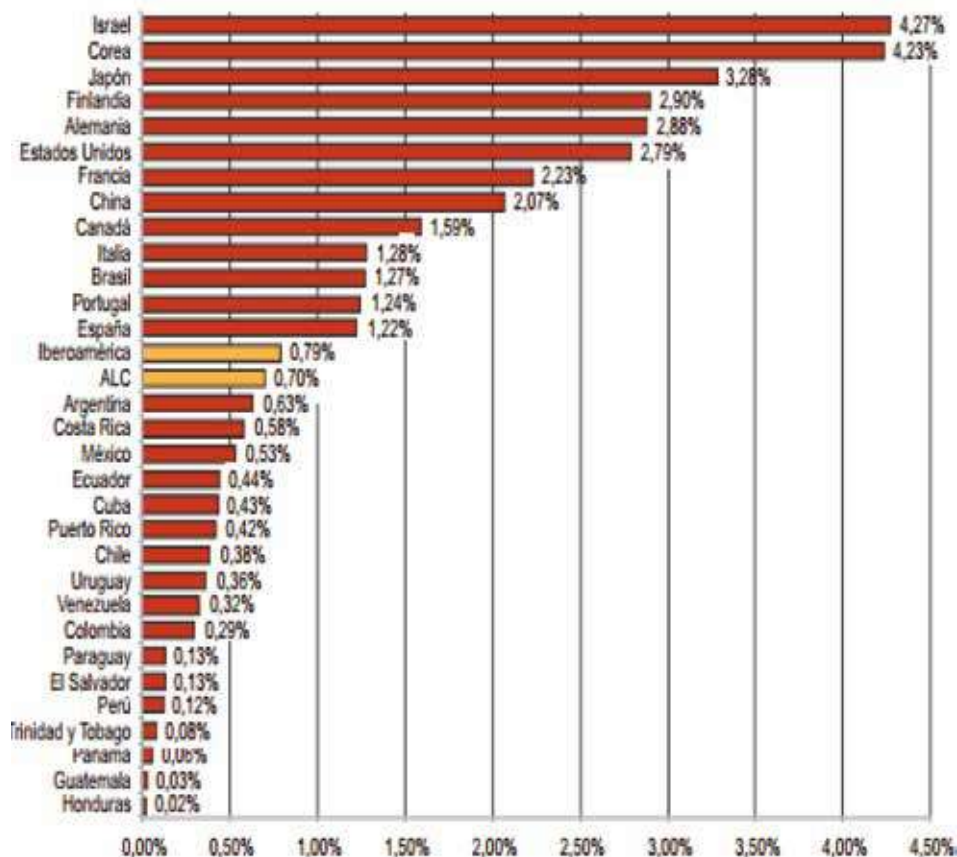
Sin embargo, el Gráfico N° 3 muestra datos actualizados hasta fines del 2017 y revelan comparativamente un descenso en la inversión en general, así el PIB en Iberoamérica representa 0.79% (*es decir 0.09% menos de inversión que en 2015*), el PIB en ALC muestra 0.70% (*es decir 0.07% menos de inversión que en 2015*).

Brasil invirtió hasta la fecha 1.27% de su PIB en I+D (*es decir 0.03% más de inversión que en 2015*); el resto de países latinoamericanos

invierten menos del 0.7%, para el caso boliviano la situación es aún más deprimente.

Luego, entre los países industrializados algunos mantienen su inversión en I+D como Francia; otros bajaron su inversión en I+D como China, Japón, Finlandia, Portugal, España; y otros en cambio incrementaron su inversión en I+D como Estados Unidos, Canadá, Alemania, Italia, Corea del Sur e Israel.

Gráfico N° 3
Porcentaje de PIB destinado a I + D (hasta fines del 2017)



Fuente: Extraído de Estado de la ciencia: principales indicadores de ciencia y tecnología iberoamericanos e interamericanos, 2017; p 21.

En efecto, los gráficos y sus cifras antepuestas son congruentes con las seis conclusiones de las seis mesas de trabajo⁵ que tuvo lugar en el “primer encuentro de 54 científicos bolivianos radicados en 33 países del exterior (Página SIETE; 2016: 28); **primera**, crear tres instancias nacionales de ciencia para la liberación tecnológica del país: “el Ministerio de Ciencia, Tecnología e Innovación (*daría la normativa*)”, “el Consejo de Ciencia, Tecnología e Innovación (*daría la regulación*)”, y “la Agencia de Ciencia, Tecnología e Innovación (*ejecutaría las acciones de hecho para iniciar la*

liberación tecnológica)”; **segunda**, contar con un fondo económico del 1% del PIB similar al de otros países (*este fondo debería ser independiente del sistema institucional, el 50% debería destinarse a investigación y el 50% restante a desarrollo e innovación*). Por unanimidad, en dicho encuentro los expertos sugirieron insistentemente que solo así podría fomentarse el avance científico y tecnológico de Bolivia, constituyéndose de este modo en el mecanismo para vincular el gobierno, la empresa y la universidad; **tercera**, el gobierno de turno actual del presidente Morales no apoya a las

5 Este primer encuentro de **científicos bolivianos** radicados en Alemania, Brasil, Bélgica, Canadá, China, Chile, España, USA, Francia, Holanda, México, Checoslovaquia, Reino unido, Suecia, Suiza, Japón, Panamá, tuvo lugar en Tiquipaya–Cochabamba con una duración de dos días disponiéndose para ello seis mesas de trabajo: **mesa N°1**: salud e industria del medicamento; **mesa N°2**: biodiversidad, medio ambiente, desarrollo agropecuario y tecnología de alimentos; **mesa N°3**: energía e hidrocarburo, minería y metalurgia; **mesa N°4**: transformación industrial o industrialización; **mesa N°5**: TIC; **mesa N°6**: desarrollo de hardware y software.

universidades estatales ni privadas que se supone deben producir el cambio, y por ende, ningún gobierno debe imponer dogmas políticos de ningún tipo a ninguna casa de educación superior de estudios tal y como se pretende; **cuarta**, el ministerio una vez creado debería trabajar de forma transversal con otros ministerios; **quinta**, implementar un programa de postgrado que garantice recursos a los centros de investigación privados y estatales para sostener programas de maestría, doctorado y postdoctorado, además de ofrecer becas; **sexta**, fomentar la formación bilingüe (*idioma inglés no así idiomas nativos*) de los estudiantes desde colegio y con énfasis en las universidades.

Surge otra cuestión ¿en los próximos años nuestro gobierno tendrá recién el interés político luego de más de una década de gestión destinar más recursos a todas las universidades de Bolivia bajo el enfoque de I+D+I = progreso?, ¿habrá resultados concretos o habrá más cumbres sociales?

Por consiguiente, las conclusiones a las que llegaron los científicos si bien son plausibles, penosamente contradicen con las ideas sociopolíticas dogmáticas de quienes nos gobiernan al presente; por ejemplo, para el actual gobierno de turno, los idiomas nativos es prioridad gubernamental nacional hace más de una década⁶ versus el idioma inglés como prioridad internacional en países industrializados y países emergentes, pues para éstos últimos se constituye como segundo idioma oficial desde hace más de media centuria.

5. PORCENTAJE QUE OCUPA EL SECTOR PRIVADO EN INVESTIGACIÓN + DESARROLLO (I+D)

¿Será que las investigaciones se efectúan solo a nivel estatal en lugar de relacionarse con la empresa privada o con ambas?, parece ser que esta es una de las razones. Según la Organización para la Cooperación y el Desarrollo Económico (OCDE) y los indicadores publicados en (RICYT; 2015:17) toda Latinoamérica invierte por el sector privado menos del 50% en I+D, de ese monto es destinado el 26% en Argentina, 41% en Brasil, 29% en Chile, 30% en México, 47% en Colombia, 10% en Venezuela. En cambio, Corea del Sur destina 74%, China 60% y EEUU 64%.

Pensamos que las Universidades *-en particular las estatales-* deberían ofrecer incentivos económicos a sus profesores para que investiguen y comercialicen nuevos productos, en consecuencia, las universidades se beneficiarían de un sistema filantrópico por el cual los egresados apoyarían económicamente a su alma mater como hábito, dotándolas de fideicomisos inmensos. Este proceder se convertiría en un círculo vicioso positivo, es decir, ingresan los mejores estudiantes a las universidades, éstas poseen los mejores docentes de una región, la universidad lanza al mercado laboral profesionales exitosos, éstos últimos hacen donaciones económicas a su alma mater y la universidad en cuestión sigue generando profesionales cualificados.

6 Cita textual de la actual Constitución Política del Estado en su Artículo 5 de su Capítulo I Modelo del Estado, pág. 5 que dice [*son idiomas oficiales del Estado el castellano y todos los idiomas de las naciones y pueblos indígenas originario campesinos que son el aymara, araona....*]

6. CONCLUSIÓN

- Los países más ricos del mundo⁷ no son aquellos ricos en fauna, agua, alimentos, minerales, petróleo o gas, sino más bien aquellos que producen más con el cerebro, aquellos que inventan nuevos productos, es decir a través de una labor creativa, innovadora e intelectual.
- A su vez, una estrategia económica que permitiera reducir la pobreza e incrementar la economía de un país⁸ sería a través de “las políticas de libre mercado” en un ambiente de estabilidad económica, esto conduciría a generar más fuentes de empleo demandando profesionales cualificados; esto por supuesto involucra gozar de la mejor educación posible.
- Las materias primas según el Banco Mundial, en el año 1960 constituían el 30% del Producto Bruto Mundial, en la década del 2000 representaba un exiguo 3%. En el año 2010, el 68% de la economía mundial radica en el sector servicios, el 29% en el sector industrial.
- Cuando un país es adicto o dependiente de sus recursos naturales -*como es el caso de México y Perú a la plata, Venezuela al petróleo, Bolivia al gas, prontamente al litio y al hierro, Chile al cobre, Argentina y Brasil a sus tierras muy fértiles y minerales*- forja inevitablemente una economía poco diversificada.
- Históricamente según el Banco Mundial el 93% de la población latina y el 97% de la actividad económica de la región reside en países exportadores netos de *Commodities* comparativamente con los países asiáticos (30%). En 2010 casi ¼ parte de los ingresos fiscales de los países latinos derivaban de las materias primas, en contraste con el 9% de los países desarrollados.
- Uno de los países que ya no es tan adicto a sus materias primas es Brasil, en efecto, en la actualidad desarrolla con éxito la industria aeronáutica, la automotriz y la tecnológica; su esfuerzo de haber invertido 1.27% de su PIB en I+D valió la pena.
- Otro país que supo salir de esa dependencia es Chile, presumiblemente debido a ocho componentes: **i)** supieron capitalizar sus recursos y resguardarlos de las sacudidas de la economía internacional; **ii)** supieron diversificar sus exportaciones; **iii)** supieron mantener un equilibrio en sus cuentas públicas y evitaron un endeudamiento externo; **iv)** supieron ahorrar en épocas de vacas gordas en un fondo de estabilización económico social lo cual derivó del fondo de estabilización

7 Esto implica poseer un PIB elevado y un PIB per cápita elevado, en consecuencia, un nivel de vida también elevado.

8 El crecimiento económico de un país puede estar limitado por el acceso al crédito; por falta de reformas energéticas, reformas tributarias y reformas de telecomunicaciones; por la inestabilidad macroeconómica; por la inestabilidad política; por los impuestos demasiados variables o elevados; por la rigidez en el mercado laboral; por la falta de coordinación en el descubrimiento de nuevas actividades productivas; por la lenta, ineficiente e ineficaz reacción de las empresas de un país “X” que comercializa ciertos productos en un mercado frente a otro país “W” que saca continuamente productos nuevos de mayor calidad y de alto valor agregado para reemplazar a los antiguos productos en el mismo mercado, ..., si el país “X” baja solo los precios a sus productos antiguos para tener alguna participación en el mercado a la larga perderá, no crecerá económicamente ni podrá competir con el país “Y” que posee mayor capacidad creativa, innovadora e intelectual ni con el país “Z” en el mismo mercado; causa: **la baja calidad de su oferta educativa**. Éste último punto tiene una relación directamente proporcional con que una empresa posea como activo “profesionales de alta calidad” (*artículo publicado por la Universidad de Harvard en 2009, p 32 en World Economic Forum*).

- del cobre; **v)** supieron crear un fondo de compensación para salvar la época de vacas flacas; **vi)** supieron estabilizar inteligentemente el tipo de cambio del dólar estadounidense; **vii)** supieron invertir en programas de reducción de pobreza y fomentar extraordinariamente la educación; **viii)** y finalmente supieron invertir en I+D+I a tiempo.
- Finalmente, le haría mucho bien a nuestra nación: **i)** invertir 4.56% del PIB destinado exclusivamente a I+D+I (y no el vergonzoso actual del 0.16%); **ii)** invertir 5.5% del PIB destinado exclusivamente a calidad educativa (y no el irrisorio actual del 0.9%), que no es lo mismo que el elevado porcentaje destinado a infraestructura y canchas con césped sintético; **iii)** invertir 25% del PIB destinado exclusivamente a inversión pública y privada (y no el fachoso actual del 17%). Debería ser así, si lo que realmente anhelamos con esperanza es salir del subdesarrollo y pobreza para siempre. Científica y tecnológicamente Bolivia ya no sería la misma nunca más; la clave está en contar con una educación de altísimo nivel, la cual no la tenemos aún.
 - Lo anterior es coherente con (La Razón; 2018:9) y el último Índice de Competitividad⁹ de 138 economías, medido en 2016-2017 por el Foro Económico Mundial que sitúa a Bolivia en competitividad en el vergonzoso puesto 121, en innovación puesto 135, en capacidad para la innovación puesto 131, en la calidad de las instituciones en investigación científica puesto 131, en el gasto económico en empresas de I+D+I puesto 134, en la colaboración universidad e industria en I+D+I puesto 136, en la adquisición gubernamental de tecnología avanzada puesto 110, en la disponibilidad de científicos e ingenieros puesto 130, en las aplicaciones de solicitudes de patente puesto 102.

9 El índice de competitividad mide la habilidad de los países de proveer prosperidad a sus ciudadanos, esta habilidad depende de cuán productivamente un país emplea sus recursos disponibles. O, en otras palabras, mide a un conjunto de instituciones, políticas y factores que definen los niveles de prosperidad económica y sostenible. Todo lo anterior se refleja en “The global competitiveness Report” cuyo informe muestra las fortalezas y debilidades de cada país con el fin de que sean sujeto de reformas políticas por sus mandatarios.

BIBLIOGRAFÍA

- La Razón. (2018, domingo 4 de marzo). El financiero: seis equipos de carrera por aportar al desarrollo de la economía. *La Razón*, p 9 B. <https://www.datosmacro.com/paises/comparar/bolivia/singapur>
- Albornoz, F.; Barrere, Rodolfo (2017). El Estado de la ciencia. Cap. I. *En Principales indicadores de ciencia y tecnología Iberoamericanos/Interamericanos RICYT*; p. 21. Buenos Aires, Argentina: Altuna impresores S.R.L. <https://datos.bancomundial.org/pais/mexico>
- Página SIETE. (2016, sábado 9 de enero). Sociedad: Científicos plantean 1% del PIB para investigación e innovación. *Página SIETE*, pág. 28. <https://datos.bancomundial.org/pais/singapur>
- Albornoz, F. & García Lembergman, E. (2015). El Estado de la ciencia. Cap. I. *En Red de indicadores de ciencia y tecnología Iberoamericanos/Interamericanos RICYT*; p. 21. Buenos Aires, Argentina: Altuna impresores S.R.L. <https://datos.bancomundial.org/pais/chile>
- Albornoz, F. & García Lembergman, E. (2015). El Estado de la ciencia. Cap. I. *En Red de indicadores de ciencia y tecnología Iberoamericanos/Interamericanos RICYT*; p. 21. Buenos Aires, Argentina: Altuna impresores S.R.L. <https://datos.bancomundial.org/pais/israel>
- Albornoz, F. & García Lembergman, E. (2015). El Estado de la ciencia. Cap. I. *En Red de indicadores de ciencia y tecnología Iberoamericanos/Interamericanos RICYT*; p. 21. Buenos Aires, Argentina: Altuna impresores S.R.L. <https://datos.bancomundial.org/pais/bolivia>
- Albornoz, F. & García Lembergman, E. (2015). El Estado de la ciencia. Cap. I. *En Red de indicadores de ciencia y tecnología Iberoamericanos/Interamericanos RICYT*; p. 21. Buenos Aires, Argentina: Altuna impresores S.R.L. http://www3.weforum.org/docs/GCR2016-2017/05FullReport/TheGlobalCompetitivenessReport2016-2017_FINAL.pdf

***Calle 27 de Cota Cota
Bloque F.C.P.N. - Primer Piso***

La Paz - Bolivia