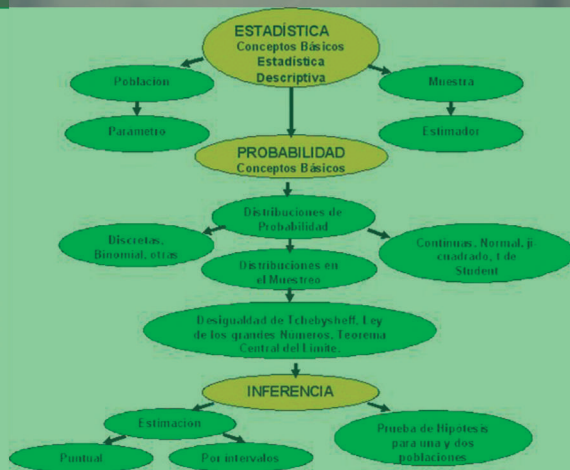




Universidad Mayor de San Andrés

Varianza

Revista de la Carrera de Estadística





Varianza

Revista de la Carrera de Estadística

*Publicación del Instituto de Estadística Teórica y Aplicada
Carrera de Estadística
Facultad de Ciencias Puras y Naturales
Universidad Mayor de San Andrés*

*Número 11
Febrero, 2015*



ISSN 9876-6789

Director a.i. Instituto de Estadística Teórica y Aplicada

Lic. Jaime Tito Pinto Ajhuacho

Diagramación y diseño

Lic. Omar Chocotea Poca

Lic. Enrique Villagra Arancibia

Redacción y edición

Msc. Fernando Oday Rivero Suguiura

Registro a Revistas Bolivianas

Sr. Alfredo Candia Chambi

Autores de artículos

Lic. Aliaga Casceres Iván

Lic. Chocotea Poca Omar

Lic. Delgado Álvarez Raúl

Lic. Flores López Juan Carlos

Lic. Flores Rivero Luis Fernando

Lic. Jiménez Careaga María Gilka

Lic. Pinto Ajhuacho Jaime Tito

Msc. Rivero Suguiura Fernando Oday

Lic. Valdez Blanco Dindo

Mg.Sc. Vargas Salazar F. Rodrigo

Lic. Vega Flores Carmen

Los artículos escritos son entera responsabilidad de los autores.

Las notas adicionales son de responsabilidad del editor.



Presentación

El presente número de la revista *varianza*, órgano oficial de difusión de la investigación en la Carrera de Estadística, muestra una vez más el interés de nuestros docentes para dar a conocer temas de interés práctico y actual de la estadística, mostrando además proyectos de mejoramiento didáctico para la enseñanza de la ciencia que nos apasiona.

Espero que la socialización del conocimiento expuesto a través de esta revista, nos permita crecer en el ámbito de la investigación mediante la respuesta de nuestros lectores, reiteramos el compromiso con nuestra casa de estudios para ofrecer próximamente una nueva versión de *Varianza*.

Lic. Raúl Delgado Álvarez
Director
Carrera de Estadística



Contenido

☑ El Uso del Tiempo. Aplicación de los Datos Composicionales Asimétricos. <i>Aliaga I.</i>	1
☑ Modelos de Regresión Binaria Bayesiana Power y Reciprocal Power. Aplicación: La Calidad del Servicio de Salud Pública en la Ciudad de La Paz. <i>Chocotea O.</i>	10
☑ La Distribución de Irwin–Hall. <i>Delgado R.</i>	16
☑ Resultados de Población y Vivienda de la Encuesta Sociodemográfica del Departamento de La Paz. <i>Flores J. C.</i>	20
☑ Clasificación mediante Redes Neuronales y Conglomerados No Jerárquicos de las Condiciones de Vida de los Hogares de Bolivia. <i>Flores L. F.</i>	25
☑ Superficies de Respuesta Mediante un Diseño Central Compuesto. <i>Jiménez M. G.</i>	31
☑ Tamaño de Muestra en Poblaciones Agropecuarias. <i>Pinto J. T.</i>	37
☑ Análisis y Clasificación de los Municipios del Departamento de La Paz mediante Escalamiento Multidimensional y Análisis Cluster. <i>Rivero F. O.</i>	43
☑ Estimación Bayesiana de los Resultados de las Encuestas Electorales. <i>Valdez D.</i>	50
☑ Implicaciones del Modelo de Guilford en la Enseñanza de Matemáticas, Física, Química y Biología. <<El pensamiento divergente>>. <i>Vargas F. R.</i>	53
☑ ¿Qué es el valor P ?. <i>Vega C.</i>	62
☑ En Realidad: ¿Qué es el Éxito en la Vida?, y ¿Qué es el Éxito Académico?. <i>Vargas F. R.</i>	63
☑ Autoevaluación y Acreditación Carrera de Estadística	67
☑ XXX Aniversario de la Carrera de Estadística	72
☑ Elecciones Dirección de Carrera	73
☑ Encuesta Sociodemográfica del Departamento de La Paz	75
☑ CLATSE – Congreso Latinoamericano de Sociedades de Estadística	79
☑ Potos de Estadística	81
☑ 11 de julio – Día Mundial de la Población	83
☑ Cronología de la Desaparición de 43 Estudiantes en Iguala, Guerrero	87



El Uso del Tiempo

Aplicación de los Datos Composicionales Asimétricos

Lic. Iván Aliaga Casceres
✉ powervan@gmail.com

Resumen. El Uso de tiempo se entiende como el tiempo que dispone cada individuo en sus labores cotidianas que se analizan en un ámbito social, económico y hasta político, en este trabajo se expone la aplicación de la teoría de pruebas de bondad y ajuste de una muestra para la Distribución Normal Asimétrica Multivariante en un espacio muestral restringido y estructurado llamado “Simplex”, que fue desarrollado por Aitchison (1986), el cual fue denominado hoy en día como Análisis de Datos Composicionales. La aplicación se desarrolla sobre datos observados del Uso del Tiempo en tres días de la semana tomadas aleatoriamente, excluyendo el día domingo, en diferentes actividades de 20 personas residentes mayores de 18 años de 2 distritos de la Ciudad de El Alto entre los meses de marzo-mayo de 2011.

1. Introducción

Los datos Composicionales son datos con información numérica real positiva que contienen una restricción subyacente natural y de orden geométrico, matemáticamente pueden ser descritos como: Datos que proceden de una relación de equivalencia del ortante positivo real \mathbb{R}^{+D} , cuyo cálculo se basa en un subconjunto de este espacio llamado “Simplex” que contiene una restricción de suma constante y que a su vez mantiene una estructura euclídea propia[1]. El uso de estos datos composicionales hoy en día hacen que muchas de las interpretaciones antes pensadas para \mathbb{R}^{+D} sean incorrectas, sin embargo, durante el último decenio surgieron grandes avances al respecto y ahora con el uso de esta metodología composicional, brinda más luces de investigación sobre un conjunto de problemas cuya aplicación va al “El uso del tiempo de las personas en las diferentes actividades diarias”[6].

De acuerdo con varios autores, el uso del tiempo de las personas en diferentes actividades, tiene un ámbito social, económico y hasta científico laboral, ya que revelan un orden y una estructura de estado de desarrollo de las fuerzas productivas y no productivas de la sociedad civil. La vida cotidiana podría ser estructurada según una rutina que puede variar con el género, la edad, la ocupación, el nivel educacional, etc. El estudio del uso del tiempo evidencia científicamente la situación de los indicadores del nivel de vida de la población y también la disponibilidad de tiempo de cada individuo en sus labores diarias[2].

Se utilizó las encuestas para investigar el uso del tiempo de la población desde el inicio del siglo pasado principalmente en Europa y los Estados Unidos. Existen antecedentes en Londres 1913, URSS 1920, USA 1920 y 1934, posteriormente después de la II Guerra Mundial comenzaron a proliferar este tipo de encuestas, entre las que se pueden resaltar el llamado estudio Szalai auspiciado por la UNESCO entre 1965 y 1966 donde participaron 13 ciudades de 11 países (Bélgica, Checoslovaquia, ex República Federal de Alemania, Francia, Hungría, Bulgaria, Polonia, ex Unión Soviética, USA, Yugoslavia y Perú). Otros países del mundo han sistematizado este tipo de encuestas como Holanda que la realiza cada 5 años y Dinamarca. Gran Bretaña y Francia llevan a cabo cada 10 años. También la llevan a cabo, aunque sin ninguna sistematicidad Bélgica, Alemania, Italia y España[7].

En las Naciones Unidas la medición del tiempo y su utilización constituye una de las prioridades de investigación. En su 31 período de sesiones en marzo del 2000, la Comisión de Estadísticas examinó el Plan de ejecución del proyecto “E/CN.3/2000/13”[8] sobre las cuestiones de género en la medición del trabajo remunerado y no remunerado, reconoció la importancia de esta labor realizada y señaló la labor conexas

adicional que llevaba a cabo diversos países y organismos en particular en relación con los estudios sobre el empleo del tiempo.

El Instituto Nacional de Estadística de Bolivia inicia investigaciones en este campo aproximadamente desde el 2001, sin embargo, es el año 2010-2011 donde se realizó encuestas pilotos por mandato ministerial del 2 de junio enmarcado en el artículo 338 de la Constitución Política del Estado que señala “El Estado reconoce el valor económico del trabajo del hogar como fuente de riqueza y deberá cuantificarse en las Cuentas Públicas”, con el propósito de visualizar esta temática y el rol protagónico para generar políticas públicas, el INE incluye dentro de sus áreas de trabajo el levantamiento de información, procesamiento y publicación de la investigación del Uso del Tiempo, sin embargo, los datos analizados en este paper fueron recolectados por el propio autor tomando para ello una muestra no probabilística en dos distritos de la ciudad de El Alto y cuya metodología de trabajo fue extraída en base a la tesis doctoral de Glória Mateu i Figueras[4].

2. Datos Composicionales

Sea el vector $w = (w_1, w_2, \dots, w_D)^t$ con D elementos o componentes cuyos elementos son positivos y la suma de todos ellos es constante, es decir, que se tiene la definición de un espacio estructurado llamado “Simplex” definido como:

$$\mathcal{S}^D = \left\{ (w_1, w_2, \dots, w_D)^t : w_1 > 0, w_2 > 0, \dots, w_D > 0; \sum_{i=1}^D w_i = K \right\} \quad (1)$$

La restricción de suma constante da una correspondencia específica a las partes, ya que los componentes de la composición $(w_1, w_2, \dots, w_D)^t$ con D partes positivas pueden ser completamente especificadas si se conocen $D - 1$ componentes. Este espacio simplex es un espacio vectorial porque se pueden definir operadores como en el cuerpo de los reales, los operadores de perturbación y potencia son las operaciones interna y externa del espacio vectorial, estos a su vez incluyen a un espacio de medida de probabilidad definiendo claro esta su medida de probabilidad y su respectiva función.

2.1. Transformaciones log-cocientes

Aitchison (1986) define la transformación logcociente Aditiva, denominado *alr*, del inglés *Additive log ratio*.

Definición 1. Dada una composición con D partes, la transformación log cociente aditiva (*alr*) de $X \in \mathcal{S}^D$ a $\Upsilon \in \mathbb{R}^{D-1}$ se define como:

$$\Upsilon = alr(X) = \left(\ln \left(\frac{x_1}{x_D} \right), \ln \left(\frac{x_2}{x_D} \right), \dots, \ln \left(\frac{x_{D-1}}{x_D} \right) \right)^t. \quad (2)$$

Aitchison Define una transformación log cociente centrada, denominada *clr*, del inglés *Centered log ratio*.

Definición 2. Dada una composición con D partes, la transformación log cociente centrada (*clr*) de $X \in \mathcal{S}^D$ a $Z \in \mathbb{R}^D$ se define como:

$$Z = clr(X) = \left(\ln \left(\frac{x_1}{g(X)} \right), \ln \left(\frac{x_2}{g(X)} \right), \dots, \ln \left(\frac{x_D}{g(X)} \right) \right)^t, \quad (3)$$

donde $g(X)$ es la media geométrica de las D componentes de X .

La transformación *alr* y *clr* son aplicaciones lineales entre los espacios vectoriales \mathcal{S}^D y \mathbb{R}^{D-1} ya que conservan las operaciones interna y externa, pero un inconveniente de la transformación *alr* es su falta de simetría porque el denominador de cada log cociente adquiere un protagonismo muy especial respecto al resto de los componentes, se podría escoger a priori otra componente como denominador.

Sin embargo la transformación clr es simétrica entre las partes, su imagen es el hiperplano V de \mathbb{R}^D que pasa por el origen y es ortogonal al vector unitario, aquí se encuentra una nueva dificultad ya que la suma de los componentes del vector transformado es igual a cero, a raíz de estas dificultades en el año 2003 Egozcue et al [5]., define una isometría entre los espacios \mathcal{S}^D y \mathbb{R}^{D-1} con objeto de superar los defectos e inconvenientes de las dos transformaciones anteriores, en donde la transformación alr no es una isometría y la transformación clr a pesar de conservar las distancias y el producto escalar, transforma la composición en vectores de un subespacio de \mathbb{R}^D con la restricción adicional que la suma de las componentes es igual a cero. Estos inconvenientes dan ciertas dificultades a la hora de interpretar resultados y provocaron una larga discusión del método propuesto por Aitchison (1986).

La transformación isométrica surge de manera natural, observando la transformación clr cuya condición $\sum z_k = 0$ que satisface las componentes de los vectores del subespacio $V = clr(\mathcal{S}^D)$ indica que el vector $(1, 1, \dots, 1)$ es ortogonal a este hiperplano. Si se escoge una base del espacio \mathbb{R}^D formada por $D - 1$ vectores ortogonales del subespacio V con un vector unitario y normal a V , es decir, $1/\sqrt{D}(1, 1, \dots, 1)$, se obtendrá que su última componente es igual a cero. A continuación aplicando una proyección sobre el hiperplano V da lugar a una isometría entre los espacios \mathcal{S}^D y \mathbb{R}^{D-1} .

Definición 3. Dada una base ortonormal del simplex \mathcal{S}^D , $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$, se define la transformación log cociente isométrica ilr de una composición $X \in \mathcal{S}^D$ a un vector $v \in \mathbb{R}^{D-1}$ como:

$$v = ilr(X) = (\langle x, \mathbf{e}_1 \rangle_a, \langle x, \mathbf{e}_2 \rangle_a, \dots, \langle x, \mathbf{e}_{D-1} \rangle_a)^t. \quad (4)$$

Con estas transformaciones y definiciones de operaciones en el simplex brindan una mejor comprensión sobre el análisis de datos composicionales cuya definición de espacios de medida y de distribuciones probabilísticas son análogas al espacio euclideo real.

3. Distribución normal asimétrica

Una distribución genérica es capaz de generar casos particulares de distribuciones de probabilidad, ciertamente la distribución normal asimétrica \mathcal{SN} es una de ellas, con relación a la ya conocida distribución normal y sus aplicaciones habituales, la \mathcal{SN} genera a la distribución normal como caso particular. La distribución normal asimétrica univariante, conocida en literatura inglesa como skew-normal, fue introducida y detallada por Azzalini (1985)[3].

Definición 4. Dada una variable aleatoria z , se dice que tiene una distribución normal asimétrica continua, cuya función de densidad es:

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad \lambda, z \in \mathbb{R} \quad (5)$$

donde $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$, $\Phi(\lambda z) = \int_{-\infty}^{\lambda z} \phi(u)du$, $z \in \mathbb{R}$ es la función de densidad y distribución normal estándar, cuya notación es la siguiente $z \sim \mathcal{SN}(\lambda)$.

El parámetro λ es un indicador de asimetría de la distribución, tal que su dominio es toda la recta real, cuando $\lambda = 0$ se obtiene la forma de una $\mathcal{N}(0, 1)$ y cuando $\lambda \rightarrow \pm\infty$ la distribución tiende a una normal truncada en el punto 0. La asimetría crece a medida que el valor absoluto del parámetro λ aumenta, de tal manera que a partir del valor $\lambda = 20$ aquel incremento es ligeramente inapreciable. Para valores λ positivos se obtiene una distribución asimétrica con una cola prolongada a la derecha y para valores λ negativos se tiene una distribución asimétrica con una cola prolongada a la izquierda.

Como el propio autor expresa que esta distribución tiene mejor ajuste en condiciones donde la muestra objeto de análisis se aproxima a una distribución normal. La definición multivariante de esta distribución es la siguiente:

Definición 5. Dado un vector aleatorio Z de dimension $D \times 1$, se dice que tiene una distribución Normal Asimétrica Multivariante continua cuya función de densidad es:

$$f_Z(z) = 2\phi_D(Z; \Omega)\Phi(\alpha^t Z), \quad \alpha, Z \in \mathbb{R}^D \tag{6}$$

donde $\phi_D(\cdot; \Omega)$ representa la función de densidad de un vector normal $(D \times 1)$ -dimensional con marginales estandarizadas y matriz de correlación Ω ; $\Phi(\cdot)$ es la función de distribución multivariante de una $\mathcal{N}(0, I_D)$ en $\alpha^t Z$; y $\alpha \in \mathbb{R}^D$, se utiliza la notación $Z \sim \mathcal{SN}^D(\Omega, \alpha)$ o simplemente $Z \sim \mathcal{SN}^D(\alpha)$, el parámetro de forma es α , cuando $\alpha = (0, 0, \dots, 0)^t$ se obtiene la densidad de una normal multivariante D -dimensional[9].

3.1. Contraste de hipótesis

El modelo normal es una caso particular de la familia Normal Asimétrica, ya que corresponde al caso de $\alpha = 0$, en la práctica solo interesa comparar el modelo normal asimétrico ajustado con el modelo normal. En ese caso tan solo se contrasta la hipótesis nula de $\alpha = 0$ contra la hipótesis alternativa de que $\alpha \neq 0$, aplicando un test de razón de verosimilitud, para esto se necesita el máximo de la función de log verosimilitud bajo la hipótesis de normalidad asimétrica $l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha})$ y el máximo de la función de log verosimilitud bajo la hipótesis de normalidad, $l(\bar{x}, \hat{S}, 0)$, donde \bar{x} y \hat{S} representan los estimadores de máxima verosimilitud de los parámetros μ y Σ bajo la hipótesis de normalidad, el estadístico de prueba es:

$$RV = 2(l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) - l(\bar{x}, \hat{S}, 0)), \tag{7}$$

el cual, bajo la hipótesis nula, sigue una distribución χ^2 con D grados de libertad, donde D es el número de componentes del vector aleatorio.

4. Aplicación

Ahora se presentará la aplicación de la teoría de Datos Composicionales y la Distribución Normal Asimétrica con datos del *Uso del tiempo de las personas en diferentes actividades diarias*, realizado en los meses de Marzo-Mayo 2011, con personas que accedieron a ser observadas en su quehacer cotidiano del tiempo en 60 días[2]. La suma de los tiempos usados de todo un conjunto de actividades por cada unidad observacional es constante, suma 24 hrs., se tiene una muestra de 20 personas, 10 hombres y 10 mujeres, tanto profesionales como no profesionales mayores a 18 años residentes en los distritos 4 y 6 de la Ciudad de El Alto, cuya medición del uso del tiempo en horas empleadas en actividades fue durante 3 días para cada persona, estos días fueron escogidos al azar de la semana excluyendo el día domingo, por un lapso de 2 meses y dos días (Marzo-Mayo, 2011), se tuvieron en total 60 días de observación, esto asegura que existan días trabajados alternados en cada semana y no exista tiempo recurrente o posibles arrastres de actividades.

Cuadro 1 ➤ Variables utilizadas

1	TTRAB	Tiempo usado en actividades de trabajo remunerado.
2	TNREM	Tiempo usado en actividades no remuneradas*.
3	TASOC	Tiempo usado en Actividades Sociales.
4	TASEOP	Tiempo empleado en actividades de aseo y cuidado personal.
5	TINV	Tiempo empleado en investigación y/o repasos para el desarrollo profesional.
6	TOAC	Tiempo empleado en otras actividades*.

De esta forma se tiene un conjunto de datos definido composicionalmente como:

$$X = [x_{k,i}; k = 1, \dots, 60; i = 1, \dots, 6] \quad X \in \mathcal{S}^6 \tag{8}$$

$$\mathcal{S}^6 = [x_{k,i} > 0; \forall k = 1, \dots, 60, i = 1, \dots, 6; \sum_i^6 x_{k,i} = 24] \tag{9}$$

La interpretación de los datos se observa de forma relativa de sus partes y no pueden justificarse interpretaciones que involucren magnitudes absolutas, por lo tanto cualquier aseveración de una composición debe hacerse en términos de los cocientes entre las partes, los cuales medirán dicha relación relativa, a continuación se observa la matriz de variación composicional.

Cuadro 2 ➤ Matriz de variación composicional de los datos observados

	1	2	3	4	5	6
1		1.215	1.589	0.348	0.632	0.236
2	0.610		0.809	1.394	1.917	1.424
3	2.633	2.023		1.727	2.269	1.735
4	0.662	0.051	-1.972		0.641	0.369
5	1.859	1.249	-0.774	1.198		0.599
6	-0.588	-1.199	-3.222	-1.250	-2.448	

El cuadro (2) muestra la matriz de variación composicional cuya diagonal inferior muestra las esperanzas de log cocientes de las partes, la diagonal superior muestra las varianzas de log cocientes de las mismas partes.

Puede observarse que la variación relativa más alta se encuentra entre los tiempos usados en: Actividades Sociales y Actividades de Investigación para el desarrollo profesional con $\hat{\tau}_{35} = 2.269$. Adicionalmente el valor negativo $\hat{\xi}_{35} = -0.774$ indica que los tiempos en actividades en investigación para el desarrollo profesional tienden a ser ligeramente más grandes que los tiempos dedicados a las actividades sociales.

Otro valor casi extremo es la existencia pequeña de variabilidad relativa entre los tiempos usados en: Actividades de Trabajo Remunerado con actividades relacionadas al Aseo y Cuidado Personal $\hat{\tau}_{14} = 0.348$, el valor positivo de $\hat{\xi}_{14} = 0.662$ muestra que los tiempos dedicados a las actividades de trabajo remunerado tienden a ser grandes que los tiempos dedicados a las actividades de aseo y cuidado personal.

Los estadígrafos centrales y de dispersión se muestran en el cuadro (3).

Cuadro 3 ➤ Resumen Estadístico de las 6 componentes

Componentes	1	2	3	4	5	6
Media	5.87	3.19	0.42	3.03	0.92	10.57
Varianza	3.29	6.21	19.09	5.70	12.47	2.03
Mínimo	2.70	0.41	0.03	0.79	0.23	2.78
1er cuartil	3.90	1.22	0.16	1.98	0.48	7.94
Mediana	5.18	4.38	0.49	2.84	0.70	9.81
3er cuartil	6.61	5.83	0.85	3.67	1.27	11.49
Máximo	8.47	12.12	3.27	6.39	4.79	17.56

Trabajando con la subcomposición TTRAB, TNREM y TASOC que corresponde al tiempo usado en: Actividades de trabajo remunerado, Actividades de trabajo no remunerado y Actividades sociales, se procede a comprobar si esta composición procede de una Distribución Normal Logística Aditiva o una Normal Asimétrica Logística Aditiva, se utilizan todas las componentes como denominadores en cada caso de la transformación alr y se aplica el test de razón de verosimilitud y las pruebas estadísticas de hipótesis de normalidad como ser: A-Darling, CV-Mises y Watson K-Smirnov.

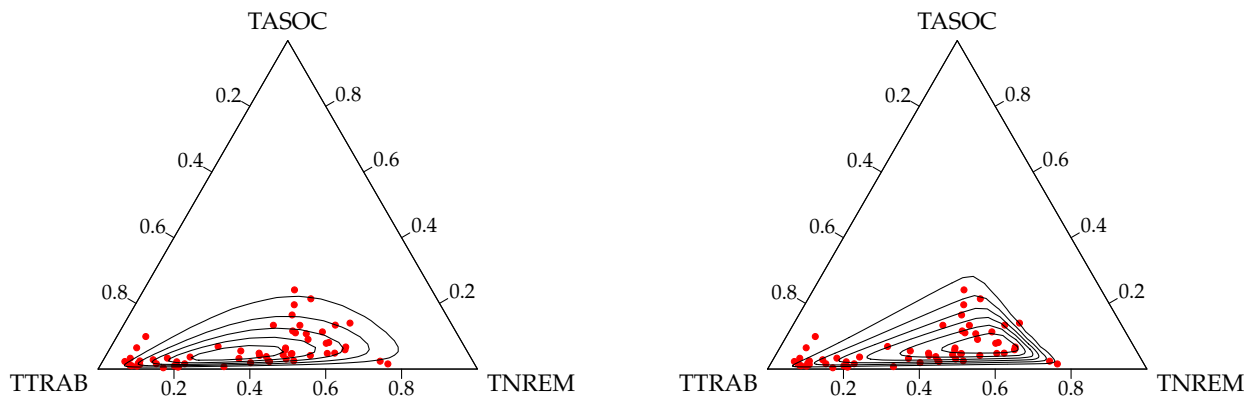


Figura 1 Diagrama Ternario, subcomposición TTRAB, TNREM y TASOC; izquierda: diagrama ternario asumiendo normalidad; derecha: diagrama ternario asumiendo normalidad asimétrica.

Estas dos facetas de representación ternaria composicional asumiendo que la muestra proceda de una Distribución Normal (izquierda) y Distribución Normal Asimétrica (derecha) se ven marcados. Aplicando la transformación logcociente aditiva con la componente TASOC como denominador se tendrán dos marginales $\ln(TTRAB/TASOC)$ y $\ln(TNREM/TASOC)$, los estimadores de Máxima Verosimilitud Asimétrica son:

$$\hat{\mu} = (1.22, 2.04)^t \quad \hat{\Sigma} = \begin{pmatrix} 3.56 & 0.56 \\ 0.56 & 0.80 \end{pmatrix} \quad \hat{\alpha} = (14.44, -5.09)^t \quad (10)$$

El valor de la función de logverosimilitud en estos puntos es: $l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) = -159.56$, luego los estimadores de Máxima Verosimilitud asumiendo normalidad en los datos son:

$$\bar{x} = (2.63, 2.02)^t \quad \hat{S} = \begin{pmatrix} 1.59 & 0.59 \\ 0.59 & 0.80 \end{pmatrix} \quad (11)$$


El valor de la función de máxima verosimilitud de estos estimadores es $l(\bar{x}, \hat{S}, 0) = -167.29$, comparando numéricamente el test de razón de verosimilitud, se obtiene que la diferencia entre estos dos ajustes es significativa $2(l(\hat{\mu}, \hat{\Sigma}, \hat{\alpha}) - l(\bar{x}, \hat{S}, 0)) = 15.46 = \chi_2^2(1)$. Por lo tanto se presume que la subcomposición tiene un buen ajuste a razón de la Distribución Normal Asimétrica, para corroborar lo mencionado, se procede a calcular los estadísticos de bondad y ajuste bajo la hipótesis nula de normalidad y normalidad asimétrica en las observaciones transformadas.

Cuadro 4 Resumen de las pruebas estadísticas de contraste bajo la hipótesis nula de normalidad; alr_i indica la transformación alr con $i = 1, 2, 3$ componentes como denominador; los números subrayados representan puntos significativos al 95 % de confianza.

		A-Darling	CV-Mises	Watson	K-Smirnov		
Marginal	i						
	1	alr_1	<u>2.212</u>	<u>0.383</u>	<u>0.357</u>	<u>1.322</u>	
		alr_2	<u>2.213</u>	<u>0.384</u>	<u>0.357</u>	<u>1.322</u>	
		alr_3	0.334	0.044	0.042	0.574	
	2	alr_1	0.334	0.045	0.043	0.575	
		alr_2	0.257	0.035	0.033	0.475	
		alr_3	0.257	0.034	0.032	0.475	
	Prueba radii	i					
		1	2	alr_1	0.959	0.159	0.098
alr_2				0.592	0.104	0.099	<u>1.033</u>
alr_3				0.592	0.103	0.098	<u>1.032</u>

Se observa que en la marginal 2 no existe evidencia en contra de la hipótesis nula de normalidad en los tres tipos de transformaciones aditivas, en la marginal 1 las pruebas de A-Darling, Watson y K-Smirnov comprueban que en los primeros dos tipos de transformación aditiva se halla la existencia en contra de la hipótesis nula de normalidad, mientras en la prueba del ángulo bivariado, la prueba de Kolmogorov-Smirnov corrobora que existen pruebas suficientes en contra de la hipótesis nula de normalidad conjunta, por lo tanto no puede aseverarse que los datos en tres partes de la subcomposición pueda provenir de una Distribución Normal.

Seguidamente se procede a calcular los mismos estadísticos para las pruebas de bondad y ajuste bajo la hipótesis nula de que los datos provienen de una Distribución Normal Asimétrica.

Cuadro 5  Resumen de las pruebas estadísticas de contraste bajo la hipótesis nula de normalidad asimétrica; alr_i indica la transformación alr con $i = 1,2,3$ componentes como denominador; (*) puntos significativos al 95 % de confianza.

		$n = 60$		A-Darling	CV-Mises	Watson	K-Smirnov		
Marginal	i	1	alr_1	$\lambda_1 = -5.248$	0.929*	0.162*	0.160*	0.869	
			alr_2	$\lambda_1 = 5.248$	0.929*	0.163*	0.161*	0.870	
			alr_3	$\lambda_1 = 1.896$	0.295	0.041	0.039	0.545	
		2	alr_1	$\lambda_2 = -1.896$	0.295	0.040	0.039	0.544	
			alr_2	$\lambda_2 = 0.623$	0.245	0.033	0.033	0.470	
			alr_3	$\lambda_2 = -0.623$	0.246	0.034	0.034	0.470	
	Ángulo Bivariado	i	j	1	alr_1	0.456	0.080	0.076	0.837
					alr_2	0.457	0.080	0.077	0.838
					alr_3	0.458	0.081	0.077	0.840

Asumiendo normalidad asimétrica solo en la marginal 1 de los dos primeros tipos de transformaciones son significativas al 95 % de confianza, dando lugar a la no existencia de pruebas en contra de la hipótesis nula de normalidad asimétrica en la marginal 2 ni en la conjunta, puede aseverarse que los datos en composición de tres partes provienen de una población con función de Distribución Normal Asimétrica, sin embargo la influencia del denominador en las transformaciones aditivas 1 y 2 hace que las conclusiones no sean tan validos del todo, porque podría usarse cualquier otra componente como denominador y llegar a las mismas conclusiones, pero al usar estas, se llegan a diferentes puntos de vista, esto es una desventaja propia de la transformación aditiva logística.

A continuación se presenta la tabla de pruebas de bondad y ajuste en todas las variables utilizando para ello la transformación isométrica con la siguiente base ortonormal,

$$U = \begin{bmatrix} 0.91 & & & & & \\ -0.18 & 0.89 & & & & \\ -0.18 & -0.22 & 0.87 & & & \\ -0.18 & -0.22 & -0.29 & 0.82 & & \\ -0.18 & -0.22 & -0.29 & -0.41 & 0.71 & \\ -0.18 & -0.22 & -0.29 & -0.41 & -0.71 & \end{bmatrix} \quad (12)$$

Que consiste en la partición binaria de las coordenadas de las partes según el orden de las mismas[5].

Según los valores de los estadísticos de prueba bajo la hipótesis nula de normalidad asimétrica respecto de la base ortonormal U , estos valores no logran superar los puntos críticos al 95 % de confianza en las marginales, a excepción de la primera marginal para la prueba de Kolmogorov-Smirnov, sin embargo,

todas las otras pruebas multivariantes incluyendo las marginales conjuntas no logran refutar la hipótesis nula de normalidad asimétrica en las observaciones de 6 partes, las observaciones del uso del tiempo en diferentes actividades proceden de una Distribución Normal Asimétrica.

Cuadro 6 ➤ Estadísticos de contraste para la prueba de bondad y ajuste bajo la hipótesis de normalidad asimétrica utilizando coordenadas según la base ortonormal U ; (*) puntos significativos al 95 % de confianza.

		$n = 60$	A-Darling	CV-Mises	Watson	K-Smirnov
Marginal	i					
	1	$\lambda = -7.255$	0.770	0.149	0.110	1.059*
	2	$\lambda = -2.971$	0.529	0.073	0.073	0.622
	3	$\lambda = -2.083$	0.427	0.067	0.067	0.771
	4	$\lambda = -2.609$	0.384	0.061	0.060	0.803
	5	$\lambda = 0.577$	0.369	0.065	0.065	0.640
Test Radii	i	j				
	1	2	1.109	0.212	0.136	1.067
	1	3	0.755	0.126	0.122	0.878
	1	4	1.201	0.224	0.163	1.021
	1	5	1.242	0.229	0.125	1.042
	2	3	0.550	0.088	0.082	0.736
	2	4	0.932	0.142	0.135	1.049
	2	5	0.332	0.045	0.033	0.525
	3	4	1.236	0.213	0.123	0.955
	3	5	0.743	0.078	0.071	0.602
	4	5	0.411	0.072	0.052	0.683
Test Radii	$D = 5$		0.468	0.054	0.050	0.542

5. Conclusiones

El uso de la metodología de datos composicionales brinda un estudio completo y detallado, cuando los datos disponibles son por naturaleza restrictivos y de suma constante, en este paper se propuso contrastar dos distribuciones en un espacio estructurado llamado simplex donde se vio que no existen pruebas en contra de la hipótesis nula de normalidad asimétrica en datos del uso del tiempo de personas en diferentes actividades, sin embargo, esto no quiere decir que en todas las muestras obtenidas del uso del tiempo tengan una misma distribución asimétrica, el hecho de generalizar distribuciones de probabilidad aumentando las operaciones de cálculo, esto debido al ámbito paramétrico, no debe tomarse a la ligera, ya que si bien ayuda a tener una buena bondad de ajuste con la SN , se tiene la desventaja en el cálculo directo de las estimaciones en función de la muestra.

Los datos obtenidos y analizados no muestran observaciones faltantes, razón por la cual los datos no han sido previamente procesados con algún método de imputación, si así fuera, el método escogido tendría que estar modificado bajo la hipótesis nula de que los datos proceden de una población con función de Distribución Normal o Distribución Normal Asimétrica y probar su ajuste bajo las mismas hipótesis mencionadas.

Adicionalmente se recomienda utilizar la distribución asimétrica cuando exista una continua evidencia en contra de la hipótesis nula de normalidad en las observaciones o datos recolectados.

Se debe tomar en cuenta con sumo cuidado en la utilización de la transformación isométrica, cuando se realiza la transformación siempre debe utilizar una base ortonormal, esto con el fin de realizar las interpretaciones adecuadas de los componentes como balances entre las partes.

Referencias

- [1] Aitchison, J. *The Statistical Analysis of Compositional Data*. The Blackburn Press, 1986.
- [2] Aliaga, I. *Aplicación del Modelo Normal Asimétrico Composicional Multivariante al Uso del Tiempo*. Universidad Mayor de San Andrés, 2011.
- [3] Azzalini, A. A note on regions of given probability of the skew normal distribution. *University of Padua, Italy* 1 (2004), 8.
- [4] I Figueras , G. M. *Models De Distribució Sobre El Simplex*. PhD thesis, U.P.C., 2003.
- [5] J. J. Egozc Ue, V. P.-G. Groups of parts and their balances in compositional data analysis. *Mathematical Geology* 37 (2005), 795–828.
- [6] Rosario Aguirre, Cristina García Sainz, C. C. *El Tiempo, Los Tiempos una Vara de Desigualdad*. CEPAL, 2005.
- [7] Teres A Lara Junc O, Neyda González Nápoles, E. M. L. D. J. E. C. M. A. A. L. D. C. C. *Encuesta Sobre el Uso del Tiempo*. Oficina Nacional de Estadísticas, División de Estadísticas de Naciones Unidas, 2001.
- [8] Unid As, N., Ed. *Informe sobre su 31 período de sesiones (29 de febrero a 3 de marzo de 2000)*, Consejo Económico y Social Suplemento No. 4. CEPAL, Comisión Estadística, 2000.
- [9] Y A. Dalla Valle, A. A. The multivariate skew-normal distribution. *Biometrika* 83 (1996), 715–726.

Modelos de Regresión Binaria Bayesiana Power y Reciprocal Power

Aplicación: La Calidad del Servicio de Salud Pública en la Ciudad de La Paz

Lic. Omar Chocotea Poca
 ✉ omarchp@outlook.com

Resumen. En este artículo, se presenta a dos nuevos modelos de regresión binaria generales con perspectiva bayesiana, al power y al reciprocal power, e incluyen al power probit y al power logit, y al reciprocal power probit y al reciprocal power logit, respectivamente —aún no disponibles en *software* comercial—, y tienen como casos especiales al probit y al logit. El resultado de la aplicación en la data de la Encuesta Calidad de Vida en la Ciudad de La Paz en su tópico salud indica que los cuatro modelos tienen un buen performance.

Palabras clave: Modelos de regresión binaria generales; perspectiva bayesiana.

1. Introducción

Los modelos de regresión binaria son utilizados para predecir la probabilidad de una respuesta binaria (0/1) en función de diversas variables explicativas. La aplicación se encuentra en un rango más amplio de escenarios de investigación que el análisis discriminante.

Bliss (1935) introduce el primer modelo de regresión binaria, el probit, aún es utilizado, sin olvidarnos de sus contendientes históricos, el logit (Berkson, 1944) y el cloglog (Gumbel, 1958), pero el investigador no debe ser estricto con su afinidad o recomendación, más bien debe ubicar alternativas para comparar y seleccionar al mejor (Chocotea, 2014).

2. Links simétricos y asimétricos

Definición 1. Sea $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ un vector de n variables aleatorias independientes binarias (0/1), $\mathbf{x}_i = (1, x_{i2}, \dots, x_{ik})^\top$ un vector de diseño y $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^\top$ un vector de coeficientes de regresión. El modelo de regresión binaria, está dado por

$$p_i = \Pr(y_i = 1) = F(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad i = 1, 2, \dots, n, \tag{1}$$

donde $F(\cdot)$ denota a la fda (función de distribución acumulada), su inverso $F^{-1}(\cdot)$ de acuerdo a la teoría del modelo lineal general es llamado link.

Un link resulta ser simétrico cuando la fda procede de una fdp (función de densidad de probabilidad) simétrica. Por supuesto, un link resulta ser asimétrico cuando la fda procede de una fdp asimétrica. También, un link asimétrico puede reducirse a un link simétrico. Chen *et al.* (1999), Bazán *et al.* (2014) y Chocotea (2014) argumentan que, cuando la probabilidad de una respuesta binaria se aproxima a 0 en una tasa diferente que cuando se aproxima a 1, los modelos con link asimétrico son adecuados.

La siguiente definición es una generalización de los resultados presentados por Bolfarine y Bazán (2010), Pewsey *et al.* (2012), Martínez–Flórez *et al.* (2013, 2014), Bazán *et al.* (2014), y Chocotea (2014), y es central para nuestro desarrollo.

Definición 2. Denotando por $F(\cdot)$ a la fda de una fdp simétrica alrededor del origen con soporte en la recta real. Sean

$$\mathcal{F}_A = \{F_1(\eta; \lambda) = [F(\eta)]^\lambda, \lambda \in \mathbb{R}_+\} \quad \text{y} \quad \mathcal{F}_B = \{F_2(\eta; \lambda) = 1 - [F(-\eta)]^\lambda, \lambda \in \mathbb{R}_+\} \tag{2}$$

dos clases de fda's.

Los modelos definidos en (2) tienen propiedades atractivas convenientes: (a) cuando $\lambda = 0$ se reducen a un modelo con link simétrico; (b) la asimetría puede ser determinada por λ ; y (c) $F_1(\eta; \lambda) + F_2(-\eta; \lambda) = 1$. El modelo de regresión binaria power o recíprocal power, es caracterizado por

$$p_i = \Pr(y_i = 1) = F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta}), \tag{3}$$

donde $F_\lambda(\cdot)$ denota a la fda $F_1(\cdot; \lambda)$ o $F_2(\cdot; \lambda)$ de (2), respectivamente.

Cuadro 1 Características de los nuevos modelos

Clase	$F(\cdot)$	$\lambda > 0$	Modelo	Link
\mathcal{F}_A	$\Phi(\cdot)$	$\lambda \neq 1$	power probit	asimétrico
		$\lambda = 1$	probit	simétrico
	$\Psi(\cdot)$	$\lambda \neq 1$	power logit	asimétrico
		$\lambda = 1$	logit	simétrico
\mathcal{F}_B	$\Phi(\cdot)$	$\lambda \neq 1$	recíprocal power probit	asimétrico
		$\lambda = 1$	probit	simétrico
	$\Psi(\cdot)$	$\lambda \neq 1$	recíprocal power logit	asimétrico
		$\lambda = 1$	logit	simétrico

$\Phi(\cdot)$ denota a la fda procedente de la fdp normal estándar

$\Psi(\cdot)$ denota a la fda procedente de la fdp logística estándar

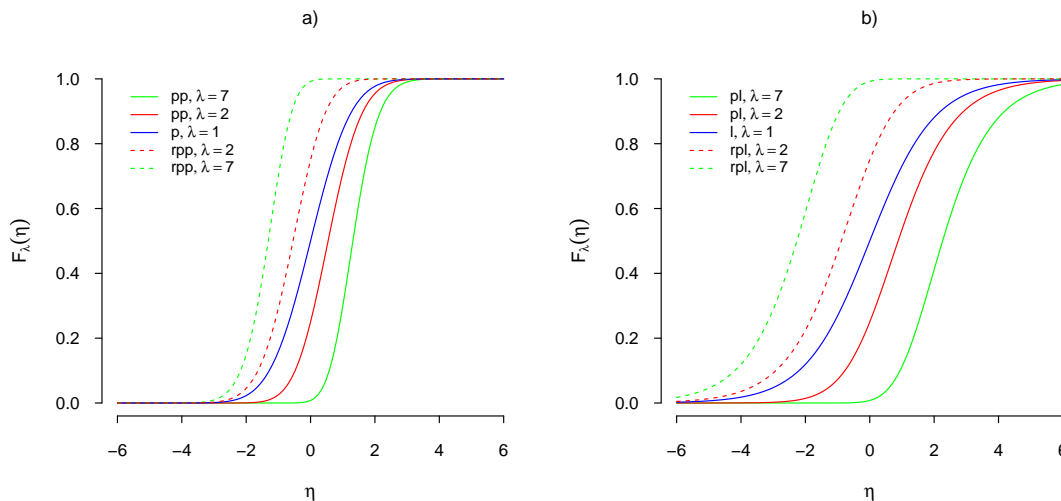


Figura 1 Curvas de probabilidad de los modelos a) power probit (pp), probit (p), recíprocal power probit (rpp), b) power logit (pl), logit (l) y recíprocal power logit (rpl).

Para la estimación e inferencia, se puede utilizar la data aumentada (Bolfarine y Bazán 2010; Chocotea, 2014; Kroese y Chan, 2014). Detalles importantes de modelos de variables latentes en el análisis de datos categóricos aparecen en Agresti y Kateri (2014). Las variables latentes evitan trabajar con la verosimilitud tipo Bernoulli (Bolfarine y Bazán, 2010; Chocotea, 2014).

3. Análisis bayesiano

Asumiendo independencia entre las previas (Bazán *et al.*, 2014), la estructura jerárquica está dada por

$$y_i | \mathbf{x}_i, \boldsymbol{\beta}, \lambda \sim \text{Ber}(F_\lambda(\mathbf{x}_i^\top \boldsymbol{\beta})) \quad \boldsymbol{\beta} \sim \pi_1(\boldsymbol{\beta}) \quad \lambda \sim \pi_2(\lambda). \tag{4}$$

Cuando se tiene varios modelos alternativos, surge la pregunta inmediata ¿cómo los comparo y selecciono al mejor? Lo primero que no se debe olvidar es que, no siempre existe un mejor modelo, y se puede tener un conjunto de modelos que ostenten un desempeño similar. El criterio principal de bondad de ajuste para comparar modelos Bayesianos es, el Criterio de Información de la Devianza (*Deviance Information Criterion*: DIC). Simplificando, un modelo es mejor que otro cuando su DIC es menor.

4. Aplicación

¿Me siento satisfecho con la calidad del servicio de salud pública en general?, es una de las preguntas que se hacen miles de paceños y paceñas, pero ¿cómo podemos analizar los motivos de satisfacción o insatisfacción respecto a un amplio espectro de situaciones que condicionan su calidad de vida y su felicidad?.

El Observatorio La Paz Cómo Vamos el 2012 realizó la segunda Encuesta de Percepción Ciudadana sobre la Calidad de Vida en la ciudad de La Paz. Ilustremos el análisis a partir del tópico salud, las variables son: *tiempo*, e indica la insatisfacción (1) o satisfacción (0) con el tiempo de demora para conseguir una consulta médica en el servicio público; *trato*, e indica la insatisfacción (1) o satisfacción (0) con la calidad del trato en los centros de salud y hospitales públicos; *infraequi*, e indica la insatisfacción (1) o satisfacción (0) con la calidad de la infraestructura y equipamiento de los centros de salud y hospitales públicos; *esfurespo*, e indica la insatisfacción (1) o satisfacción (0) con el esfuerzo y responsabilidad de los paceños para cuidar la salud de la familia; y *servicio*, e indica la insatisfacción (1) o satisfacción (0) con la calidad del servicio de salud pública en general. Las variables explicativas serán *tiempo*, *trato*, *infraequi* y *esfurespo*, y la variable dependiente será *servicio*. Debido a la no respuesta en algunas preguntas, la data se reduce a 955.

Cuadro 2 ➔ Sumario descriptivo de variables

Variable	Insatisfecho	Satisfecho	Media	DE
tiempo	783 (82.0 %)	172 (18.0 %)	0.820	0.384
trato	781 (81.8 %)	174 (18.2 %)	0.818	0.386
infraequi	782 (81.9 %)	173 (18.1 %)	0.819	0.385
esfurespo	730 (76.4 %)	225 (23.6 %)	0.764	0.424
servicio	776 (81.3 %)	179 (18.7 %)	0.813	0.390

DE: Desviación Estándar

Comparemos a siete modelos de regresión binaria, al probit (\mathcal{M}_1), al logit (\mathcal{M}_2), al cloglog (\mathcal{M}_3), al power probit (\mathcal{M}_4), al reciprocal power probit (\mathcal{M}_5), al power logit (\mathcal{M}_6) y al reciprocal power logit (\mathcal{M}_7). Las fdp's a posteriori son difíciles de obtener, por lo tanto las buenas aproximaciones son dadas bajo los métodos de Monte Carlo vía Cadenas de Markov (*Markov Chain Monte Carlo*: MCMC). Sus sintaxis pueden ser implementadas en los softwares RStudio, R, OpenBUGS, WinBUGS, o SAS, o también online, la nueva interfaz se llama WebBUGS y solo es necesario registrarse, la dirección de la web es <http://WebBUGS.psychstat.org>.

Cuadro 3 ➔ Sumario de la inferencia para los parámetros de regresión y valores del DIC

Parámetro	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_5	\mathcal{M}_6	\mathcal{M}_7
β_1	-1.705	-3.076	-2.280	-6.560	-1.008	-8.698	-3.502
β_2	1.046	1.895	1.028	2.088	1.813	3.167	1.490
β_3	0.512	0.906	0.585	1.148	0.898	1.841	0.784
β_4	1.003	1.763	1.054	2.091	1.736	3.304	1.420
β_5	1.223	2.275	1.081	2.364	2.212	3.369	1.732
λ				0.129	0.264	0.311	2.474
DIC	416.5	421.1	417.4	411.5	398.2	417.0	403.6

El valor del DIC determina que, el modelo con mejor desempeño o el más apropiado es el recíproco power probit (\mathcal{M}_5).

Cuadro 4  Sumario de la inferencia para los parámetros de regresión de \mathcal{M}_5

Parámetro	A posteriori			Intervalo HPD 95 %	
	Media	Mediana	DE	Inferior	Superior
β_1	-1.008	-1.008	0.381	-1.720	-0.259
β_2	1.813	1.777	0.496	0.968	2.832
β_3	0.898	0.869	0.376	0.214	1.718
β_4	1.736	1.719	0.455	0.940	2.621
β_5	2.212	2.204	0.534	1.261	3.200
λ	0.264	0.197	0.187	0.100	0.847

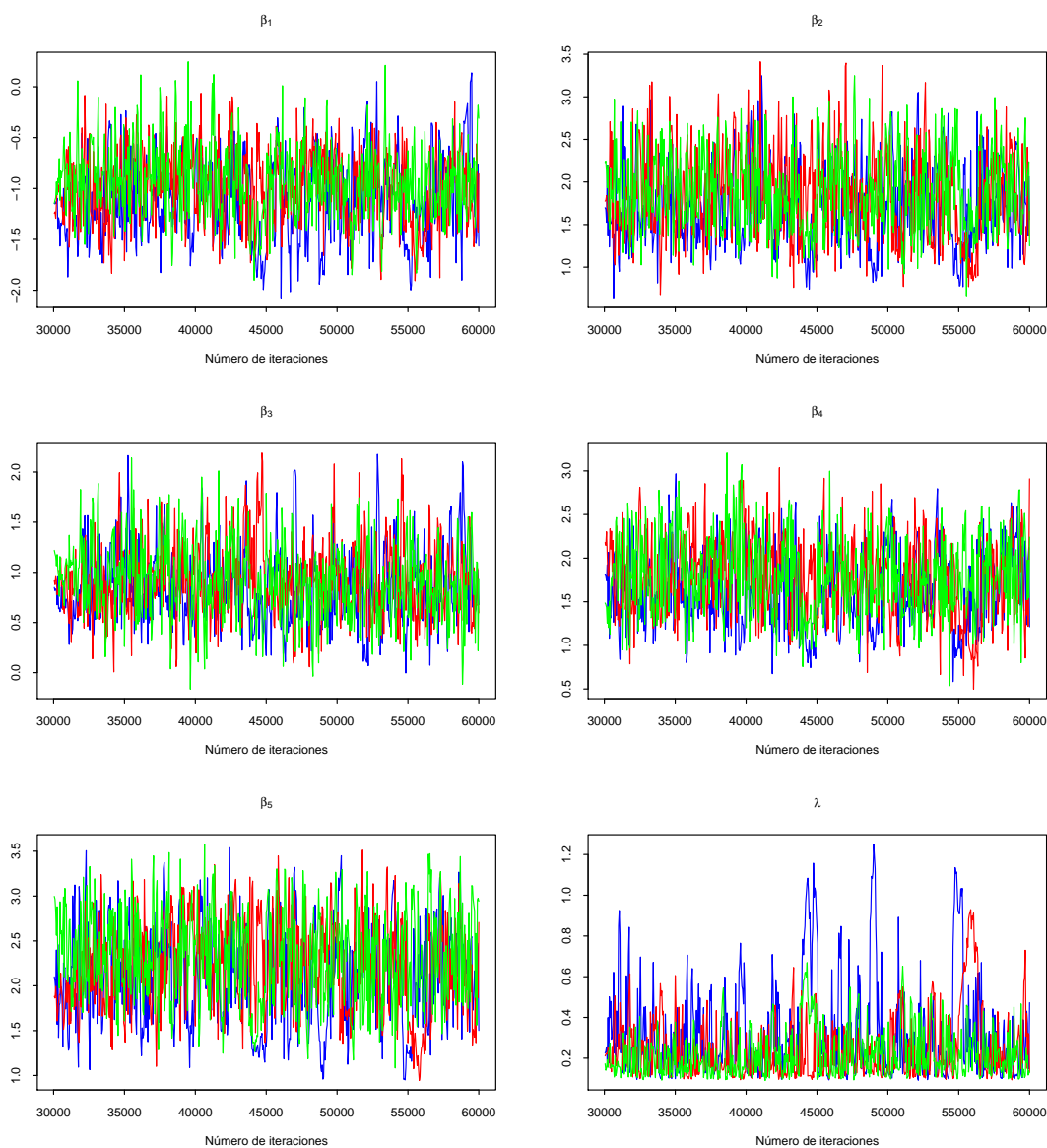



Figura 2  Historial de las cadenas de Markov de las fdp's a posteriori de \mathcal{M}_5 (1.^a línea azul, 2.^a línea roja y 3.^a línea verde).

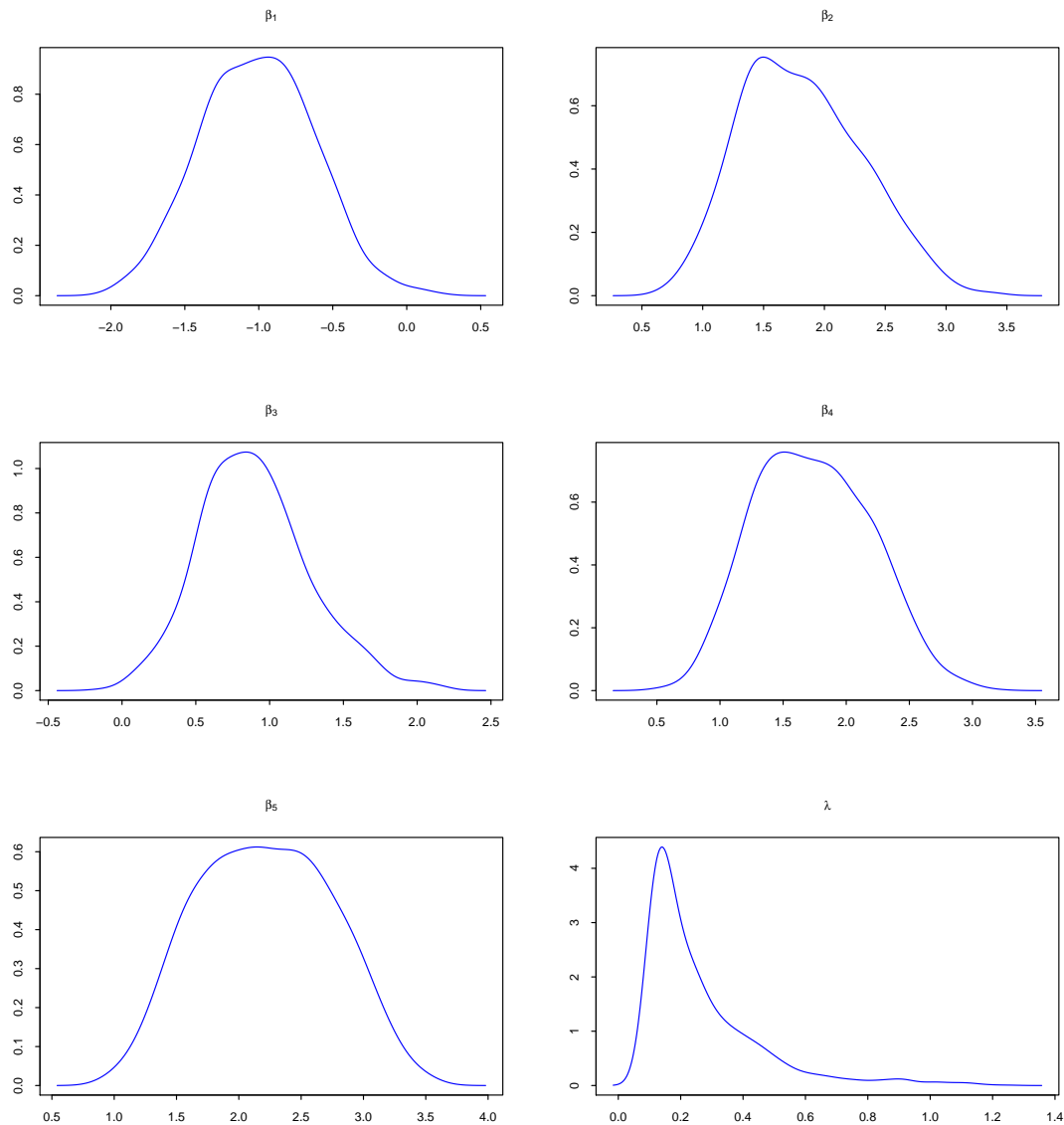


Figura 3 ▶ Plots de las fdp's a posteriori de \mathcal{M}_5 (1.^a cadena).

El modelo seleccionado para el análisis, está dado por

$$\Pr(\text{servicio}_i = 1) = 1 - [\Phi(1.008 - 1.813\text{tiempo}_i - 0.898\text{trato}_i - 1.736\text{infraequi}_i - 2.212\text{esfurespo}_i)]^{0.264},$$

donde $\Phi(\cdot)$ denota a la fda de la $\mathcal{N}(0,1)$.

La insatisfacción con la calidad del servicio de salud pública en general se debe en primer lugar a la insatisfacción con el esfuerzo y responsabilidad de los pacaños para cuidar la salud de la familia y en segundo lugar a la insatisfacción con el tiempo de demora para conseguir una consulta médica en el servicio público.

5. Conclusiones

Está claro que, la actualización del investigador va coherente con los nuevos aplicativos estadísticos para tomar mejores decisiones con los mejores modelos. Por ahora los cuatro modelos resaltaron por tener mejor desempeño.

Un estudio fortificante para una buena teoría, es el de las fdp's de las distribuciones power normal y power logística. Las extensiones van hacia los modelos de teoría de la respuesta al ítem (TRI).

Referencias

- [1] Agresti, A., y Kateri, M. (2014) Some Remarks on Latent Variable Models in Categorical Data Analysis. *Communications in Statistics – Theory and Methods* **43**, 801–814.
- [2] Bazán, J. L., Romeo, J. S., y Rodrigues, J. (2014) Bayesian Skew–Probit Regression for Binary Response Data. *Brazilian Journal of Probability and Statistics* **28**(4), 467–482.
- [3] Berkson, J. (1944) Application of the Logistic Function to Bio–Assay. *Journal of the American Statistical Association* **39**, 357–365.
- [4] Bliss, C. I. (1935) The Calculation of the Dosage–Mortality Curve. *Annals of Applied Biology* **22**, 134–167.
- [5] Bolfarine, H., y Bazán, J. L. (2010) Bayesian Estimation of the Logistic Positive Exponent IRT Model. *Journal of Educational and Behavioral Statistics* **35**, 693–713.
- [6] Chen, M.–H., Dey, D. K., y Shao, Q.–M. (1999) A New Skewed Link Model for Dichotomous Quantal Response Data. *Journal of the American Statistical Association* **94**, 1172–1186.
- [7] Chocotea, O. (2014) *Modelos de Regresión Binaria Bayesiana Skew Probit*. Tesis de Licenciatura, Universidad Mayor de San Andrés, La Paz.
- [8] Gumbel, E. J. (1958) *Statistics of Extremes*. Columbia University Press, New York.
- [9] Kroese, D. P., y Chan, J. C. C. (2014) *Statistical Modeling and Computation*. Springer, New York.
- [10] Martínez–Flórez, G., Bolfarine, H., y Gómez, H. W. (2013) The Alpha–Power Tobit Model. *Communications in Statistics – Theory and Methods* **42**(4), 633–643.
- [11] ——— (2014) An Alpha–Power Extension for the Birnbaum–Saunders Distribution. *Statistics* **48**(4), 896–912.
- [12] Pewsey, A., Gómez, H. W., y Bolfarine, H. (2012) Likelihood–Based Inference for Power Distributions. *Test* **21**, 775–789.

La Distribución de Irwin–Hall

Lic. Raúl Delgado Álvarez
 ✉ dea_5@hotmail.com

1. Introducción

Una manera de entender, por qué el teorema de límite central tiene una importancia gravitante en el estudio de la Estadística, es observar el comportamiento de la suma de variables aleatorias independientes.

Es así que una distribución nada acampanada como es la distribución uniforme nos permite observar la rápida convergencia de la suma de variables aleatorias hacia la distribución Normal en forma muy didáctica en base a los gráficos de las sumas de variables aleatorias independientes con distribución Uniforme.

En este trabajo primero observamos la distribución de la suma de dos variables aleatorias distribuidas uniformemente y se construye su gráfico, posteriormente se estudia la distribución de la suma de tres variables aleatorias independientes con distribución uniforme, de esta manera la distribución de Irwin-Hall se constituye en una generalización del problema citado y por último se construye un programa para la suma de un número finito de variables aleatorias con distribución uniforme.

Problema. Sean x y y variables aleatorias independientes, cada una con distribución uniforme en el intervalo $(0,1)$. Hallar la distribución de $z = x + y$.

Solución. Si

$$x \sim U[0,1) \Rightarrow f(x) = 1, \quad 0 \leq x \leq 1 \quad y \sim U[0,1) \Rightarrow f(y) = 1, \quad 0 \leq y \leq 1$$

Siendo x e y independientes la densidad conjunta será:

$$f(x,y) = f(x)f(y) = 1, \quad 0 \leq x \leq 1, 0 \leq y \leq 1$$

definida en el cuadrado.

Sea $z = x + y$ y la variable aleatoria auxiliar $w = y$

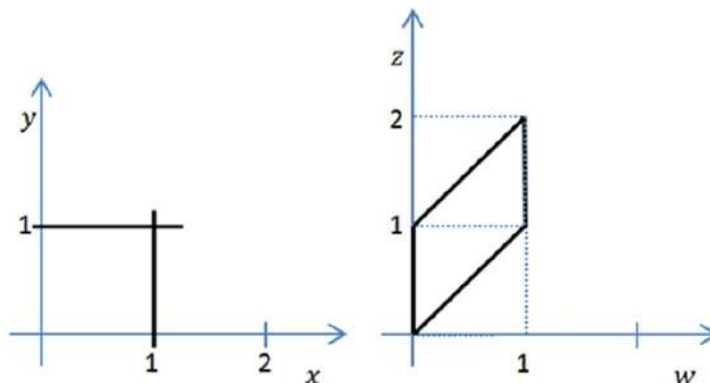


Gráfico 1 ➤ Región inicial y región transformada

Se tiene $\begin{matrix} x = z - w \\ y = w \end{matrix}$ luego el Jacobiano de la transformación es $J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$. Luego la función conjunta de z y w será:

$$K(z,w) = 1: \quad 0 \leq z - w \leq 1, 0 \leq w \leq 1 \quad K(z,w) = 1: \quad w \leq z \leq w + 1, 0 \leq w \leq 1$$

Es decir $k(z,w)$ se define en el paralelogramo.

Para hallar la distribución de la suma, se halla la distribución de la marginal

$$g_Z(z) = \int_{R_w} k(z,w)dw = \begin{cases} \int_0^z dw; & 0 \leq z \leq 1 \\ \int_z^1 dw; & 1 \leq z \leq 2 \end{cases} = \begin{cases} z; & 0 \leq z \leq 1 \\ 2-z; & 1 \leq z \leq 2 \end{cases}$$

Sea $z = x_1 + x_2 + x_3$, donde x_1, x_2 y x_3 son variables aleatorias independientes cada una con distribución uniforme en el intervalo $(0,1)$. Hallar la función de densidad de z .

Solución. Sea $z = (x_1 + x_2) + x_3$ y $y = x_1 + x_2$ del problema anterior $w = x_3$

$$f(y) = \begin{cases} y; & 0 \leq y \leq 1 \\ 2-y; & 1 \leq y \leq 2 \end{cases}$$

$f(w) = 1; 0 \leq w \leq 1$, la densidad conjunta de y y w será:

$$f(y,w) = \begin{cases} y; & 0 \leq y \leq 1; 0 \leq w \leq 1 \\ 2-y; & 1 \leq y \leq 2; 0 \leq w \leq 1 \end{cases}$$

Luego

$$z = y + w \Rightarrow y = z - h \quad H = w \Rightarrow w = h$$

Cuyo $J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$.

Luego la densidad conjunta de z, H será:

$$k(z,H) = \begin{cases} -h; & 0 \leq z-h \leq 1; 0 \leq h \leq 1 \\ 2-z+h; & 1 \leq z-h \leq 2; 0 \leq h \leq 1 \\ z-h; & h \leq z \leq h+1; 0 \leq h \leq 1 \\ (2+h)-z; & h+1 \leq z \leq h+2; 0 \leq h \leq 1 \end{cases}$$

Definida en el paralelogramo:

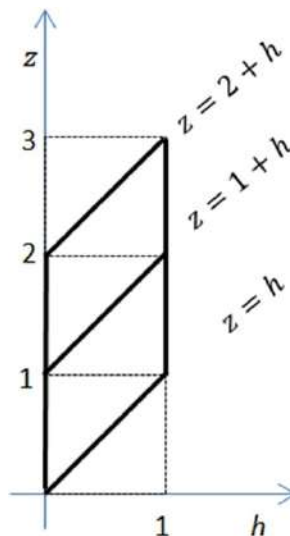


Gráfico 2 ➤ Región transformada para tres variables

Para hallar la distribución de z , hallamos la marginal:

$$g(z) = \int_{Rw} K(z,h)dh = \begin{cases} \frac{z^2}{2}; & 0 \leq z \leq 1 \\ -z^2 + 3z - \frac{3}{2}; & 1 \leq z \leq 2 \\ \frac{z^2}{2} - 3z + \frac{9}{2}; & 2 \leq z \leq 3 \end{cases}$$

2. La distribución de Irwin–Hall

La *distribución de Irwin-Hall*, llamada así por Joseph Irwin y Phillip Hall, es la distribución que rige el comportamiento la suma de variables aleatorias independientes, cada una con la distribución uniforme estándar. También se conoce como la *distribución uniforme suma*. También sirve como un buen ejemplo del teorema del límite central, observando la convergencia de esta suma como veremos a continuación.

Sea $Z = \sum_{i=1}^N x_i$, donde x_i se distribuye uniformemente en el intervalo (0,1), entonces se distribuye de acuerdo a:

$$f(z) = \frac{1}{(n-1)!} \sum_{j=0}^k \binom{N}{j} (-1)^j (z-j)^{N-1}; \quad k \leq z < k+1, \quad k = 0,1,2,\dots,N-1.$$

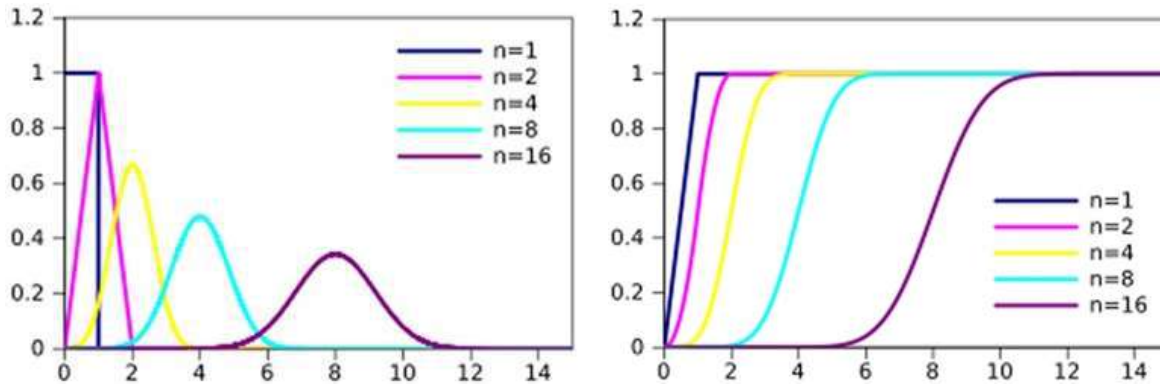


Gráfico 3 ➤ Evolución de la función de densidad y función acumulada. Extraído de Wikipedia

Para evidenciar el comportamiento de Z , se realiza un programa que permite generalizar La distribución de Z , además de mostrar el gráfico resultante para distintos valores de N , para $N = 8$, se muestra la indicada distribución y su gráfico, obsérvese que prácticamente se comporta como una normal, lo mismo se puede hacer para distintos valores de N .

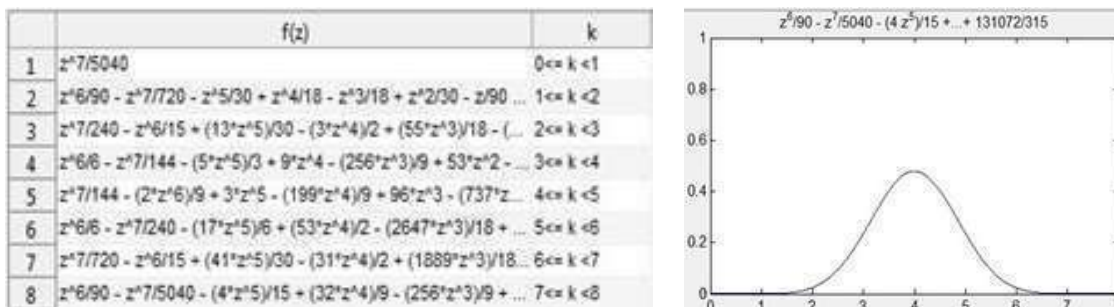


Gráfico 4 ➤ Resultado del programa

Programa

```
%programa que genera la funcion de Irwin-Hall en Matlab
%definimos variables simbolicas
syms z c;
%preguntamos si n es mayor a 1
n=input('Ingrese el valor de N=');
% inicializamos el vector donde se guardaran las funciones
A=zeros(n,1);
%creamos un vector simbolico a partir de A
B=sym(A);
%inicializamos las variables auxiliares
f=0;
%definimos la constante simbolica
c=1/factorial(n-1);
% comenzamos el calculo de las variables
for j=0:n-1
    %guardamos en el vector f la suma de los
    f=f+nchoosek(n,j)*(-1)^j*(z-j)^(n-1);
    %comenzamos el calculo de las variables
    f=collect(f);
    %desplegamos el calculo de las variables multiplicada por la constante definida antes
    disp(c*f);
    %almacenamos cada valor impreso en el vector simbolico
    B(j+1,1)=c*f;
End
%graficamos las variables simbolicas halladas en el vector
for i=1:n
    ezplot(B(i,1),[i-1,i]);
%mientras no terminemos el vector cada variable debe ser trazada en el mismo lienzo
if i<n
    hold on
end
end
hold off
%delimitamos el eje de x entre 0 y n y y entre 0 y 1
axis([0 n 0 1]);
```

Resultados de Población y Vivienda de la Encuesta Sociodemográfica del Departamento de La Paz

Lic. Juan Carlos Flores López
 ✉ caarloslopez1@gmail.com

1. Introducción

El Instituto de Estadística Teórica y Aplicada (IETA) de la Carrera de Estadística dependiente de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés (UMSA), realizó la Encuesta Socio Demográfica para el departamento de La Paz del Estado Plurinacional de Bolivia, a través del convenio Gobernación La Paz UMSA.

Estamos conscientes de que esta información es de mucha importancia para el departamento La Paz. Pensamos que es necesario seguir investigando las características del departamento.

2. Metodología

En esta oportunidad se utiliza la metodología estadística como es el análisis factorial de correspondencias para describir las características más importantes de población y vivienda de la encuesta en consideración.

Tabla 1 ➤ ¿Qué idiomas o lenguas habla?

	Porcentaje	Acumulado
Castellano	84.91	84.91
Quechua	1.11	86.03
Aymara	11.77	97.80
Guarani	0.01	97.81
No habla	2.15	99.96
Extranjero	0.04	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 1, muestra que el 84.91 % de la población de La Paz, habla castellano, en segundo lugar el idioma más hablado es el aymara con un 11.77 %, de acuerdo a estos resultados en la población de La Paz el idioma que más se habla es el castellano.

Tabla 2 ➤ ¿Cuál es su estado civil o conyugal actual?

	Porcentaje	Acumulado
Soltero(a)	42.70	42.70
Casado(a)	42.54	85.24
Conviviente/concubino(a)	8.58	93.82
Separado(a)	1.57	95.39
Divorciado(a)	0.78	96.17
Viudo(a)	3.83	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 2, nos muestra que aproximadamente el 42.00 % de las personas son solteros(as) y también casados(as), se puede ver también que el 8.58 % viven en concubinato.

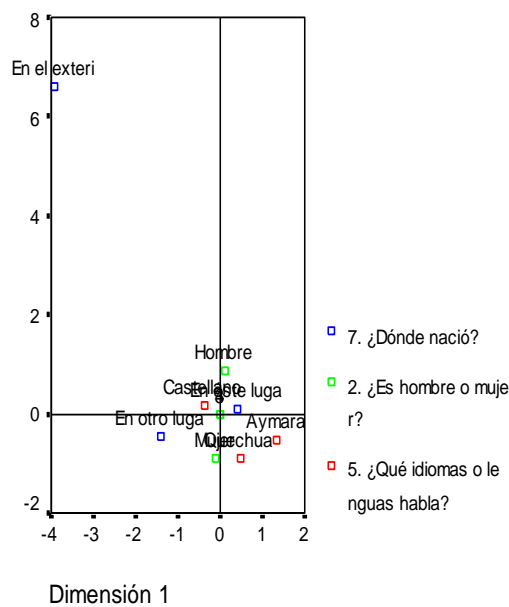
Tabla 3 ➤ ¿Dónde nació?

	Porcentaje	Acumulado
En éste lugar	75.76	75.76
En otro lugar del país	23.97	99.73
En el exterior	0.27	100.00
Total	100.00	

Fuente: Carrera de Estadística

Por otro lado la Tabla 3 nos muestra que el 75.76 % de la población de La Paz nació en este departamento, el 23.97 % nació en otro lugar del país y un 0.27 % nació en el exterior.

Gráfico 1



Aplicando el método de análisis factorial de correspondencias el Gráfico 1, nos muestra que efectivamente hablan más castellano entre hombres y mujeres y que nacieron en el departamento de La Paz.

Otra de las características poblacionales que se analiza es la variable empleo, donde la pregunta es: Durante la semana pasada, ¿trabajó al menos una hora?, y la pregunta: En esta ocupación ¿usted trabaja como . . .?. Los resultados fueron los siguientes.

Tabla 4 ➤ Durante la semana pasada, ¿trabajó al menos una hora?

	Porcentaje	Acumulado
Si	53.74	53.74
No	46.25	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 4, nos muestra que el 53.74 % si trabajo durante la anterior semana, trabajó al menos una hora y el 46.25 % no trabajo al menos una hora durante la semana pasada.

Se ve que en el departamento de La Paz solamente un poco más del 50 % trabaja, lo que llama la atención respecto a la ocupación en el departamento de La Paz.

Tabla 5 ➤ En esta ocupación ¿usted trabaja como ...?

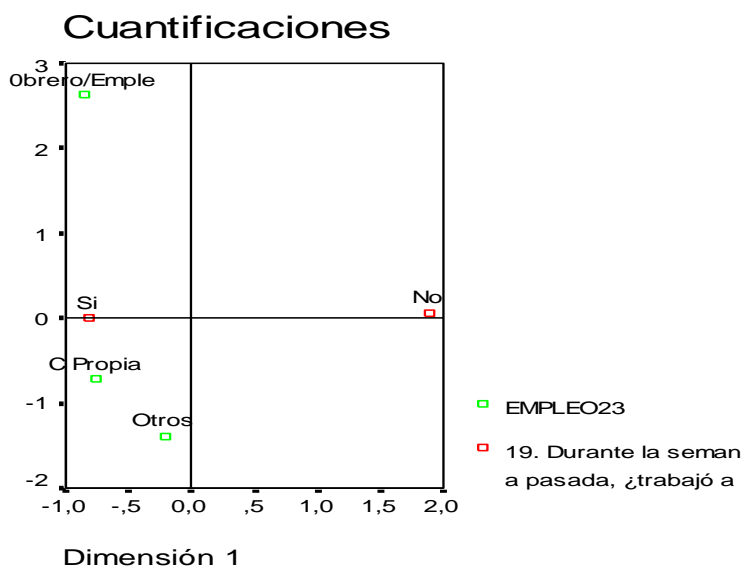
	Porcentaje	Acumulado
Obrero/empleo	34.72	34.72
Cuenta propia	58.17	92.89
Otros	7.11	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 5 nos muestra que el 58.17 % trabaja por cuenta propia como segunda ocupación es obrero/empleo con el 34.72 % total de la población.

Después de los resultados obtenidos se aplica el método de análisis factorial de correspondencias, el Gráfico 2, nos muestra que la mayoría trabajo por cuenta propia.

Gráfico 2 ➤ Representación de municipios en el espacio bidimensional



A continuación hacemos la descripción de la parte vivienda, los resultados fueron los siguientes:

Tabla 6 ➤ La vivienda que ocupa es ...?

	Porcentaje	Acumulado
Propia	73.55	73.55
Alquilada, anticrético, mixto	17.23	90.79
Cedida por servicios	1.94	92.73
Prestada por parientes o amigos	7.23	99.96
Otra	0.04	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 6 nos muestra que los habitantes de La Paz en un 73.55 % tienen vivienda propia y en un 17.23 % tiene alquilada, anticrético o mixto.

La Tabla 7 nos muestra que las viviendas de La Paz en un 81.64 % obtienen agua para beber y cocinar por cañería de red, el 8.12 % obtiene agua de la pileta pública y el resto tiene agua para su uso de carro repartidor, pozo, río/vertiente/acequia, lago y otros.

Tabla 7 ➤ Principalmente el agua para beber y cocinar la obtienen

	Porcentaje	Acumulado
Cañería de red	81.64	81.64
Pileta pública	8.12	89.76
Carro repartidor	0.12	89.88
Pozo o noria con bomba	2.06	91.94
Pozo o noria sin bomba	4.80	96.74
Río/vertiente/acequia	2.94	99.69
Lago/laguna/curiche	0.15	99.83
Otro	0.17	100.00
Total	100.00	

Fuente: Carrera de Estadística

En la Tabla 8 un 80.67 % de las viviendas tienen baño wáter o letrina y el restante 19.33 % no lo tiene.

Tabla 8 ➤ Tiene baño, water o letrina?

	Porcentaje	Acumulado
Si	80.67	80.67
No	19.33	100.00
Total	100.00	

Fuente: Carrera de Estadística

La Tabla 9 nos muestra que el 93.21 % de las viviendas usa energía eléctrica para alumbrar.

Tabla 9 ➤ Usa energía eléctrica para alumbrar esta vivienda?

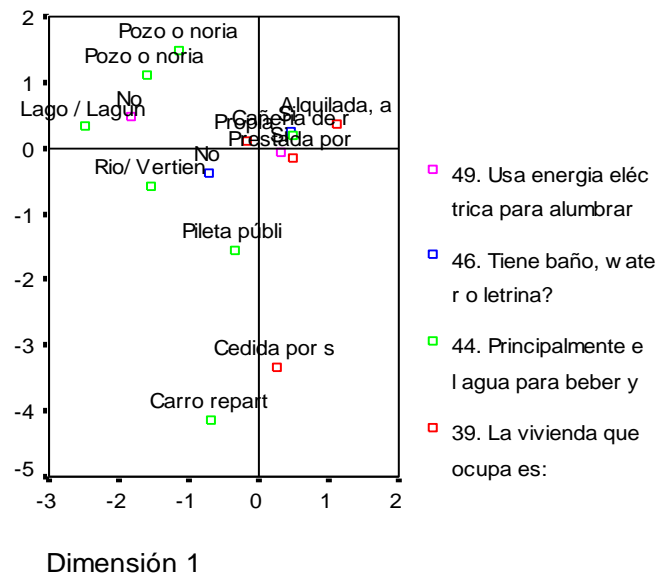
	Porcentaje	Acumulado
Si	93.21	93.21
No	6.79	100.00
Total	100.00	

Fuente: Carrera de Estadística

Una vez descrito los resultados, aplicamos el método de análisis factorial de correspondencias donde el Gráfico 3, nos muestra que la mayoría tiene casa propia y que tiene agua por cañería y cuenta con luz eléctrica y además tiene baño.

En resumen podemos decir de acuerdo a los resultados de la encuesta sociodemográfica realizada a los habitantes del departamento de La Paz, la mayoría habla castellano, que nacieron en el mismo municipio, que trabajan por cuenta propia, que la vivienda que ocupan es propia, que tiene agua para beber a través cañería de red y que tienen luz eléctrica.

Gráfico 3



Referencias

[1] Hair, Joseph F., Anderson, Rolph E., Tatham, Ronald L. & Black William C. (2000). Análisis Multivariante. 5ta. Ed. Madrid: Prentice Hall, 799p.

[2] Ho Robert. (2014). Handbook of Univariate and Multivariate Data Analysis whit IBM SPSS. New York: CRC Press, 561p.

[3] Johnson, Richard A., Wichern Dean W. (2007). Applied Multivariate Statistical Analysis.6ta. Ed. United States of America, New Jersey: PEARSON Prentice Hall, 773p.

[4] Mardia K.V., Kent J.T., Bibby J.M.(2003). Multivariate Analysis. California: Academic Press, 518p.

[5] Peña Daniel. (2002). Análisis de Datos Multivariante. España: McGraw Hill, 539p.

[6] Perez, Cesar. (2013). Análisis Multivariante de Datos. España, Madrid: Garceta, grupo editorial, 563p.

Clasificación mediante Redes Neuronales y Conglomerados No Jerárquicos de las Condiciones de Vida de los Hogares de Bolivia

Lic. Luis Fernando Flores Rivero
✉ luisfernandofloresr@gmail.com

1. Introducción

En los últimos años se han desarrollado metodologías para medir condiciones de vida de los hogares bolivianos, mediante indicadores de pobreza, tales como, el NBI, la incidencia de pobreza y la brecha de pobreza, que son indicadores descriptivos que resumen la información sin mostrar un panorama con la extensión de los datos como lo hace el análisis de conglomerados o las redes neuronales.

En este sentido, se pretende investigar cuál de los dos métodos (conglomerados no jerárquicos o redes neuronales) es óptimo.

Las redes neuronales artificiales son sistemas conexionistas dentro del campo de la inteligencia artificial, las cuales, dependiendo del tipo de arquitectura neuronal, pueden tener diferentes aplicaciones. Pueden utilizarse para el reconocimiento de patrones, la comprensión de información, el modelaje y la reducción de la dimensionalidad, el agrupamiento, la clasificación, etc.

El análisis de conglomerados reúne un conjunto de técnicas del análisis multivariante, cuya principal aplicación reside en la clasificación u obtención de tipologías que no están previamente identificadas.

2. Objetivo

El presente trabajo tiene como objetivo clasificar las condiciones de vida en los hogares en Bolivia con ambas metodologías, comparar sus resultados y analizar sus diferencias.

3. Metodología

Se utilizará al indicador de Necesidades Básicas Insatisfechas (NBI) como variable control o variable tipo. Dicho indicador servirá como punto de referencia para la evaluación de cada método. Además de lo anterior, el Indicador NBI permitirá evitar los problemas de variables con diferentes unidades o escalas de medida (proceso de estandarización).

3.1. Metodología NBI

El indicador NBI está conformado por cuatro componentes: Vivienda, Insumos y Servicios Básicos, Educación y Salud. Los pasos para el cálculo del indicador son los siguientes: 1) se determinan normas mínimas para cada uno de los componentes, 2) en relación a la norma, se elabora un índice de carencia que refleje la necesidad de un hogar para cada componente, 3) se procede a agregar todos estos componentes en un sólo índice de carencia (NBI) para cada hogar, 4) se agregan todos los hogares en cinco estratos de pobreza según su valor NBI (Figura 1).

3.2. Metodología de conglomerados

El análisis de conglomerados se divide en los grandes partes: El método jerárquico y el método no jerárquico. Por fines propios del trabajo se eligió el algoritmo no jerárquico de k-medias. Este procedimiento está dividido en cuatro etapas: 1) Establecer el número de conglomerados, 2) Seleccionar los centros iniciales, 3) Asignar los objetos a los grupos y 4) Optimizar o validar los resultados (Figura 2).

Figura 1 Metodología NBI

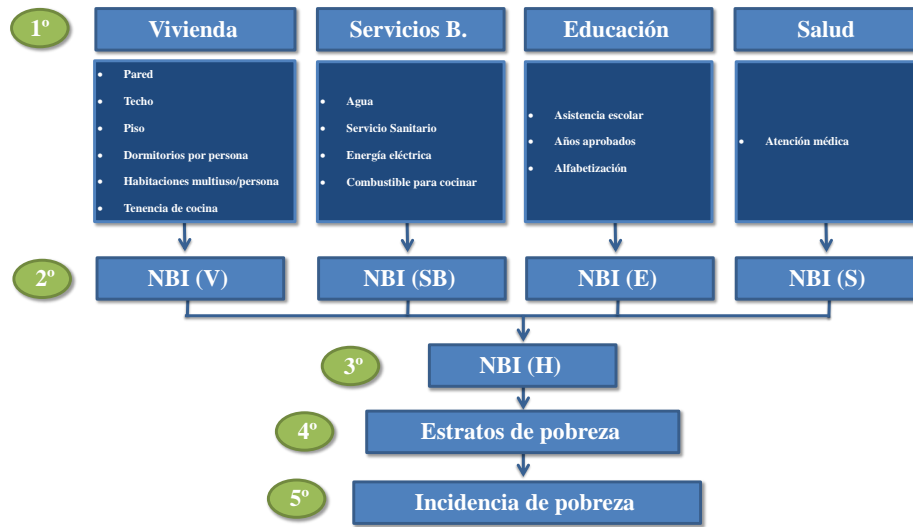
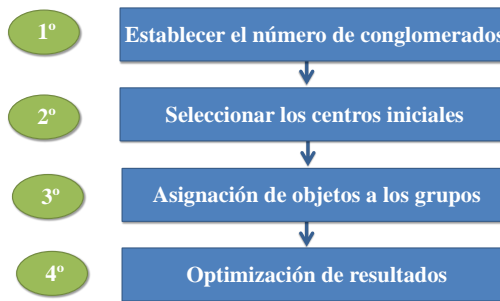


Figura 2 Metodología de conglomerados no jerárquicos



3.3. Metodología de la red neuronal

El modelo de red neuronal elegido para satisfacer las necesidades del trabajo, es la Red Neuronal Bacpropagation (BPN).

En una BPN existe una capa de entrada con “n” neuronas y una capa de salida con “m” neuronas y al menos una capa oculta de neuronas internas. Cada neurona de una capa (excepto las de entrada) recibe entradas de todas las neuronas de la capa anterior y envía su salida a todas las neuronas de la capa posterior (excepto las de salida). La estructura de la red BPN se muestra en la Figura 3.

La aplicación del algoritmo tiene dos fases, una hacia delante y otra hacia atrás. Durante la primera fase el patrón de entrada es presentado a la red y propagado a través de las capas hasta llegar a la capa de salida.

Obtenidos los valores de salida de la red, se inicia la segunda fase y se comparan los resultados con la salida esperada para así obtener el error. Se ajustan los pesos de la última capa proporcionalmente al error. Se pasa a la capa anterior con una retropropagación del error, ajustando los pesos y continuando con este proceso hasta llegar a la primera capa.

El proceso se repite hasta que el error sea nulo o se considere mínimo.

4. Análisis comparativo de resultados

Con el objetivo de verificar la precisión de cada método en la clasificación de las condiciones de vida de los hogares bolivianos, se compararon los resultados obtenidos de cada metodología y se analizaron las

posibles diferencias por estratos de pobreza.

Figura 3 ➤ Red Backpropagation

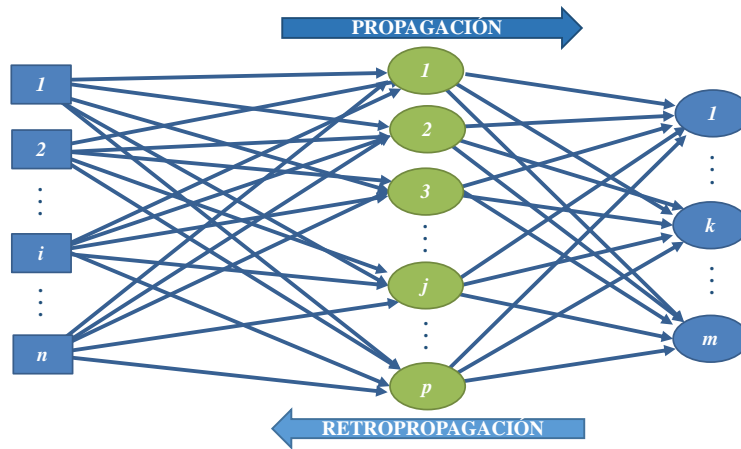


Figura 4 ➤ Metodología Red BPN

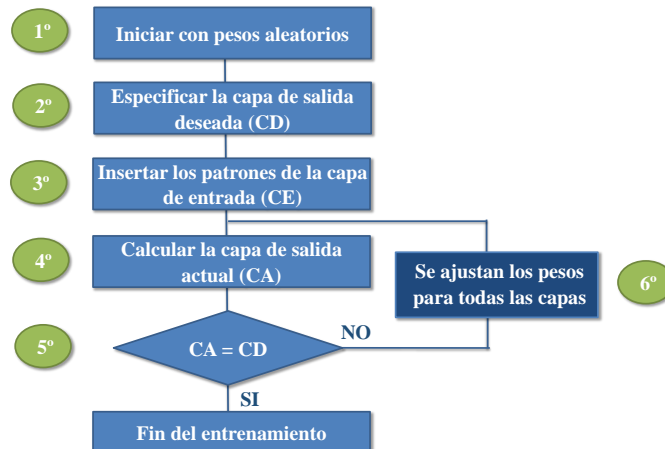
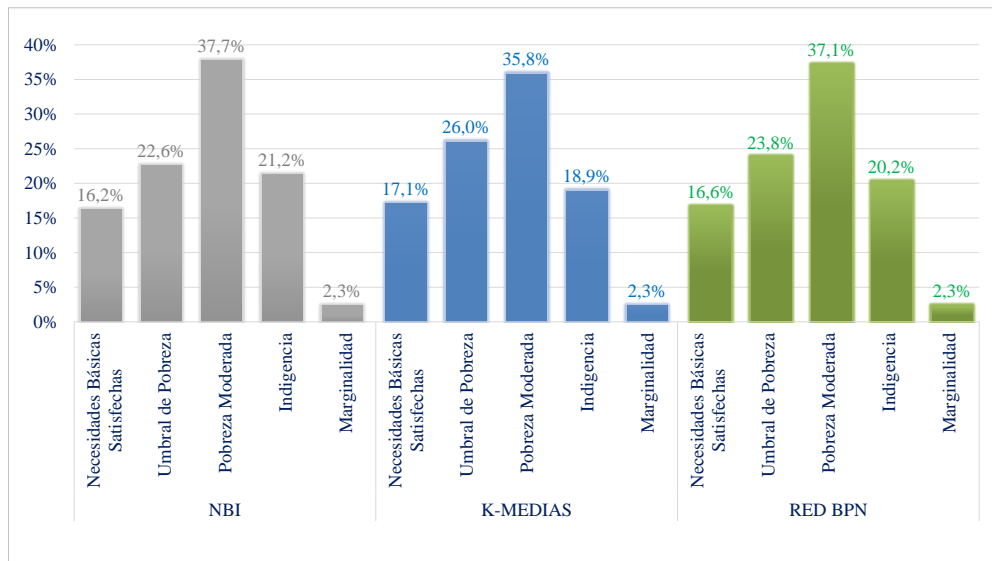


Figura 5 ➤ Estratos de pobreza según metodología de clasificación

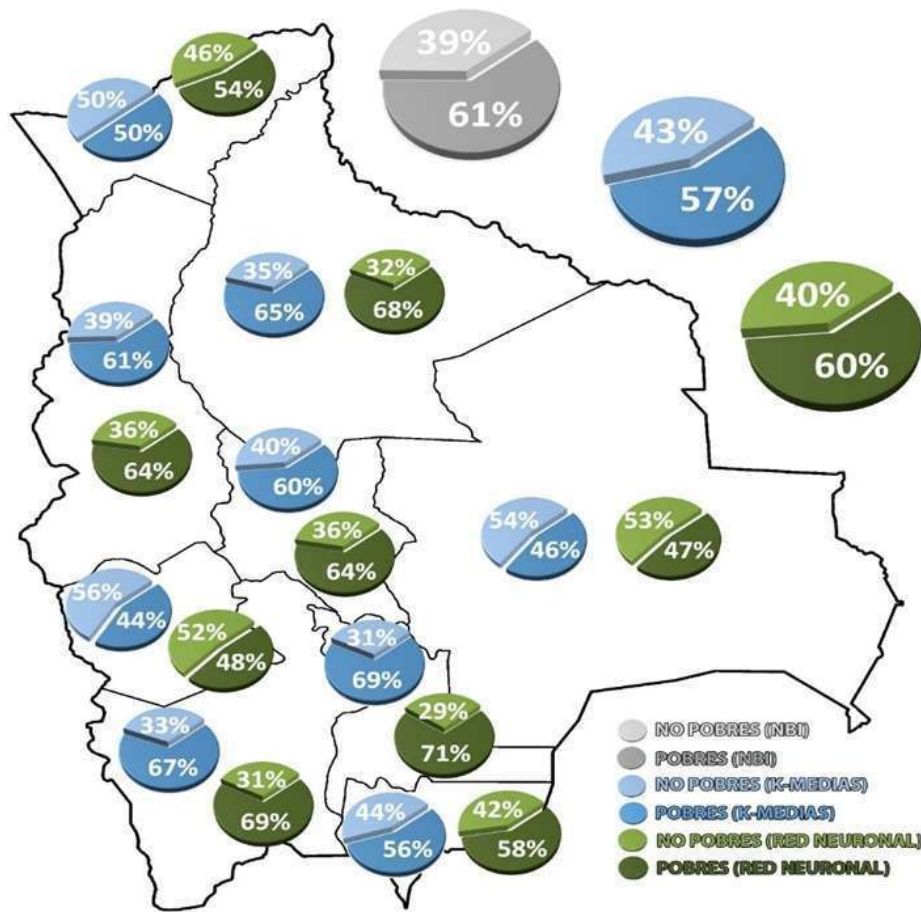


Luego, se contrastaron los resultados por condición de pobreza y a nivel departamental; y finalmente, se realizó el análisis de concordancia a través del estadístico Kappa, con el objetivo de verificar la optimalidad de cada método.

Según la Figura 5, la variable de control (NBI) muestra que la mayoría (37,7 %) de los hogares en Bolivia tienen pobreza moderada, seguido de un 22,6 % de hogares que cumplen con lo necesario para satisfacer sus necesidades básicas. Un 21,2 % está compuesto por hogares indigentes, tan solo el 16,2 % del total tiene satisfacción plena y el 2,3 % vive en condiciones marginales. Se observa también que los resultados obtenidos son bastante similares a los de la red neuronal BPN.

En relación a la variable control, el resultado de la clasificación por el método de conglomerados (k-medias) presenta menos hogares con pobreza moderada e indigencia, y aumenta considerablemente el umbral de pobreza en más de tres puntos porcentuales. Bajo esta clasificación, existirían menos hogares pobres en el territorio nacional.

Figura 6 ► Pobreza en los hogares de Bolivia según metodología de clasificación



Por otro lado y en referencia a la Figura¹ 6, se puede observar que según la clasificación NBI, el 61,0 % de los hogares en Bolivia son pobres y el 39,0 % no sufren de pobreza. Situación similar ocurre con los resultados de la red neuronal (40,0 % no pobres y 60,0 % pobres). Sin embargo la clasificación por conglomerados muestra mayor cantidad de hogares no pobres (43,0 %) y menor porcentaje de hogares pobres (57,0 %).

¹Dado que el diseño muestral de la Encuesta de Hogares no incluye en su planteamiento la desagregación por departamento, se calcularon los errores de muestreo relativos para este nivel de desagregación y se obtuvieron resultados dentro del rango de aceptación.

Las mayores diferencias de clasificación se presentan en los departamentos de Cochabamba y Pando, seguidos de Oruro, La Paz, Beni y Tarija.

Según la metodología BPN, Chuquisaca es el departamento más pobre del país, con un 71,0 % de hogares que no cubren sus necesidades básicas; seguido de Potosí y Beni con niveles de pobreza del 69,0 % y 68,0 % respectivamente. Aproximadamente la mitad de los hogares del departamento de Oruro (48,0 %) son no pobres.

De acuerdo a las gráficas y al análisis anterior todo parece indicar que las redes neuronales logran clasificar de mejor forma las condiciones de vida de los hogares de Bolivia.

Con el fin de contrastar ambas metodologías y analizar sus diferencias se presenta la siguiente tabla:

Tabla 1 Comparación de resultados por estrato de pobreza, según método de clasificación

MÉTODO	ESTRATOS DE POBREZA	Necesidades Básicas Satisfechas	Umbral de Pobreza	Pobreza Moderada	Indigencia	Marginalidad
K-MEDIAS	Necesidades Básicas Satisfechas	386.679	0	0	0	0
	Umbral de Pobreza	19.702	517.655	0	0	0
	Pobreza Moderada	0	101.006	796.977	0	0
	Indigencia	0	0	55.536	449.780	0
	Marginalidad	0	0	0	0	55.164
TOTAL		406.381	618.661	852.513	449.780	55.164
RED BPN	Necesidades Básicas Satisfechas	386.679	0	0	0	0
	Umbral de Pobreza	8.933	528.424	0	0	0
	Pobreza Moderada	0	37.716	860.267	0	0
	Indigencia	0	0	23.770	481.546	0
	Marginalidad	0	0	0	0	55.164
TOTAL		395.612	566.140	884.037	481.546	55.164

Fuente: Encuesta de Hogares. Elaboración propia.

Los valores de la diagonal principal, representan los casos clasificados de forma correcta y los valores fuera de la diagonal principal (resaltados en color celeste) representan los errores de clasificación.

Tabla 2 Test de concordancia de kappa, según método de clasificación

Método	Valor	Significación aproximada	Número de casos válidos
K-Medias	0,90	0	2.382.499
Red BPN	0,96	0	2.382.499

Fuente: Encuesta de Hogares. Elaboración propia.

De acuerdo al estadístico Kappa, el método de conglomerados no jerárquicos alcanza un 90,0 % de objetos correctamente clasificados, mientras que la clasificación por redes neuronales muestra un 96,0 % de éxito. Por lo que se puede concluir que la técnica de clasificación por redes neuronales es 6 puntos porcentuales más efectiva.

La superioridad de las redes neuronales viene acompañada de recursos temporales y computacionales considerables. Por el contrario, y con menor nivel de precisión en la clasificación (10,0 %); la fácil accesibilidad, los pocos recursos informáticos y temporales, benefician el método por conglomerados no jerárquicos.

Finalmente y habiendo descrito brevemente los pros y contras de cada procedimiento, se deja a consideración del lector la elección de un método ante otro y se recomienda la utilización de ambos para un mayor

enriquecimiento y complementación en el análisis de resultados.

Referencias

- [1] Choque Aspiazu, G., Loza Herrera, R. y Mendez Quintanilla, R. (2009). *Redes Neuronales Artificiales, Aplicaciones en MATLAB*. La Paz, Bolivia.
- [2] Freeman, J.A. and Skapura, D.M. (1993). *Redes Neuronales. Algoritmos, aplicaciones y técnicas de programación*. Addison Wesley/Díaz de Santos. USA.
- [3] Hecht-Nielsen, R. (1988). Neurocomputing: picking the human brain. *IEEE. Spectrum*, 25, No.3, 36-41.
- [4] Hiler, J. y Martínez, V. (1995). *Redes Neuronales Artificiales. Fundamentos, Modelos y aplicaciones*. Rama. Madrid, España.
- [5] Johnson, D.E. (2000). *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores.
- [6] Luque Martínez, T. (2000). *Técnicas de análisis de datos en investigación de mercados*. Ediciones Pirámide. Madrid, España.
- [7] Martín, B. y Sanz, A. (2002). *Redes Neuronales y Sistemas Difusos*, 2ª ed. Alfaomega Rama. Madrid, España.
- [8] Peña, D. (2002). *Análisis de datos multivariantes*. McGRAW-HILL. Madrid, España.
- [9] Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos, Aplicaciones con SPSS*. Pearson Educación S. A. Madrid, España.

Superficies de Respuesta mediante un Diseño Central Compuesto

Lic. María Gilka Jiménez Careaga
✉ gilka.jimenez@gmail.com

1. Introducción

La Metodología de Superficies de Respuesta (MSR) es una combinación de análisis de Regresión y Diseño Experimental que fue introducida por Box y Wilson en 1951. Es una estrategia de experimentación secuencial y modelación que conduce a la localización de los valores óptimos de las variables independientes que maximizan, minimizan o cumplen ciertas restricciones en la variable respuesta.

2. Construcción de una superficie de respuesta

La investigación de una superficie de respuesta en primera instancia aborda la localización de la región en la cual se encuentra la respuesta óptima Y , en las siguientes etapas, luego de comprobar experimentalmente, va reduciendo la superficie de respuesta en estudio, en algunos casos hasta obtener la convergencia en un punto óptimo.

El proceso se inicia con la planificación de un diseño experimental exploratorio que permite la aproximación a la superficie mediante la representación matemática de los modelos más simples, entre estos tenemos:

Un modelo de primer orden (lineal) sin interacciones o productos cruzados:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (1)$$

Modelo de primer orden (lineal) con interacciones o productos cruzados:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=2}^k \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \varepsilon, \quad (2)$$

donde ε representa el ruido o error observado en la respuesta Y , x_i son factores cuantitativos observados. En esta primera fase, se utilizan experimentos factoriales completos 2^k o fraccionarios 2^{k-p} , con el fin de estimar las respuestas medias para un modelo lineal o de primer orden, como el de la ecuación (2). Se recomienda generalmente agregar dos o más observaciones en el nivel medio de cada uno de los factores para estimar el error experimental y tener un mecanismo de evaluación para saber si el modelo lineal es apropiado.

En la siguiente etapa, una vez identificada la región de respuesta óptima, los diseños factoriales completos o fraccionarios a dos niveles no son suficientes, pues se requieren al menos tres niveles para cada factor y el diseño debe de tener $1 + 2k + k(k-1)/2$ puntos distintos para estimar los parámetros de un modelo de regresión cuadrática.

Modelo cuadrático o de segundo orden:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=2}^k \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \varepsilon \quad (3)$$

Entre los diseños experimentales que pueden ser utilizados en esta fase están los experimentos factoriales 3^k , sin embargo son considerados poco prácticos por el elevado número de combinaciones de tratamientos que requiere. Existen otros diseños desarrollados para la aproximación a una superficie de segundo orden, que no requieren tantas combinaciones de tratamientos, entre estos están los diseños centrales compuestos propuestos por Box y Wilson, y los diseños Box-Behnken.

3. Análisis de la superficie de respuesta

La naturaleza secuencial e iterativa de la metodología de Superficie de Respuesta requiere que en cada fase se identifiquen los factores significativos, se realice un análisis de adecuación del modelo y se determine la dirección en la cual se encuentran las condiciones óptimas de experimentación.

Para identificar los factores que tienen mayor influencia en la respuesta y examinar la bondad de ajuste del modelo se utiliza el análisis de varianza. Mediante el método de máxima pendiente se determina la dirección en la cual se encuentra la región óptima. El análisis de la superficie ajustada se realiza mediante el análisis canónico, el cual permite determinar la naturaleza del punto estacionario.

En el caso de un modelo de segundo orden, la ecuación escrita en forma matricial es:

$$\hat{Y} = \hat{\beta}_0 + x'b + x' Bx \tag{4}$$

El punto crítico, si es que existe, será un conjunto de condiciones en x tal que las derivadas parciales $\frac{\partial \hat{Y}}{\partial x_i}$ son simultáneamente cero, es decir:

$$\begin{aligned} \frac{\partial \hat{Y}}{\partial x} &= \frac{\partial}{\partial x}(x'b + x' Bx) = 0 \\ b + 2Bx &= 0, \end{aligned}$$

siendo el punto estacionario

$$x_0 = \frac{-B^{-1}b}{2}. \tag{5}$$

El punto estacionario puede ser un punto máximo, un punto mínimo o un punto de ensilladura (minimax), la forma canónica de la función permite establecer su condición.

Reemplazando (5) en (4)

$$\hat{Y} = \hat{\beta}_0 + x'_0 b/2.$$

Forma canónica:

$$\hat{Y} = \hat{Y}_0 + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \dots + \lambda_k w_k^2,$$

donde \hat{Y}_0 es la respuesta estimada en el punto x_0 , λ_i son las raíces características de la matriz B y $w = M'z$ (siendo M la matriz de los vectores propios asociados a λ_i).

Si todos los λ_i son negativos, x_0 representa un punto de máxima respuesta; si los λ_i son todos positivos, x_0 es un mínimo de la superficie observada. En el caso de que los λ_i difieran en signo, el punto estacionario es un punto de ensilladura. En general si algún λ_i es cero o un valor muy cercano a cero, el punto estacionario no es único, se encuentra en una loma estacionaria.

4. Diseño central compuesto

Para comprender los diseños centrales compuestos, que se utilizarán en el ejemplo que se abordará, es necesario mencionar que los factores generalmente se codifican, pues facilita los cálculos, estandariza las variables eliminando las unidades de medida de las variables originales y se incrementa la precisión al estimar los coeficientes del modelo.

Los niveles codificados de los factores de un diseño factorial 2^k se obtienen mediante la siguiente fórmula: $x_i = 2(A_i - \bar{A})/R_i$, donde A_i es el i -ésimo nivel del factor A , \bar{A} es el nivel promedio del factor A y R_i es el rango entre el valor más alto y el más bajo de del factor A . Con este procedimiento, se asigna valores ± 1 a las variables asociadas al arreglo factorial 2^k .

Los diseños centrales compuestos son diseños de tratamientos factoriales 2^k con $2k$ combinaciones adicionales llamadas puntos axiales y n_c puntos centrales. Las coordenadas de los puntos axiales de los ejes del factor codificado son $(\pm\alpha, 0, 0, \dots, 0)$, $(0, \pm\alpha, 0, \dots, 0)$, \dots , $(0, 0, 0, \dots, \pm\alpha)$, y los puntos centrales son de

la forma $(0,0,0,\dots,0)$. Dependiendo de la elección de α en los puntos axiales, el diseño central compuesto puede tener diferentes propiedades como ortogonalidad, rotabilidad y uniformidad. Se considerará solamente una propiedad deseable en estos diseños consistente en que la varianza de los valores estimados sea constante en puntos equidistantes del centro del diseño. Esta propiedad llamada rotabilidad se logra estableciendo $\alpha = (2^k)^{1/4}$. Así, el valor de α para un diseño con dos factores es $\alpha = 1,414$ y para tres factores $\alpha = 1,682$. La fórmula para α cambiará si se realizan réplicas del diseño o si se utiliza un diseño factorial fraccionario.

5. Ejemplo de aplicación

Para ejemplificar el procedimiento de superficies de respuesta se utilizarán los datos del ejercicio propuesto 5.6 del libro *Response Surfaces* de André I. Khuri, John A. Cornell.

Se aplicaron a parcelas experimentales de maní dos tipos de fertilizantes, uno estándar de la combinación de Nitrógeno-Fósforo-Potasio (Factor A) y el otro un suplemento nutricional (Factor B), se observará los efectos en el rendimiento del maní medido en libras por parcela. Los niveles de cantidad (libras/parcela) de cada fertilizante aplicado fueron determinados de acuerdo a lo requerido por el diseño central compuesto rotatable.

En el siguiente cuadro se presenta los datos utilizados, el procedimiento se realizó aplicando el software Statgraphics, en el cual se eligió el diseño compuesto 2^k estrella y el modelo cuadrático para la estimación del rendimiento (Y).

Cuadro 1 ◉ Niveles de los factores, variables codificadas y variable respuesta

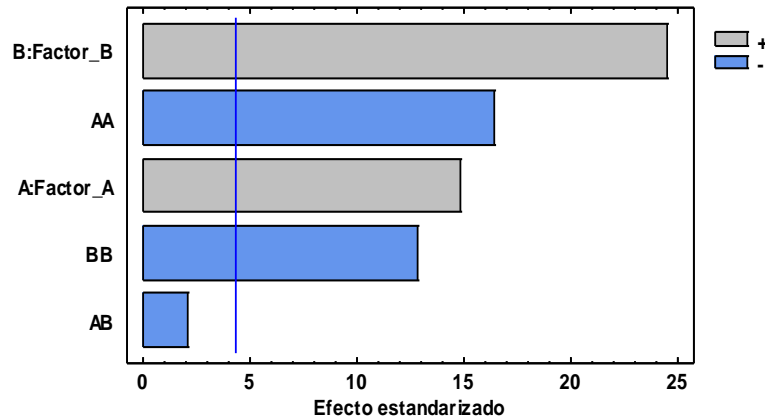
Factor 1	Factor 2	x_1	x_2	Rendimiento (Y) (lb/parcela)	
				Réplica 1	Réplica 2
50	15	-1	-1	7,52	8,12
120	15	1	-1	12,37	11,84
50	25	-1	1	13,55	12,35
120	25	1	1	16,48	15,32
35,5	20	-1,414	0	8,63	9,44
134,5	20	1,414	0	14,22	12,57
85	12,9	0	-1,414	7,9	7,33
85	27,1	0	1,414	16,49	17,4
85	20	0	0	15,73	17
85	20	0	0	16,3	16,3

Cuadro 2 ◉ Efectos estimados para rendimiento (lb)

Efecto	Estimado	Int. Confianza	V.I.F.
promedio	16,33250	$\pm 0,97102$	
A: Factor A	3,35025	$\pm 0,97102$	1,000
B: Factor B	5,52991	$\pm 0,97102$	1,000
AA	-4,89439	$\pm 1,28454$	1,225
AB	-0,66750	$\pm 1,37323$	1,000
BB	-3,82938	$\pm 1,28454$	1,225
bloque	-0,15200	$\pm 0,44010$	1,000

Intervalos de confianza del 95.0 de confianza (intervalos) basados en el error puro con 2 g.l. ($t = 4.30265$).

Gráfico 1 Diagrama de Pareto estandarizada para VRen



El Cuadro 2 muestra las estimaciones para cada uno de los efectos estimados y las interacciones. Igualmente se muestran los intervalos de confianza del 95,0 % para los estimados. Note también que el factor de inflación de varianza (V.I.F.) más grande, es igual a 1,225. Para un diseño perfectamente ortogonal, todos los factores serían igual a 1.

Cuadro 3 Análisis de varianza para el rendimiento (Y)

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
A: Factor A	44,8965	1	44,8965	220,38	0,0045
B: Factor B	122,319	1	122,319	600,41	0,0017
AA	54,754	1	54,754	268,76	0,0037
AB	0,891112	1	0,891112	4,37	0,1716
BB	33,5179	1	33,5179	164,52	0,0060
bloques	0,11552	1	0,11552	0,57	0,5300
Falta de ajuste	9,90654	11	0,900595	4,42	0,1988
Error puro	0,40745	2	0,203725		
Total (corr.)	241,688	19			

R-cuadrada = 95,7325 por ciento

Error absoluto medio = 0,606823

R-cuadrada (ajustada por g.l.) = 93,7629 por ciento

Estadístico Durbin-Watson = 1,88927 ($P=0,3267$)

Error estándar del est. = 0,451359


Autocorrelación residual de Lag 1 = 0,0523126

El cuadro 3 (ANOVA) particiona la variabilidad de Y (rendimiento) en valores separados para cada uno de los efectos, prueba la significancia estadística de cada efecto comparando su cuadrado medio contra un estimado del error experimental. En este caso, 4 efectos tienen una valor-P menor que 0,05 indicando que son significativamente diferentes de cero con un nivel de confianza del 95,0 %.

La prueba de falta de ajuste está diseñada para determinar si el modelo seleccionado es adecuado para describir los datos observados o si se debería usar un modelo más complicado. La prueba se realiza comparando la variabilidad de los residuos del modelo actual con la variabilidad entre observaciones obtenidas en condiciones repetidas de los factores. Dado que el valor-P para la falta de ajuste en la tabla ANOVA es mayor que 0,05, el modelo cuadrático parece ser adecuado para los datos observados al nivel de confianza del 95,0 %.

El estadístico R-cuadrado indica que el modelo, así ajustado, explica 95,7325 % de la variabilidad en Y. El estadístico R-cuadrado ajustado, que es más adecuado para comparar modelos con diferente número de variables independientes, es 93,7629 %. El error estándar del estimado muestra que la desviación

estándar de los residuos es 0,451359. El error medio absoluto (MAE) de 0,606823 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) prueba si en los residuos hay alguna correlación significativa basada en el orden en que se presentan los datos en el archivo. Puesto que el valor- P es mayor que 5,0 %, no hay indicación de autocorrelación serial en los residuos con un nivel de significancia del 5,0 %.

Cuadro 4  Coeficientes de regresión para VRen

Coeficiente	Estimado
constante	16,3325
A: Factor A	1,67512
B: Factor B	2,76495
AA	-2,44719
AB	-0,33375
BB	-1,91469

La ecuación del modelo ajustado es:

$$\text{Rendimiento} = 16,3325 + 1,67512 \times \text{Factor}_A + 2,76495 \times \text{Factor}_B - 2,44719 \times \text{Factor}_A^2 - 0,33375 \times \text{Factor}_A \times \text{Factor}_B - 1,91469 \times \text{Factor}_B^2$$

Gráfico 2  Interacción para VRen

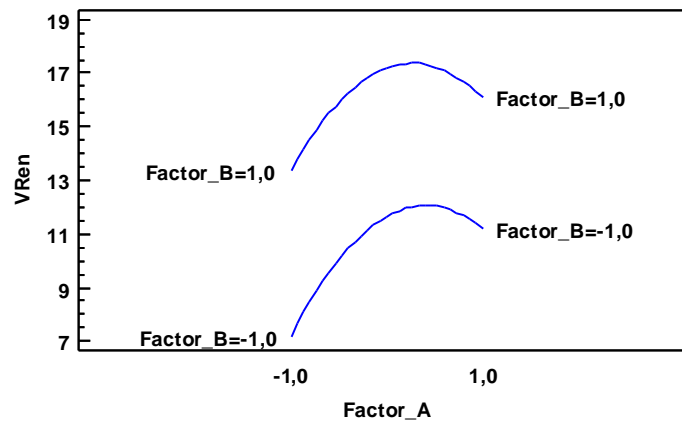
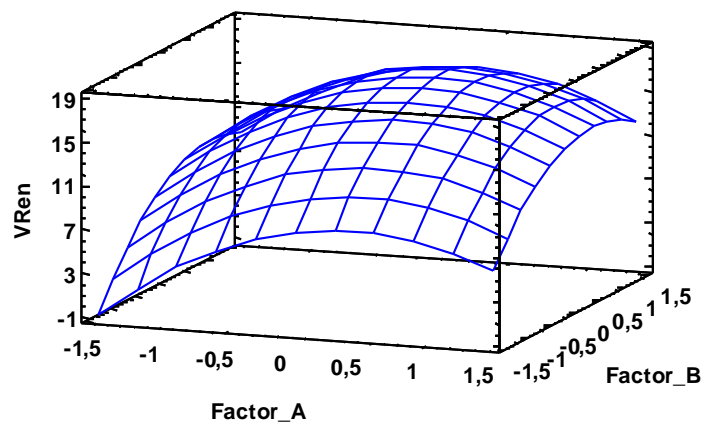


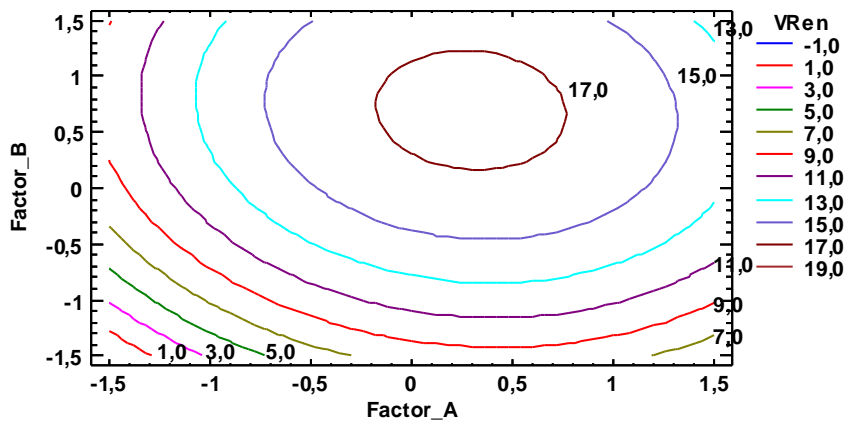
Gráfico 3  Superficie de respuesta estimada



Cuadro 5 Camino de máximo ascenso para rendimiento (lb)

Factor A	Factor B	Predicción para Rendimiento (Y) (lb)
0,0	0,0	16,3325
1,0	0,913353	16,1837
2,0	1,21709	9,61052
3,0	1,51843	-2,40338
4,0	1,81851	-19,8535
5,0	2,11775	-42,7373

Gráfico 4 Contornos de la superficie de respuesta estimada



Para esta superficie el punto estacionario es:

$$x_0 = \frac{-B^{-1}b}{2} = -\frac{1}{2} \begin{bmatrix} -2,447190 & -0,166875 \\ -0,166875 & -1,914690 \end{bmatrix}^{-1} \begin{bmatrix} 16,7512 \\ 2,76495 \end{bmatrix} = \begin{bmatrix} -3,3934690 \\ -0,4262777 \end{bmatrix}$$

$$\hat{Y}_0 = 19,5422679$$

Se observa que el Rendimiento (Y) es mayor para este punto.

Este es el trayecto, desde el centro de la región experimental actual, a través del cual la respuesta estimada (Y) cambia más rápidamente con un cambio menor en los factores experimentales. Indica buenas características para ejecutar experimentos adicionales si el objetivo es incrementar el Rendimiento. Se nota que la mayor respuesta (16,3325) se encuentra próxima al origen por tanto los nuevos puntos experimentales deberían estar en esa dirección.

Referencias

- [1] Box, George E. P.; Draper, Norman R. Empirical Model – Building and Response Surfaces.
- [2] Cochran, William G.; Cox, Getrude M. Diseños Experimentales.
- [3] Draper, N. R. Center Points in Second Order Response Surface Design. Technometrics Vol.18.
- [4] Heiberger, Richard M. Computation for the Analysis of Designed.
- [5] Khuri, André I.; Cornell, John A. Response Surfaces Designs and Analyses.

Tamaño de Muestra en Poblaciones Agropecuarias

Lic. Jaime T. Pinto Ajuacho
✉ titojaime_pinto@yahoo.com

1. Población, muestra y marco de muestreo

Población, denominaremos a un conjunto de unidades del que se desea obtener cierta información, limitándose las observaciones a una parte que denominamos, muestra. Los datos obtenidos a partir de ella nos permiten inferir unos valores aproximados de la población en su totalidad, a estos valores aproximados se les denomina “estimaciones” las cuales vendrán afectadas de un error que denominaremos “error debido al muestreo”.

La estimación, considera un “Marco de muestreo” que es el conjunto de información que puede ser útil en cualquier momento del diseño de la encuesta, en el que debemos distinguir el listado y la información complementaria.

El listado, es la información de la lista de unidades a muestrear que también se conoce con el nombre de “marco” en un sentido restringido, es una de las operaciones más delicadas en el diseño muestral.

Un marco por lista, como su nombre lo indica, es una lista de los elementos de la población, junto con los datos de identificación adecuados, como parte del marco, puede incluirse información suplementaria que permita la clasificación de los elementos.

Un marco por lista que consista de los nombres y direcciones de agricultores se adapta muy bien para la recolección de informes agrícolas, el bajo costo de la recolección de datos por este medio constituye una de las principales ventajas; una desventaja es que hay dificultad para ajustar la lista con los cambios que ocurre en el universo o población.

2. Estratificación y submuestreo

Con la estratificación o división de la población en subpoblaciones o estratos, obteniéndose posteriormente muestras independientes de cada estrato, pueden perseguirse diversos fines, como por ejemplo.

- Dar estimaciones separadas para ciertas subpoblaciones.
- Agrupar unidades de muestreo homogéneas entre si en estratos, con objeto de mejorar la precisión de las estimaciones globales.
- Utilizar métodos diferentes de muestreo en los distintos estratos (según el marco de que se disponga).

En la estratificación se da lugar a la llamada “Afijación óptima”, con este criterio se deberán muestrear con más intensidad los estratos que presentan mayor variabilidad interna, menor costo o un tamaño superior al medio. También en este caso existen expresiones matemáticas que determinan el tamaño óptimo de la muestra en cada estrato.

El “Muestreo estratificado”, es un recurso empleado para reducir el error en el muestreo comprende la subdivisión del universo entre un número de estratificaciones o niveles y posteriormente asignar a cada estrato una porción de la muestra.

El “Submuestreo” ó muestreo en dos etapas, consiste en agrupar primero las unidades individuales de muestreo en conglomerados de unidades primarias de muestreo (a veces, muy grandes), obteniendo una muestra de conglomerados y posteriormente sacando una muestra de las unidades individuales de cada conglomerado muestral.

Muchas encuestas sobre unidades agropecuarias que se hacen por medio de submuestreo, han usado a los segmentos agropecuarios como unidades de muestreo. En la primera etapa se selecciona una muestra de segmentos (Conglomerado de Unidades Agropecuarias) y en la segunda etapa se seleccionan las muestras de las unidades agropecuarias en estos segmentos seleccionados, donde un segmento es un conglomerado de unidades agropecuarias.

Es así que el país en el sector agropecuario se divide en áreas que son unidades primarias de muestreo (UPM) usando mapas de escala pequeña para este propósito.

Respecto de la segmentación, la mayoría de los diseños de muestra requieren que las UPM (Unidades Primarias de muestreo) sean divididas en unidades más pequeñas (sectores) con límites claramente definidos; éstas, a su vez, suelen dividirse en segmentos y se escogen uno o más segmentos para la muestra.

Hablando de la delimitación del segmento de muestra, uno de los usos principales de un mapa es el de indicar los límites del área terrestre que debe abarcar la muestra. Si el diseño de muestra se basa en datos para un área definida con exactitud, entonces el mapa debe indicar claramente los límites de dicha área.

3. Cobertura

La población objetivo de estudio está constituida por todos los productores agropecuarios del país, agrupados en cada uno de los nueve departamentos.

4. La unidad de investigación

Es la “Unidad Agropecuaria del Productor”.

5. Marco muestral

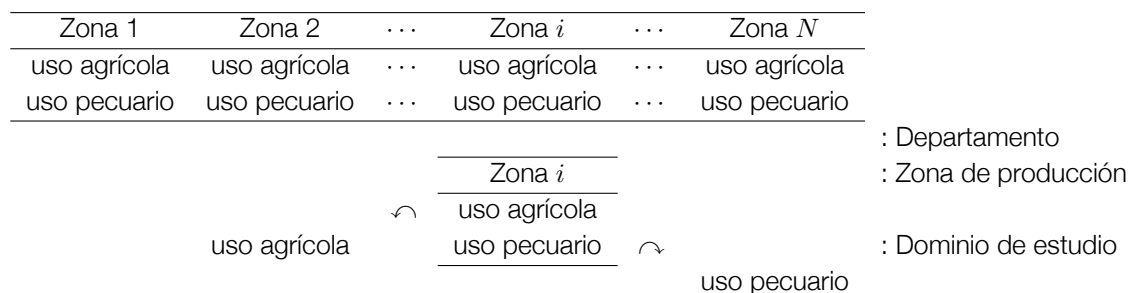
La construcción del marco muestral, utilizado en la Encuesta Nacional Agropecuaria, es realizada a partir de una revisión y adecuación de los marcos existentes en el Instituto Nacional de Estadística, a fin de obtener el marco estratificado. Los insumos utilizados en la generación del Marco Estratificado, son:

- o Marco elaborado en el Censo Agropecuario.
- o Estratificación del país, en Zonas de Producción Agropecuaria a nivel de cantones.
- o Cartografía a nivel cantonal y de segmento.
- o Tipo de actividad Agrícola o Pecuario.

El marco censal agropecuario, tiene la división político administrativa del país, departamentos particionados en provincias y cantones; luego las unidades de desagregación estadísticas como los sectores y segmentos agropecuarios, este último definido como un conglomerado en promedio 100 unidades agropecuarias en la zona del altiplano y valles, 70 unidades agropecuarias en las zonas tropicales.

Los subuniversos armados, se considera conformados por la intersección de Zonas de producción agrícola o Zonas de producción pecuaria, denominados “Dominios de Estudios” (Ver esquema 1).

Esquema 1 ➤ Subuniverso (Dominio de estudio)



A nivel nacional, el número de segmentos por tipo de actividad (agrícola o pecuario), según zonas de producción y departamento.

Para determinar el subuniverso de tipo de actividad agrícola, se considero la superficie cultivada y para los de uso pecuario se convertía el número de especies ganaderas a “unidades animal”.

El Listado de segmentos agropecuarios de los subuniversos, ya sea Agrícola o Pecuario, tiene las siguientes características:

Cuadro 1 ➤ Lista de segmentos agropecuarios. Subuniverso: xxxxx

Nro. de segmento	Identificadores geográficos	Nro. unidades de observación	Superficie agrícola	Existencia ganadera
1	—	A	—	—
2	—	B	—	—
3	—	C	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
<i>N</i>	—	—	—	—

El Marco Muestral está conformado por Identificadores geográficos y Variables Agropecuarias.

6. Identificadores geográficos

Departamento – Provincia – Zona de Producción – Cantón –Sector – Segmento.

7. Variables agropecuarias

SUPERFICIE AGRÍCOLA: Superficie Cultivada – Superficie de Pastos Naturales – Superficie de Pastos Cultivados – Superficie Resto – Superficie Total.

EXISTENCIA GANADERA: Total de Bobinos – Total de Ovinos – Total de Porcinos – Total Camelidos – Total de Equinos – Total aves de corral.

8. Estratificación

En cada subuniverso, tomando en cuenta su participación pequeña, media o grande en la actividad agrícola o pecuaria, los estratos son definidos:

ESTRATO 1: Segmentos con participación agropecuaria pequeña.

ESTRATO 2: Segmentos con participación agropecuaria media.

ESTRATO 3: Segmentos con participación agropecuaria grande.

9. Tamaño de la muestra

A nivel del subuniverso y tomando en cuenta el muestreo estratificado con asignación optima en el tamaño de la muestra se utiliza la siguiente expresión:

$$n = \frac{\left(\sum_{h=1}^L W_h S_h\right)^2}{\left(\frac{e}{t}\right)^2 + \sum_{h=1}^L \frac{N_h}{N^2} S_h^2}, \quad h = 1, 2, 3, \dots, L. \quad (1)$$

$e = e' \bar{Y}$. En la que

n =Tamaño de la muestra (en segmentos)

e =Error respecto al promedio.

e' =Error (%) de medición relativa.

t_α =Nivel de significación para un α dado

$W_h = N_h/N$: Ponderación del estrato h

$S_h^2 = \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2 / (N_h - 1)$: Varianza de la variable “Superficie Cultivada” o “Número abstracto de Cabezas”, por segmento del h -ésimo estrato.

N_h =Total de segmentos del estrato h .

N =Total de segmentos del Subuniverso.

L =Número total de estratos.

Y_{ih} =Total de la “Superficie Cultivada” o “Numero abstracto de Cabezas” del i -ésimo segmento en el h -ésimo estrato. $i = 1, \dots, N_h$. $h = 1, \dots, L$.

$n_h = nN_h S_h / (\sum_{h=1}^L N_h S_h)$: Tamaño de la submuestra por estrato.

10. Procedimiento del cálculo de tamaño de muestra

En el diseño muestral, se utiliza las fórmulas que corresponden al muestreo estratificado con asignación optima; diseñándose una hoja electrónica para generalizar el cálculo del tamaño de la muestra para los distintos subuniversos.

$N = \sum_{h=1}^L N_h$: Total de elementos de la población.

$W_h = N_h/N$: Ponderación por estrato

$Y = \sum_{h=1}^L Y_h$: Total de la variable Y

$Y = \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{ih}$: Total de la variable Y por elemento y por estrato.

$h = 1, 2, \dots, L$: Estratos.

$i = 1, 2, \dots, N_h$: Elementos.

$\bar{Y} = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{ih} = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} N_h \bar{Y}_h$: Promedio general o poblacional de Y .

$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Varianza

$\sigma^2 = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Desviación

$S_h^2 = \frac{1}{N-1} \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Cuasivarianza

$S_b^2 = \frac{1}{N-1} \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2$: Intervarianza (Entre)

$S_w^2 = \frac{1}{N-1} \sum_{h=1}^L S_h^2 (N_h - 1)$: Intravarianza (Dentro).

11. Ejemplo

Para la Zona de Producción Llanos de Chaco, uso agrícola, departamento de Chuquisaca, el cálculo de parámetros y tamaño de muestra es:

Departamento: Chuquisaca
 Zona de Producción: Llanos del Chaco –110
 Uso: Agrícola

Muestreo Estratificado
 Tipo de Afijación: Optima

Población Agropecuaria

Detalle	Descrip	Estrato 1	Estrato 2	Estrato 3	Totales		
1	Y_1	59	196	500	331	620	—
2	Y_2	66	199	502	333	657	—
3	Y_3	77	200	513	340	660	—
4	Y_4	104	201	519	343	997	—
5	Y_5	109	210	401	373	1273	—
6	Y_6	109	211	403	379	2113	—
7	Y_7	119	212	405	389	542	—
8	Y_8	125	213	406	393	543	—
9	Y_9	135	216	410	328	548	—
10	Y_{10}	142	222	438	327	551	—
11	Y_{11}	148	224	441		560	—
12	Y_{12}	154	231	444			—
13	Y_{13}	157	236	485			—
14	Y_{14}	160	250	306			—
15	Y_{15}	165	255	307			—
16	Y_{16}	165	275	307			—
17	Y_{17}	172	284	310			—
18	Y_{18}	174	288	323			—
19	Y_{19}	177	299	326			—
20	Y_{20}	193	305	331			—
Número total de segmentos	N_h	40.00	30.00	11.00			81.00
Ponderación	W_h	0.49	0.37	0.14			1.00
Total Variable Y	Y_h	7437.00	11613.00	9064.00			28114.00
Promedio por estratos	\bar{Y}_h	185.93	387.10	824.00			347.09
Varianza σ^2	σ^2	3856.30	4438.30	215303.40			77093.91
Desviación σ	σ	62.10	66.60	464.01			277.66
Cuasivarianza S^2	S^2	3955.20	4591.40	236833.80			78057.58
Desviación S	S	62.89	67.76	486.66			279.39
Intervarianza (Entre)	S_b^2	—	—	—	55.47 %	(44860.81)	
Intravarianza (Dentro)	S_w^2	—	—	—	44.53 %	(33196.77)	
Coef. de var. (%)	CV_h	33.40 %	17.21 %	56.31 %			80.00 %

El cálculo de tamaño de la muestra, para un error del 5 % y un nivel de confianza del 95 %, es:

$$n = \frac{\left(\sum_{h=1}^L W_h S_h\right)^2}{\left(\frac{e}{t}\right)^2 + \sum_{h=1}^L \frac{N_h}{N^2} S_h^2} \quad (2)$$

Tomando en cuenta un error del 5 %.

$$e' = 0.05 \quad e = e' \bar{Y} = 0.05(347.09) = 17.3545 \quad n = 29.54 \cong 29 \quad (3)$$

Distribución de la muestra por estrato:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

$$n_1 = 29 \frac{2515.6}{9901.66} = 7 \quad n_2 = 29 \frac{2032.8}{9901.66} = 6 \quad n_3 = 29 \frac{5353.26}{9901.66} = 16 \quad (4)$$

Calculando tamaños de muestra para distintos errores, tendremos:

Departamento: Chuquisaca
 Zona de Producción: Llanos del Chaco –110
 Uso: Agrícola

Muestreo Estratificado
 Tipo de Afijación: Optima

Tamaños de muestra para diferentes errores
 ($t = 1.96$ para un nivel de confianza del 95 %.)

Error (%)	n_1	n_2	n_3	n
1	9	7	18	34
2	8	7	18	33
3	8	7	17	32
4	8	6	17	31
5	7	6	16	29
6	7	5	15	27
7	7	5	14	26
8	6	5	13	24
9	6	4	12	22
10	5	4	11	20

Análisis y Clasificación de los Municipios del Departamento de La Paz mediante Escalamiento Multidimensional y Análisis Cluster

M. Sc. Fernando Oday Rivero Suguira
✉ fors2004@yahoo.com.ar

1. Introducción

En el presente artículo se hace una aplicación de los métodos estadísticos multivariantes *Escalamiento Multidimensional* y *Análisis Cluster* a un conjunto de indicadores sociodemográficos de la Encuesta Sociodemográfica del Departamento de La Paz¹. Se pretende dar una visión general del funcionamiento del escalado multidimensional comparándolo con otra técnica multivariante como es el análisis cluster, de modo que pueda servir como alternativa y complemento de investigación.

El escalamiento multidimensional es una técnica multivariante de interdependencia que trata de representar datos en un espacio geométrico o mapa perceptual de pocas dimensiones mediante proximidades existentes entre un conjunto de objetos. Mientras que el análisis cluster jerárquico clasifica los objetos mediante cierto nivel de similitud.

2. Información estadística

La información empleada para esta investigación, corresponde a catorce indicadores sociales y económicos de setenta y siete municipios del departamento de La Paz, estos son:

- DM0102 Población en la niñez con respecto a mujeres en edad reproductiva. Es el cociente de la población menor de cinco años entre la población femenina en edad reproductiva.
- DM0506 Relación de dependencia demográfica. Es el cociente de la población menor de 15 años y la población mayor de 64 años sobre la población de 15 a 64 años.
- DM0507 Relación de masculinidad. Es el cociente de la población masculina sobre la población femenina.
- ED0101 Tasa de asistencia en edad escolar. Es el cociente entre la población de 6 a 19 años de edad que asiste a una unidad educativa y la población total del mismo grupo de edad.
- ED0401 Tasa de analfabetismo. Es el cociente entre la población de 15 años y más que no sabe leer ni escribir y la población total del mismo grupo de edad.
- ED0402 Promedio de escolaridad. Expresa el número promedio de años de escolaridad aprobados por la población de 19 años y más de edad.
- EI0101 Tasa de oferta laboral potencial. Es la relación porcentual de la población en edad de trabajar con respecto a la población total.
- EI0102 Tasa global de participación. Es la relación de la población económicamente activa entre la población en edad de trabajar.
- EI0105 Tasa de desempleo abierto. Es la relación porcentual de la población desocupada abierta con respecto a la población económicamente activa.

¹Encuesta realizada por el Instituto de Estadística Teórica y Aplicada de la Carrera de Estadística de la UMSA, en coordinación con el Gobierno Departamental de La Paz.

EI0107	Índice de carga económica. Es la relación porcentual entre la población económicamente inactiva y la población económicamente activa.
WV0101	Distribución de hogares según disponibilidad de dormitorios por persona. Es la relación entre el número de personas o miembros en el hogar y el número de dormitorios existentes en las viviendas ocupadas por el hogar. Bajo este criterio, se identifican tres tipos de calidad de la vivienda: viviendas donde habitan hasta 2 personas por dormitorio, más de 2 hasta 3 personas por dormitorio y más de 3 personas por dormitorio.
WV0203	Viviendas con servicio sanitario. Es el cociente entre el número de viviendas particulares que tienen algún tipo de servicio sanitario (baño, letrina) y el total de viviendas particulares.
WV0204	Viviendas con energía eléctrica. Es la relación entre el número de viviendas particulares que cuentan con energía eléctrica para uso doméstico y el total de viviendas particulares.
NBI2001	Proporción de población pobre según el método NBI ² .

Antes de utilizar los resultados del cuadro N° 1 siguiente, se hace una interpretación de estos para el municipio de La Paz en la primera fila y así comprender mejor la información disponible para los otros municipios.

- o DM0102 igual a 0,19. Significa que existe 19 niños menores de 5 años por cada 100 mujeres en edad reproductiva en el municipio de La Paz.
- o DM0506 igual a 0,37. Relación de dependencia demográfica, es decir, existe un 37 % de la población del municipio de La Paz que es menor de 15 años o mayor de 64 años en relación a la población entre 15 y 64 años. Los primeros dependen en diferentes aspectos social, económico del segundo grupo de personas.
- o DM0507 igual a 0,97. Relación de masculinidad, existen 97 hombres por cada 100 mujeres dentro del municipio de La Paz.
- o ED0101 igual a 0,92. Tasa de asistencia en edad escolar, el 92 % de la población del municipio de La Paz entre 6 y 19 años de edad asiste a una unidad educativa, respecto a la población total del mismo grupo de edad.
- o ED0401 igual a 0,03. Tasa de analfabetismo, es decir, el 3 % de la población del municipio de La Paz de 15 años y más no sabe leer ni escribir respecto a la población del mismo grupo de edad.
- o ED0402 igual a 12,83. Promedio de escolaridad. Se interpreta como que son aproximadamente 13 años promedio de estudio en personas de 19 años y más de edad en el municipio de La Paz.
- o EI0101 igual a 0,88. Tasa de oferta laboral potencial. El 88 % de la población del municipio de La Paz está en edad de trabajar, considerando 10 años y más la edad de trabajar.
- o EI0102 igual a 0,58. Tasa global de participación, se interpreta como que el 58 % de las personas del municipio de La Paz, están dentro de la población económicamente activa respecto a la población en edad de trabajar.
- o EI0105 igual a 0,03. Tasa de desempleo abierto. El 3 % de la población está desocupada respecto a la población económicamente activa.
- o EI0107 igual a 0,74. Índice de carga económica. El 74 % de la población del municipio de La Paz está dentro de la población económicamente inactiva respecto a la población económicamente activa.

²Indicador de Pobreza, fuente INE 2001.

- W0101 igual a 1,31. Distribución de los hogares según disponibilidad de dormitorios por persona. Se tienen 131 personas por cada 100 dormitorios promedio en el municipio de La Paz.
- V0203 igual a 0,95. Viviendas con servicio sanitario. El 95 % de las viviendas particulares tienen algún tipo de servicio sanitario (baño, letrina).
- V0204 igual a 0,99. Viviendas con energía eléctrica. El 99 % de las viviendas particulares cuentan con energía eléctrica para uso doméstico en el municipio de La Paz.
- NBI2001 igual a 0,35. El 35 % de las personas que habitan el municipio de La Paz son pobres, según el método NBI.

Los datos de las variables anteriores a nivel municipio, se presenta en el siguiente cuadro:

Cuadro 1 Encuesta sociodemográfica departamento de La Paz. Municipios del Departamento de La Paz

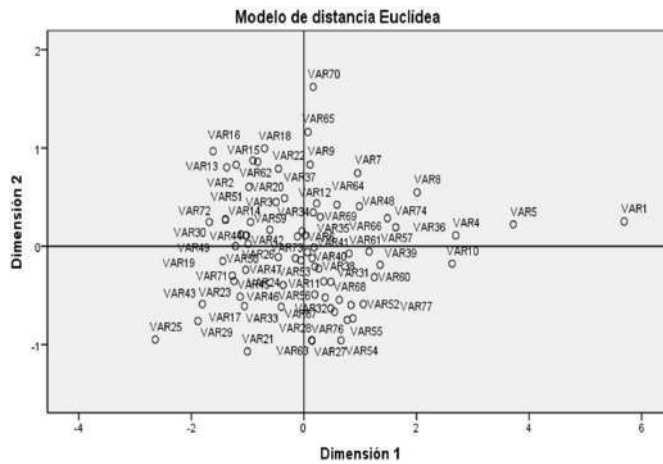
NRO	MUNICIPIO	DM0102	DM0506	DM0507	ED0101	ED0401	ED0402	EI0101	EI0102	EI0105	EI0107	VV0101	VV0203	VV0204	NBI2001
1	LA PAZ	0.19	0.37	0.97	0.92	0.03	12.83	0.88	0.58	0.03	0.74	1.31	0.95	0.99	0.35
2	PALCA	0.44	0.66	1.13	0.90	0.12	6.29	0.80	0.60	0.01	0.67	1.34	0.37	0.94	0.99
3	MECAPACA	0.34	0.70	0.99	0.94	0.17	6.99	0.81	0.62	0.01	0.60	1.64	0.41	0.96	0.82
4	ACHOCALLA	0.25	0.50	0.92	0.91	0.05	10.08	0.84	0.57	0.03	0.75	1.72	0.64	0.99	0.99
5	EL ALTO	0.21	0.42	0.98	0.92	0.04	11.00	0.85	0.54	0.04	0.84	1.52	0.93	0.99	0.67
6	ACHACACHI	0.25	0.58	1.10	0.82	0.13	7.52	0.85	0.55	0.03	0.80	2.13	0.83	0.96	0.93
7	ANCORAÍMES	0.13	0.46	1.05	0.78	0.08	8.34	0.89	0.66	0.03	0.49	1.27	0.68	0.92	0.98
8	SANTIAGO DE HUATA	0.18	0.42	1.34	0.99	0.10	9.29	0.89	0.47	0.00	1.14	1.44	0.98	1.00	0.98
9	CHUA COCANI	0.16	0.47	1.15	0.89	0.13	7.56	0.87	0.54	0.05	0.84	1.30	0.20	0.88	0.98
10	HUATAJATA	0.10	0.37	0.87	0.85	0.04	9.99	0.90	0.49	0.05	1.05	2.04	0.80	0.98	0.98
11	HUARINA	0.16	0.55	1.06	0.92	0.12	7.84	0.87	0.59	0.01	0.64	2.40	0.76	0.92	0.98
12	CORO CORO	0.23	0.58	0.96	0.88	0.07	7.70	0.83	0.60	0.02	0.66	1.62	0.40	0.61	0.90
13	CAQUIAVIRI	0.29	0.81	0.98	0.94	0.13	6.16	0.85	0.63	0.00	0.59	1.46	0.70	0.32	0.98
14	CALACOTO	0.26	0.73	1.00	0.88	0.07	6.66	0.83	0.66	0.02	0.51	1.98	0.62	0.19	0.99
15	COMANCHE	0.22	0.65	1.01	0.92	0.07	6.65	0.81	0.54	0.04	0.87	1.30	0.43	0.62	0.98
16	NAZACARA DE PACAJES	0.35	0.89	1.00	0.90	0.19	5.89	0.81	0.60	0.02	0.66	1.24	0.55	0.64	0.98
17	SANTIAGO DE CALLAPA	0.18	0.62	1.15	0.90	0.06	6.61	0.84	0.68	0.05	0.48	2.64	0.42	0.25	1.00
18	PUERTO ACOSTA	0.23	0.57	1.07	0.89	0.10	6.75	0.81	0.54	0.03	0.85	1.15	0.41	0.71	0.97
19	MOCOMOCO	0.27	0.60	1.04	0.87	0.17	6.16	0.83	0.74	0.01	0.36	2.30	0.36	0.82	0.99
20	PUERTO CARABUCO	0.26	0.60	1.06	0.89	0.09	6.54	0.84	0.60	0.03	0.68	1.56	0.32	0.52	0.96
21	HUMANATA	0.33	0.92	0.94	0.90	0.16	6.67	0.45	0.56	0.00	0.78	3.14	0.36	0.73	0.96
22	ESCOMA	0.20	0.67	1.12	0.96	0.13	7.01	0.83	0.54	0.04	0.87	1.35	0.31	0.96	0.96
23	CHUMA	0.31	0.68	0.95	0.86	0.19	6.34	0.82	0.64	0.01	0.57	2.51	0.44	0.95	0.99
24	AYATA	0.42	0.73	1.17	0.93	0.21	6.56	0.80	0.57	0.03	0.77	2.37	0.28	0.90	1.00
25	AUCAPATA	0.45	0.75	1.09	0.89	0.27	5.10	0.79	0.49	0.05	1.04	3.20	0.19	0.85	1.00
26	SORATA	0.35	0.57	0.97	0.86	0.15	7.05	0.81	0.61	0.03	0.64	2.21	0.61	0.92	0.97
27	GUANAY	0.41	0.79	0.99	0.85	0.04	7.72	0.73	0.60	0.02	0.67	2.99	0.65	0.84	0.82
28	TACACOMA	0.44	0.55	1.00	0.83	0.09	8.09	0.78	0.60	0.02	0.67	2.69	0.63	0.83	0.95
29	QUIABAYA	0.47	0.85	1.11	0.88	0.17	5.80	0.76	0.69	0.00	0.44	2.95	0.34	0.86	0.99
30	COMBAYA	0.27	0.73	1.13	0.78	0.23	5.90	0.87	0.72	0.02	0.38	1.94	0.35	0.91	1.00
31	TIPUANI	0.31	0.69	1.01	0.91	0.05	7.97	0.78	0.52	0.01	0.93	2.38	0.56	0.95	0.73
32	MAPIRI	0.44	0.71	1.12	0.86	0.04	8.02	0.75	0.52	0.03	0.91	2.63	0.47	0.97	0.93
33	TEOPONTE	0.29	0.65	1.10	0.92	0.05	7.14	0.81	0.51	0.02	0.97	2.68	0.69	0.80	0.89
34	APOLO	0.47	0.81	1.02	0.92	0.04	7.64	0.74	0.49	0.06	1.04	1.77	0.57	0.70	0.98
35	PELECHUCO	0.43	0.55	1.17	0.85	0.09	7.53	0.79	0.54	0.06	0.84	1.96	0.38	0.84	0.99
36	VIACHA	0.29	0.62	0.95	0.92	0.10	9.06	0.84	0.61	0.04	0.65	1.72	0.71	0.89	0.79
37	GUAQUI	0.28	0.73	0.94	0.91	0.11	7.09	0.80	0.58	0.02	0.74	1.61	0.72	0.95	0.91
38	TIHUANACU	0.17	0.64	0.91	0.92	0.08	7.66	0.79	0.55	0.02	0.82	2.26	0.79	0.94	0.96
39	DESAGUADERO	0.39	0.67	0.88	0.92	0.05	8.82	0.77	0.62	0.02	0.61	2.13	0.65	0.95	0.78
40	TARACO	0.25	0.65	1.00	0.90	0.07	7.78	0.80	0.58	0.02	0.72	2.27	0.67	0.50	0.99
41	LURIBAY	0.29	0.59	1.02	0.91	0.10	7.64	0.83	0.60	0.03	0.67	2.19	0.60	0.93	0.97
42	SAPAHQUI	0.25	0.58	0.98	0.85	0.13	6.96	0.86	0.78	0.00	0.28	1.96	0.33	0.88	0.99
43	YACO	0.41	0.82	1.13	0.90	0.14	5.87	0.79	0.67	0.02	0.50	2.77	0.14	0.87	1.00
44	MALLA	0.55	0.84	1.13	0.88	0.13	6.59	0.77	0.65	0.01	0.54	2.01	0.15	0.91	1.00
45	CAIROMA	0.38	0.77	1.04	0.94	0.10	7.21	0.79	0.57	0.01	0.75	2.46	0.25	0.96	0.96
46	INQUISIVI	0.36	0.75	1.03	0.90	0.14	6.53	0.79	0.66	0.02	0.52	2.60	0.30	0.35	0.98
47	QUIME	0.44	0.73	1.03	0.92	0.10	7.37	0.79	0.48	0.02	1.08	2.18	0.40	0.87	0.88
48	CAJUATA	0.35	0.59	1.12	0.84	0.09	8.43	0.83	0.66	0.02	0.51	1.57	0.61	0.93	0.95
49	COLQUIRI	0.51	0.76	1.04	0.89	0.17	6.37	0.78	0.54	0.03	0.84	2.15	0.24	0.90	0.91
50	ICHUCA	0.51	0.90	1.12	0.92	0.11	6.61	0.75	0.58	0.02	0.72	2.10	0.24	0.68	0.99
51	LICOMA PAMPA	0.40	0.68	1.13	0.85	0.12	6.17	0.82	0.72	0.01	0.38	1.90	0.44	0.65	0.88
52	CHULUMANI	0.29	0.64	1.02	0.90	0.03	8.55	0.78	0.72	0.01	0.38	2.51	0.71	0.97	0.74
53	IRUPANA	0.31	0.62	1.09	0.92	0.05	7.44	0.81	0.63	0.01	0.58	2.21	0.70	0.95	0.89
54	YANACACHI	0.32	0.69	1.00	0.92	0.03	8.18	0.77	0.56	0.03	0.77	2.93	0.77	0.99	0.65
55	PALOS BLANCOS	0.38	0.76	1.15	0.93	0.04	8.27	0.76	0.54	0.01	0.84	2.72	0.90	0.89	0.91
56	LA ASUNTA	0.31	0.74	1.06	0.94	0.04	7.75	0.77	0.68	0.00	0.47	2.50	0.60	0.59	0.95
57	PUCARANI	0.27	0.63	1.06	0.91	0.07	8.60	0.81	0.50	0.02	1.00	2.04	0.74	0.85	0.98
58	LAJA	0.29	0.64	0.98	0.91	0.18	6.48	0.83	0.65	0.02	0.53	2.03	0.61	0.92	0.99
59	BATALLAS	0.15	0.54	1.09	0.87	0.09	7.42	0.86	0.52	0.03	0.92	1.95	0.70	0.93	0.96
60	PUERTO PEREZ	0.10	0.40	0.97	0.86	0.08	8.71	0.83	0.56	0.09	0.78	2.28	0.52	0.99	0.99
61	SICA SICA	0.24	0.52	0.97	0.78	0.09	8.30	0.84	0.61	0.04	0.63	2.08	0.42	0.87	0.99
62	UMALA	0.28	0.85	1.02	0.91	0.07	6.55	0.78	0.53	0.01	0.88	1.31	0.68	0.74	0.97
63	CALAMARCA	0.27	0.60	1.00	0.82	0.06	7.74	0.80	0.61	0.05	0.65	2.99	0.37	0.78	0.98
64	PATACAMAYA	0.30	0.58	1.02	0.80	0.07	8.02	0.80	0.58	0.04	0.72	1.58	0.68	0.95	0.97
65	COLOQUENCHA	0.34	0.66	1.08	0.88	0.06	7.47	0.79	0.56	0.02	0.79	0.90	0.54	0.91	1.00
66	COLLANA	0.20	0.63	0.98	0.82	0.12	7.66	0.85	0.63	0.03	0.59	2.07	0.74	0.86	0.94
67	COROICO	0.27	0.64	1.01	0.92	0.04	7.88	0.80	0.62	0.03	0.60	2.55	0.73	0.97	0.76
68	CORIPATA	0.20	0.63	1.06	0.95	0.06	8.14	0.80	0.67	0.01	0.49	2.54	0.72	0.93	0.89
69	WAMAS	0.49	0.82	0.98	0.80	0.06	7.75	0.75	0.53	0.01	0.90	1.77	0.70	0.71	0.91
70	SAN BUENAVENTURA	0.54	0.78	0.97	0.87	0.04	7.49	0.73	0.53	0.02	0.89	0.52	0.89	0.79	0.85
71	CHARAZANI	0.46	0.73	1.10	0.91	0.16	6.33	0.79	0.59	0.03	0.71	2.45	0.24	0.88	0.98
72	CURVA	0.39	0.62	1.09	0.85	0.13	6.15	0.78	0.68	0.02	0.47	1.89	0.42	0.75	0.99
73	COPACABANA	0.19	0.61	0.93	0.85	0.09	7.34	0.82	0.63	0.02	0.60	1.98	0.80	0.90	0.88
74	SAN PEDRO DE TIQUINA	0.30	0.67	0.85	0.88	0.09	8.92	0.82	0.55	0.05	0.82	1.66	0.46	0.93	0.91
75	TITO YUPANQUI	0.24	0.59	0.85	0.76	0.21	6.45	0.86	0.71	0.03	0.41	2.03	0.34	0.90	0.99
76	CARANAVI	0.37	0.74	1.15	0.92	0.05	8.39	0.73	0.63	0.02	0.58	2.67	0.87	0.64	0.87
77	ALTO BENI	0.42	0.65	1.15	0.91	0.06	8.33	0.81	0.74	0.00	0.36	2.51	0.66	0.87	0.89

Fuente: Encuesta Sociodemográfica Departamento de La Paz – UMSA – Instituto de Estadística Teórica y Aplicada

3. Resultados

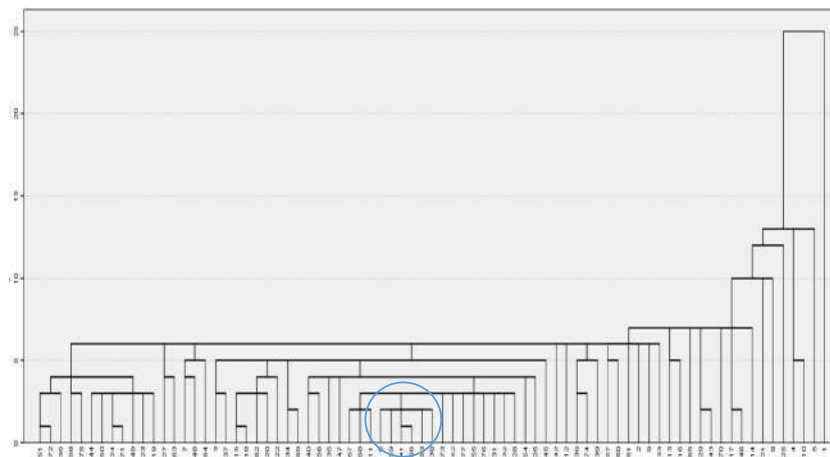
El siguiente gráfico muestra la distribución por similitud (distancia euclidiana) de los setenta y siete municipios estudiados en la encuesta, mediante el método de escalado multidimensional.

Gráfico 1 ➔ Representación de municipios en el espacio bidimensional



El gráfico N° 2 presenta un dendograma de clasificación de los setenta y siete municipios del departamento de La Paz, de acuerdo a los catorce indicadores utilizados en el análisis.

Gráfico 2 ➔ Dendograma



El cuadro N° 2 muestra los resultados del anterior gráfico N° 2, es decir, presenta el historial de la clasificación de los municipios y los porcentajes de clasificación en diferentes niveles.

4. Interpretación

Los municipios más similares en el nivel 11,0 % para los catorce indicadores sociodemográficos presentados en el cuadro N° 2, son los de Licoma Pampa y Curva. En DM0102 están entre 0,39 y 0,40 en DM0506 fluctúan entre 0,62 y 0,68; en DM0507 entre 1,09 y 1,13; en ED0101 son idénticos (0,85), en el indicador ED0401 van entre 0,12 y 0,13, en ED0402 están entre 6,15 y 6,17. La tasa de oferta laboral potencial (EI0101) va entre 0,78 y 0,82; en EI0102 alcanza a 0,68 y 0,72. Para EI0105 es 0,01 y 0,02; para EI0107 cambia de 0,38 a 0,47. Finalmente para indicadores de vivienda VV0101 pasa de 1,89 a 1,90, para VV0203 de 0,42 a 0,44 y para VV0204 de 0,65 a 0,75, y el indicador de pobreza NBI2001 de 0,88 a 0,99 (mayor diferencia).

Cuadro 2 Encuesta sociodemográfica departamento de La Paz. Municipios del Departamento de La Paz. Análisis cluster

COEFICIENTE	ETAPA	CLUSTER	MUNICIPIO	DM0102	DM0206	DM0507	ED0101	ED0401	ED0402	EI0101	EI0102	EI0105	EI0107	VW0101	VW0203	VW0204	NB0201
11.0%	1	51	LICOMA PAMPA	0.40	0.68	1.13	0.85	0.12	6.17	0.82	0.72	0.01	0.38	1.90	0.44	0.65	0.88
	2	72	CURVA	0.39	0.62	1.09	0.85	0.13	6.15	0.78	0.68	0.02	0.47	1.89	0.42	0.75	0.99
12.4%	2	15	COMANCHE	0.22	0.65	1.01	0.92	0.07	6.66	0.81	0.54	0.04	0.87	1.30	0.43	0.62	0.98
	3	18	PUERTO ACOSTA	0.23	0.57	1.07	0.89	0.10	6.75	0.81	0.54	0.03	0.85	1.15	0.41	0.71	0.97
14.0%	3	41	LURIBAY	0.29	0.59	1.02	0.91	0.10	7.64	0.83	0.60	0.03	0.67	2.19	0.60	0.93	0.97
	4	24	COLLANA	0.20	0.63	0.98	0.82	0.12	7.66	0.85	0.63	0.03	0.59	2.07	0.74	0.86	0.94
14.2%	4	66	AYATA	0.42	0.73	1.17	0.93	0.21	6.56	0.80	0.57	0.03	0.77	2.37	0.28	0.90	1.00
	4	71	CHARAZANI	0.46	0.73	1.10	0.91	0.16	6.33	0.79	0.59	0.03	0.71	2.45	0.24	0.86	0.98
14.7%	5	34	APOLO	0.47	0.81	1.02	0.92	0.04	7.64	0.74	0.49	0.06	1.04	1.77	0.57	0.70	0.98
	5	69	IXAMAS	0.49	0.82	0.98	0.80	0.06	7.75	0.75	0.53	0.01	0.90	1.77	0.70	0.71	0.91
14.9%	6	(41, 66)	LURIBAY - COLLANA	0.24	0.61	1.00	0.87	0.11	7.66	0.84	0.61	0.03	0.63	2.13	0.67	0.89	0.95
	6	53	IRUPANA	0.31	0.62	1.09	0.92	0.05	7.44	0.81	0.63	0.01	0.58	2.21	0.70	0.95	0.89
15.9%	7	6	ACHACACHI	0.25	0.58	1.10	0.82	0.13	7.52	0.85	0.55	0.03	0.80	2.13	0.83	0.96	0.93
	7	59	BATALLAS	0.15	0.54	1.09	0.87	0.09	7.42	0.86	0.52	0.03	0.92	1.95	0.70	0.93	0.96
16.1%	8	29	QUIJUYA	0.47	0.85	1.11	0.88	0.17	5.80	0.76	0.69	0.00	0.44	2.95	0.34	0.86	0.99
	8	43	YACAO	0.41	0.82	1.13	0.90	0.14	5.97	0.79	0.67	0.02	0.50	2.77	0.54	0.87	1.00
16.8%	9	(41, 66, 53)	LURIBAY - COLLANA - IRUPANA	0.28	0.62	1.06	0.89	0.08	7.54	0.83	0.62	0.02	0.61	2.17	0.58	0.92	0.92
	9	38	TIHUJANACU	0.17	0.64	0.91	0.92	0.08	7.66	0.79	0.55	0.02	0.82	2.26	0.79	0.94	0.96
	10	(6, 59)	ACHACACHI - BATALLAS	0.20	0.56	1.10	0.84	0.11	7.47	0.86	0.54	0.03	0.86	2.04	0.77	0.94	0.95
17.2%	10	(41, 66, 53, 38)	LURIBAY - COLLANA - IRUPANA - TIHUJANACU	0.22	0.63	0.98	0.91	0.08	7.60	0.81	0.59	0.02	0.71	2.21	0.74	0.93	0.94
17.5%	11	17	SANTIAGO DE CALLAPA	0.18	0.62	1.15	0.90	0.06	6.61	0.84	0.68	0.05	0.48	2.64	0.42	0.25	1.00
	11	46	INOUSIVI	0.36	0.75	1.03	0.90	0.14	6.53	0.79	0.66	0.02	0.52	2.60	0.30	0.35	0.98
17.6%	12	67	COROICO	0.27	0.64	1.01	0.92	0.04	7.88	0.80	0.62	0.03	0.60	2.55	0.73	0.97	0.76
	12	68	CORIPATA	0.20	0.63	1.06	0.95	0.06	8.14	0.80	0.67	0.01	0.49	2.54	0.72	0.93	0.88
17.9%	13	(67, 68)	COROICO - CORIPATA	0.23	0.64	1.04	0.93	0.05	8.01	0.80	0.65	0.02	0.55	2.54	0.72	0.95	0.82
	13	11	HUARINA	0.16	0.55	1.06	0.92	0.12	7.84	0.87	0.59	0.01	0.64	2.40	0.76	0.92	0.98
18.2%	14	44	MALLA	0.55	0.84	1.13	0.88	0.13	6.59	0.77	0.65	0.01	0.54	2.01	0.15	0.91	1.00
	14	50	ICHOCA	0.51	0.90	1.12	0.92	0.11	6.61	0.75	0.58	0.02	0.72	2.10	0.24	0.68	0.99
18.7%	15	(24, 71)	AYATA - CHARAZANI	0.44	0.73	1.14	0.92	0.19	6.44	0.80	0.58	0.03	0.74	2.41	0.26	0.89	0.99
	15	49	COLQUIRI	0.51	0.76	1.04	0.89	0.17	6.37	0.78	0.54	0.03	0.84	2.15	0.24	0.90	0.91
18.7%	16	(6, 59, 41, 66, 53, 38)	ACHACACHI - BATALLAS - LURIBAY - COLLANA - IRUPANA - TIHUJANACU	0.21	0.59	1.04	0.88	0.10	7.53	0.83	0.56	0.02	0.79	2.13	0.75	0.94	0.94
	16	73	COPACABANA	0.19	0.61	0.93	0.85	0.09	7.34	0.82	0.63	0.02	0.60	1.98	0.80	0.90	0.86
18.9%	17	(6, 59, 41, 66, 53, 38, 73)	ACHACACHI - BATALLAS - LURIBAY - COLLANA - IRUPANA - TIHUJANACU - COPACABANA	0.20	0.60	0.98	0.86	0.09	7.44	0.83	0.59	0.02	0.69	2.05	0.78	0.92	0.91
	17	(67, 68, 11)	COROICO - CORIPATA - HUARINA	0.19	0.59	1.05	0.93	0.08	7.93	0.84	0.62	0.01	0.59	2.47	0.74	0.94	0.90
19.1%	18	(24, 71, 49)	AYATA - CHARAZANI - COLQUIRI	0.47	0.75	1.09	0.90	0.18	6.41	0.79	0.56	0.03	0.79	2.28	0.25	0.89	0.95
	18	23	CHUMA	0.31	0.68	0.95	0.86	0.19	6.34	0.82	0.64	0.01	0.57	2.51	0.44	0.95	0.99
19.5%	19	(15, 18)	COMANCHE - PUERTO ACOSTA	0.23	0.61	1.04	0.90	0.08	6.70	0.81	0.54	0.03	0.86	1.23	0.42	0.66	0.97
	19	62	UMALA	0.28	0.85	1.02	0.91	0.07	6.55	0.78	0.53	0.01	0.88	1.31	0.68	0.74	0.97
20.1%	20	58	LAJA	0.29	0.64	0.98	0.91	0.18	6.48	0.83	0.65	0.02	0.53	2.03	0.61	0.92	0.99
	20	75	TITO YUPANQUI	0.24	0.59	0.85	0.76	0.21	6.45	0.86	0.71	0.03	0.41	2.03	0.34	0.90	0.99
20.1%	21	52	CHULUMANI	0.29	0.64	1.02	0.90	0.03	8.55	0.78	0.72	0.01	0.38	2.51	0.71	0.97	0.74
	21	77	ALTO BENI	0.42	0.65	1.15	0.91	0.06	6.33	0.81	0.74	0.00	0.36	2.51	0.86	0.87	0.89
20.2%	22	31	TIPUANI	0.31	0.69	1.01	0.91	0.05	7.97	0.78	0.62	0.01	0.33	2.38	0.56	0.95	0.73
	22	32	MAPIRI	0.44	0.71	1.12	0.86	0.04	8.02	0.75	0.52	0.03	0.91	2.63	0.47	0.97	0.93
20.4%	23	(6, 59, 41, 66, 53, 38, 73, 67, 68, 11)	ACHACACHI - BATALLAS - LURIBAY - COLLANA - IRUPANA - TIHUJANACU - COPACABANA - CHULUMANI - ALTO BENI	0.20	0.60	1.01	0.89	0.09	7.68	0.83	0.61	0.02	0.64	2.26	0.76	0.93	0.91
	23	(52 - 77)	COROICO - CORIPATA - HUARINA	0.36	0.65	1.09	0.91	0.04	8.44	0.80	0.73	0.00	0.37	2.51	0.79	0.92	0.81
20.6%	24	(15, 18, 62)	COMANCHE - PUERTO ACOSTA - UMALA	0.25	0.73	1.03	0.91	0.08	6.62	0.80	0.53	0.02	0.87	1.27	0.55	0.70	0.97
	24	20	PUERTO CARABUCO	0.26	0.60	1.06	0.89	0.09	6.54	0.84	0.60	0.03	0.68	1.56	0.32	0.52	0.96
20.6%	25	36	VIACHA	0.29	0.62	0.95	0.92	0.10	9.06	0.84	0.61	0.04	0.65	1.72	0.71	0.89	0.79
	25	74	SAN PEDRO DE TIQUINA	0.30	0.67	0.85	0.88	0.09	8.92	0.82	0.55	0.05	0.82	1.66	0.46	0.93	0.91
20.8%	26	3	MECAPACA	0.34	0.70	0.99	0.94	0.17	6.99	0.81	0.62	0.01	0.60	1.64	0.41	0.96	0.82
	26	37	GUJUVI	0.28	0.73	0.94	0.91	0.14	7.09	0.80	0.58	0.02	0.74	1.61	0.72	0.95	0.91
20.9%	27	(51, 72)	LICOMA PAMPA - CURVA	0.39	0.65	1.11	0.85	0.13	6.16	0.80	0.70	0.02	0.43	1.89	0.43	0.70	0.93
	27	30	COMBAYA	0.27	0.73	1.13	0.78	0.23	5.90	0.87	0.72	0.02	0.38	1.94	0.35	0.91	1.00
21.2%	28	40	TARACO	0.25	0.65	1.00	0.90	0.07	7.78	0.80	0.58	0.02	0.72	2.27	0.67	0.50	0.99
	28	56	LA ASUNTA	0.31	0.74	1.06	0.94	0.04	7.75	0.77	0.68	0.00	0.47	2.50	0.60	0.59	0.95
21.3%	29	55	PALOS BLANCOS	0.38	0.76	1.15	0.93	0.04	8.27	0.76	0.54	0.01	0.84	2.72	0.90	0.89	0.91
	29	76	CARANAVI	0.37	0.74	1.15	0.92	0.05	8.39	0.73	0.63	0.02	0.58	2.67	0.87	0.64	0.87
21.4%	30	(6, 59, 41, 66, 53, 38, 73, 67, 68, 11, 52, 77)	ACHACACHI - BATALLAS - LURIBAY - COLLANA - IRUPANA - TIHUJANACU - COPACABANA - COROICO - CORIPATA - HUARINA - CHULUMANI - ALTO BENI	0.28	0.62	1.05	0.90	0.07	8.06	0.81	0.67	0.01	0.51	2.39	0.77	0.92	0.86
	30	(55, 76)	PALOS BLANCOS - CARANAVI	0.38	0.75	1.15	0.93	0.04	8.33	0.75	0.59	0.01	0.71	2.69	0.89	0.96	0.89
21.4%	31	(24, 71, 49, 23)	AYATA - CHARAZANI - COLQUIRI - CHUMA	0.39	0.71	1.02	0.88	0.19	6.37	0.80	0.60	0.02	0.68	2.39	0.34	0.72	0.97
	31	(44, 50)	MALLA - ICHOCA	0.53	0.97	1.13	0.90	0.12	6.60	0.76	0.62	0.01	0.63	2.06	0.19	0.90	0.99
21.6%	32	(31, 32)	TIPUANI - MAPIRI	0.37	0.70	1.06	0.89	0.05	8.00	0.77	0.62	0.02	0.92	2.51	0.77	0.96	0.93
	32	28	TACACOMA	0.44	0.55	1.00	0.83	0.09	8.09	0.78	0.60	0.02	0.67	2.69	0.63	0.83	0.95
21.6%	33	(

Como se puede notar, son los municipios más parecidos respecto a los indicadores mencionados. Estos dos municipios aparecen juntos a la derecha del segundo cuadrante del gráfico N° 1 y en el primer nivel del dendograma del gráfico N° 2.

En segundo nivel o segunda etapa con un 12,4 % de similitud aparecen los municipios de Comanche y Puerto Acosta. Difieren mayormente en los indicadores ED0402 de 6,65 a 6,75 y en el indicador de vivienda VV0204 de 0,62 a 0,71. En los demás son altamente similares o iguales. Respecto a la identificación de ambos municipios en el gráfico N° 1, se encuentran en el segundo cuadrante en la parte superior muy juntos, como era de esperar.

En el tercer nivel de similaridad de los indicadores (14,0 %) se encuentran juntos los municipios de Luribay y Collana. Se parecen menos en el indicador DM0102 de 0,20 a 0,29 y en VV0203 de 0,60 a 0,74. Son iguales en EI0105 con 0,03. En el gráfico N° 1 del escalado multidimensional aparecen unidos en el primer cuadrante muy cerca al punto de origen.

En el cuarto nivel de parentesco con un 14,2 % están los municipios de Ayata y Charazani con igualdad en el indicador DM0506 (0,73), EI0105 (0,03) y con leve diferencia en ED0101, EI0101, VV0204 y NBI2001. En el gráfico N° 1 se encuentran juntos en el tercer cuadrante sector medio.

Con 14,7 % de parentesco porcentual, en el quinto nivel, se encuentran los municipios de Apolo e Ixiamas. Aparecen juntos en el primer cuadrante, muy cercanos al origen del gráfico N° 1. Difieren en el indicador ED0101, ED0402, EI0107, VV0203 y NBI2001. En el dendograma del gráfico N° 2, hacen un cluster por el sector izquierdo medio.

En el sexto nivel con 14,9 % se produce un conglomerado similar de tres municipios: Luribay, Collana (que ya se unieron en la tercera etapa) y Irupana. Los indicadores más similares son DM0506, DM0507, EI0101, EI0102 y EI0105. En el gráfico N° 1 del escalado multidimensional aparecen Luribay y Collana en el primer cuadrante, como se indicó anteriormente, pero unen formando un cluster de tres municipios ubicando a Irupana entre el tercero y cuarto cuadrante muy cerca al origen formando una especie de línea.

En el nivel o etapa siete están unidos los municipios de Achacachi y Batallas con un 15,9 % de similitud. Según el gráfico N° 1 corresponden al cluster 6 y 59, respectivamente y se encuentran Achacachi casi sobre la línea del origen y el cluster 59 de Batallas en el segundo cuadrante muy cerca al origen. En el gráfico N° 2 ambos municipios se anidan cerca al medio del dendograma y comparten el cluster con otros municipios.

Se continúa el análisis hasta llegar, por ejemplo a la etapa 10 que con un nivel porcentual del 17,2 % se unen en un cluster los municipios de Achacachi, Batallas, Luribay, Collana, Irupana y Tiahuanacu. Estos municipios aparecen agrupados en el centro del dendograma donde anteriormente, en la etapa 3 a nivel 14,0 % se unieron los municipios de Luribay y Collana. Si se observan los indicadores, el DM0102 cambia de 0,20 a 0,22, el DM0506 de 0,56 a 0,63. El indicador DM0507 (relación de masculinidad) varía de 0,98 a 1,10. ED0401 (tasa de analfabetismo) cambia de 8 % a 11 %, el ED0402 va de 7,47 a 7,60, EI0101 de 0,81 a 0,86. El indicador EI0102 va de 0,54 a 0,59; EI0105 de 0,02 a 0,03; EI0107 de 0,71 a 0,86; VV0101 de 2,04 a 2,21; VV0203 de 0,74 a 0,77; VV0204 de 0,93 a 0,94; y finalmente NBI2001 de 0,94 a 0,95.

Así sucesivamente se continua el análisis concluyendo que los municipios vistos en el gráfico N° 1 cambiarían de posición si se aumentasen o disminuyesen indicadores sociodemográficos, pero no cambiarían de estructura, es decir, continuarían esparcidos al medio de los cuatro cuadrantes del gráfico. La estructura de la posición de los municipios no cambiará, a excepción de los municipios VAR1 (La Paz) y VAR5 (El Alto) separados de los demás.

Los municipios de La Paz y El Alto se separan del grupo compacto, así como se observa en el gráfico N° 2. Esto ocurre especialmente porque recogen información diferenciada urbana de lo rural, que son la mayoría de los municipios, datos influyentes en los indicadores para las ciudades que las componen.

Si se observa el gráfico N° 2, el municipio de El Alto es el que cierra el dendograma en penúltima instancia, cerrando el municipio de La Paz con todos los demás municipios.

El análisis de clasificación mediante estos dos métodos es complementario y ayudan a confirmar la clasificación siempre y cuando se utilicen los mismos métodos de distancia.

Referencias

- [1] T. W. Anderson (1984). An Introduction to Multivariate Statistical Analysis. Ed. John Wiley & Sons.
- [2] Richard A. Johnson y Dean W. Wichern (2007). Applied Multivariate Statistical Analysis.
- [3] Ezequiel Uriel y Joaquín Aldás (2005). Análisis Multivariante Aplicado.
- [4] César Pérez (2004). Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS.

Estimación Bayesiana de los Resultados de las Encuestas Electorales

Lic. Dindo Valdez Blanco
 ✉ dindovaldez@hotmail.com

Resumen. En el presente artículo se revisan los elementos básicos del paradigma bayesiano y se realiza una estimación bayesiana puntual y de intervalo de la proporción de votación a favor del actual presidente Evo Morales basado en los resultados de la encuesta realizada por la empresa Ipsos en el mes de agosto de 2014.

Palabras clave: Inferencia Bayesiana, encuestas electorales, función a priori, función a posteriori.

1. Introducción

El presidente del Estado Evo Morales, domina con amplia ventaja en las encuestas seis semanas antes de las elecciones generales del 12 de octubre. En la encuesta realizada por la encuestadora privada Ipsos, Morales se encuentra a la cabeza de intención electoral con un caudal del 59 %, y 42 % de voto seguro. La ficha técnica de dicho sondeo indica que se entrevistó a 3000 ciudadanos entre el 1 y 16 de agosto, la muestra tiene una confianza de 95 % y un error máximo de estimación de 1.79 %, la encuesta fue efectuada en 10 ciudades capitales, 10 localidades urbanas y 88 localidades rurales de Bolivia.

Estos resultados son estimaciones del tipo frecuentista denominado estimación clásica, sin embargo es posible obtener una estimación bayesiana y calcular la precisión que tienen ambos estimadores.

2. El paradigma bayesiano

Si definimos el parámetro π como la proporción de electores de la población que tienen la intención de votar por Evo Morales, definimos también una variable binaria X , tal que $x = 0$ si una persona no piensa votar por el actual presidente, y $x = 1$ si tiene la intención de votar por el candidato oficialista. Entonces la función de probabilidad de X condicionada en el parámetro π es Bernoulli, tal que su función de probabilidad paramétrica es:

$$f(x|\pi) = \pi^x(1 - \pi)^{1-x}; \quad x = 0,1., 0 \leq \pi \leq 1. \tag{1}$$

El estimador máximo verosímil de π basado en una muestra aleatoria de tamaño n de X está definido por: $\hat{\pi}_F = \sum_{i=1}^n x_i/n$, el mismo basado en los datos de la encuesta electoral resulta igual a $\hat{\pi}_F = 0.59$. En este caso se considera a π como un parámetro desconocido pero fijo.

Sin embargo desde el punto de vista bayesiano, el parámetro π es una variable aleatoria. Este aspecto tiene sentido porque la intención de voto de los electores es cambiante, debido a muchos factores como la información que emiten los medios de comunicación y las propuestas de los candidatos. La estimación bayesiana se basa en la función de probabilidad condicional denominada función de probabilidad a posteriori definida por:

$$f(\pi|muestra) = \frac{f(\pi) \times f(muestra|\pi)}{f(muestra)} \tag{2}$$

3. Estimación bayesiana tomando una función a priori Beta

Si se supone una función a priori $Beta(\alpha, \beta)$ para π con α y β conocidos:

$$f(\pi) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1 - \pi)^{\beta-1}; \quad 0 \leq \pi \leq 1 \tag{3}$$

La función de probabilidad a posteriori de π dada la información de la muestra resulta:

$$f(\pi|muestra) = \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + \sum x_i)\Gamma(\beta + n - \sum x_i)} \pi^{\alpha + \sum x_i - 1} (1 - \pi)^{\beta + n - \sum x_i - 1}; \quad 0 \leq \pi \leq 1. \quad (4)$$

Claramente se reconoce que la distribución posterior corresponde a una ley $Beta(\alpha' = \alpha + \sum x_i, \beta' = \beta + n - \sum x_i)$. Basado en este hecho el estimador de Bayes se define como la media condicional de la función posterior:

$$\hat{\pi}_B = E(\pi|muestra) = \frac{\alpha + \sum x_i}{\alpha + \beta + n}. \quad (5)$$

Para la elección de los valores iniciales α y β , se puede utilizar un sistema de ecuaciones basados en la media y varianza de la distribución a priori de π , si se considera una media igual a 0.5 y una varianza de 0.01, el sistema resulta:

$$E(\pi) = \frac{\alpha}{\alpha + \beta} = 0.5 \quad V(\pi) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = 0.01 \quad (6)$$

De donde se obtienen los valores iniciales $\alpha = 12, \beta = 12$. Tomando la información de la muestra $n = 3000, \sum x_i = 1770$, el estimador bayesiano resulta $\hat{\pi}_B = 0.5893$. De la misma manera se puede calcular un intervalo de confianza al 95 % para la estimación bayesiana basado en los percentiles 2.5 y 97.5 de la distribución $Beta(\alpha' = 1782, \beta' = 1242)$. De donde resulta el intervalo (0.5717; 0.6068).

4. Estimación bayesiana tomando una función a priori Uniforme

Si no se tiene ninguna suposición sobre la distribución del parámetro π se puede asumir una función de probabilidad uniforme $U(0,1)$, es decir: $f(\pi) = 1, 0 \leq \pi \leq 1$. De igual manera se obtiene la función de densidad posterior de π dada la información de la muestra:

$$f(\pi|muestra) = \frac{\Gamma(n + 2)}{\Gamma(\sum x_i + 1)\Gamma(n - \sum x_i + 1)} \pi^{\sum x_i} (1 - \pi)^{n - \sum x_i}; \quad 0 \leq \pi \leq 1. \quad (7)$$

La expresión anterior muestra que la distribución a posteriori también sigue una ley $Beta(\alpha' = \sum x_i + 1, \beta' = n - \sum x_i + 1)$. Nuevamente reemplazando las observaciones de la muestra, se tiene que la distribución a posteriori es $\pi|muestra \sim Beta(\alpha' = 1770, \beta' = 1231)$. Por tal razón el estimador puntual bayesiano de la proporción de la población que piensa apoyar al presidente en las próximas elecciones es igual a la media condicional de la densidad posterior.

$$\hat{\pi}_B = E(\pi|muestra) = \frac{\alpha'}{\alpha' + \beta'} = 0.5898 \quad (8)$$

Finalmente el intervalo de confianza al 95 % para el parámetro π desde el enfoque bayesiano es (0.5626; 0.5976).

5. Medida de precisión de los estimadores

El estimador frecuentista clásico es insesgado, en cambio los estimadores bayesianos por lo general tienen un sesgo, la diferencia radica en que el Error Cuadrático Medio de las estimaciones es menor en el caso Bayesiano en comparación del estimador frecuentista. El error cuadrático medio (ECM) de un estimador $\hat{\theta}$ está definido como:

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = (E[\hat{\theta}] - \theta)^2 + Var[\hat{\theta}] \quad (9)$$

En el caso del estimador frecuentista $\hat{\pi}_F$, su media y varianza son: $E[\hat{\pi}_F] = \pi$, $V[\hat{\pi}_F] = \pi(1 - \pi)/n$. Por lo que su ECM resulta:

$$ECM(\hat{\pi}_F) = \frac{\pi(1 - \pi)}{n} \tag{10}$$

Con $n = 3000$ y $\pi = 0.5$, se tiene un error cuadrático medio igual a $ECM(\hat{\pi}_F) = 0.000083$.

De igual manera para el caso bayesiano la media y varianza de $\hat{\pi}_B$ es igual a:

$$E(\hat{\pi}_B) = \frac{\alpha + n\pi}{n + \alpha + \beta} \quad V(\hat{\pi}_B) = \frac{n\pi(1 - \pi)}{(n + \alpha + \beta)^2} \tag{11}$$

De tal forma que su error cuadrático medio es igual a:

$$ECM(\hat{\pi}_B) = \left(\frac{\alpha - (\alpha + \beta)\pi}{n + \alpha + \beta} \right)^2 + \frac{n\pi(1 - \pi)}{(n + \alpha + \beta)^2} \tag{12}$$

Con $n = 3000$, $\pi = 0.5$, $\alpha = 12$ y $\beta = 12$, se obtiene $ECM(\hat{\pi}_B) = 0.000082$. Esto confirma que el estimador bayesiano tiene menor error en comparación con el estimador frecuentista.

La siguiente tabla muestra las estimaciones del parámetro π con sus respectivos intervalos de confianza.

Tabla 1  Estimaciones del parámetro π

Método	Función a priori	Estimación puntual	Intervalo de confianza	Error cuadrático medio
Frecuentista	—	0.5900	(0.5724; 0.6076)	0.000083
Bayesiano	<i>Beta</i> (12,12)	0.5893	(0.5717; 0.6068)	0.000082
Bayesiano	<i>Uniforme</i> (0,1)	0.5898	(0.5626; 0.5976)	0.000083

6. Conclusión

Mediante esta aplicación se concluye que los estimadores bayesianos tienen igual o mejor precisión que el estimador frecuentista, adicionalmente la estimación bayesiana es muy parecida en los dos casos, no importando la función a priori inicial que se tome.

Referencias

- [1] Bernardo, J. M., Bayarri, M. J., Berger, *Bayesian Statistics*. Oxford: University Press, 2003.
- [2] Lee, P. M. *Bayesian Statistics: An Introduction* (3rd. ed.) London University Press, 2004.

Implicaciones del Modelo de Guilford en la Enseñanza de Matemáticas, Física, Química y Biología «El pensamiento divergente»

MgSc. F. Rodrigo Vargas Salazar

✉ cipacohc@gmail.com

Resumen. Se plantea esta investigación, para evaluar el nivel de pensamiento divergente. El enfoque es de tipo cuantitativo y diseño longitudinal de tipo panel, efectuándose un monitoreo a 39 estudiantes de ambos géneros que aprobaron el curso preuniversitario a la Carrera de Bioquímica de la Facultad de Ciencias Farmacéuticas y Bioquímicas de la UMSA en el segundo semestre del 2013. Así, se pudo determinar en el presente trabajo, que las tres dimensiones o factores del pensamiento divergente estudiadas [DMU, DSU, DMS] son bajísimas. Por tal razón, el presente artículo nos muestra una visión reflexiva en torno al modelo de Joy Paul Guilford aludiendo a tres factores del pensamiento divergente, y a sus implicaciones en el ámbito pedagógico. Así mismo, una de las fases más importantes de la educación es que los educadores empleen estos u otros dispositivos que favorezcan el desarrollo de comportamientos creativos en los estudiantes haciendo que éstos se interesen más por lo que aprenden. No es difícil darnos cuenta que la estructura del intelecto (EI) de Guilford simplificará nuestra comprensión sobre cómo fomentar la creatividad dentro y fuera del aula; y tener en cuenta que a mayor desarrollo del pensamiento divergente será mayor nuestra creatividad.

Palabras clave: Pensamiento divergente.

Abstract. We make this investigation, to evaluate the level of divergent thinking. The focus is of quantitative type and longitudinal design of type panel, being made to 39 students male and female that approved the course of preparation university to the Career of Biochemistry of the School of Pharmaceutical and Biochemical Sciences of the UMSA in the second semester of the 2013. So, you could determine work presently, that the three dimensions or factors of the studied divergent thinking [DMU, DSU, DMS] they are low. For such a reason, the present article shows us a reflexive vision around the pattern of Joy Paul Guilford mentioning to three factors of the divergent thinking, and to its implications in the pedagogic environment. Likewise, one of the most important phases in the education is that the educators use these or other devices that favor the development of creative behaviors in the students making that these are interested more in what they learn. It is not difficult to realize that the structure of the intellect (EI) of Guilford it will simplify our understanding on how to foment the creativity inside of and outside of the classroom; and to keep in mind that to bigger development of the divergent thinking will be bigger our creativity.

Key words: Divergent thinking.

1. Introducción

Dada la importancia indudable que poseen estas ciencias al menos en las áreas de biología, ingenierías y ciencias de la salud, hay también la necesidad de emplear por parte del educador un proceso creativo en el aula; esta afirmación es válida también para cualquier otra asignatura sea en Colegio o en la Universidad, hablamos de un proceso que motive la expresión creativa en los estudiantes.

La divergencia en las matemáticas¹ incluso más que las otras ciencias, puede resultar a priori algo molesta para el docente, pues como bien sabemos los resultados en términos matemáticos son únicos, por ejemplo

¹La divergencia en las matemáticas significa <<elaborar preguntas de matemáticas para estimular los factores del pensamiento divergente que quisiéramos, ya sea DMU, DMS, DSU, etc>>.

$2+2=$ es 4, sin embargo, no son únicos los cálculos y las diversas maneras de llegar al resultado, como tampoco es única la aplicabilidad de logaritmos o derivadas a solo problemas matemáticos, ya que se las puede aplicar a la economía, a la microbiología, a la fabricación de materiales, al diseño de automóviles. Por ello, el que el docente conozca y aplique el modelo del intelecto de Guilford en aula fomentaría el desarrollo de habilidades creativas de sus estudiantes.

Y dado que el pensamiento es una actividad constante en todo ser humano y durante toda la vida de éste, asumimos también que el pensamiento es lo que pensamos. (Merani, A.; 1979:117) nos indica que se entiende el pensamiento, como la secuencia de procesos mentales de carácter simbólico, estrechamente relacionados entre sí, que comienzan con una tarea o un problema, en general por grados, y llegan a una conclusión o a una solución.

1.1. Clasificación del pensamiento

Merani y Dorsch, F. (2008) de acuerdo a lo expresado líneas arriba hacen una distinción precisa de los significados de algunos tipos de “pensamiento”, y los clasifica como: Pensamiento elementales; Pensamiento asociativo; Pensamiento directo, también llamado pensamiento intencional; Pensamiento autístico, también llamado pensamiento desiderativo, ilusorio, mágico; Pensamiento deréístico, es el pensamiento ilógico, de tipo fantástico; Pensamiento discursivo, también llamado pensamiento conceptivo; Pensamiento de grupo; Pensamiento animista; Pensamiento convergente, también llamado producción convergente; Pensamiento divergente, también llamado producción divergente; Pensamiento abstracto; Pensamiento crítico; Pensamiento aparentemente dialéctico.

1.2. El pensamiento divergente

Ken Robinson PhD, (2010) hoy en día un reconocido erudito mundial en reformas educativas internacionales en temas de educación, puntualiza que el pensamiento divergente es “la capacidad de ver montones de respuestas posibles a una cuestión”, también nos dice que son “los montones de maneras distintas de interpretar una cuestión”. Joy Paul Guilford PhD en su teoría de la estructura del intelecto menciona claramente que <<el pensamiento divergente es el mecanismo mental usado para la resolución de problemas que admiten soluciones distintas todas ellas igualmente válidas, aunque puedan ser de distinto nivel o calidad. El sujeto no queda condicionado por alternativas concretas sino que puede sacar sus propias conclusiones frente a los problemas o cuestiones planteados pudiendo obtenerse las respuestas más imprevisibles>>.

1.3. Origen del pensamiento divergente

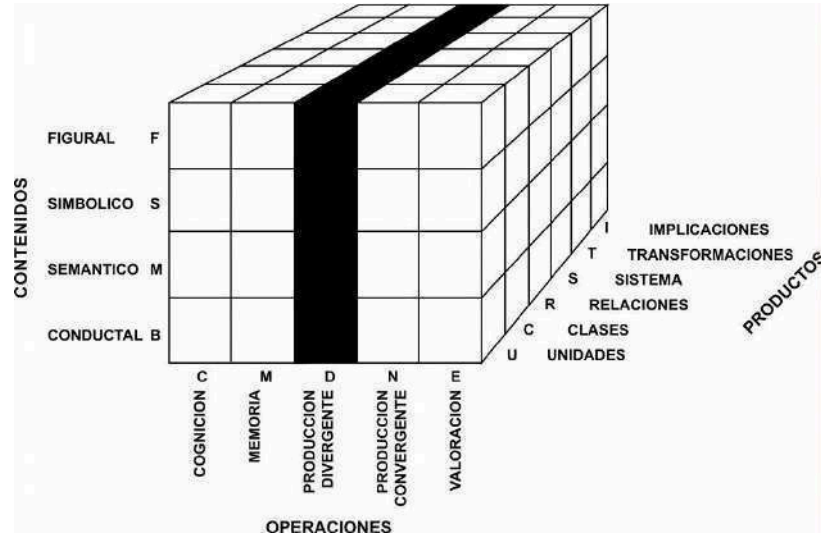
Este modelo o teoría sobre la estructura del intelecto (EI) también conocida como teoría general de la inteligencia planteada por Joy Paul Guilford PhD, quien es además el primer científico quien formuló el término de producción divergente o pensamiento divergente, siendo así, la base de incontabilísimas publicaciones científicas sobre creatividad. Este célebre científico postuló que “nuestro intelecto humano esta compuesto de 120 aptitudes intelectivas”; a diferencia de él Howard Gardner PhD propone la teoría sobre la estructura de la mente llamada “las 7 inteligencias múltiples (IM)” como ser la Verbal; Lógico-Matemática; Espacial; Musical; Cinestésica; interpersonal; intrapersonal” aunque también en posteriores investigaciones complementa a su propia teoría con otras 4 operaciones mentales de orden superior como la “creatividad, el sentido común, la capacidad metafórica, la sabiduría”.

1.4. Definición y clasificación del pensamiento divergente

La geometría parte indispensable de la ciencia matemática es el artificio para poder explicar la (EI) con las 120 aptitudes intelectivas de Guilford, el cual se grafica en forma de cubo de forma tridimensional (3D); éste cubo se muestra a continuación y hace énfasis en el pensamiento divergente en color negro.

Tres son las variables constitutivas básicas del cubo tridimensional de la teoría de Guilford (operaciones u operaciones mentales, contenidos y productos).

Diagrama 1 ➔ Cubo tridimensional de Guilford



Fuente: Extraído de Vargas, S. R. (2014). Tesis doctoral sobre evaluación predictiva del éxito académico en base a la correlación de tres factores de selección con el puntaje final del test electrónico como contribución al actual sistema de admisión universitaria. (pp. 71-81). UMSA. La Paz, Bolivia.

1. OPERACIONES,² se entiende como lo que el organismo hace con la información, y pueden ser de 5 tipos: Cognición o Conocimiento (C); Memoria (M); Producción divergente (D); Producción convergente (N); Valoración o Evaluación (E).
2. CONTENIDOS,³ se entiende como la generalidad de la información, y pueden ser de 4 tipos: Figural o Figurativo (F); Simbólico (S); Semántico (M); Conductual o Comportamental o Conductista (B).
3. PRODUCTOS,⁴ se entiende como la organización de la información después de procesarla en la mente, y pueden ser de 6 tipos: Unidades (U); Clases (C); Relaciones (R); Sistemas (S); Transformaciones (T); Implicaciones (I),

Dado que puede resultar complejo concebir el cubo mostrado en tres dimensiones (3D), entonces nos sería más cómodo verlo en dos dimensiones (2D) apreciándose así dichos perfiles intelectivos.

En el diagrama N° 2 mostrado abajo, se observa cinco matrices, cada una de ellas representa una operación mental. En cada matriz cada columna representa un tipo de contenido (de antemano conocemos que hay 4 tipos de contenido), y cada fila representa un tipo de producto (de antemano conocemos que hay 6 tipos de producto). Por ende, esto nos quiere decir que en cada matriz hay 24 casillas x 5 matrices u operaciones mentales y tendríamos 120 aptitudes intelectivas de Guilford.

La forma de emplear el cubo de Guilford se expresaría de la siguiente manera: cuando un estudiante respondiese a una interrogante cualquiera, se pondría la nota correspondiente en la casilla adecuada dentro de la matriz a la que corresponde la operación mental.

Los pasos son simples, por ejemplo si quisiéramos conocer la fluidez verbal, es decir el DSU en correspondencia a la designación Guilforiana, tendríamos que:

se coloca la "operación mental" correspondiente al pensamiento divergente ⇒ D
 luego se coloca el "contenido" ⇒ S
 y finalmente se coloca el "producto" ⇒ U

²(Davis, G.;1992:32).

³(Davis, G.;1992:32).

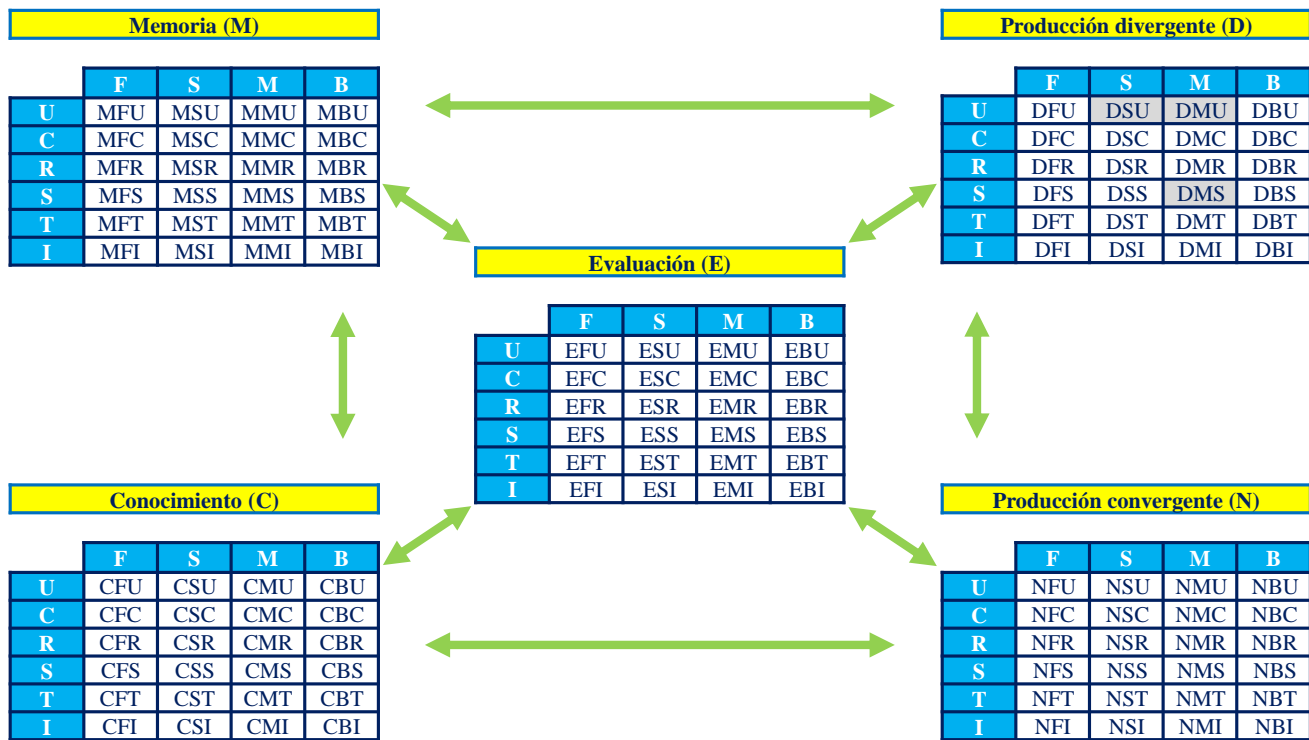
⁴(Davis, G.;1992:32).

Ahora bien, para efectos de congruencia y elección, en este artículo nos enfocaremos en la matriz de producción divergente (D) también conocida como matriz de pensamiento divergente; obviamente vemos que posee 24 casillas, o sea 24 modalidades de pensamiento divergente; el investigador fijó su mirada en 3 modalidades de pensamiento divergente: DSU, DMU, DMS.

1.5. Indicadores o factores del pensamiento divergente

- o Fluidez verbal (DSU), es la producción divergente de unidades simbólicas. Por ejemplo, enumerar o escribir la mayor cantidad posible de palabras que empiecen por una letra dada; o escribir palabras que contengan una letra determinada; o escribir la mayor cantidad de definiciones posibles que se le puede dar a un objeto determinado.
- o Fluidez asociativa (DMR), es la producción divergente de relaciones semánticas. Por ejemplo, proponer diferentes adjetivos para completar similes o semejantes.
- o Fluidez ideativa o ideacional (DMU), es la producción divergente de unidades semánticas. Por ejemplo, escribir la mayor cantidad de usos posibles que se le puede dar a un objeto determinado.
- o Fluidez de expresión (DMS), es la producción divergente de sistemas semánticos. Por ejemplo, explicar de formas distintas oraciones que contienen similes; o escribir la mayor cantidad posible de interpretaciones que se le puede dar a un objeto determinado.

Diagrama 2 ➔ 120 aptitudes intelectivas de Guilford



Fuente: Elaboración propia en base a la literatura revisada. En color gris se muestra las 3 aptitudes intelectivas que se estudió en este trabajo de investigación.

2. Método

El enfoque de esta investigación es de tipo cuantitativo y diseño longitudinal de tipo panel, efectuado en estudiantes que aprobaron el curso preuniversitario a la Carrera de Bioquímica de la Facultad de Ciencias Farmacéuticas y Bioquímicas de la UMSA en el segundo semestre del 2013. Ingresaron al estudio

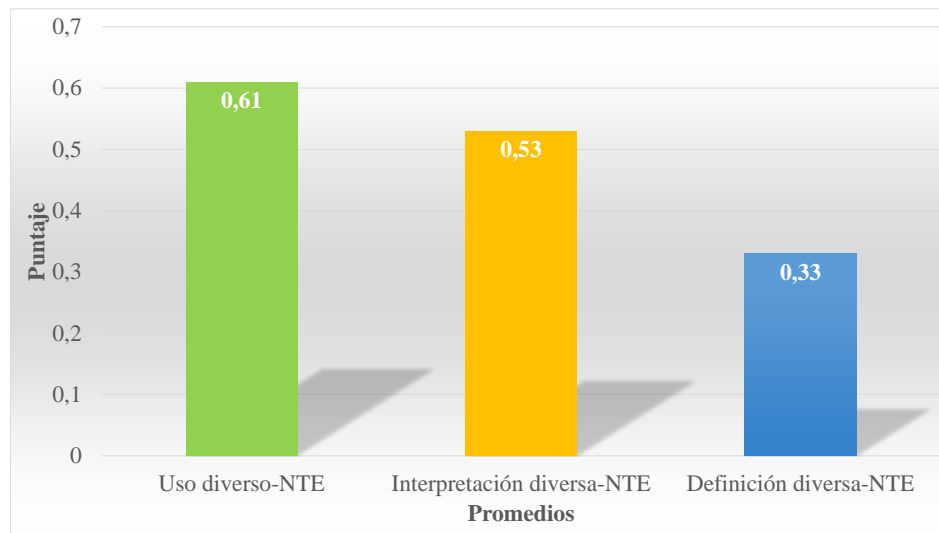
los estudiantes que cumplían los siguientes criterios: estudiantes regulares, estudiantes que están inscritos legalmente, estudiantes que aprobaron dicho curso prefacultativo, estudiantes que no abandonaron la Carrera en el primer año curricular. Y se excluyeron del trabajo a: estudiantes inscritos a la prueba de suficiencia académica y que aprobaron, estudiantes inscritos a la prueba de suficiencia académica y que reprobaron, estudiantes inscritos a otras Facultades y/o Carreras de la UMSA u otras Universidades. Para ello, se efectuó el siguiente procedimiento: Previamente, se diseñó y validó un nuevo instrumento de admisión universitaria de tipo electrónico, capaz de medir o evaluar el pensamiento divergente en los estudiantes. Luego, desde un inicio hubo 411 estudiantes inscritos, de ellos desertaron 119, lo que nos lleva a contar con solo 292 estudiantes que cursaron el prefacultativo. Desafortunadamente en el proceso reprobaron 179, y en contraparte aprobaron 101, de éstos, 39 estudiantes son de bioquímica y 62 de farmacia. A estas cantidades netas hay que añadir los 12 estudiantes que vencieron el examen de suficiencia. Lo que nos da un total de 113 estudiantes que ahora cursan el primer año para ambas Carreras. Por tanto, el investigador monitoreó exclusivamente a los 39 estudiantes de bioquímica dada las características del diseño de investigación de tipo panel y según lo planificado de antemano. El acta de calificaciones finales de aprobación –al cual se tuvo acceso previa solicitud- comprobó la cantidad de los estudiantes aprobados al curso prefacultativo como también si fueron precisamente los que figuraban en la base de datos del software. Este proceder fue exhaustivo en los dos momentos de la investigación, con el fin de eliminar a los desertores, a los reprobados.

3. Resultados

El estudio efectuado en el curso preuniversitario de la Carrera de Bioquímica-UMSA en el segundo semestre del 2013, mostró que de una población heterogénea compuesta por 39 estudiantes de ambos géneros sus promedios y puntajes en pensamiento divergente son muy bajos.

En el gráfico 1, se muestra los datos obtenidos de varones y mujeres, en cambio en el gráfico 2 y 3 se hace respectivamente una diferenciación por género, con el fin de efectuar una comparación más minuciosa.

Gráfico 1 ➔ Promedios y puntajes en pensamiento divergente NTE (varones y mujeres)



Fuente: Excel – Elaboración propia.

Referencia: NTE notas del test electrónico

Uso diverso es la sección en que mejor se desempeñaron todos los participantes del test. Interpretación diversa y definición diversa como lo prueba el gráfico le sobrevienen correspondientemente a Uso diverso. Este panorama en comparación con el “Gráfico N° 2 – Promedios y puntajes en pensamiento divergente NTE (varones)” mostrado líneas abajo nos indica que uso diverso es la sección en que mejor desempeño

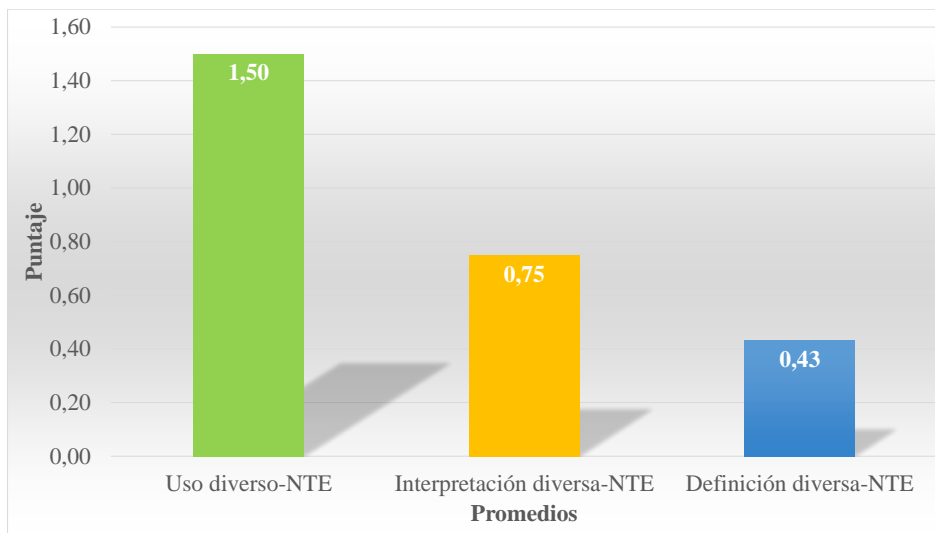
poseen los varones. De igual manera sucede con las mujeres, pues uso diverso es la sección en que mejor desempeño poseen, esto es expresado en el “Gráfico N° 3 – Promedios y puntajes en pensamiento divergente NTE (mujeres)”.

Estas tres dimensiones o factores del pensamiento divergente, nos lleva a inferir que en general el desempeño de la Fluidez ideativa o ideacional (DMU) correspondiente a “escribir la mayor cantidad de usos posibles que se le puede dar a un objeto determinado” es muy baja en ambos géneros, dado que [0.61] equivale decir que escribieron en 3 de 140 casillas (se prepararon 7 preguntas para esta sección; cabe hacer notar que cada pregunta comprende 20 casillas a llenar o escribir la respuesta por el estudiante) y esto a su vez representa tan solo un 2 % sobre el 100 % esperado.

Respecto al desempeño en general de la Fluidez de expresión (DMS) correspondiente a “escribir la mayor cantidad posible de interpretaciones que se le puede dar a un objeto determinado” es muy bajísima en ambos géneros, dado que [0.53] equivale decir que escribieron en 2 de 120 casillas (se prepararon 6 preguntas para esta sección; cabe hacer notar que cada pregunta comprende 20 casillas a llenar o escribir la respuesta por el estudiante) y esto a su vez representa tan solo un 1.7 % sobre el 100 % esperado.

Finalmente el desempeño en general de la Fluidez verbal (DSU) correspondiente a “escribir la mayor cantidad de definiciones posibles que se le puede dar a un objeto determinado” es imperceptible en ambos géneros, dado que [0.33] equivale decir que escribieron en 1 de 140 casillas (se prepararon 7 preguntas para esta sección; cabe hacer notar que cada pregunta comprende 20 casillas a llenar o escribir la respuesta por el estudiante) y esto a su vez representa tan solo un 0.7 % sobre el 100 % esperado.

Gráfico 2 ➔ Promedios y puntajes en pensamiento divergente NTE (varones)



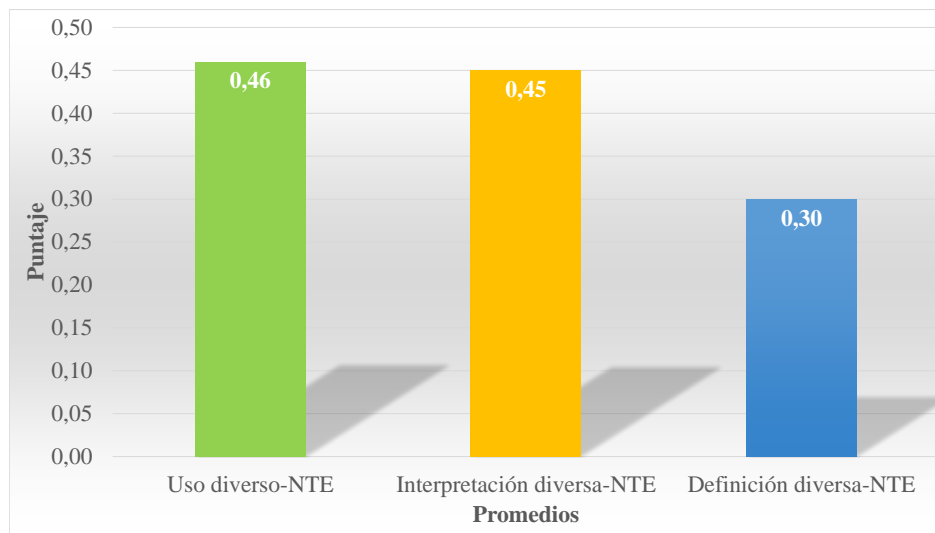
Fuente: Excel – Elaboración propia.

El primer mejor puntaje para ambos géneros es el de usos diversos. El segundo mejor puntaje para ambos géneros es el de interpretación diversa. El tercer mejor puntaje para ambos géneros es el de definición diversa.

La Fluidez ideativa o ideacional (DMU) correspondiente a “escribir la mayor cantidad de usos posibles que se le puede dar a un objeto determinado” es insignificante tanto en varones como en mujeres, dado que el puntaje mínimo aceptable por casilla es de [0.25] el cual a su vez equivale escribir en una de 140 casillas, y como se muestra en el gráfico los puntajes de ambos géneros apenas supera este puntaje mínimo.

Respecto a la Fluidez de expresión (DMS) y a la Fluidez verbal (DSU) no se los tomará en cuenta por su insignificancia investigativa.

Gráfico 3 ▶ Promedios y puntajes en pensamiento divergente NTE (mujeres)



Fuente: Excel – Elaboración propia.

4. Discusión

El desarrollo de los tres factores del pensamiento divergente [DMU, DMS, DSU] fortalecería de cierta manera la creatividad. La sobre estimulación divergente recaería a su vez en el desarrollo de habilidades como la sensibilidad a los problemas, fluidez de pensamiento, flexibilidad de pensamiento, originalidad, capacidad de redefinición, capacidad de análisis o capacidad de abstracción, capacidad de síntesis, coherencia en organización -8 cualidades o rasgos o indicadores de la persona creativa⁵-, tal y como lo propusieron Guilford PhD y Víctor Lowenfeld PhD en la Universidad de California y la de Pennsylvania-EEUU respectivamente.

5. Ejemplos de aplicación del pensamiento divergente

A continuación se muestran seis ejemplos (según sea la asignatura) que el docente podría formular en sus exámenes o controles de lectura a sus estudiantes con el fin de desarrollar en éstos el pensamiento divergente y por ende la creatividad.

Manera 1: Para estimular el DMU

Ej. Nº 1. Indica aplicaciones que se te ocurran de: a) El segundo principio de termodinámica. b) Las leyes de los gases.

Ej. Nº 2. Indica 22 usos de un clip de plástico, y luego responde qué es el plástico?

Manera 2: Para estimular el DMS

Ej. Nº 3. Cómo le explicarías a tu hermanito de 6 años la diferencias o similitudes entre aceleración, velocidad e ir de prisa?

Ej. Nº 4. De los siguientes objetos que se muestran a continuación elige objetos a tu criterio e idea 2 experimentos distintos con ellos (dibuja y explica cada procedimiento).

Manera 3: Para estimular el DSU

Ej. Nº 5. En tus palabras, cómo defines parábola? (emplea 88 palabras en tu definición).

Ej. Nº 6. Escribe la mayor cantidad de definiciones posibles acerca de la gravedad.

⁵(Davis, G.; Scott, J.; 1992:22-23).

6. Conclusiones

La enseñanza de las ciencias no debería ser difícil, mas bien debería inducirnos más curiosidad, debería ayudarnos a saber leer y comprender definiciones, a conocer cómo funcionan las cosas, a explicarlas en nuestras propias palabras y no solo describirlas mecánicamente, a disfrutar de poseer conocimiento teórico dándoles utilidad práctica, a aprender haciendo, ..., y el camino más corto y efectivo es desarrollando el pensamiento divergente. Por tanto, el pensamiento o producción divergente pertenecen a la generación de las ideas, el resolver problemas en donde la alta variedad de respuestas es muy importante. Esto quiere decir que los estudiantes creativos dependen del medio en el cual se hallan estudiando o trabajando. Finalmente, este estudio nos ayudó a obtener el perfil creativo de los estudiantes de manera electrónica (ver gráficos 1, 2, 3).

Referencias

- [1] DAVIS, GARY PhD, et al. 1992. *Training creative thinking*. Argentina: Paidós Educador.
- [2] DORSCH, FRIEDRICH. 2008. *Diccionario de psicología*. (8ª ed.). Barcelona, España: Herder.
- [3] GUILFORD, J. P. PhD. 1950. *Creativity*. p. 444-454. American Psychologist.
- [4] GUILFORD, J. P. PhD. 1952. *A factor analytic study of creative thinking*. Laboratory of Psychologist. University Southern California.
- [5] GUILFORD, J. P. PhD. 1952. *General psychology*. Princeton.
- [6] GUILFORD, J. P. PhD. 1956. *The structure of intellect*. p. 267-293. Psychological Bulletin.
- [7] GUILFORD, J. P. PhD. 1959. *Intellectual resources and their value as seen by creative scientists*. p. 128-149. University of Utah Press.
- [8] GUILFORD, J. P. PhD. 1959. *Three faces of intellect*. p. 469-479. American Psychologist.
- [9] GUILFORD, J. P. PhD. 1961. *Creative thinking in children at the junior high level*. Laboratory of Psychologist. University Southern California.
- [10] GUILFORD, J. P. PhD. 1961. *Creative thinking at the junior high school level*. University Southern California.
- [11] GUILFORD, J. P. PhD. 1961. *Factorial angles to psychology*. p. 1-20
- [12] GUILFORD, J. P. PhD. 1961. *The structure of intellect model: its uses and implications*. University Southern California.
- [13] GUILFORD, J. P. PhD. 1962. *Creativity: its measurement and development*. p. 151-168. New York
- [14] GUILFORD, J. P. PhD. 1966. *Instructional media and creativity*. p. 71-103. New York
- [15] GUILFORD, J. P. PhD. 1967. *Creativity: yesterday, today and tomorrow*. The journal of creative behavior.
- [16] GUILFORD, J. P. PhD. 1967. *The nature of human intelligence*. New York.
- [17] GUILFORD, J. P. PhD. 1968. *Intelligence, creativity, and their educational implications*. p. 187 California, San Diego.
- [18] MERANI, ALBERTO PhD. (Ed). 1979. *Diccionario de psicología*. (1ª ed.). D. F., México: Grijalbo.
- [19] ROBINSON, KEN PhD. 2010. *El elemento*. D.F., México: Random House Mondadori.
- [20] ROBINSON, KEN PhD. 2010. *Cahanging Paradigms* [DVD]. Inland.: RSA company.
- [21] ROBINSON, KEN PhD. 2010. *A iniciar la revolución del aprendizaje* [DVD]. Longbeach, California: TED company.

-
- [22] STERNBERG, R. J. PhD. 1993. *A guide to scientific writing for students and researchers*. Cambridge University:Press.
- [23] STERNBERG, R. J. PhD. 1999. *A triarchic approach to the understanding and assessment of intelligence in multicultural populations*. *Journal of school psychology*, 37(2), 145-159.
- [24] STERNBERG, R. J. PhD. 2003. *Our research program validating the triarchic theory of successful intelligence: Reply to gotfredson intelligence*, 31, 399-413.

¿Qué es el Valor P ?

Lic. Carmen Vega Flores
 ✉ c.vegaflores@hotmail.com

Una alternativa a las pruebas de hipótesis clásicas se puede realizar calculando una probabilidad denominada valor P , en algunos textos también se denomina como el valor de probabilidad o la probabilidad asociada o la probabilidad significativa.

El valor P se define como la probabilidad de obtener un resultado tan extremo como el estadístico de prueba observado (en la dirección de la hipótesis alternativa) cuando la hipótesis nula es verdadera. Esta probabilidad puede ser calculada para los valores observados del test estadístico o para alguna función de este como el estimador muestral del parámetro de la hipótesis. Por ejemplo, si se desea docimar las hipótesis:

$$\begin{aligned} H_0 : \mu &= 20 \\ H_1 : \mu &> 20. \end{aligned} \tag{1}$$

Y se tiene el estimador muestral observado $\bar{x} = 22$. Entonces el valor P es calculado como:

$$\text{valor } P = \Pr(\bar{x} \geq 22 / \mu = 20). \tag{2}$$

Si el valor P es pequeño, esto quiere decir que nuestra muestra produjo un resultado que tiene muy poca probabilidad de ocurrir cuando la hipótesis nula es cierta. En vista que el resultado de la muestra es una evidencia que contradice en probabilidad a la hipótesis nula, se toma la decisión de rechazar la hipótesis nula a favor de la hipótesis alterna. En cambio si el valor P es una probabilidad alta, esto indica que el resultado de la muestra es consistente con la hipótesis nula, por tanto no puede ser rechazada.

Si se deseara utilizar el valor P para tomar una decisión sobre la hipótesis nula, se debería seleccionar un nivel de significación α , si el valor P es mayor o igual que dicho nivel de significación entonces se debería rechazar porque la probabilidad de que H_0 sea verdadera es muy baja, en caso contrario no se debería rechazar porque el riesgo de que H_0 sea verdadera es mayor al nivel de significación determinado como riesgo máximo por el investigador. Es así como se puede interpretar al valor P como el nivel de significación mínimo para el cual la hipótesis nula es rechazada.

El valor P no solo se utiliza para la toma de decisión en una prueba de hipótesis, también se aplica para determinar el grado de la evidencia muestral a favor de dicha hipótesis. Por ejemplo, si se tiene una muestra aleatoria con un estadístico de T_1 el cuál resulta en un valor P igual a 0.012, mientras que para otra muestra aleatoria con un estadístico T_2 , se tiene un valor P igual a 0.045. Este hecho es una evidencia que la primera muestra contradice más fuertemente a la hipótesis nula que la segunda muestra porque el estadístico observado es menos probable en la primera muestra.

Finalmente se puede definir al valor P como una variable aleatoria, al ser una función del estadístico muestral, por ejemplo para la dócima de la media poblacional se utiliza el estadístico \bar{x} el cual tiene función de distribución acumulada $F(\bar{x})$ bajo H_0 y una función de distribución acumulada $G(\bar{x})$ bajo H_1 digamos de una cola derecha, entonces el valor P estaría definido por:

$$\text{valor } P = 1 - F(\bar{x}). \tag{3}$$

Este hecho implica que debemos tener presente que el cálculo del valor P puede variar y que es necesario analizar sus propiedades como ser: media, varianza, función de distribución, etc.

En Realidad: ¿Qué es el Éxito en la Vida?, y ¿Qué es el Éxito Académico?

MgSc. F. Rodrigo Vargas Salazar
✉ cipacohc@gmail.com

Resumen. Este artículo nos sumerge en una reflexión, e inicialmente nos señala que el éxito literalmente puede ser interpretado de diversas maneras, aunque la ciencia del éxito nos la define de forma muy concisa. Para ello, se postula un análisis del significado del éxito enfocado en la vida y en lo académico valiéndonos de literatura muy escrupulosa sobre este tema.

Palabras clave: Éxito en la vida, éxito en lo académico.

Abstract. This article submerges us in a reflection, and initially it points out us that the success literally can be interpreted in diverse ways, although the science of the success is defined it in a very concise way. For it, an analysis of the meaning of the success is postulated focused in the life and in the academic thing being worth us of very scrupulous literature on this topic.

Key words: Success in the life, success in the academic.

1. Introducción

Es muy importante para nosotros inicialmente, asimilar la terminología de la cual se hace uso en todo este trabajo, con la intención de no caer en mal interpretaciones semánticas que a posteriori nos confunda lo que realmente se desea decir.

Según lo dicho, el <<rendimiento académico>> se ve reflejado en el éxito académico del estudiante.

Y si intentásemos desentrañar al pie de la letra el término <<éxito académico>> tendríamos que apoyarnos tanto en definiciones efectuadas por expertos en la “ciencia del éxito”¹ como las efectuadas por “personas exitosas en diferentes rubros”, en caso contrario tendríamos dificultades en manejar dicho término para vincularlas con otras variables. Entonces, lo haremos sin ahondar la temática, sino mas bien de manera muy escueta.

1.1. Significado del éxito académico

Muchos expertos en negocios han definido el <<éxito>> como la capacidad de conseguir logros, la cual es diferente a la <<suerte>> que es el sentido de oportunidad. Lo que se trata de decir, es que en realidad la suerte en sí misma no existe. Tampoco existe la mala suerte.

Empero, si quisiéramos o intentásemos asimilar y definir la palabra suerte señalaríamos que es el producto de la siguiente ecuación: “suerte = oportunidad + tiempo + preparación”. Esto es así, dado que las oportunidades aparecen en cualquier momento o tiempo en nuestras vidas, y solo hay que estar preparados

¹La ciencia del éxito se inició con el investigador Napoleón Hill, quien trabajó para el magnate estadounidense Andrew Carnegie. Carnegie encontró a un joven de 19 años de edad en una fiesta, y le encomendó una tarea muy simple: debía investigar por el lapso de 25 años qué tenían en común las personas con éxito. Hill meditó la propuesta por unas semanas, y finalmente decidió aceptar la oferta. Transcurrido esas dos décadas Hill sabía cuales eran las características de las personas con éxito. Llegó a la conclusión de que hay varios factores para que una personas tenga éxito, el más importante de ellos es la actitud mental positiva u optimista, es decir el software (programación mental) y no el hardware (cerebro humano). Los otros 5 componentes del éxito son la autoestima, es decir quererse a sí mismo; la comunicación o sea la habilidad de expresar y mostrar al mundo lo que conocemos, lo que tienes dentro; las metas, es decir poner fecha límite de tiempo o de ejecución a nuestros sueños para que se hagan realidad; el trabajo, que se refiere a que el verdadero trabajo no es hacer lo que te guste, sino que te guste lo que haces; y la ambición, que es el deseo, las ganas de conseguir algo. Éstos se complementan, no es posible influir en una sin influir en las demás.

para identificarlas y aprovecharlas. Por ejemplo: Si no estuviéramos preparados y se presentaría alguna entrevista y oferta de trabajo en un momento dado digamos mañana a las 18:00, entonces diríamos que tuvimos mala suerte por no haber conseguido dicho trabajo, y cualquier otra persona opinaría lo mismo. En realidad esto se debería al simple hecho de no estar preparados; estar preparados es la clave!!!!. E inclusive, con esas definiciones en lenguaje lato diríamos que para tener éxito académico hay que tener suerte??. De acuerdo a esa ecuación, la respuesta es sí.

Ahora bien, retomando nuestro análisis, la gente triunfadora o exitosa supera los miedos, siempre son optimistas empedernidos. La gente exitosa se arriesga. Según William James Brian el éxito es como el destino, no es asunto de casualidad sino de decisión.

La definición de “éxito en la vida” según muchos expertos en la temática, por ejemplo Al Ries PhD² et al se marca a continuación: <<El éxito depende de tu reacción veloz ante las oportunidades inesperadas. Es estar en el lugar indicado en el momento indicado y hacer lo indicado. El éxito no llega cuando se necesita, sino cuando se encuentra, en donde uno lo busca. El éxito es conocer a la persona indicada; es decir, uno no puede alcanzar el éxito solo, los demás nos brindan éxito. Muchas veces necesitamos el reconocimiento del otro para que nos conduzca al éxito. No obstante, para muchísimas personas el éxito es aquello que uno quiere que sea, por ejemplo: tener dinero, poder, status, reconocimiento. En cambio para otras personas es lograr algo con sus vidas. Por ende, existen personas orientadas a las metas, otras orientadas al trabajo arduo, y otras orientadas al éxito; éstas últimas aprovechan los accidentes³, no se encasillan en algo, se arriesgan. Por consiguiente, quienes lograron el éxito son quienes dedicaron su vida a una idea>>.

Según Lair Ribeiro PhD (2000) el éxito es conseguir lo que se quiere en la vida, concepto que es diferente de <<felicidad>> la cual la define como querer lo que se ha conseguido en la vida; Tomando la palabras de Riberiro “Para ser feliz no se precisa de muchas cosas, basta con estar satisfecho con lo que tienes, es una cuestión de aceptación mental; no se debe asociar la felicidad con el éxito, pues la felicidad al igual que elegir nuestro destino es cuestión de decisión”.

El diccionario común (cf. Librairie Larousse, p 450) describe al éxito como sinónimo de acierto, ventaja, victoria, triunfo, y si bien no es imperfecta esta apreciación la cual fue analizada líneas arriba, su significado tiene una connotación o alcance particular cuando nos referimos al ámbito académico o educativo. En otras palabras, alcanzar el éxito académico es sutilmente diferente al referimos a lograr éxito en la vida aunque muchos principios rijan a ambos.

Entonces, esas 5 características de Ribeiro respecto a la ciencia del éxito antes mencionadas, son aplicables al ámbito académico o es que no lo son?. Para responder a esta cuestión, debemos remontarnos a los estudios de tipo longitudinal pioneros en su tipo efectuados por Lewis Terman (1921-1959) aplicados a mil quinientos estudiantes, él afirmaba que el cociente intelectual (CI) tenía una relación directamente proporcional con el éxito o rendimiento académico y con el éxito profesional.

Bloom confirmó estos datos más tarde acotando que también el medioambiente de aprendizaje el cual es subvalorado cumple aun un rol más substancial en el cociente intelectual y en el rendimiento académico.

Sucesivos estudios de Burt (1985), Eysenk (1986) y de Jenks (1986) atribuían que se podía determinar el éxito académico analizando exclusivamente los factores familiares desde el nacimiento. Hoy vemos que estos testimonios son insostenibles.

Posteriormente Terman (1959) se retractó, pues los estudios de Bloom le sirvieron para dejar de lado sus posturas unilaterales, mencionando que otras variables de tipo emocional e interpersonal, y el medioambiente de aprendizaje podían intervenir más que el CI de cada estudiante en el éxito académico.

²(Horse sense: the key to success is finding a horse to ride, 1992).

³Para efectos de comprensión en este trabajo, definiremos “accidentes” como aquellas situaciones [de probabilidad] inesperadas positivas o negativas en la vida como por ejemplo momentos de crisis económica financiera, y que deberíamos aprovecharlas al máximo. Es ver oportunidades donde otros vislumbran caos.

Los estudios experimentales emprendidos por Robert Sternberg (1999) concluyen demostrando que solo el 20 % del éxito académico esta atribuido al CI y que mas bien hay todo un cúmulo de factores que determinan el éxito o el fracaso académico: efecto pigmalión⁴ positivo en el plano familiar (factores familiares), efecto pigmalión negativo en el plano escolar (factores institucionales), factores personales como autonomía, capacidad intelectual, creatividad, interés por el conocimiento, además del medioambiente, y lo que Gardner llama inteligencia emocional, intrapersonal e interpersonal.

En otros estudios sucesivos, Sternberg da profusa importancia a la creatividad en correlación al éxito académico. Luego, comprobó también que la perseverancia hasta culminar tareas, la capacidad de relacionar objetivos, la autoconfianza y la versatilidad son de la misma manera factores de éxito académico cuando se desarrollan en el estudiante como deberían desarrollarse. Esto implica claro que el profesor de Colegio y Universidad también debería estar muy capacitado, es decir contar con elevados niveles de estudios y escolaridad para crear las condiciones necesarias que favorezcan alcanzar un levado rendimiento académico por sus estudiantes.

Las investigaciones (De Zubiría y Galindo, 2002) llegaron a la misma conclusión que Sternberg, demostrando que evidentemente dicha correlación es altísima.

2. Discusión

En las escuelas de primaria japonesas (Carl Simons, 2006) existe el juku que en lengua castellana significa “escuela intensiva”, la cual comprende tres horas de trabajo extra por día tres veces por semana. Para esta cultura el “éxito” es sinónimo de competencia y competitividad. Les enseñan desde el jardín de niños a entender que competir y ganar es lo que debe hacer todo buen japonés. Aquí vemos que lo que señala Al Ries et al es evidente, para los japoneses y según su paradigma por ejemplo, el éxito “es lograr algo con sus vidas, es competir y ganar”. Si tomamos las palabras de este investigador <<para muchísimas personas el éxito es aquello que uno quiere que sea>>, cabría cuestionarnos si en nuestro medio el éxito tiene acaso algún significado?.

3. Conclusiones

Las anteriores reflexiones nos revelan que para alcanzar el éxito académico no debemos someterlo a una causalidad exclusiva de una sola variable, por ejemplo atribuyendo alcanzar el éxito académico si se contase solo con un elevado CI, pues como lo vimos, se trataría de una conclusión confusa. Mas al contrario, y dado que el ser humano es un ser complejo y socio-afectivo hay variados factores que conducen a lograr mayor rendimiento académico.

Referencias

- [1] BUNGE, MARIO PhD. 2007. *Diccionario de Filosofía*. (5ª ed.). Buenos Aires, Argentina: Siglo XXI editores.
- [2] CHOMSKY, NOAM PhD. 2012. *Purpose of education, Impact of technology, Cost or investment, Assessment vs Autonomy*. [Interview in MIT]. EEUU: Learning without frontiers.
- [3] COVEY, STEPHEN PhD. 1996. *Los siete hábitos de la gente altamente efectiva*. [DVD]. EE.UU.: Covey company.
- [4] Dorland (Ed). 1993. *Diccionario médico* (24ª ed.). Madrid, España: Mc Graw Hill.
- [5] DORSCH, FRIEDRICH. 2008. *Diccionario de psicología*. (8ª ed.). Barcelona, España: Herder.
- [6] FRANÇOISE, RAYNAL et al. 2010. *Diccionario de Pedagogía*. (1ª ed.). Madrid, España: Popular.

⁴El efecto pigmalión permite determinar el papel de la familia y el de los Colegios en el éxito o en el fracaso de los estudiantes. Según Terrasier (1994) el rendimiento académico del estudiante está en mayor o en menor medida determinado por las expectativas que sobre él tienen los profesores de Colegio y los padres de familia que por las propias capacidades. En palabras de Sternberg “estas expectativas deberían ser siempre positivas, pues si fuesen negativas obstaculizarían el desarrollo de la inteligencia del estudiante”.

- [7] Larousse ilustrado. (Ed). 1964. *Diccionario pequeño*. Paris: Librairie Larousse.
- [8] MERANI, ALBERTO PhD. (Ed). 1979. *Diccionario de psicología*. (1ª ed.). D. F., México: Grijalbo.
- [9] PRIETO, M. & CASTEJÓN, J. 2000. *Los superdotados: Esos alumnos excepcionales*. Barcelona, España.
- [10] PUNSET, EDUARD. (Periodista y Divulgador científico). 2011. *El sistema educativo es anacrónico*. [REDES-Programa de divulgación científica N° 87]. España: Televisión española, Antena 2.
- [11] PUNSET, EDUARD. (Periodista y Divulgador científico). 2011. *La creatividad en la vida cotidiana*. [REDES-Programa de divulgación científica N° 291]. España: Televisión española, Antena 2.
- [12] PUNSET, EDUARD. (Periodista y Divulgador científico). 2011. *La inteligencia creativa*. [REDES-Programa de divulgación científica N° 217]. España: Televisión española, Antena 2.
- [13] PUNSET, EDUARD. (Periodista y Divulgador científico). 2011. *Los secretos de la creatividad*. [REDES-Programa de divulgación científica N° 89]. España: Televisión española, Antena 2.
- [14] RIBEIRO, LAIR PhD. 2000. *O sucesso não ocorre por acaso*. São Paulo, Brasil: Moderna.
- [15] RIES, AL PhD et al. 1992. *Horse sense: the key to success is finding a horse to ride*. USA: Mac Graw Hill.
- [16] ROBINSON, KEN PhD. 2010. *A iniciar la revolución del aprendizaje* [DVD]. Longbeach, California: TED company.
- [17] ROBINSON, KEN PhD. 2010. *Cahanging Paradigms* [DVD]. Inland.: RSA company.
- [18] ROBINSON, KEN PhD. 2010. *El elemento*. D.F., México: Random House Mondadori.
- [19] STERNBERG, R. J. PhD. 1993. *A guide to scientific writing for students and researchers*. Cambridge University: Press.
- [20] STERNBERG, R. J. PhD. 1999. *A triarchic approach to the understanding and assessment of intelligence in multicultural populations*. *Jornal of school psychology*, 37(2), 145-159.
- [21] STERNBERG, R. J. PhD. 2003. *Our research program validanting the triarchic theory of successful intelligence: Reply to gotfredson intelligence*, 31, 399-413.

Autoevaluación y Acreditación Carrera de Estadística

La Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés inició el proceso de autoevaluación y acreditación a fines del año 2010 y concluyó este 2014. Este año se realizó una primera autoevaluación con pares nacionales y en fechas 24, 25 y 26 de noviembre una comisión de pares evaluadores internacionales y nacionales efectuaron la segunda y última evaluación a nuestra carrera.

La fase de acreditación comprende el reconocimiento público de la calidad de nuestra unidad académica por parte de una institución acreditadora autorizada, en este caso el SNEA¹ que es parte del CEUB². Es el resultado de la verificación que la carrera debe cumplir con un conjunto de estándares mínimos previamente establecidos.

El proceso de acreditación implica dos fases fundamentales: la autoevaluación en la que la Carrera de Estadística se evalúa y entrega un informe y la acreditación misma en la que el CEUB mediante un Comité de Pares Evaluadores revisa el informe y verifica si la unidad académica cuenta con lo que dice el informe y cumple con los estándares mínimos.



Equipo de evaluadores

El día lunes 24 de noviembre a horas 9:30 de la mañana, se dio por inaugurado el proceso de autoevaluación externa en el Salón de Honor del Monoblock Central de la UMSA, en presencia de autoridades universitarias, facultativas y de carrera, el comité de pares evaluadores, el coordinador técnico del CEUB, personal del DEAGEC³, docentes, administrativos y estudiantes de la Carrera de Estadística. En esta oportunidad el Sr. Rector Dr. Waldo Albarracín, la Sra. Vicedecana de la Facultad de Ciencias Puras y Naturales Dra. María Eugenia García y el Sr. Decano de la Facultad Msc. Franz Cuevas, resaltaron el rol que cumple la Carrera de Estadística en beneficio de la Universidad y de la sociedad; y en particular destacaron el proyecto de la Encuesta Sociodemográfica del departamento de La Paz desarrollado por el Instituto de Estadística Teórica y Aplicada bajo el convenio Gobernación del Departamento de La Paz y la Universidad Mayor de San Andrés.

Los días lunes y martes fueron de arduo trabajo, los evaluadores reunieron a docentes, a la comisión de autoevaluación, a administrativos y estudiantes para conocer sus aptitudes y dificultades. Visitaron las instalaciones del Monoblock Central y de Cota Cota, ambientes que le corresponde a la carrera. Revisaron los documentos y los trabajos de investigación. Finalmente visitaron la biblioteca y las oficinas de kardex.

El acto de clausura del proceso de autoevaluación externa se llevó a cabo en el Salón de Honor de la UMSA el día miércoles 26 de noviembre del presente año con la presencia de autoridades del CEUB, de

¹(SNEA) Sistema Nacional de Evaluación y Acreditación.

²(CEUB) Comité Ejecutivo de la Universidad Boliviana.

³(DEAGEC). Comité de Autoevaluación y Acreditación de la UMSA.

la Facultad de Ciencias Puras y Naturales, del DEAGEC, del director de carrera, docentes, administrativos y estudiantes de la carrera.



Reunión con estudiantes



Lectura del informe

El Ing. Ernesto Valdivia Baldomar, Secretario Nacional de Evaluación y Acreditación del CEUB dio lectura al informe de la comisión de pares evaluadores, en el cual la Carrera de Estadística de la UMSA logra un puntaje satisfactorio de 88,79 % con funcionamiento óptimo, alcanzando la acreditación por el periodo de seis años.

El informe también describe las fortalezas y debilidades de la carrera y da a conocer un conjunto de recomendaciones que deben ser tomadas en cuenta para la mejora de la unidad académica, entre ellas se tiene:

- Incrementar la admisión de nuevos estudiantes.
- Aumentar y promover la titulación de profesionales en estadística.
- Incrementar el número de docentes titulares.
- Mejorar las condiciones de trabajo de los docentes en espacios físicos, calidad en la enseñanza y ampliar los ámbitos de la investigación.
- Incentivar a los docentes en publicaciones en revistas internacionales indexadas y en el desarrollo de proyectos de investigación concursables.
- Incentivar a los estudiantes a seguir estudios de postgrado mediante becas y convenios con universidades nacionales y extranjeras.
- Participar en publicaciones de revistas nacionales e internacionales.

Así como existen debilidades también hay fortalezas identificadas por el equipo de evaluadores. Estas son algunas de ellas nombradas en la lectura del informe:

- Identificación y compromiso con la carrera de docentes y estudiantes.
- Plan de estudios coherente y pertinente.
- Facilidad de acceso al mercado laboral.
- Impacto social de los trabajos de investigación y de titulación.

- Equipamiento e infraestructura adecuada para el proceso de enseñanza – aprendizaje.

Las recomendaciones son:

- Realizar cursos de postgrado para alcanzar al menos 50 % de docentes con grado de maestría y/o doctorado.
- Superar el 40 % de docentes a tiempo completo para el próximo semestre I/2015.
- Proponer proyectos de investigación e interacción social en el área de la estadística a fondos concursables para el próximo año 2015.
- Gestionar la suscripción de al menos a tres revistas científicas del área de la estadística durante el 2015.
- Aumentar los recursos económicos para incentivar el desarrollo académico.
- Mejorar los espacios físicos para los docentes en el monoblock central.

Finalmente el informe indica que la Carrera de Estadística debe atender las recomendaciones del equipo de evaluación externa en el entendido de que la Acreditación estará sujeta al cumplimiento de estas recomendaciones, la misma que será verificada en un año por la instancia correspondiente.

El equipo de evaluación estuvo conformado por:

Lic. Mario Soto

Coordinador

Universidad Autónoma Tomás Frías

Pares

Dr. Carlos Arturo Navarrete

Universidad de la Serena – Chile

Ms.c. Ing. Wilson Orlando Trujillo A.

Universidad Mayor de San Simón

Lic. Porfidia Ajata Ramos

Instituto Nacional de Estadística

Lic. Blanca Franco Martínez

Consultor Externo

Nuestro reconocimiento y agradecimiento por todo su esfuerzo de trabajo a este equipo de profesionales.



Estudiantes guía

Beneficios de la Acreditación

- Lograr la estandarización del proceso de formación profesional (calidad académica).
- Brindar al profesional de la Carrera de Estadística ventajas competitivas en el mercado laboral internacional.
- Garantizar el mejoramiento continuo de la calidad en la formación profesional.
- Desarrollar proyectos de investigación científica y de interacción social.
- Lograr la inserción sostenida en el contexto nacional e internacional.
- Acceder a financiamiento externo para la ejecución de programas, proyectos y trabajos de investigación relacionados con la academia y el desarrollo regional y nacional.

Socialización y difusión en Coroico

Estudiantes, docentes y administrativos de la Carrera de Estadística se trasladaron a la localidad de Coroico el día viernes 19 de septiembre de 2014, para conocer aspectos de detalle del proceso de autoevaluación y acreditación.

La Directora interina de la Carrera de Estadística Lic. Nilda Flores S. y la Comisión de Autoevaluación, organizaron el viaje con el objetivo de socializar, difundir y presentar los documentos de trabajo de la autoevaluación. Los documentos presentados fueron:

- Informe de Autoevaluación
- Plan de Desarrollo Institucional
- Plan de Estudios 2012
- Anexos con Documentación Respaldataoria

El evento se realizó en el Hotel Arcángel de esta localidad entre viernes 19 y sábado 20 de septiembre. El programa se inició con un almuerzo, el registro de participantes y entrega de material, y posteriormente la directora procedió con la inauguración del evento. Posteriormente se realizó la exposición de cada una de las áreas, el plan de desarrollo institucional y el plan de estudios 2012 a cargo de docentes y estudiantes de la Comisión Evaluadora con presencia de autoridades del DEAGEC, que también dieron sus recomendaciones y apreciaciones al proceso. Posteriormente se conformaron grupos de trabajo entre docentes, administrativos y estudiantes para comprender mejor cada una de las partes de los documentos.



➤ Presentaciones



➤ Comisiones de trabajo



Comisiones de trabajo



Participantes, Comunidad de la Carrera de Estadística

En conclusión, este proceso lo han realizado todos los que son parte de esta gran institución llamada Carrera Estadística. Docentes, administrativos, profesionales en estadística, egresados y estudiantes, todos aportaron de una u otra manera. Algunos hicieron menos otros más, lo importante es el logro alcanzado en esta primera instancia.

Felicidades Carrera de Estadística Acreditada en tus XXX Aniversario!!

XXX Aniversario de la Carrera de Estadística

La Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés celebró sus 30 años de vida. Los festejos se iniciaron el día lunes 13 y concluyeron el viernes 17 de octubre del presente año.

Se programaron diferentes actividades, el lunes tuvo lugar el acto de inauguración de la Semana de Estadística con un vino de honor, la presencia del Sr. Decano de la Facultad Msc. Franz Cuevas Q., el Director interino de la Carrera Msc. Fernando Rivero S., docentes, estudiantes y administrativos.

Durante los días de la semana se programaron charlas y exposiciones de docentes y estudiantes, así también actividades deportivas y juegos de mesa.



El día miércoles 15 de octubre por la tarde se realizó el acto central recordando el día de la fundación de la Carrera, después de separarse del Departamento de Matemática, del cual dependía hasta el año 1984. En este solemne acto participaron autoridades de la facultad, director, docentes, estudiantes y administrativos de nuestra carrera. En esta oportunidad se entregó un reconocimiento al Lic. Luis Zapata Escobar como uno de los promotores y fundadores de esta unidad académica; a los docentes eméritos por su servicio en beneficio de la educación universitaria, y a los estudiantes sobresalientes con mejores promedios por su esfuerzo de mejorar en sus estudios. El acto concluyó con un vino de honor, el sople de velas y el canto de cumpleaños para nuestra querida Carrera de Estadística.



Elecciones Dirección de Carrera

Las elecciones para elegir al nuevo Director de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la UMSA, se llevo a cabo el día 23 de octubre del presente año. Se presentaron dos frentes denominados “FE” (Fuerza Estadística) del Lic. Juan Carlos Flores López y “EUC” (Estadísticos Unidos por el Cambio) del Lic. Raúl Delgado Álvarez.

Ambos candidatos se pusieron en campaña con anticipación de una semana y media al día de las elecciones. Votaron un total de 108 estudiantes y 13 docentes, del estamento docente sufragaron todos aquellos habilitados para el voto, sin embargo, en estudiantes participaron el 66 % aproximadamente. Los resultados de ambos estamentos y porcentaje ponderado se presentan en el siguiente cuadro.

Cuadro 1 ➤ Resultados elección carrera de estadística. Gestión 2014-2017

Fórmula	Votos de docentes		Votos de estudiantes		Porcentaje ponderado
	Número	Porcentaje	Número	Porcentaje	
FE	7	29.2 %	40	19.2 %	48.4 %
EUC	5	10.8 %	63	30.3 %	51.1 %
Voto en blanco	0	0.0 %	1	0.5 %	0.5 %
Validos	12	50.0 %	104	50.0 %	100.0 %
Nulos	1	—	4	—	—
Emitidos	13	—	108	—	—

Como se puede observar, el frente EUC es el ganador de la elección con un 51.1 % ponderado, con una diferencia de apenas 2.7 % respecto al frente FE. Se ha dado un voto en blanco en el estamento estudiantil y ninguno en el de docentes; cinco votos nulos (1 de docentes y 4 de estudiantes). Los votos válidos alcanzan a 12 en docentes y 104 en estudiantes, haciendo un total de 116 votos válidos de 121 emitidos, es decir, el 95.9 %.

Gráfico 1 ➤ Voto docente

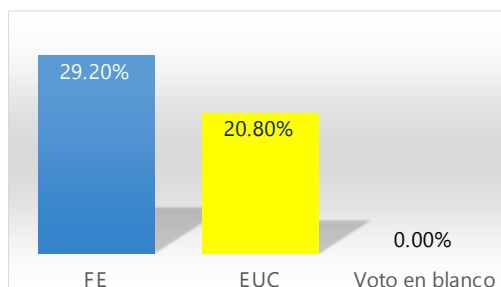
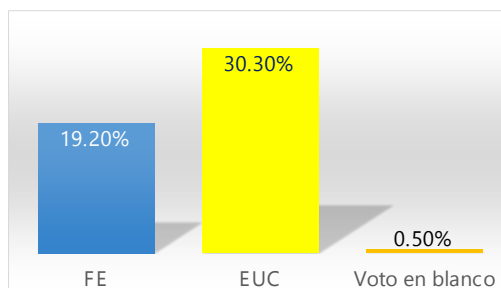
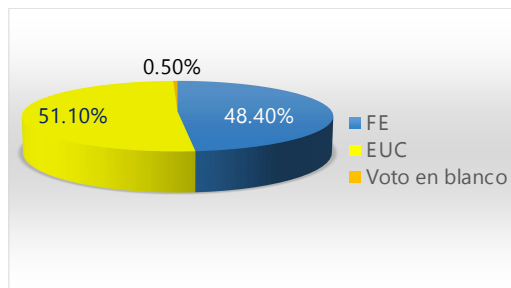


Gráfico 2 ➤ Voto estudiantil



Los gráficos 1 y 2 muestran los porcentajes obtenidos de ambos frentes en docentes y estudiantes con resultados inversos en cada uno de ellos. Los resultados de las elecciones presentan un panorama dividido en la votación. El frente EUC alcanza un porcentaje mayor al de FE en estudiantes y a la inversa en docentes. El porcentaje ponderado entre ambos frentes es tendiente al 50 %.

Gráfico 3 ➤ Voto ponderado



Ambos candidatos presentaron sus propuestas a docentes y estudiantes el día martes 21 de octubre en el aula 109 de la carrera, dos días antes de la votación. Las presentaciones se realizaron dentro del marco del respeto y responsabilidad.

El nuevo Director de la Carrera de Estadística Lic. Raúl Delgado Alvarez fue posesionado en su cargo el día jueves 6 de noviembre del presente año en un acto donde estuvieron presentes autoridades de la Facultad y miembros de Comité Electoral.

Desde este medio de comunicación y difusión, hacer augurios de que el nuevo Director emprenda mejoras en lo académico, investigación e interacción social en beneficio de la carrera y especialmente de los estudiantes.

Encuesta Sociodemográfica del Departamento de La Paz

El convenio firmado entre el Gobierno Departamental de La Paz y la Universidad Mayor de San Andrés establece que nuestra universidad se constituye en el brazo derecho de la gobernación del departamento de La Paz en materia de ciencia y tecnología y la elaboración conjunta de proyectos de desarrollo e investigación social y económica en beneficio del departamento.

Al inicio de este convenio, las facultades y carreras de nuestra universidad que participarían conjuntamente la gobernación en proyectos de desarrollo departamental, solicitaron a las autoridades universitarias el acceso a contar con la base fundamental de decisión que es la información estadística. Es allí que uno de los proyectos fundamentales de inicio fue la Encuesta Sociodemográfica del Departamento de La Paz, donde se obtendría información estadística primaria, básica para establecer un diagnóstico para la toma de decisiones.

Es así que a comienzos del año 2010 se empieza la ardua labor de planificación de la encuesta, encargada al Instituto de Estadística Teórica y Aplicada (IETA). Ese mismo año, a fines del mes de octubre y comienzos del mes de noviembre arranca la primera encuesta en la región del Altiplano Sur cubriendo en ella 13 municipios de los 20. La falta de cobertura se debió a que muchos de los dirigentes municipales no fueron comunicados oportunamente de la importancia de este operativo. Participaron aproximadamente 500 personas entre docentes, profesionales, profesores rurales, dirigentes municipales, funcionarios de la gobernación y estudiantes de la UMSA. Los resultados de la investigación se plasman en el primer volumen (figura 1).



Figura 1

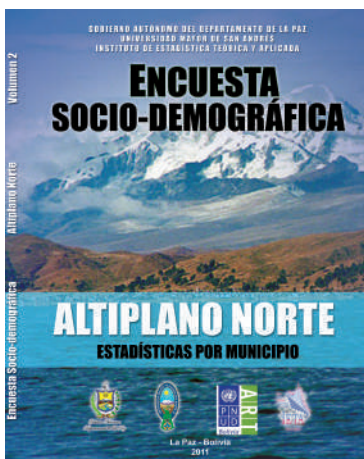


Figura 2

Paralelamente al primer operativo de encuesta, se realiza la planificación del segundo operativo, esta vez en la región del Altiplano Norte, una zona extensa que comprende 20 municipios. Esta vez la cobertura es total y el apoyo de autoridades locales y municipales se intensifica y así se logra un gran trabajo donde participan aproximadamente 700 personas. El mes de marzo de 2011 se efectúa la encuesta donde profesores rurales y estudiantes de San Andrés visitan a un total de 3.894 viviendas de la muestra en los 20 municipios. La figura 2 presenta el documento de resultados estadísticos de la encuesta en la región.

El rectorado de ese entonces dirigido por la Dra. Teresa Rescala M. y el Departamento de Planificación y Coordinación de la UMSA dirigido por el Msc. Julio Alvarado hacen entrega de estos dos primeros documentos al Gobierno Departamental de La Paz precepto por el Dr. Cesar Cocarico Yana.

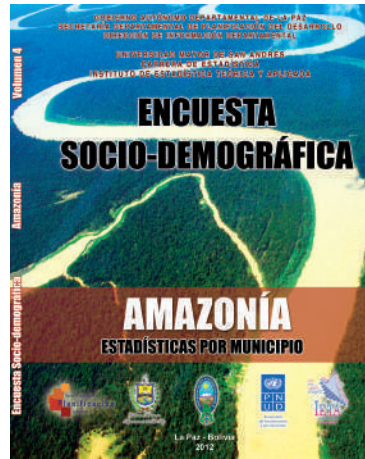
El mes de octubre de 2011 se realiza el tercer operativo de campo con docentes, técnicos de la gobernación, profesores rurales, autoridades de los municipios y estudiantes universitarios de la UMSA en la región de los Yungas de La Paz. Se visitan los 7 municipios que la componen: Coripata, Coroico, Yanacachi, La Asunta, Palos Blancos, Chulumani e Irupana. Participan aproximadamente 250 personas en todo el proceso. De este trabajo se obtiene un tercer documento de resultados estadísticos (figura 3).

Quedaban todavía por investigar 4 regiones más: Amazonía, Valles Norte, Valles Sur y Metropolitana. La región de la Amazonía es la más extensa del departamento en términos de superficie y dificultad de acceso,

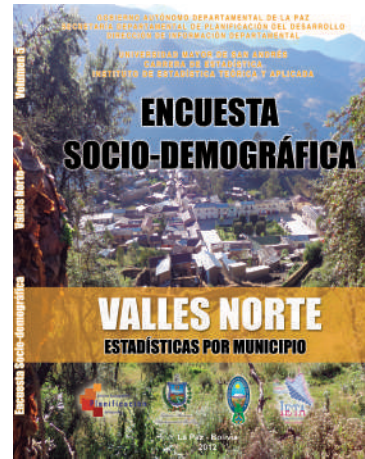
comprende 9 municipios. El compromiso asumido por la UMSA y la Gobernación se tornó aún más grande. El IETA asume la responsabilidad de seguir trabajando en beneficio de la sociedad, es así que entre noviembre de 2011 y marzo de 2012, y por los grandes problemas de acceso al lugar se realiza el cuarto operativo de campo en la región con el apoyo de 350 personas y un total de 2.624 viviendas visitadas. El documento resultado de este trabajo es el que se presenta en la figura número 4.



▶ Figura 3



▶ Figura 4



▶ Figura 5

Las autoridades de la UMSA hacen nuevamente entrega oficial de estos dos documentos más al Gobierno Departamental de La Paz en un acto realizado en instalaciones de la Gobernación (GADLP) donde participaron las autoridades del departamento, autoridades de la UMSA, autoridades de los municipios, estudiantes y docentes de la Carrera de Estadística, personal de la Gobernación y público en general.



▶ Entrega de documentos al GADLP

Las oficinas del IETA en los predios de Cota Cota, sirvieron para el montaje de un gran operativo. Día a día el trabajo era permanente de los estudiantes becarios que llegaban a un total de 30 universitarios de la Carrera de Estadística en la labor de supervisores, críticos, codificadores, transcritores y validadores de datos. La jornada se iniciaba a las 8:00 de la mañana y concluía muchas veces a las 21:00 horas. Los consultores contratados por el IETA hacían la labor de apoyo en la planificación junto al Director del Instituto y los docentes investigadores. Los consultores contratados para este fin, hacían el trabajo de procesamiento y análisis de resultados logrando así concluir con los documentos de resultados de las 7 regiones de nuestro departamento.

El mes de mayo de 2012 nuevamente se desplazan docentes y estudiantes de la UMSA para el levantamiento de campo en 11 municipios de la región de Valles Norte del departamento de La Paz. Se visitan un

total de 3.456 viviendas de la muestra con el apoyo logístico del Gobierno Autónomo Departamental de La Paz y las autoridades campesinas municipales. Este trabajo fue un éxito pues toda la población de los municipios colaboraba con los encuestadores a convocatoria de sus dirigentes. Con gran satisfacción se obtiene un nuevo manuscrito de resultados de la investigación en su volumen número 5 (figura 5).

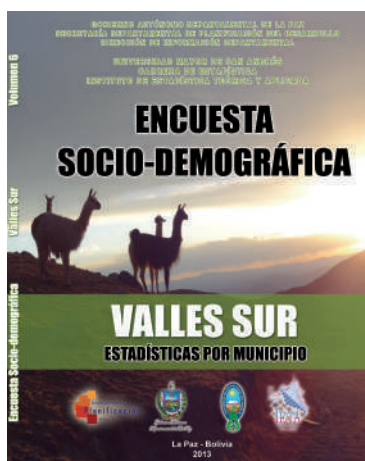
Es la primera vez que un departamento de Bolivia, como La Paz, logra resultados sociales y económicos a nivel de municipios. Particularmente, los datos estadísticos oficiales con estas características, tienen desagregación regional o departamental. Es así que el esfuerzo realizado no tiene límites, la información estadística a nivel municipal es hoy en día prioritaria pues todos los planes y programas proyectados están a nivel municipal. La Carrera de Estadística mediante su instituto de investigación impulso esta labor sin atenuantes.

La penúltima región por estudiar fue la región de los Valles Sur con sus 11 municipios que la componen (Luribay, Sapahaqui, Yaco, Malla, Cairoma, Inquisivi, Quime, Cajuata, Colquiri, Ichoca y Licoma Pampa). La muestra de viviendas alcanzó un total de 3.743 visitadas por estudiantes de la UMSA en su calidad de encuestadores y supervisores agrupados en brigadas de trabajo de campo, bajo la coordinación de los docentes de la Carrera de Estadística. Han participado un total de 200 personas entre docentes y estudiantes. La figura 6 muestra el documento de resultados de la Encuesta Sociodemográfica de la región de Valles Sur. El levantamiento de información estadística se realizó el mes de octubre de 2012.

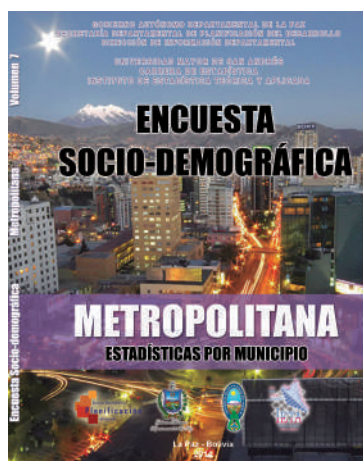
El trabajo culminó con la región Metropolitana donde sus principales municipios lo constituyen La Paz y El Alto. El operativo de campo se realizó seguidamente al de Valles Sur en noviembre de 2012. La muestra de viviendas alcanzó un total de 4.779 esto debido a la gran concentración de población que tiene la región.

La zona comprende un total de 7 municipios. Esta última región ha sido el epílogo de un trabajo sacrificado que la UMSA ha impulsado por medio de sus principales autoridades en bien de la sociedad. El escrito de resultados ha sido concluido, la figura 7 muestra su presentación.

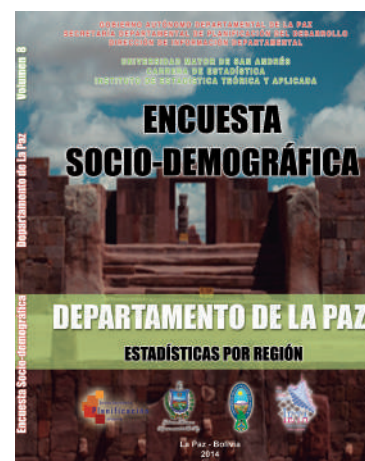
El resultado final consta de 7 volúmenes de información estadística, uno por región. Se ha elaborado un documento resumen denominado “Consolidado del Departamento de La Paz”. Este documento será entregado oficialmente el próximo año a las autoridades departamentales y municipales del departamento de La Paz (figura 8).



➤ Figura 6



➤ Figura 7



➤ Figura 8

No queda más que agradecer el impulso y la confianza depositada por parte de las autoridades de la Universidad Mayor de San Andrés con la Carrera de Estadística por intermedio de su Instituto para el logro de este proyecto.

Agradecer también a los docentes y estudiantes de la Carrera de Estadística por considerarse parte de este trabajo. A los consultores del IETA por su labor encomendada. A las autoridades municipales y profesores

rurales de los municipios, al Gobernador del Departamento de La Paz y a los técnicos de la Gobernación mi gratitud.

Msc. Fernando O. Rivero Suguiura
RESPONSABLE DE PROYECTO
CARRERA DE ESTADÍSTICA – UMSA

CLATSE - Congreso Latinoamericano de Sociedades de Estadística

El XI Congreso Latinoamericano de Sociedades de Estadística se llevó a cabo del 20 al 23 de octubre de 2014 en la ciudad de Serena Chile, organizado por la Universidad de la Serena de esa ciudad. Este evento de sigla CLATSE¹ se viene realizando desde el año 1991 constituido conjuntamente por la Sociedad Chilena de Estadística (SOCHE), la Sociedad Uruguaya de Estadística (SUE) y la Sociedad Argentina de Estadística (SAE). Cada dos o tres años se reúnen para contribuir a la integración de las sociedades y asociaciones de estadística iberoamericanas.

Profesionales académicos de doce universidades de Chile participaron del Congreso, con la misión de potenciar el conocimiento y la divulgación de la estadística como aporte al progreso regional y nacional, fortaleciendo los lazos de cooperación para el desarrollo conjunto de la investigación científica no sólo de la estadística, sino también del amplio espectro de disciplinas científicas que la aplican y contribuyen a su perfeccionamiento.

Un grupo de 10 estudiantes de la Carrera de Estadística de la UMSA, viajaron con el fin de intercambiar experiencias con estudiantes de otros países participantes y conocer el nivel que tienen los profesionales académicos en el área de la estadística de la región.

Los conferencistas participantes fueron:

- María José García. Doctora en Estadística de la Pontificia Universidad Católica de Chile.
- Adelchi Azzalini, trabaja en el Departamento de Estadística de la Universidad de Padova en Italia.
- Alejandra Martínez. Doctora en Ciencias Matemáticas de la Universidad de Buenos Aires Argentina.
- Marta Quaquino. Doctora en Matemática de la Universidad Politécnica de Valencia España.
- Juan José Goyeneche. Doctor en Estadística de la Iowa State University. Uruguayo.



➤ Poster del CLATSE



➤ Conferencia – Dr. Azzalini

¹(CLATSE) Congreso Latinoamericano de Sociedades en Estadística.

Se dictaron los siguientes minicursos:

- Modelos de regresión para datos censurados con aplicaciones en R.
- Tópicos en estadística y data mining en big data.
- Estadística genética: herramientas y modelos para el análisis de información genética de alta dimensionalidad.
- Taller de enseñanza de la estadística en educación media.

Felicitar a los estudiantes que participaron en este evento internacional, demostrando responsabilidad e interés de conocer otro tipo de nivel, percepción, enfoque y experiencia de personas profesionales dedicadas a la estadística en Latinoamérica. Que puedan también hacer comparaciones entre lo que tenemos en nuestro país y lo que ve afuera, no sólo en lo que respecta a los recursos humanos sino también en los medios tecnológicos y científicos.



Estudiantes de la carrera de estadística de la UMSA

Los profesionales en Estadística de Bolivia, tenemos el reto de organizarnos para conformar la Sociedad de Estadísticos y a partir de ella, participar en congresos y seminarios nacionales e internacionales para el bien de nuestra patria y de la sociedad.

Potos de Estadística

El pasado sábado 2 de agosto de 2014 se llevó a cabo la XXVII versión de la Entrada Folklórica Universitaria de la Universidad Mayor de San Andrés. Más de 12 mil estudiantes en 65 fraternidades participaron y deleitaron al público durante su recorrido por las calles paceñas.

En esta oportunidad la fraternidad Potos de Estadística participó nuevamente de la entrada ingresando en la posición número 17. Desde este medio de difusión que tiene nuestra Carrera, hacemos llegar una gran felicitación a la fraternidad Potos de Estadística, pues alcanzó el tercer lugar en la calificación general y un primer lugar en la calificación de su categoría. Los resultados se presentan en el siguiente cuadro:



Cuadro 1 ➤ XXVII Entrada folklórica universitaria 2014. Puntaje Final de los primeros 10 grupos

Puesto	Categoría	Fraternidad	Danza	Dependencia	Monografía 40%	Presentación y danza 50%	Disciplina 10%	Puntaje total
1	C	Diablada de Medicina	diablada	Facultad de Medicina	31,6	46,8	8,7	87,1
2	C	Verdaderos Caporales de ciencias Económicas y Financieras	caporal	Facultad de Ciencias Económicas y Financieras	31,6	44,8	9,0	85,4
3	B	Potos de Estadística	potolo	Carrera de Estadística	33,2	41,8	9,7	84,7
4	B	Huayna Lisos de Electromecánica	tinku	Carrera de Electromecánica	30,8	45,2	7,7	83,7
5	B	Taller Cultural Tinkus Wistus	tinku	Facultad de Ciencias Económicas y Financieras	28,8	46,2	8,0	83,0
6	B	Tinkus Puros y Naturales	tinku	Facultad de Ciencias Puras y Naturales	37,2	36,2	8,7	82,1
7	C	Llamerada de San Andrés	llamerada	Universidad Mayor de San Andrés	30,8	40,6	10,0	81,4
8	C	Mineros de Sociología	mineritos	Carrera de Sociología	25,0	45,8	8,3	79,1
9	C	Chaqueños de Artes	chacarera	Carrera de Artes	27,0	42,2	9,7	78,9
10	C	Caporales de Ingeniería	caporal	Facultad de Ingeniería	27,2	44,0	6,7	77,9

Fuente: Comisión de Cultura H.C.U. - Universidad Mayor de San Andrés



Los datos muestran como ganador de la Entrada, a la Diablada de la Facultad de Medicina con un puntaje total de 87,1 %, considerando que el puntaje máximo en la presentación de la monografía es de 40,0 %, presentación y danza 50,0 % y disciplina durante el recorrido de 10,0 %, haciendo un total de 100,0 %. La fraternidad Potos de Estadística alcanzó un puntaje de 84,1 % total desagregado en monografía igual a 33,2 %, presentación y danza 41,8 % y disciplina 9,7 %.

Como premio a su gran actuación, la fraternidad Potos de Estadística participó el día sábado 15 de noviembre en la V Entrada Universitaria Nacional llevada a cabo en la ciudad de Trinidad - Beni. Este evento congrega a fraternidades representantes de las 11 universidades públicas de nuestro país. Participaron también grupos invitados como los caporales de San Simón, la llamerada de San Andrés, el ballet de la Universidad Gabriel René Moreno y la banda Imperial de Oruro.



Felicidades una vez más a los estudiantes y docentes frateros de la Carrera de Estadística por llevar en alto la representación de la UMSA y de Bolivia entera en cultura y folklore.

11 de Julio - Día Mundial de la Población

La propuesta de observar el Día Mundial de la Población partió del Consejo de Administración del Programa de las Naciones Unidas para el Desarrollo (PNUD) en 1989, que la plasmó en su decisión 89/46 recomendando la fecha del 11 de julio para su celebración¹.

Según estimaciones recientes, la población de este planeta alcanza a 7.500 millones de habitantes y según expertos de la ONU (World Population Prospects) ya dentro de 11 años, en el 2025, habrá 1.000 millones de habitantes más, es decir, 8.500. A comienzos del próximo siglo, el número de personas podría superar los 16.600 millones.

La población del planeta aumentará de manera espectacular, especialmente en las regiones más pobres del mundo, advierte la ONU. A partir del año 2011 la India se convirtió en el país más poblado del mundo superando a China. Un estudio señala que está previsto que la mitad de todo el crecimiento de la población entre 2013 y 2100 se concentre en sólo 8 países: India, Nigeria, Tanzania, Congo, Níger, Uganda, Etiopía y Estados Unidos. Aunque el crecimiento de la población mundial ha disminuido en su conjunto, este estudio nos recuerda que algunos países en desarrollo, especialmente en África todavía crecen con rapidez, explica el subsecretario general de la ONU.

World Population Prospects también presta atención al rápido crecimiento del número de personas mayores y no sólo en las regiones de nivel económico alto. En las partes más desarrolladas del mundo el 23 % de la población ya es mayor de 60 años. Se prevé que el porcentaje ascienda hasta un 32 % en el año 2050 y al 34 % en 2100. A pesar de que la población va aumentando, el estudio de la ONU prevé que la fertilidad en el mundo descienda de 4,5 a 2,9 hijos por mujer en el periodo 2045 – 2050, y a 2,1 hijos en el quinquenio 2095 – 2100.

El cuadro siguiente muestra la población en miles de los principales 15 países más poblados del mundo con tasa de crecimiento anual en el periodo 2000 – 2010.

Cuadro 1 ► Población total y tasa de crecimiento anual para el periodo 2000 - 2010

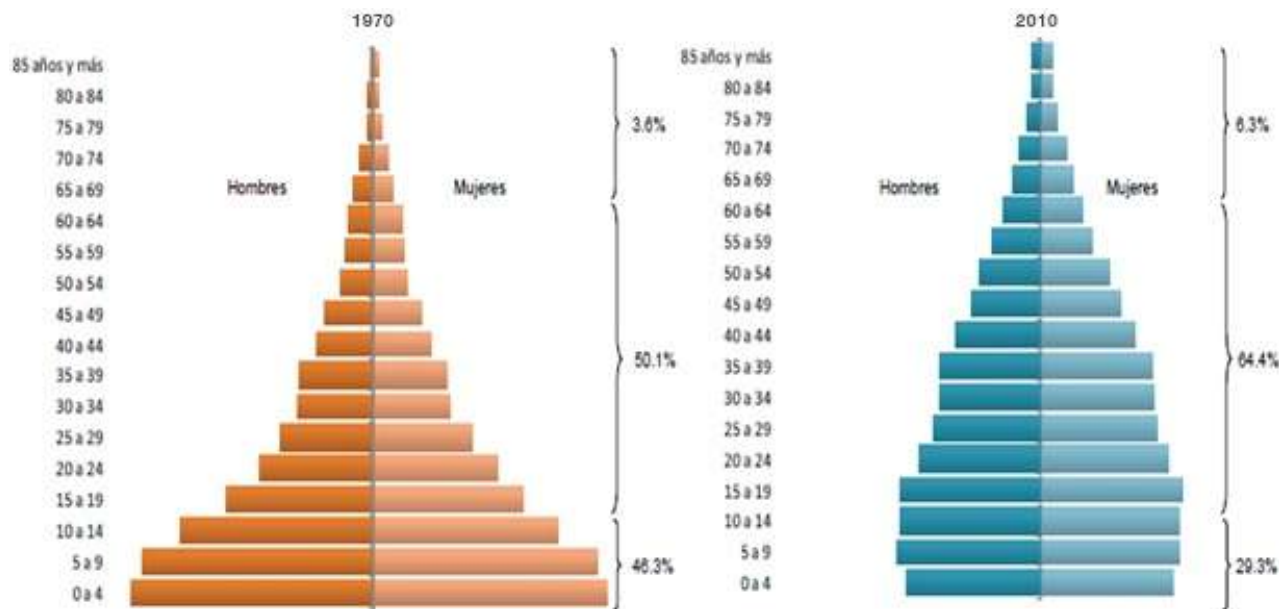
Puesto	País	Población (miles)	Tasa de crecimiento
1	China	1.341.335	0,55
2	India	1.224.614	1,51
3	Estados Unidos	310.384	0,95
4	Indonesia	239.871	1,18
5	Brasil	194.946	1,12
6	Paquistán	173.593	1,85
7	Nigeria	158.423	2,51
8	Bangladesh	148.692	1,38
9	Federación de Rusia	142.958	-0,26
10	Japón	126.536	0,06
11	México	112.336	1,38
12	Filipinas	93.261	1,89
13	Vietnam	87.848	1,10
14	Etiopía	82.950	2,38
15	Alemania	82.302	-0,01

Fuente: Organización de las Naciones Unidas (ONU), World Population Prospects 2010.

¹<http://www.un.org/es/events/populationday/>

La figura muestra la estructura piramidal de la población mundial por edad y sexo de los años 1970 y 2010. El cambio demuestra la pérdida de la población porcentual en edades menores.

Gráfico 1 ➤ Distribución de la población por edad y sexo, 1970 y 2010



Fuente: Censo general de población y vivienda 1970 y censo de población y vivienda 2010.

Algunas cifras propuestas por el Programa de Naciones Unidas para el Desarrollo

- Más de 1.000 millones de seres humanos viven con menos de un dólar por día.
- 2.800 millones de personas, es decir, cerca del 40 % de la población mundial, viven con menos de 2 dólares por día.
- 448 millones de niños sufren de bajo peso.
- 876 millones de adultos son analfabetos, de los cuales dos terceras partes son mujeres.
- Cada día, 30.000 niños menores de cinco años mueren de enfermedades que podrían haber sido evitadas.
- Más de 1.000 millones de personas no tienen acceso al agua potable.
- El 20 % de la población mundial posee el 90 % de las riquezas.

La FAO y el Programa Mundial de Alimentos de la ONU (PMA) anunciaron que el número de personas que sufren hambre en el mundo es inaceptablemente alto, a pesar de los recientes avances esperados, que han situado esa cifra por debajo de los 1.000 millones. La nueva estimación sobre el número de personas que sufrirán hambre crónica este año es de 925 millones: 98 millones menos respecto a los 1.023 millones calculados en 2009.

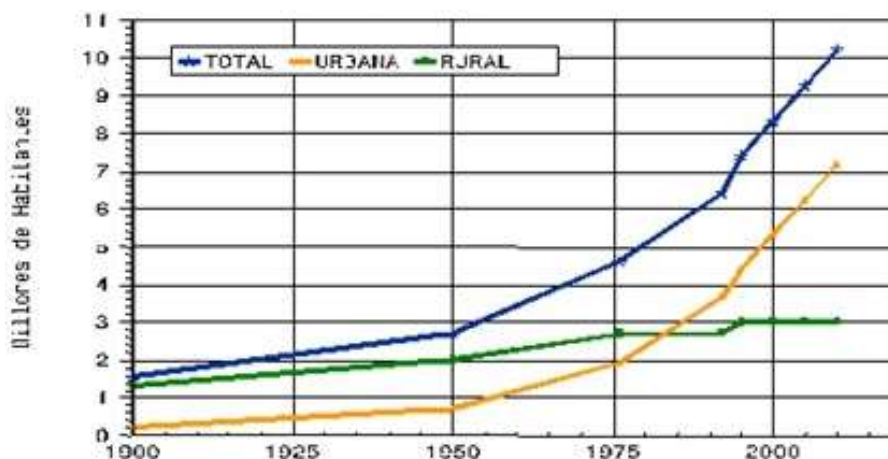
Un niño muere cada seis segundos debido a problemas relacionados con la desnutrición, el hambre sigue siendo la mayor tragedia y el mayor escándalo del mundo, aseguró el Director General de la FAO, Jacques Diouf. "Es algo absolutamente inaceptable", añadió. La permanencia de un elevado nivel de hambre en el mundo "hace muy difícil alcanzar no sólo el primer Objetivo de Desarrollo del Milenio (ODM), sino también el resto de los ODMs", alertó Diouf. "El logro de la meta internacional de reducción del hambre se encuentra

seriamente amenazado”, añadió el responsable de la FAO, señalando además que los recientes aumentos de los precios alimentarios, de persistir, pueden perjudicar los esfuerzos para nuevas reducciones en la cifra de personas hambrientas en el mundo.

”La actuación enérgica y urgente por parte de los países y del mundo ha resultado eficaz para detener la escalada en la cifra de personas que sufren hambre”, subrayó la Directora Ejecutiva del PMA. ”Pero no es el momento de bajar la guardia. Tenemos que mantener a raya el hambre para garantizar la estabilidad y proteger las vidas y la dignidad de las personas, añadió.

En el contexto nacional, las cifras censales describen y permiten estimar los retos que hay que enfrentar en el presente y el futuro de Bolivia. Los procesos de cambio demográfico (figura 1), se han dado en Bolivia con desfases en el tiempo y en el espacio. La población que habitaba el territorio boliviano hacia 1900, apenas llegaba al millón y medio de habitantes, luego a 3 millones de habitantes en 1950; y 4.6 millones en 1976. En la actualidad el país cuenta con alrededor de 11 millones de habitantes de los cuales 50,1 % son mujeres y 49,9 % hombres.

Gráfico 2 ◊ Bolivia. Evolución demográfica, según área de residencia, 1900 - 2010



El censo poblacional realizado el año 2001 reflejó que Bolivia tenía 8.274.325 habitantes. Con el censo del 2012 se llegó a 10.389.913 habitantes, lo que muestra una tasa de crecimiento intercensal de 2,0 % y un crecimiento en la población de 26 %, esto implica que la población se incrementó en 2.115.588 personas. La tasa de crecimiento ha decrecido en los últimos 15 años de un nivel del 2,8 % entre 1950-1976 al 2,3 % en 1992 y 2,0 % para el año 2012. Dicha reducción se habría debido sobre todo al efecto combinado del descenso de la fecundidad (de 6,2 hijos por mujer alrededor de 1975, una fuerte reducción en los últimos 15 años a 4,8 hijos por mujer alrededor de 1991 y de 3,3 para el 2012), y la reducción de la mortalidad (sobre todo la infantil, se obtuvo logros importantes en la esperanza de vida al nacer de la población, se han ganado alrededor de 12 años entre 1976 y 1992, de una esperanza de vida al nacer de alrededor de 47 años en 1976 se pasó a 59 años en 1992 y a 67 años en 2012). También tuvo un impacto importante las migraciones internas que han generado un acelerado proceso de redistribución espacial de la población.

En términos absolutos las zonas urbanas, en lo que va del presente siglo han incrementado su importancia demográfica en más de 17 veces, en cambio las zonas rurales apenas si se duplicaron. Para el presente año (2014) la población boliviana habrá llegado aproximadamente a los 12 millones de habitantes de los cuales 8 millones vivirán en las zonas urbanas y sólo alrededor de una tercera parte de los bolivianos vivirán en las zonas rurales.

Uno de los factores gravitantes en el patrón de desigualdad vigente en Bolivia, se explica por las profundas brechas económicas, sociales, políticas, ambientales, espaciales y culturales existentes. El acceso a oportunidades de contar con una educación de calidad y orientada al mundo del trabajo es sólo accesible a un

grupo reducido de bolivianos. Lo que existe en el país es un amplio plano de privaciones, insatisfacciones y carencias que han contribuido en la ampliación de las brechas existentes y han generado la exclusión de cerca de la mitad de los pobladores bolivianos, es decir, se hallan en este grupo los pobladores rurales y las mujeres.

La mayor jerarquía urbana se ha producido en sólo tres ciudades principales: La Paz, Cochabamba y Santa Cruz, dichas ciudades acogen a dos terceras partes de los habitantes urbanos de Bolivia, a cuatro de cada cinco económicamente activos. El tamaño demográfico en cada ciudad es superior a las 600 mil personas. Todas ellas crecen a un ritmo demográfico por encima del 5 % anual y en sólo 15 años, han superado su duplicación. En dichas ciudades también se produjeron las mayores densidades netas de población y de la oferta de bienes y servicios.

La explicación del crecimiento urbano en Bolivia se debe entonces en el fuerte peso del crecimiento urbano de tres ciudades: El Alto y La Paz del departamento de La Paz, Cochabamba y Santa Cruz de los departamentos de Cochabamba y Santa Cruz de la Sierra, así como el de las ciudades de tamaño intermedio.

El crecimiento de la población guarda estrecha relación con el dinamismo económico y social de cada región, de tal manera que los patrones regionales de desarrollo son los que en última instancia permiten entender el ritmo de su crecimiento.

La redistribución de población en Bolivia, expresa la pérdida de importancia relativa del sector agrícola y la expansión del sector terciario de base esencialmente urbana. El panorama descrito en párrafos precedentes, nos permite imaginar, que bajo cualquier modelo de crecimiento, el escenario social boliviano va a estar caracterizado por muchos años por la presencia masiva de la pobreza rural y urbana.

Cronología de la Desaparición de 43 Estudiantes en Iguala, Guerrero

La pregunta más grande en la plaza principal de México, sigue sin respuesta: ¿dónde están los 43 estudiantes de la Normal de Ayotzinapa?

El 26 de septiembre del presente año ha sido una fecha que pasó a la historia de México y el mundo, tras la muerte de seis jóvenes y la desaparición de 43 estudiantes de la Escuela Normal Rural de Ayotzinapa, en la localidad de Iguala, Guerrero, al sur de México.

A continuación, una cronología de los hechos.

Septiembre 26. Mueren tres estudiantes de la Escuela Normal de Ayotzinapa al ser atacados por policías locales y presuntos grupos armados. Hubo otras tres víctimas que no pertenecían a la institución. Además, 57 personas quedaron en calidad de desaparecidas.



Septiembre 29. Veintidós policías municipales de Iguala son trasladados a un penal en Acaapulco tras pasar dos días en el Ministerio Público del Fuero Común. Se les señalaba por ser probables responsables de los seis homicidios.



Septiembre 30. Blanco Cabrera, procurador (fiscal) General de Guerrero, dice que 14 de los 57 desaparecidos fueron ubicados. Además corroboró la versión de que los policías municipales dispararon en contra de los estudiantes y también contra un autobús en el que se trasladaba el equipo de fútbol de tercera división los Avispones de Chilpancingo. Ese mismo día, el alcalde José Luis Abarca pidió licencia al cargo por 30 días. Sostuvo que la intención era permitir que las investigaciones avanzaran.

Octubre 2. Jesús Martínez Garnelo, secretario General de Gobierno de Guerrero, amplió el radio de búsqueda de los 43 desaparecidos hasta el municipio de Taxco de Alarcón. En esa fecha

había ya más de 1.800 personas buscando a los desaparecidos, entre policías estatales, el Ejército y la Armada de México.

Octubre 3. La Organización de las Naciones Unidas (ONU) en México condenó la desaparición de 43 estudiantes y la violencia de que fueron víctima, además de considerar este hecho como uno de los “más terribles de los tiempos recientes”. Por su parte, la Comisión Interamericana de Derechos Humanos (CIDH) requirió al gobierno de México que tomara medidas de protección en favor de los desaparecidos. En la misma fecha, la Secretaría de Gobernación (Segob, ministerio de Interior) ofrece a la familia de los estudiantes ausentes “coadyuvar” en las tareas de búsqueda. Mientras, en Guerrero se solicita un juicio de procedencia en contra del alcalde Abarca para que le fuera retirado el fuero constitucional.

Octubre 4. Policías federales y de la entidad de Guerrero encuentran seis fosas clandestinas en Iguala, Guerrero, en una zona conocida como Pueblo Viejo. En ellas se encuentra un número indeterminado de cadáveres y se presume que podrían corresponder a los desaparecidos. Por su parte, la Procuraduría General de la República (PGR, fiscalía general) informa que atrajo la investigación, además de que envía a un grupo de expertos forenses a la zona para realizar los exámenes pertinentes y determinar la identidad de las víctimas halladas en las fosas clandestinas.



Octubre 5. Lázaro Mazón Alonso, secretario de Salud de Guerrero, confirma que nueve cuerpos calcinados localizados en las fosas de Pueblo Viejo se encontraban en las instalaciones del Servicio Médico Forense de Chilpancingo, a fin de ser identificados. El mismo día, la Comisión Nacional de Derechos Humanos (CNDH) advierte que en el caso se cometieron graves violaciones a los derechos humanos, como ejecuciones extrajudiciales, desapariciones forzadas, atentados contra la dignidad de las personas y omisiones en el cumplimiento de la función de seguridad pública.

También el 5 de octubre, Blanco informa que dos delincuentes afirmaron que, junto con policías de Iguala, asesinaron a 17 de los 43 estudiantes. Las declaraciones aseguran que Francisco Salgado, director de Seguridad Pública de Iguala, ordenó la detención de los jóvenes, mientras que alias El Chucky, jefe del cártel Guerreros Unidos, habría dado la orden de matarlos.

Octubre 6. Un informe del Centro de Investigación y Seguridad Nacional mexicano (CISEN) vincula a Abarca alcalde de Iguala y a su esposa, María de los Ángeles Pineda Villa, con la delincuencia organizada. Supuestamente él ordena matar a los 43 estudiantes en México. En tanto, el comisionado Nacional de Seguridad, Monte Alejandro Rubido, informa que elementos de la recién creada división de Gendarmería de la Policía Federal y del Ejército mexicano asumen el control absoluto de la seguridad en el municipio de Iguala, Guerrero.



Octubre 7. El Consejo Nacional del Partido de la Revolución Democrática (PRD, izquierda y que presentó a Abarca a la elección) se disculpa con los habitantes de Iguala en Guerrero, por haber postulado a un candidato externo y que llegara a la alcaldía. Mientras, el pleno de la Cámara de Diputados mexicana aprueba la creación de la Comisión Especial para dar seguimiento a las investigaciones al respecto.

Octubre 8. Se realiza una marcha multitudinaria del monumento Ángel de la Independencia en Ciudad de México hacia la plancha del Zócalo. Se hace eco de la manifestación en otras 20 entidades del país y otros nueve países del mundo. La exigencia es la aparición de los 43 estudiantes.

También la Organización de Estados Americanos (OEA), así como el gobierno de Estados Unidos demandan al gobierno mexicano una investigación exhaustiva para encontrarlos. En la misma fecha, los diputados locales de Guerrero turnan el juicio de procedencia contra el alcalde para quitarle el fuero a Abarca y permitir que fuera juzgado conforme a derecho. Ese día arranca el 42 Festival Internacional Cervantino

(FIC) en Guanajuato, al centro de México. En la ceremonia de inauguración, el director del FIC, Jorge Volpi, lamenta lo sucedido en Iguala.

Octubre 9. El fiscal de Guerrero informa del traslado de los 22 policías municipales al Centro de Rehabilitación Social de Tepic, en Nayarit. José Luis Abarca es acusado de omisiones, lesiones y desaparición forzada. El mismo día la PGR informa de la captura de cuatro personas que dieron información sobre otras cuatro fosas cerca de Iguala. En ellas se halló un número indeterminado de cuerpos calcinados. Suman en total 34 detenidos por los hechos, mientras el edil con licencia y su esposa son buscados para que declaren. También blanco de las críticas, Ángel Aguirre, gobernador del Estado de Guerrero, informa su intención de solicitar al Congreso local someter a revocación su mandato.



Octubre 10. En exigencia de que se encontrase con vida a los 43 desaparecidos, maestros de la Coordinadora Estatal de Trabajadores de la Educación de Guerrero (CETEG) toman dos sucursales bancarias, sacan a los clientes y realizan saqueos.

Octubre 12. Se realiza una marcha en Chilpancingo, capital de Guerrero, entre líderes empresariales, comerciantes, estudiantes y ciudadanos. Piden seguridad, el esclarecimiento de los hechos y la remoción del gobernador Aguirre. La delegación de la Unión Europea se suma a la preocupación y exigencia de justicia en el caso de los estudiantes de Ayotzinapa.

Octubre 13. Se realizan quemas, saqueos y destrozos en edificios públicos, resultando el más afectado el palacio de gobierno local. Se responsabiliza de los actos vandálicos a estudiantes de la Normal Rural de Ayotzinapa, miembros de la CETEG y familiares de las víctimas.



El mismo día, estudiantes de más de 10 escuelas y facultades de la Universidad Nacional Autónoma de México (UNAM) y la Universidad Autónoma Metropolitana (UAM) convocan a realizar un paro de 48 horas en solidaridad con los alumnos de Ayotzinapa.

Octubre 14. Suman casi 50 detenidos, luego de que la PGR informase la detención de 14 policías de la localidad de Cocula, vecina a Iguala. Se les señala de participar en la detención, traslado y entrega de los 43 alumnos a criminales. Jesús Murillo Karam, el titular de la PGR, afirma que ninguno de los cuerpos hallados en las primeras fosas corresponde a la identidad de los alumnos desaparecidos. Estudiantes de la Escuela Normal

Indígena de Cherán marchan al centro de Morelia (en la entidad de Michoacán, vecina de Guerrero) en solidaridad con sus compañeros de Ayotzinapa. En los hechos fueron vandalizados varios edificios, entre ellos el palacio de gobierno.

Octubre 15. La Policía Comunitaria de la Costa Chica, que se sumó a la búsqueda de los estudiantes, hallan seis fosas clandestinas más. Suman 19 fosas encontradas en los alrededores de Iguala.

Octubre 17. Fue arrestado en un punto de revisión en la autopista México - Toluca Sidronio Casarrubias Salgado, líder del grupo criminal Guerreros Unidos, al que se atribuye la desaparición de los jóvenes.

Octubre 18. La Policía Federal también se hizo de la seguridad del municipio de Cocula. El síndico de Iguala, Óscar Chávez Pineda, es presentado ante la fiscalía general como parte de las investigaciones.

Octubre 19. La fiscalía general mexicana ofrece una recompensa de 64.5 millones de pesos para quien brinde información que ayude a dar con el paradero de los estudiantes.



RECOMPENSA
 A quien o quienes proporcionen información veraz e útil para el esclarecimiento, detección, aprehensión y juzgamiento de los delincuentes que se combata para la localización de los personas reportadas en esta página.

 Acuerdo A/218/12 HASTA \$1,500,000	 Acuerdo A/233/11 HASTA \$1,500,000
 Acuerdo A/252/12 HASTA \$1,500,000	 Acuerdo A/326/11 HASTA \$1,500,000

Octubre 20. El sacerdote y defensor de derechos humanos mexicano, Alejandro Solalinde, trata de entregar a la PGR una declaración que contiene testimonios que recabó de varias personas que cuentan su versión.

Octubre 21. Maestros de la CETEG incendiaron la sede estatal del PRD.

Octubre 22. La PGR imputa a José Luis Abarca y María de los Ángeles Pineda, así como al director de Seguridad Pública, Felipe Flores, por la desaparición de los normalistas. El mismo día participaron unas 50 mil personas en una marcha en Ciudad de México.

Octubre 23. Ante la presión ejercida y en una conferencia en la que estuvo solo, el gobernador Aguirre

pidió licencia a su cargo por seis meses.

Octubre 24. La fiscalía detalla que han hallado a la fecha 11 fosas clandestinas y 38 cuerpos encontrados en Guerrero desde que comenzó la búsqueda de los 43 estudiantes.

Octubre 26. El Congreso de Guerrero designa al académico Rogelio Ortega como gobernador interino.

Octubre 29. Estudiantes de escuelas normales toman radiodifusoras en la entidad de Guerrero. Convocan a participar el 5 de noviembre en una jornada nacional por los desaparecidos. El mismo día se da la primera reunión entre el presidente Peña Nieto y los familiares de los desaparecidos, donde firman un acuerdo de 10 puntos y el mandatario se compromete a investigar a fondo.



Octubre 30. La subsecretaria de Asuntos Jurídicos de la Secretaría de Gobernación, Lía Limón, encabeza una delegación que asiste a la 153 sesión ordinaria de la Comisión Interamericana de Derechos Humanos, para presentar avances en las investigaciones.

Noviembre 3. Parte de Iguala con destino a Ciudad de México, el contingente de la caminata “43 x 43 Ni un desaparecido más”.

Noviembre 4. Fue detenido por la madrugada el ex alcalde José Luis Abarca junto con su esposa en una vivienda en Ciudad de México. Fue detenida también Noemí Berumen Rodríguez, por haber facilitado la evasión de Abarca.

Noviembre 5. José Luis Abarca es trasladado a una prisión de máxima seguridad.

Noviembre 6. En la reunión entre los familiares de los estudiantes desaparecidos en Iguala y el presidente de México, Enrique Peña Nieto, uno de los padres le dice al mandatario que considere su renuncia por la falta de resultados. En la misma fecha, la situación de los derechos humanos en México es “crítica” y crímenes como la desaparición de 43 estudiantes tras ataques de policías de Iguala, Guerrero, confirman la profundidad de la crisis que vive el país, alertó la ONG Human Rights Watch (HRW).

Noviembre 7. El procurador general de la República Mexicana, Jesús Murillo Karam, anunció que un grupo de nuevos detenidos confesó haber matado a un grupo grande, de más de 40 personas, que serían los 43 jóvenes desaparecidos en Iguala, el 26 de septiembre pasado.



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

La Paz - Bolivia