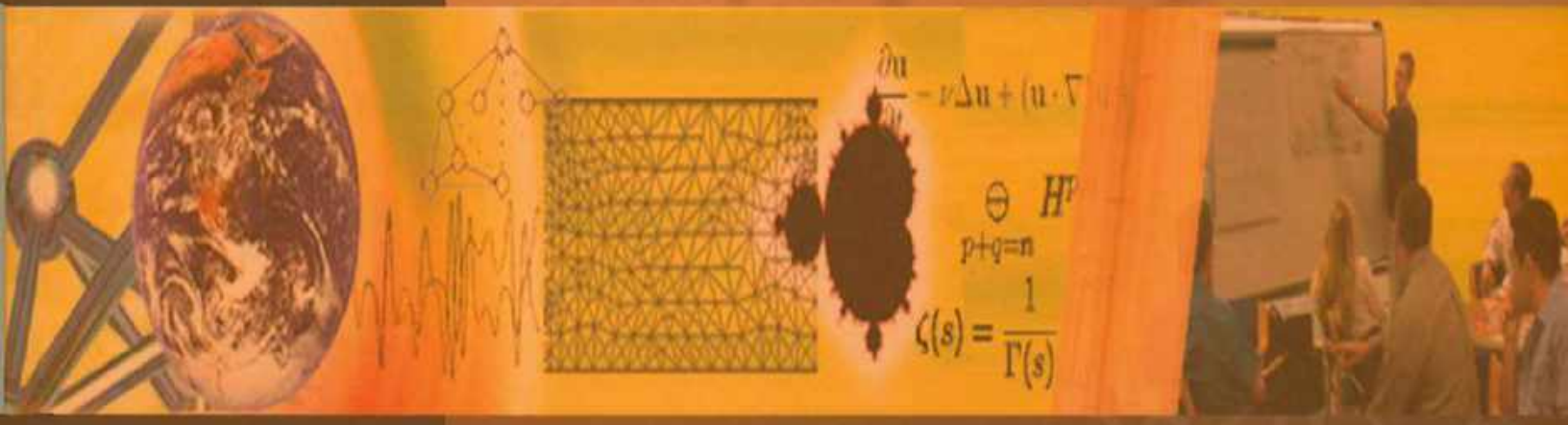
Universidad Mayor
de San Andrés

Varianza

Revista de la Carrera de Estadística



UMSA
FCPN
CARRERA
ESTADÍSTICA
IETA
Instituto de Estadística
Teórica y Aplicada

Varianza

Revista de la Carrera de Estadística

Publicación del Instituto de Estadística Teórica y Aplicada

Carrera de Estadística

Facultad de Ciencias Puras y Naturales

Universidad Mayor de San Andrés

Número 7

Noviembre de 2010

Revista Varianza
N° 7 - Noviembre, 2010

Dirección
Fernando O. Rivero Suguiura
Director del I.E.T.A.

Diagramación, Diseño y Fotografía
Mauricio J. Gamarra Urquidi

Diseño Logotipo IETA
Alvaro A. Merlo Ramos

Redacción y Edición
Consuelo B. Barrios Guzmán
Graciela M. Heredia Catacora
Beatriz W. Vallejos Mamani
Fernando O. Rivero Suguiura

Colaboradores
Nilda Flores Salinas
Dindo Valdez Blanco
Juan Carlos Flores López
Nicolás Chávez Quisbert
José Flores Aguilar
Lizzeth Arana Cuadros

Impresion
Stigma Tel. 2204403

Presentación

Este es un medio de comunicación que tiene la Carrera de Estadística dentro y fuera del entorno universitario, puesto que ésta revista no sólo llega a la comunidad universitaria de San Andrés, sino también, a instituciones y a la sociedad en sí. Esta edición permite la difusión de artículos científicos, historia, investigación, curiosidades de la estadística y las distintas actividades que se realizan, con el objeto de mostrar que la Carrera de Estadística en la UMSA, no sólo es enseñanza académica, sino también es investigación y actividad social. La revista que se edita este año, es especial a las ya presentadas anteriormente. Primero porque se la vuelve a publicar después de tres años de receso (la última edición fue del año 2007), y segundo porque ésta es diferente en su contenido, haciendo honor a su nombre "Varianza", que la hace más accesible a distinto público lector. A los lectores pedirles respetuosamente sus opiniones y sugerencias para mejorar la revista en las próximas ediciones.

Lic. Fernando O. Rivero Suguiura
DIRECTOR a.i. CARRERA DE ESTADISTICA

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

La Paz – Bolivia
Edificio Antiguo – Planta Baja
Telefax: 2442100

Dedicada a:

Rubén Belmonte, Luis Zapata y Raúl Marquiegui.
Por la creación de nuestra Carrera,
por su entrega y dedicación durante muchos años.
Algo importante que también nos han brindado:
calidad humana, amistad y respeto en favor de
quienes somos parte de ésta Carrera.

Gracias por lo que han hecho...

Contenido

El Modelo de Vectores Autorregresivos VAR

Autor: M.Sc. Nicolas Chavez Quisbert

1. Introducción a los Procesos Multivariados Autorregresivos

Los modelos autorregresivos fueron planteados inicialmente por Christopher Sims en un artículo publicado en 1980 en ECONOMETRICA, bajo el título de "Macroeconomía y la Realidad"

En el modelo VAR todas las variables son consideradas como endógenas, pues cada una de ellas se expresa como una función lineal de sus propios valores rezagados y de los valores rezagados de las restantes variables del modelo. Lo anterior permite capturar más apropiadamente los conocimientos de las variables y la dinámica de sus interrelaciones de corto plazo, lo cual no es detectable con modelos univariantes como los ARIMA. El VAR es también una técnica poderosa para generar pronósticos confiables en el corto plazo, aunque se le señalan ciertas limitaciones[1].

2. Proceso Multivariado Autorregresivo de orden p con intercepto VAR (p)

Definición.- Sea $\{\mu_t\}$ un proceso de ruido blanco multivariado con vector de medias $E(\mu_t)=0$ y matriz de varianzas y covarianzas Σ_μ, entonces un vector $\{Y_t\}$ se dice que es un proceso multivariado autorregresivo de orden p con intercepto, si: [LUTKEPOHL: 1991].

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots\ldots + \phi_p Y_{t-p} + \mu_t \qquad (1)$$

para

$$t = 0, \pm1, \pm2, \pm3, \pm4, \pm5, \pm6, \pm7, \ldots\ldots\ldots$$

donde:

$$Y_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \cdot \\ \cdot \\ \cdot \\ Y_{Kt} \end{bmatrix}$$

Es un vector de orden K × 1 de variables aleatorias.

Donde ϕ_i es una matriz de coeficiente o parámetros autorregresivos de orden K × K para (i=1, 2, 3,……., p denominados mecanismos de propagación del modelo).

$$C = \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_K \end{bmatrix}$$

Vector de términos interceptos de orden K × 1 existe siempre que $E(Y_t) \neq 0$

[1] Entre otros problemas, los VAR omiten la probabilidad de considerar relaciones no lineales entre las variables y no se toma en cuenta problemas de heterocedasticidad condicional ni cambio estructural en los parámetros estimados.

$$\mu_t = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ . \\ . \\ . \\ \mu_{Kt} \end{bmatrix}$$

Es el vector K dimensional de proceso de innovación o de ruido blanco multivariado.

Análisis del Vector de constantes C de interceptos

En el caso en que el proceso multivariado Y_t tiene un vector de medias μ constante se tiene que:

$$E(Y_t) = \mu$$

Por lo tanto el modelo definido en (1) se puede denotar en función del polinomio de retraso multivariante como:

$$(I_K - \phi_1 B - \phi_2 B^2 - ..... - \phi_p B^p) Y_t = C + \mu_t$$

Aplicando esperanzas matemáticas a la expresión anterior se tiene que:

$$C = (I_K - \phi_1 - \phi_2 - ..... - \phi_p)\mu$$

Lo cual implica que la ecuación (1) es equivalente a;

$$Y_t = (I_K - \phi_1 - ..... - \phi_p)\mu + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \mu_t$$

Si $E(Y_t) = 0$ se tiene el modelo reducido sin intercepto:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \mu_t$$

3. Proceso Multivariado Autorregresivo de orden p sin intercepto VAR (p)

Definición.- Sea $\{\mu_t\}$ un proceso de ruido blanco multivariado con vector de medias $E(\mu_t) = 0$ y matriz de varianzas y covarianzas Σ_μ, entonces un vector $\{Y_t\}$ se dice que es un proceso multivariado autoregresivo de orden p sin intercepto si $E(Y_t) = 0$, y este se define de la siguiente manera:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu_t \quad (2)$$

El cual puede escribirse en las siguientes formas equivalentes:

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - ..... - \phi_p Y_{t-p} = \mu_t \quad (3)$$

$$(I_K - \phi_1 B - \phi_2 B^2 - ..... - \phi_p B^p) Y_t = \mu_t \quad (4)$$

donde:

$$\phi(B) = I_K - \phi_1 B - \phi_2 B^2 - ..... - \phi_p B^p$$

Se denomina polinomio autorregresivo multivariado y μ_t es un proceso de ruido blanco multivariado.

Estacionariedad

El proceso autorregresivo multivariado definido en (4) es estacionario, si las raíces de la ecuación característica del polinomio de retraso multivariante en x.

$$\left| I_K - \phi_1 x - \phi_2 x^2 - ..... - \phi_p x^p \right| \quad (5)$$

4. Estimación de parámetros iniciales de un modelo VAR (p)

Definición.- Dado el modelo autorregresivo VAR (p) definido en (1).

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu_t \quad (6)$$

Donde C es vector de parámetros interceptos o constantes μ_t es un proceso de ruido blanco multivariado.

Si se definen los vectores:

$$W_t = \left[1\ Y_{t-1}\ \ Y_{t-2}\ldots\ldots\ Y_{t-p}\right]' \qquad (7)$$

De orden $(np+1)\times 1$

El número 1 correspondiente a la constante de cada ecuación del modelo VAR.

y

$$\Pi = \left[C\phi_1\phi_2\ldots\ldots\ldots\ldots\ldots\phi_p\right] \quad (8)$$

de orden $\left[n\times(np+1)\right]$

Donde $Y_{t-1},\ldots\ldots\ldots,Y_{t-p+1}$, se distribuyen normales y la combinación lineal de dichas variables se distribuye según una normal multivariada.

Entonces

$$Y_t / Y_{t-1}Y_{t-2}\ldots\ldots\ldots\ldots\ldots,Y_{t-p+1}$$

Se distribuye según una normal multivariada con vector de medias ΠW y matriz de varianzas y covarianzas Σ_μ, es decir $N\left(\Pi W,\sum_\mu\right)$ la función condicional de la t – esima observación es:

$$fY_t / Y_{t-1},Y_{t-2},\ldots.X_1\left(y_t / y_{t-1}y_{t-2},\ldots,y_1,\vartheta\right) \qquad (9)$$

$$=(2\Pi)^{-n/2}\left|\sum{}^{-1}\right|^{1/2}Exp\left[-1/2(Y_t-\Pi W_t)'\sum{}_\mu^{-1}(Y_t-\Pi W_t)\right] \qquad (10)$$

Donde ϑ es un vector que tiene como elementos a las matrices C, ϕ_1, ϕ_2, ……,ϕ_p y Σ_μ. La función de verosimilitud para la estimación de T periodos se define como:

$$L_{Y_t,Y_{t-1},\ldots.Y_1/Y_0,Y_{-1},\ldots.Y_{-p+1}}\left(y_t,y_{t-1},\ldots.y_1/y_0,y_1,\ldots,y_{-p+1}\vartheta\right)$$

$$=\prod_{t=1}^{T}(2\Pi)^{-n/2}\left|\sum{}_\mu^{-1}\right|^{1/2}\left[Exp-1/2(Y_t-\Pi W_t)'\sum{}_\mu^{-1}(Y_t-\Pi W_t)\right] \qquad (11)$$

Abreviando la notación y aplicando logaritmos neperiano para linealizar la ecuación (11) se tiene que.

$$\ln L\left(\Pi,\sum{}_\mu\right)=-\frac{nT}{2}\ln(2\pi)+\frac{T}{2}\ln\left|\sum{}_\mu^{-1}\right|-\frac{1}{2}\sum_{t=1}^{T}(Y_t-\Pi W_t)'\sum{}_\mu^{-1}(Y_t-\Pi W_t) \qquad (12)$$

Estimación del Vector Π

Derivando (12) con respecto del vector Π se tiene que:

$$\frac{\partial \ln L\left(\Pi,\sum_\mu\right)}{\partial\Pi}=\left[-\frac{1}{2}\sum_{t=1}^{T}\left[(Y_t-\Pi W_t)'\sum{}_\mu^{-1}(Y_t-\Pi W_t)\right]\right]'=0 \qquad (13)$$

Derivando parcialmente por regla de la cadena con respecto a Π[2].

$$-\frac{1}{2}\sum_{t=1}^{T}2\sum{}_\mu^{-1}(Y_t-\Pi W_t)(-W_t')=0$$

$$\sum_{t=1}^{T}Y_tW_t'-\sum_{t=1}^{T}\Pi W_tW_t'=0$$

$$\sum_{t=1}^{T}\Pi W_tW_t'=\sum_{t=1}^{T}Y_tW_t'$$

$$\Pi\sum_{t=1}^{T}W_tW_t'=\sum_{t=0}^{T}Y_tW_t'$$

$$\hat{\Pi}=\left[\sum_{t=0}^{T}Y_tW_t'\right]\left[\sum_{t=0}^{T}W_tW_t'\right]^{-1} \qquad (14)$$

[2] Si A es una matriz simétrica y X^t AX es una forma cuadrática, la derivada parcial con respecto a X es $\frac{\partial(X'AX)}{\partial X}=2AX$

Dicho estimador inicial es el de mínimos cuadrados ordinarios [DOORNIK: 1994].

Estimación de Σ_μ

Derivando (12) con respecto Σ_μ por la regla de la cadena se tiene que:

$$\frac{\partial \ln(\Pi, \sum_\mu)}{\partial \sum_\mu} = \frac{T}{2}\frac{\partial \ln\left|\sum_\mu^{-1}\right|}{\partial\left|\sum_\mu^{-1}\right|} - \frac{1}{2}\sum_{t=0}^{T}\frac{\partial(\mu_t' \sum_\mu^{-1} \mu_t)}{\partial \sum_\mu^{-1}} = 0 \quad (15)$$

$$\frac{T}{2}\sum_\mu - \frac{1}{2}\sum_{t=0}^{T}\mu'\mu = 0$$

$$\hat{\sum}_\mu = \frac{\sum_{t=0}^{T}\mu_t'\mu_t}{T} = \frac{\sum_{t=0}^{T}\mu_{it}^2}{T} \quad (16)$$

Dicho estimador representa al error cuadrático medio de la regresión para la i – esima variable [DOORNIK: 1994].

5. Bibliografía

[1] *Amisano, Gianni & Giannini, Carlo.* Topies in Structural Var Econometrics. Second Edition, Springer, 1997

[2] *Doornik, J. A. & Hansen,* H. "An Omnibus Test For Univariante And Multivariante Normality". Nuffield College, Oxford, 1994

[3] *Fernandez – Corugedo* "Curso Modelos Macroeconómicos para la Política Monetaria" Banco Central de la República de Argentina, Mimeo – Argentina , 2003

[4] *Hamilton, James D.* "Time Series Analysis"; Princeton University Press; Princeton. 1994

[5] *Harvey A. C.* "The Econometric Analysis Of Time Series"; Mit Press; Cambridge, Mass: 1990

[6] *Lutkepohl, H.* "Introduction To Multiple Time Series Analysis"; Springer-Verlag. Heildelgerg. Second Edition, 1991

[7] *Maddala, G. S.* "Econometrics"; McGRAW – Hill ; New York. 1990

Índice

"Los fundamentos de la estadística están cambiando, no sólo en el sentido en que ellos fueron y continuarán evolucionando, sino también en el sentido idiomático de que ningún sistema es absolutamente estable."

L. J. Savage

Aplicación del Análisis de Escalamiento Multidimensional

Autor: Lic. Juan Carlos Flores López

Para el desarrollo de esta investigación se tomo en cuenta el estudio realizado en la investigación sobre las causas del abandono de las materias de los estudiantes de la Facultad de Ciencias Puras y Naturales, realizado para la gestión anterior.

La teoría del análisis de escalamiento multidimensional, demanda el conocimiento de distancias y las medidas de similaridad, por ejemplo, la distancia euclidea, distancia euclidea al cuadrado, Distancia de Minkowski, Distancia city block o «Manhattan», Medidas de similaridad para datos binarios, etc.

El análisis de escalamiento multidimensional (MDS)-*multidimensional scaling*- es una técnica de reducción de datos como otras como es el análisis factorial o análisis de componentes principales, por ejemplo. El objetivo principal del MDS es representar *N* objetos en un espacio dimensional reducido (q dimensiones, siendo q < *N),* de tal forma que la distorsión causada por la reducción de la dimensionalidad sea la menor posible, es decir, que las distancias entre los objetos representados en el espacio q dimensional, sean lo más parecidas posible a las distancias en el espacio *N* dimensional.

Dado que será difícil que las distancias coincidan, el objetivo del MDS es conseguir que ambas configuraciones dimensionales sean lo más parecidas posible. Para ello será necesario construir un indicador de esa proximidad que se denomina *stress y* otro similar el *s-stress*, cuya interpretación de resultados se presenta en el siguiente cuadro.

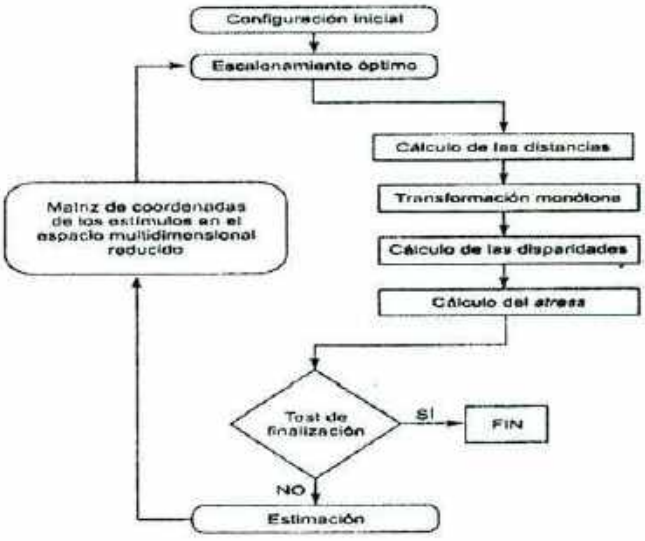
Cuadro. Interpretación de Stress en términos de bondad de ajuste del MDS.

Stress	Bondad de ajuste
0,2	Malo
0,1	Mínimo Razonable
0,05	Bueno
0,025	Excelente
0	Perfecto

El indicador *S-Stress*, esta dado por:

$$S - Stress = \sqrt{\frac{\sum_{i \neq j} \left(d^2_{ij} - \delta^2_{ij}\right)^2}{\sum_{i \neq j} d^4_{ij}}}$$

A continuación se muestra el Síntesis del algoritmo básico del MDS, para el calculo del indicador.



1. Estimación del modelo

En el caso en que no se haya concluido el proceso iterativo, en esta segunda fase se procede a la estimación del modelo. Con el algoritmo ALSCAL, como todos los diseñados para resolver el problema que se plantea el MDS es iterativo, debiéndose determinar en cada iteración dos variables fundamentales: las *coordenadas de los estímulos* y los *pesos de los individuos,* en el caso de que se trate de un *escalonamiento multidimensional ponderado.*, los pesos y las coordenadas no pueden obtenerse simultáneamente. Por ello la estimación del modelo se hace en dos etapas: (a) estimación de los pesos, y (b) estimación de las coordenadas de los estímulos. En un modelo no ponderado, lógicamente, no se aplica la primera fase.(el proceso teórico puede ser visto en la Investigación de la gestión 2009 de titulo Aplicación del Método Escalamiento Multidimensional).

2. Aplicación del análisis de Escalamiento Multidimensional

La teoría del análisis de Escalamiento Multidimensional, se toma en cuenta en el estudio realizado sobre la investigación de las causas del abandono de las materias de los estudiantes de la Facultad de Ciencias Puras y Naturales.

En la actualidad la Facultad de Ciencias Puras y Naturales (FCPN), es parte de las 13 Facultades que tiene la Universidad Mayor de San Andrés. La FCPN tiene 6 Carreras que son las siguientes: ESTADISTICA, BIOLOGIA, INFORMATICA, FISICA, MATEMATICA y QUIMICA

De acuerdo a estudios realizados, el abandono ocurre por diversos motivos, estos pueden deberse a entorno universitario, como el entorno social, cultural y familiar en el que se mueve el alumno. Según el informe elaborado por la Comisión Europea, los factores que provocan el abandono de los estudios temprano se pueden clasificar del siguiente modo: Características individuales propias de los alumnos, Razones educativas, Razones familiares, Comunidad y amigos.

La deserción universitaria es un problema educativo que afecta al desarrollo de la sociedad, y se da principalmente por falta de recursos económicos y por una desintegración familiar.
Es un fenómeno social ocasionado por diversas causas ya sean políticas, económicas, familiares, etc. Lo cual debe ser estudiado detenidamente para determinar las posibles soluciones, así como también su prevención. El abandono temporal o definitivo de los estudios de un individuo es motivo de varios factores, además de otros elementos ya sean externos o internos que motivan a tomar esta decisión. Los factores externos, por ejemplo pueden ser por presiones económicas, influencia negativa de padres, amigos, familiares, embarazos a temprana edad, maestros, complejidad de las materias, etc.

Y los factores internos pueden ser por desinterés personal, no tener motivación en la vida, desagrado por la universidad, carrera, materia (desgano de salir adelante y sobresalir entre las demás personas, cruce de horarios, etc.).

Para que un estudiante abandone una carrera universitaria se combinan aspectos como el lugar en donde reside, el nivel de ingresos, el nivel educativo de los padres de familia, la necesidad de trabajar para mantenerse o contribuir a los ingresos familiares y el propio ambiente familiar,

incluso de violencia en el que se vive. "Esta situación es la que afecta con mayor fuerza a los jóvenes de menores ingresos, por lo que el tema financiero y la eficiencia en el gasto se hace más crítico".

La "deficiente preparación previa (en el bachillerato)" es otra de las causas del abandono de los estudios universitario; la carencia de mecanismos de financiamiento o becas estudiantiles; la prevalencia de políticas de "ingreso irrestricto, selectivo sin cupo fijo o selectivo con cupo"; el desconocimiento de lo que es la profesión, el ambiente escolar y la carencia de lazos afectivos con la universidad, también impactan en los jóvenes para que dejen la Universidad.

3. Variables en consideración de la investigación

En esta investigación se ha considerado la encuesta sobre el abandono de materias de los estudiantes de la Facultad de Ciencias Puras y Naturales de la Universidad de Mayor de San Andrés. La muestra fue de tamaño 269 Estudiantes. Se realizo un muestreo aleatorio simple considerando dominios de grupo, además dentro de cada dominio se distribuye la muestra de manera proporcional a los tamaños de los dominios. Considerando que la carrera de Informática es muy numerosa, se ha visto por conveniente tomar la mitad de la muestra en informática y la otra mitad se distribuyo proporcionalmente al tamaño de cada carrera, con el objeto de tener una muestra más representativa. El resultado del muestreo es el siguiente Carrera de Estadística 22 estudiantes, Biología 32 estudiantes, Informática 134 estudiantes seleccionados, Física 16 estudiantes seleccionados, Matemáticas 43 estudiantes seleccionados, Química 22 estudiantes seleccionados.

Las variables más importantes, que se toma en cuenta en este estudio son:

¿asiste regularmente a clases?, ¿considera cómodos sus horarios de clases?, ¿considera cómodos sus horarios de clases?, ¿dispone de tiempo suficiente para realizar trabajos y/o prácticas universitarias?, ¿alguna vez abandonó una o varias materias?.

Si usted considerara abandonar una o varias materias, ¿qué lo motivaría a abandonarlas?

1) Falta de tiempo ( )
2) Flojera ( )
3) Bajo rendimiento ( )
4) Falta de interés ( )
5) Por trabajo ( )
6) Docente ( )
7) Choque de horario ( )
8) Desorganizado ( )
9) Otros_____

Si usted considerara abandonar sus estudios universitarios, ¿por qué lo haría?

1) Falta de tiempo ( )
2) Falta de dinero ( )
3) Bajo rendimiento ( )
4) Otros_____

Para el análisis, se ha pedido a los estudiantes, que nos valoren el por qué un estudiante abandona una o más materias bajo la consideración de los siguientes criterios, atendiendo a la similitud con que las perciben. Para ello utiliza una escala de 1 (totalmente diferentes) a 7 (idénticas).

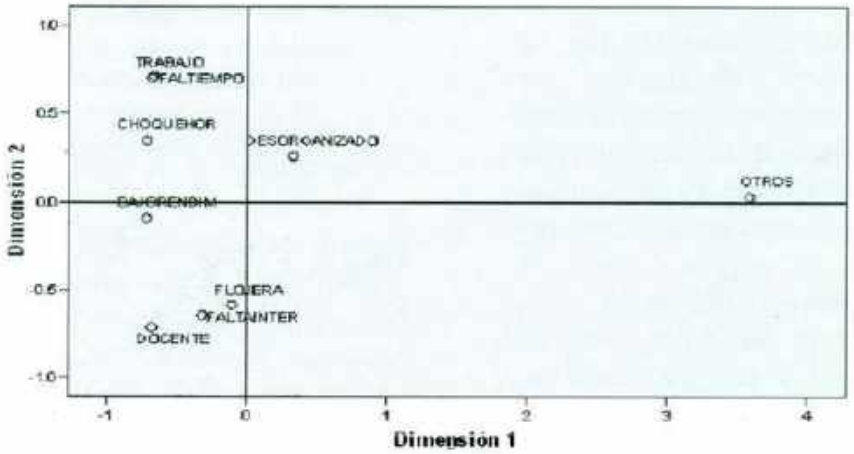
La siguiente matriz de *disparidades originales* - o *proximidades* - nos muestra las medias de las puntuaciones ofrecidas por los estudiantes de la muestra de la Facultad de Ciencias Puras y Naturales.

	falta de tiempo	Flojera	Bajo rendimiento	Falta de interés	Por trabajo	Docente	Choque de horario	Desorganizado	otros
falta de tiempo	7	3,26	3,85	3,19	4,56	3,2	4,07	3,51	2,38
Flojera	3,26	7	4,25	3,92	3,15	3,27	3,3	3,46	2,32
Bajo rendimiento	3,85	4,25	7	4,21	4,21	4	4	3,62	2,53
Falta de interés	3,19	3,92	4,21	7	3,19	3,91	3,33	3,38	2,38
Por trabajo	4,56	3,15	4,21	3,19	7	3,19	4,1	3,19	2,42
Docente	3,2	3,27	4	3,91	3,19	7	3,73	3,02	2,19
Choque de horario	4,07	3,3	4	3,33	4,1	3,73	7	3,46	2,3
Desorganizado	3,51	3,46	3,62	3,38	3,19	3,02	3,46	7	2,6
Otros	2,38	2,32	2,53	2,38	2,42	2,19	2,3	2,6	7

Nótese que en la diagonal de la matriz aparecen sietes porque es una percepción de similitud de cada variable, siempre ha de ser idéntica a sí misma. Se puede observar que esta matriz es simétrica por construcción. Si creamos un mapa en dos dimensiones para ilustrar mejor la percepción de los estudiantes por que abandonan su(s) materias en un semestre, este mapa debería representar como puntos cercanos a las variables ¿por que abandonaría una materia un estudiante?. De acuerdo a los resultados de la matriz las variables de abandono seria la falta de tiempo y por trabajo, porque la disparidad entre ellas es alta (4,56), y así sucesivamente.

Considerando las variables anteriormente mencionadas y utilizando el software SPSS se procede a la utilización del método de análisis de escalonamiento multidimensional, cuyos resultados son los siguientes:

Figura 1. Modelo de distancia euclidea



La Figura 1 ofrece el mapa que se obtiene al representar las coordenadas bidimensionales resultantes de aplicar a la matriz anterior uno de los algoritmos que existen para efectuar un MDS, el ALSCAL. Puede comprobarse que los estudiantes de la Facultad de Ciencias Puras y Naturales abandonan su(s) materia(s), por trabajo, por falta de tiempo y choque de horario, o por falta de interés, flojera y/o por su docente.

El indicador de calidad de método de escalamiento MDS *(stress* o *s-stress)* refleja este hecho, como podemos ver la salida del software

Iteration history for the 2 dimensional solution (in squared distances)
Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement

Iterations stopped because
S-stress improvement is less than ,001000

Una medida de las discrepancias entre las matrices de disparidades y distancias de acuerdo a la teoría la calidad de la representación lograda por el MDS, es el estadístico denominado S- *Stress,* que fue propuesto por Takane, Young y De Leeuw (1977), autores del algoritmo ALSCAL y que, por ello, es la función que se minimiza en este algoritmo:

El valor del *s-stress* está siempre comprendido entre 0 y 1 y cualquier valor inferior a 0,1 indica que la solución obtenida es una buena representación de los objetos de la solución *N* dimensional inicial.

En nuestro caso el S-Stress es 0.001 lo que implica una excelente representación de los objetos, es decir la similitud de la percepción de los estudiantes por que abandonan sus materias de los estudiantes de la muestra es excelente, bajo el criterio del método en consideracion.

Stress and squared correlation (RSQ) in distances
For matrix
Stress = ,05644 RSQ = ,99210

Se puede ver también que el indicador de bondad de ajuste que nos proporciona el software RSQ = 0.99210. Es muy buena, es decir la solución dimensional reducida es una excelente representación de la solución N dimensional, si la ordenación de las distancias entre los objetos de la primera mantiene la ordenación de las disparidades originales de la segunda. En el caso ideal, el grafico de dispersión que representa distancias y disparidades debería ser una línea recta como indica la teoría, lo cual se cumple en esta investigación.

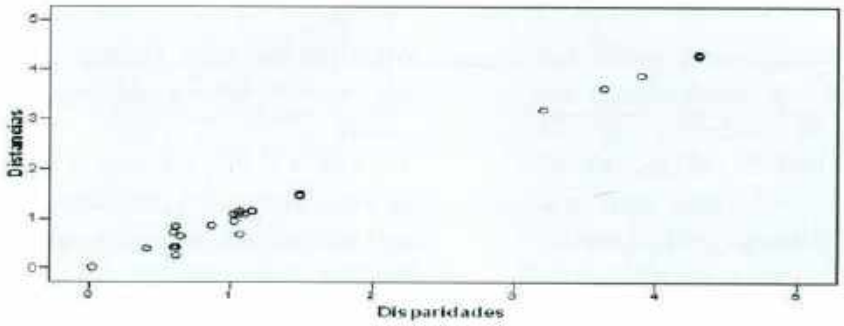
Los siguientes datos muestran la solución bidimensional final resultante de la aplicación del análisis de escalamiento multidimensional a los datos de nuestra investigación. En este cuadro aparecen las coordenadas de cada objeto (percepción de los estudiantes del por qué abandonan los estudiantes sus materias). A partir de estas coordenadas se derivan las matrices de distancias euclideas.

Configuration derived in 2 dimensions
Stimulus Coordinates

		Dimension	
Stimulus Number	Stimulus Name	1	2
1	FALTIEMP	-,6918	,7502
2	FLOJERA	-,0324	-,5889
3	BAJOREND	-,7376	-,1124
4	FALTAINT	-,2837	-,6747
5	TRABAJO	-,7044	,7293
6	DOCENTE	-,6977	-,7580
7	CHOQUEHO	-,7455	,3262
8	DESORGAN	,3490	,2891
9	OTROS	3,5440	,0391

La Figura 2 muestra la dispersión que representa distancias y disparidades de nuestros datos, el cual efectivamente se ajusta a una línea recta, como indica la teoría MDS.

Figura 2 Ajuste lineal modelo de distancia euclidea



Como resultado de la investigación, utilizando el método de Escalamiento Multidimensional, se puede ver que los estudiantes de la Facultad de Ciencias Puras y Naturales abandonan sus materias por trabajo, por falta de tiempo y choque de horario, y un segundo motivo por falta de interés, flojera y/o por su docente.

Para la comprobación de los resultados se ha utilizado el indicador de calidad de método de escalamiento MDS *(stress o s-stress)* propuesto por Takane, Young y De Leeuw autores del algoritmo ALSCAL que es la función que se minimiza en este algoritmo, cuyo resultado fue de 0.001 lo que implica una excelente representación de los objetos. Es decir la similitud de la percepción de los estudiantes del por qué abandonan sus materias de los estudiantes de la muestra por el método Escalamiento Multidimensional es excelente.

Índice

"Conseguimos obtener así la fórmula estadística para conocer aproximadamente la posición de un electrón en un instante determinado. Pero, personalmente, no creo que Dios juegue a los dados."

Albert Einstein

Marcos de Muestreo Imperfectos

Autor: Lic. Fernando O. Rivero Suguiura

1. Planteamiento de los problemas de Marcos Muestrales

En la teoría de poblaciones finitas se supone que el marco de lista de la que se selecciona la muestra, coincide con las unidades que son parte de la población objeto de estudio. En la práctica no siempre esto es cierto, ya que las listas presentan con cierta frecuencia algunos defectos que pueden dar lugar a la aparición de sesgos y a la alteración de las varianzas de los estimadores. Los problemas más frecuentes en los marcos muestrales de lista, son:

- *La sobrecobertura*
- *La subcobertura*
- *La duplicación*

Las dos últimas limitaciones de marco muestral no son tratados en esta investigación, pero vale la pena nombrarlos:

La subcobertura se presenta cuando existen unidades que están en la población objetivo y no en el marco, luego el marco muestral se convierte en un subconjunto de la población. Los siguientes ejemplos describen esta situación:

- Suponer que existen nuevas viviendas ocupadas en zonas marginales de una ciudad que son parte de la población, pero por falta de actualización del marco muestral estas no están presentes en dicho marco, dándoles probabilidad igual a cero de ser seleccionadas en la muestra. En este caso se convierten en unidades faltantes.

- Se muestrea de una lista de establecimientos económicos manufactureros del censo 2000 sin considerar establecimientos de nueva creación que son parte de la población. También son unidades faltantes.

La duplicación de unidades muestrales es cuando el marco muestral contiene elementos repetidos.

Siguiendo los dos ejemplos anteriores:

- Suponer que una vivienda se registra en el marco muestral como ocupada con tres personas, tal cual ocurre con la población objetivo en el periodo de tiempo **1.** Suponer que las tres personas dejan dicha vivienda y son ocupadas por otras cinco personas distintas a las anteriores. Esta vivienda es nuevamente registrada en el marco como ocupada en el periodo de tiempo **2.** Si no se elimina del marco la vivienda del periodo de tiempo 1, se presenta el problema de unidad duplicada.

  El establecimiento económico de nombre "El Buen Productor" está

registrado en el marco como parte de la población objetivo. Suponer el fallecimiento del dueño del establecimiento económico y el cierre temporal. Después de un periodo de tiempo, se abre nuevamente con el nombre "Carlos Andrade" y éste es registrado en el marco muestral como nuevo establecimiento económico. Es un caso de duplicación.

La existencia de *unidades extrañas* en el marco muestral es tema de esta investigación. Por unidad extraña se entiende una unidad de muestreo que, incluida en el marco, no pertenece a la población que se desea estudiar, o que no es una unidad del colectivo que se desea seleccionar para ser parte de la muestra.

Como ejemplos consideremos los siguientes:

- Para una encuesta se desea seleccionar una muestra aleatoria de viviendas ocupadas del marco muestral disponible. Si muchas de las especificadas en el marco están como ocupadas y en la población objetivo son desocupadas, se pueden muestrear unidades extrañas.

- Si para estimar alguna característica de producción en empresas manufactureras del país, se muestrea de una lista de establecimientos económicos manufactureros que en la actualidad cerraron sus puertas por falta de utilidad. Se puede seleccionar unidades extrañas.

Los ejemplos anteriores ponen de manifiesto que la presencia de unidades extrañas en el marco ocurre principalmente en dos tipos de situaciones prácticas:

a) Por no tener actualizado el marco muestral. La lista incluye unidades que han dejado de pertenecer a la población objetivo.

b) La población que se desea muestrear se convierte en una subpoblación del marco muestral generándose el problema de la *sobrecobertura*.

2. El problema de las unidades extrañas en el marco

Ante la situación de sobrecobertura de marcos muestrales, pueden adoptarse diferentes reglas de solución, no todas igualmente adecuadas, cuya puesta en práctica depende de los recursos disponibles.

a) Depuración del marco de muestreo

Eliminar del marco las unidades extrañas sabiendo el número total de inclusión. Con este proceso queda solucionado el problema de sobrecobertura puesto que luego se procede a seleccionar la muestra del marco depurado. En muchos casos esto no será posible con los recursos dados, bien sea porque el marco disponible no contiene información acerca de cuáles unidades son extrañas (siendo necesario un trabajo de campo de carácter exhaustivo o la construcción de un nuevo marco muestral) o por limitaciones de tiempo, presupuesto, personal no disponible, etc.

b) Reemplazo de las unidades extrañas en la muestra

Otra solución consiste en seleccionar una muestra aleatoria del marco disponible no depurado y sustituir las unidades que resulten ser extrañas por otras aleatoriamente seleccionadas de entre las restantes del marco, hasta completar el

tamaño de la muestra planificada con unidades todas no extrañas. Lamentablemente esta solución da lugar a estimaciones sesgadas, como se verá luego.

De acuerdo a los dos puntos anteriores, se considera algunos procedimientos para la estimación del total.

3. Relación de parámetros poblacionales en marcos depurados y no depurados

Suponer que de las N unidades de muestreo incluidas en el marco, N^* son unidades no extrañas, y por lo tanto $N-N^*$ son unidades extrañas. Entonces se puede enumerar de 1 a N^* las unidades que no son extrañas y de N^*+1 a N las unidades extrañas del marco. Es decir que el marco de muestreo disponible no depurado se constituye en el conjunto:

$$\Omega=\{U_1,U_2,\ldots,U_N\} \qquad (1)$$

y el conjunto

$$\Omega^*=\{U_{1^*},U_{2^*},\ldots,U_{N^*}\} \qquad (2)$$

es el marco de muestreo depurado.

La proporción de unidades extrañas en Ω es:

$$W=\frac{N-N^*}{N}=1-\frac{N^*}{N} \qquad (3)$$

Sea Y_i el valor de una variable que se desea investigar de la i-ésima unidad U_i, se atribuye el valor de Y_i igual a cero cuando U_i es una unidad extraña. Por tratarse de una unidad extraña, el valor de Y_i puede ser realmente cero o no estar definido. Ya que la unidad muestral no pertenece a la población objeto de estudio cuyo total se desea estimar, su contribución a dicho total es nula, justificando así el valor de Y_i igual a cero.

Los valores totales de ambos marcos son entonces coincidentes, es decir:

$$T=\sum_{i=1}^{N}Y_i=\sum_{i=1}^{N^*}Y_i \qquad (4)$$

al igual que:

$$\sum_{i,j}^{N}Y_iY_j=\sum_{i,j}^{N^*}Y_iY_j$$

$$\sum_{i=1}^{N}Y_i^2=\sum_{i=1}^{N^*}Y_i^2 \qquad (5)$$

sin embargo las medias en uno u otro marco no son iguales:

$$\overline{Y}=(1-W)\overline{Y}^* \qquad (6)$$

puesto que

$$\overline{Y}=\frac{\sum_{i=1}^{N}Y_i}{N}=\frac{N^*}{N^*}\frac{\sum_{i=1}^{N^*}Y_i}{N}=$$

$$=\frac{N^*}{N}\frac{\sum_{i=1}^{N^*}Y_i}{N^*}=(1-W)\overline{Y}^*$$

y la varianza queda como:

$$\sigma^2=(1-W)\sigma^{*2}+\frac{W}{1-W}\overline{Y}^2 \qquad (7)$$

La relación (7) entre varianzas en el marco no depurado σ^2 y en el marco

depurado σ^{*2} se obtiene de la siguiente manera:

$$N\sigma^2 = \sum_{i=1}^{N} Y_i^2 - N\left(\frac{\sum_{i=1}^{N} Y}{N}\right)^2 =$$

$$= \sum_{i=1}^{N} Y_i^2 - \frac{T^{*2}}{N^*} + \frac{T^{*2}}{N^*} - \frac{T^2}{N}$$

$$N\sigma^2 = N^*\sigma^{*2} + T^2\left(\frac{1}{N^*} - \frac{1}{N}\right) =$$

$$= N^*\sigma^{*2} + T^2\left(\frac{N - N^*}{N^* N}\right) \qquad (8)$$

Despejando σ^2 y sustituyendo $\frac{N - N^*}{N}$ por W y $\frac{N^*}{N}$ por $1 - W$, resulta la relación propuesta (7).

Si se admite la aproximación de $N - 1 \cong N$ y $N^* - 1 \cong N^*$, la relación de cuasivarianzas de ambos marcos, es:

$$S^2 = (1 - W)S^{*2} + \frac{W}{1 - W}\bar{Y}^2 \qquad (9)$$

4. Estimador del total $\hat{T}^*$ cuando se muestrea con el marco depurado

Si se considera un muestreo aleatorio simple sin reemplazo con probabilidades iguales de selección, se tiene que el estimador del total $\hat{T}^*$, es:

$$\hat{T}^* = \frac{N^*}{n} t \qquad (10)$$

con n tamaño de muestra y $t = \sum_{i=1}^{n} Y_i$ total muestral. Es un estimador insesgado de T, cuya varianza es:

$$V(\hat{T}^*) = N^{*2}(1 - f^*)\frac{S^{*2}}{n} \qquad (11)$$

donde $f^* = \frac{n}{N^*}$ es la fracción de muestreo y S^{*2} denota la cuasivarianza en el marco depurado.

Estos dos resultados forman parte de la teoría elemental de muestreo de poblaciones finitas, por ello, se aceptan sin demostración.

5. Estimador del total $\hat{T}$ cuando se muestrea con el marco no depurado

Se señalan dos aspectos en la estimación, estos son:

5.1. No se sustituyen las unidades extrañas que aparecen en la muestra

En estas condiciones, si se considera un muestreo aleatorio simple sin reemplazo con probabilidades iguales de selección, se tiene que el estimador del total $\hat{T}_1$, es:

$$\hat{T}_1 = \frac{N}{n} t \qquad (12)$$

la expresión (12) es un estimador insesgado y su varianza está dada por:

$$V(\hat{T}_1) = N^2(1 - f)\frac{S^2}{n} \qquad (13)$$

donde $f = \frac{n}{N}$ y S^2 es la cuasivarianza en el marco no depurado.

La varianza del total estimado con marco muestral no depurado de la expresión (13), es mayor a la varianza del total estimado con marco muestral depurado de la expresión (11). Para probar que $V(\hat{T}_1) > V(\hat{T}^*)$ se determina la diferencia entre ambas, es decir:

$$d = N^2(1-f)\frac{S^2}{n} - N^{*2}(1-f^*)\frac{S^{*2}}{n} \qquad (14)$$

Para verificar que $d > 0$, se debe tener en cuenta que:

$$N^2(1-f) > N^{*2}(1-f^*) \qquad (15)$$

puesto que $N > N^*$ y que

$$f^* = \frac{n}{N^*} \geq f = \frac{n}{N}$$

Luego, analizar diferencias entre cuasivarianzas de ambos marcos muestrales, es decir:

$$S^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} Y_i^2 - \frac{T^2}{N}\right) =$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} Y_i^2 - \frac{T^2}{N(N-1)}$$

$$S^{*2} = \frac{1}{N^*-1}\left(\sum_{i=1}^{N^*} Y_i^2 - \frac{T^2}{N^*}\right) =$$

$$= \frac{1}{N^*-1}\sum_{i=1}^{N^*} Y_i^2 - \frac{T^2}{N^*(N^*-1)} \qquad (16)$$

de acuerdo con (11) y (13) se obtiene:

$$d = \frac{1}{n}\left\{\sum_{i=1}^{N} Y_i^2\left[\frac{N(N-1)}{N-1} - \frac{N^*(N^*-1)}{N^*-1}\right] - T^2\left(\frac{N-n}{N-1} - \frac{N^*-n}{N^*-1}\right)\right\} \qquad (17)$$

Puesto que $T^2 = \sum_{i=1}^{N} Y_i^2 + \sum_{i\neq j}^{N} Y_i Y_j$,

reemplazando en (17) y simplificando, se tiene:

$$d = \frac{N-N^*}{(N-1)(N^*-1)n}$$

$$\left[(N-1)(N^*-1)\left(\sum_{i=1}^{N} Y_i^2\right) - (n-1)\left(\sum_{i\neq j}^{N} Y_i Y_j\right)\right] \qquad (18)$$

el primer factor de d es positivo, se debe probar que también lo es el segundo factor. En efecto, por ser:

$$N\sigma^2 = \sum_{i=1}^{N}(Y_i - \bar{Y})^2 > 0 \qquad (19)$$

se cumple que

$$(N-1)\sum_{i=1}^{N} Y_i^2 > \sum_{i=1}^{N} Y_i Y_j \qquad (20)$$

Por tanto, si se multiplica el primer miembro de la relación (20) por N^*-1 y el segundo por $n-1$, dado que $N^*-1 > n-1$ y ambos positivos, también es:

$$(N-1)(N^*-1)\sum_{i=1}^{N} Y_i^2 > (n-1)\sum_{i=1}^{N} Y_i Y_j \qquad (21)$$

lo cual completa la demostración:

$$V(\hat{T}_1) > V(\hat{T}^*)$$

La comparación de las varianzas de $\hat{T}_1$ y $\hat{T}^*$ también se puede realizar mediante el cociente de forma muy simple con el grado de aproximación que permita la sustitución de $N \cong N-1$, $N^* \cong N^*-1$. Bajo el criterio anterior, las varianzas son iguales a las cuasivarianzas, y de acuerdo a (6) y (9) se obtiene:

$$\frac{V(\hat{T}_1)}{V(\hat{T}^*)} = \frac{N-n}{N^*-n}\left(1 + W\frac{\bar{Y}^{*2}}{S^{*2}}\right)$$

(22)

Relación que también pone de manifiesto la mayor varianza de $\hat{T}_1$, por ser $N > N^*$ cuando existen unidades extrañas.

5.2 Se sustituyen aleatoriamente las unidades extrañas que aparecen en la muestra

Así, si se selecciona del marco no depurado sin reemplazo y con probabilidades iguales, el estimador del total $\hat{T}_2$, es:

$$\hat{T}_2 = \frac{N}{n}t \qquad (23)$$

La expresión (23) es un estimador sesgado. Para determinar el sesgo y la varianza de $\hat{T}_2$ basta tener en cuenta que la sustitución aleatoria de las unidades extrañas de la muestra hasta obtener n no extrañas, equivale a efectuar el muestreo en el marco depurado; así resulta que para las variables indicadoras $e_i = 1$, si U_i pertenece a la muestra; y $e_i = 0$, si U_i no pertenece a la muestra, se cumple:

$$E(e_i) = 0 \qquad \text{si } U_i \text{ es extraña}$$

$$E(e_i) = \frac{n}{N^*} \qquad \text{si } U_i \text{ es no extraña}$$

(24)

por tanto:

$$E(\hat{T}_2) = \frac{N}{n}E\left(\sum_{i=1}^{n} Y_i\right) = \frac{N}{n}E\left(\sum_{i=1}^{N} Y_i e_i\right) =$$

$$= \frac{N}{n}\sum_{i=1}^{N} Y_i E(e_i) = \frac{N}{n}\sum_{i=1}^{N^*} Y_i \frac{n}{N^*}$$

$$E(\hat{T}_2) = \frac{N}{N^*}T = \frac{1}{1-W}T \neq T$$

(25)

en consecuencia el sesgo, es:

$$B(\hat{T}_2) = T\left(\frac{1}{1-W} - 1\right) = \frac{W}{1-W}T$$

(26)

luego la varianza sería:

$$V(\hat{T}_2) = N^2(1-f^*)\frac{S^{*2}}{n}$$

(27)

donde $f^* = \frac{n}{N^*}$ y S^{*2} es la cuasivarianza similar al del marco depurado.

Teniendo en cuenta el resultado de (11):

$$V(\hat{T}_2) = \frac{1}{(1-W)^2}V(\hat{T}^*)$$

(28)

La relación anterior (28), prueba que la varianza de $\hat{T}_2$ es mayor que la de $\hat{T}^*$, puesto que cuando existen unidades extrañas se comprueba que $0 < (1-W)^2 < 1$.

De (27) y (28) se obtiene el error cuadrático medio de $\hat{T}_2$, que es igual a:

$$ECM\left(\hat{T}_2\right)=\frac{1}{(1-W)^2}\left[V\left(\hat{T}^*\right)+W^2T^{*2}\right]$$

6. Conclusiones y recomendaciones

- Si no se conoce el número de unidades extrañas incluidas en el marco, que es lo más probable, los estimadores son insesgados pero con varianza mayor al de un marco muestral con unidades depuradas. Si se identifican unidades extrañas seleccionadas en la muestra y se sustituyen por otras unidades no extrañas de un marco no depurado, entonces los estimadores son sesgados con varianza mayor al de un marco muestral depurado.

- Lo ideal es realizar actualizaciones de marcos de muestreo en periodos cortos de tiempo (año) antes de levantar una encuesta eliminando unidades extrañas. El costo y los recursos, impiden dicha labor, especialmente en los países subdesarrollados, tal es el caso nuestro. Una alternativa de solución al problema, es estimar la proporción de unidades extrañas en el marco ($\hat{W}$), mediante los listados de actualización de una muestra de unidades primarias de muestreo que arrojen un resultado estimativo y hagan que la varianza de los estimadores a obtenerse sea menor y así aumentar la precisión y disminuir el error de muestreo de los estimadores que se obtengan.

7. Bibliografía

[1] *Pandurang V. Sukhatme,* Teoría de Encuestas por Muestreo con Aplicaciones

[2] *Sharon L. Lohr,* Muestreo – Diseño y Análisis

[3] *Carl – Erik Sarndal,* Model Assisted Survey Sampling

[4] *Azorin Poch,* Curso de Muestreo y Aplicaciones

[5] *Des Raj,* La Estructura de las Encuestas por Muestreo

[6] *Leslie Kish,* Muestreo de Encuestas

Índice

"El tranquilo ha cambiado nuestro mundo, no tanto descubriendo nuevos hechos o desarrollos técnicos, sino cambiando los modos de razonar, de experimentar y de formar nuestras opiniones acerca de él."

Hacking

Regresión por Mínimos Cuadrados Parciales

Autor: Lic. Dindo Valdéz Blanco

1. Introducción

La regresión por mínimos cuadrados parciales, denominado regresión PLS (partial least squares), es una técnica que combina dos técnicas del análisis multivariante; el análisis de componentes principales y la regresión lineal múltiple.

La regresión PLS se utiliza generalmente en dos situaciones: cuando se tiene un gran número de variables predictoras, el número de variables independientes puede ser incluso mayor al número de observaciones, y/o cuando existe multicolinealidad entre las variables predictoras.

2. El Análisis Multivariante

Se considera el caso de un modelo lineal con una variable dependiente y m variables independientes, representados por la ecuación.

$$Y_{n\times 1} = X_{m\times n}\cdot \beta_{m\times 1} + E_{n\times 1} \qquad (1)$$

Donde Y representa la variable dependiente, X representa la matriz de variables independientes, β es el vector de coeficientes y E el vector de errores o residuos.

Respecto ala ecuación (1) se pueden considerar dos situaciones: Si $n > m$, entonces por lo general existe una única solución de mínimos cuadrados para la ecuación (1). Si $n < m$. Entonces no existe solución para la ecuación (1), en vista que la matriz X puede ser singular. O en su defecto existen infinitas soluciones de mínimos cuadrados.

3. El método de Mínimos Cuadrados Parciales (PLS)

La suposición básica de la regresión PLS es que el sistema depende de un número pequeño de variables instrumentales llamadas variables latentes. Este concepto es similar al de componentes principales. Las variables latentes son estimadas como combinaciones lineales de las variables observadas, como se explica más adelante. En los modelos PLS, se establece una representación de la matriz **X** en término de dichas variables latentes:

$$X_{n\times m} = \mathrm{T}_{\mathrm{n\times a}}\mathrm{P}^{\mathrm{T}}_{\mathrm{a\times n}} + E_{n\times m} \qquad (2)$$

donde **T** representa los "scores" (término que puede ser traducido como "resultados"); y la matriz **P** es denominada "loadings" (término que puede ser traducido como "cargas"). De esta manera la matriz **X** queda descompuesta en un número de "variables latentes", cada una caracterizada por un vector **t** y un vector $\mathbf{p}^{\mathrm{T}}$.

De esta forma, es posible representar la matriz **X** por una matriz **T** con un número menor de columnas. Esta descomposición se muestra en la ecuación (3).

$$\begin{aligned} X_{n\times m} &= \mathrm{T}_{\mathrm{n\times a}}\mathrm{P}^{\mathrm{T}}_{\mathrm{a\times n}} + E_{n\times m} \\ &= \mathrm{t(1)}_{\mathrm{n\times 1}}\cdot \mathrm{p(1)}^{\mathrm{T}}_{\mathrm{1\times n}} + \mathrm{t(2)}_{\mathrm{n\times 1}}\cdot \mathrm{p(2)}^{\mathrm{T}}_{\mathrm{1\times n}} + \cdots \\ &\cdots + \mathrm{t(a)}_{\mathrm{n\times 1}}\cdot \mathrm{p(a)}^{\mathrm{T}}_{\mathrm{1\times n}} + E_{n\times m} \quad ; \quad \mathrm{a < m} \end{aligned} \qquad (3)$$

Si se incluyen todas las variables latentes, el error es cero (**E** = **0**).

El modelo PLS se desarrolla de modo que las primeras variables latentes ($\mathbf{t}_{(1)}$, $\mathbf{t}_{(2)}$,...) sean las más importantes para explicar el vector **Y** en la muestra. El número de variables latentes necesarias para explicar la matriz **X** es una medida de la complejidad del modelo. Otros vectores calculados durante la etapa de construcción del modelo son el vector **w** (llamado "pesos" de **X**), y el vector **b** (denominado "sensibilidades"), La relación entre el vector **Y** y la matriz **T** es:

$$Y_{n\times 1} = \mathrm{T}_{n\times a}\,\mathrm{b}_{a\times 1} + F_{n\times m} \qquad (4)$$

Donde **b** se calcula para minimizar los errores ***F***. El vector **Y** es estimado usando los coeficientes de **b** previamente estimados por mínimos cuadrados:

$$\hat{Y}_{n\times 1} = \mathrm{T}_{n\times a}\,\mathrm{b}_{a\times 1} \qquad (5)$$

Si se toman en cuenta todas las variables latentes (a=m), los coeficientes del vector **b** son idénticos a los coeficientes del modelo de regresión lineal múltiple:

$$\mathrm{b}_{a\times 1} = \hat{\beta} = (X^T X)^{-1} X^T Y \qquad (6)$$

4. Ventajas y desventajas del método PLS

Como se ha indicado, el método PLS obtiene a partir de la matriz **X**, una matriz **T** cuyos vectores son linealmente independientes, definiendo un sistema ortogonal. De tal forma que, en los casos en que existan un número mayor de variables independientes en relación al número de observaciones (m>n) se produce una reducción del modelo. Por otro lado, en los casos en que exista colinealidad o redundancia entre las variables, la matriz **T** se usa para reducir dichas variables o sintetizarlas. Por consecuencia es posible minimizar el riesgo de cometer un error éstadístico al descartar información importante.

Una desventaja es que la regresión PLS es un modelo correlativo y no causal, en el sentido de que los modelos obtenidos no ofrecen información fundamental acerca del fenómeno estudiado, puesto que no se trabaja con las variables originales.

5. Regresión por Componentes Principales

La regresión por componentes principales consiste de dos etapas. Primero se realiza el análisis de componentes principales de la matriz de datos **X**, y luego se utilizan estos componentes principales como las variables independientes de la función de regresión final que se construye utilizando la técnica de mínimos cuadrados entre los datos proyectados y la variable respuesta **Y**. El hecho que los componentes principales son ortogonales resuelve el problema de multicolinealidad.

La desventaja de este método radica en que los componentes principales son calculados para explicar a **X** y no toman en cuenta a la variable dependiente, puesto que estas se calculan solo con la matriz de datos de **X**. Por lo tanto nada garantiza que los componentes principales los cuales "explican" **X**, también sean relevantes para explicar a **Y**.

6. Ejemplo de aplicación

Para ilustrar las diferencias entre la regresión PLS y la regresión por componentes principales utilizaremos datos simulados con 9 observaciones, 1 variable dependiente y 9 variables independientes. La matriz de correlaciones entre las variables se muestra a continuación:

Cuadro 1. Matriz de Correlaciones

	y	x1	x2	x3	x4	x5	X6	x7	x8	x9
y	1.00	0.27	-0.55	0.67*	-0.20	-0.08	-0.17	0.07	0.32	0.27
x1		1.00	-0.39	0.59	-0.67	-0.22	-0.03	0.42	0.39	1.00**
x2			1.00	-0.42	0.27	-0.12	0.14	0.11	-0.03	-0.39
x3				1.00	-0.21	0.24	0.10	0.47	0.52	0.59
x4					1.00	0.79*	0.61	0.30	0.14	-0.67*
x5						1.00	0.75*	0.54	0.41	-0.22
x6							1.00	0.67*	0.63	-0.03
x7								1.00	0.88**	0.42
x8									1.00	0.39
x9										1.00

*La correlación es significativa al nivel 0,05 (bilateral).
*La correlación es significativa al nivel 0,01 (bilateral).

Se observa la presencia de multicolinealidad entre las variables explicativas y al mismo tiempo se tienen pocas observaciones. Aplicando la técnica de componentes principales para reducir la matriz X, se tienen los siguientes resultados en el Cuadro 2.

Cuadro 2. Resultados del Método de Componentes Principales

Componente	Autovalores iniciales	% de la varianza	% acumulado	Autovalores seleccionados	% de la varianza	% acumulado
1	3.617	40.19	40.2	3.617	40.2	40.2
2	3.305	36.72	76.9	3.305	36.7	76.9
3	1.095	12.17	89.1	1.095	12.2	89.1
4	0.468	5.19	94.3			
5	0.280	3.11	97.4			
6	0.185	2.06	99.4			
7	0.051	0.56	100.0			
8	0.000	0.00	100.0			
9	0.000	0.00	100.0			

Por tal razón se eligen los primeros 3 componentes y se aplica la regresión lineal múltiple para explicar a la variable respuesta en función de estas tres variables sintéticas, el cuadro siguiente muestra los resultados de la regresión múltiple con los tres componentes:

Cuadro 3. Resultados del modelo regresión lineal por componentes principales

Modelo	Coeficientes B	Error típico	T calculado	Significancia
(Constante)	8.9207	10.618	0.8402	0.4391
c1	47.668	27.825	1.7132	0.1474
c2	-22.88	19.172	-1.193	0.2863
c3	9.7288	15.757	0.6174	0.564

ANOVA					
Fuente	Suma de cuadrados	Grados de Libertad	Media cuadrática	F	Sig.
Regresión	561.9	3	187.3	0.9983	0.4655
Residual	938.1	5	187.62		
Total	1500	8			

R	R cuadrado	R cuadrado corregida	Error típico de la estimación
0.612	0.3746	-6E-04	13.697

Ahora se aplica la técnica de regresión PLS en el paquete estadístico Minitab 15.0 para windows, para determinar el número de componentes se utiliza la técnica de la validación cruzada (crossvalidation), llegando a resumir las 9 variables independientes en un solo componente artificial, los resultados de la regresión PLS se dan en el siguiente cuadro.

Cuadro 4. Resultados del modelo regresión PLS

Modelo	Coeficientes B	Error típico	T calculado	Significancia
(Constante)	25.00	3.76	6.65	0.0003
t1	5.35	2.45	2.19	0.0649

ANOVA					
Fuente	Suma de cuadrados	Grados de Libertad	Media cuadrática	F	Sig.
Regresión	609.06	1	609.06	4.7853	0.0649
Residual	890.94	7	127.28		
Total	1500	8			

R	R cuadrado	R cuadrado corregida	Error típico de la estimación
0.6372	0.406	0.3212	11.282

7. Conclusiones

En conclusión se observa que la regresión PLS brinda un mejor ajuste, y en comparación con la regresión por componentes principales la regresión PLS en este caso ha sintetizado la matriz X en una sola componente a diferencia de los tres componentes sintéticos de la regresión por componentes principales.

8. Bibliografía

[1] *Mardia, K.V.* (1997). "Análisis multivariante", Academic Press, London.

[2] *M. Barker*. (2003). "Partial least squares". Revista de Quimiometría, 17:166–173.

[3] *Vega Carmen* (2008). "Regresión por Mínimos Cuadrados Parciales con Aplicación en Regresión Logística". Tesis UMSA, Carrera de Estadística.

Índice

Un conejo estaba sentado delante de una cueva escribiendo, cuando aparece un zorro.

- Hola, conejo, que haces?

- Estoy redactando mi tesis doctoral sobre como los conejos comen zorros.

- Ja, ja, pero que dices?

- No te lo crees?. Anda, ven conmigo dentro de la cueva...

Entran los dos y al cabo de un ratito sale el conejo con la calavera del zorro y se pone a escribir. Al cabo de un rato llega un lobo.

- Hola, conejo, que haces?

- Estoy escribiendo mi tesis sobre como los conejos comen zorros y lobos.

- Ja, ja, que bueno, que chiste más divertido!

- Que no te lo crees?. Anda, ven dentro de la cueva, que te voy a enseñar algo!

Al cabo de un rato sale el conejo con la calavera de lobo, y empieza otra vez a escribir. Después llega un oso.

- Hola, conejo, que estás haciendo?

- Estoy acabando de escribir mi tesis sobre como los conejos comen zorros, lobos y osos.

- No te lo crees ni tú

- Bueno, a que no te metes en la cueva conmigo?

De nuevo se meten los dos en la cueva, y como era de esperar, un león enorme se tira encima del oso y se lo come. El conejo recoge la calavera del oso, sale fuera y acaba su tesis.

Moraleja: *Lo más importante no es el contenido de tu tesis, sino tu asesor.*

Métodos de Predicción en Situaciones Límite

Autor: Univ. José Alberto Flores Aguilar

1. Pasado, presente y futuro sobre las limitaciones para relacionar variables

El problema del ajuste de un conjunto de puntos representados en un sistema de ejes coordenados, por una recta o mas generalmente por una curva, era objeto de estudio desde mediados del siglo XVIII (Leonhard Euler, 1749; Johan Tobías Mayer, 1750). Sin embargo la primera mención al *método de mínimos cuadrados*, fue atribuida a Adrián-Marie Legendre (1805). En dicho estudio, se consideró este método como: "el más adecuado para relacionar variables de forma lineal" señalándose, además la conveniencia de la eliminación de individuos atípicos para optimizar el establecimiento de dichas interrelaciones.

Por último, merece la pena destacar la introducción de mínimos cuadrados, realizado por Robert Adrián (1808), quien aportó un punto de vista de gran interés, complementario al de los trabajos realizados por sus antecesoras. Sin embargo, dicho método no pudo ser justificado hasta la llegada de la ley de Laplace-Gauss, "bautizada" por Kart Pearson, a finales del XIX (1893), como "ley normal".

Se ha constatado en numerosas ocasiones que la presencia de la multicolinealidad conlleva a situaciones de "inestabilidad" de los coeficientes de regresión y que estos pueden ser "no significativos".

Cuando las variables explicativas están muy correlacionadas con la variable a explicar, produciendo dificultades de interpretación de la ecuación de regresión linéal a causa de signos erráticos en los coeficientes de regresión. Por ello, la aplicación del método de "mínimos cuadrados" conduce a resultados en ocasiones poco comprensibles para los investigadores que se dedican a las ciencias experimentales.

Es interesante no solo detectar al multicolinealidad (Belsley, Kuh y Welsh, 1980), sino también tomar medidas para atenuarla, sin embargo, la ecuación de predicción lineal bajo estas medidas sigue siendo desgraciadamente en ocasiones poco comprensible para el investigador:

Otras dos situaciones límite son:

1) Número menor de individuos que variables y
2) Datos ausentes.

En cuanto a la primera situación, que contempla menos individuos que variables, conlleva sistemáticamente a que el determinante de la matriz X^tX -que hay que resolver para la obtención de los coeficientes de regresión- "sea nulo" y, por tanto, no haya modo de encontrar tales coeficientes.

De todos estos resultados concluimos que: el método de mínimos cuadrados -intensamente para relacionar variables- no funciona bien en las situaciones límite tales como:

- La multicolinealidad,
- Menor número de individuos que variables y
- Datos ausentes.

Es aconsejable sustituirlas, en esas circunstancias, por el *método de mínimos cuadrados parciales* (PLS)

2. Una reflexión en cuanto a la normalización de los datos

Vamos a presentar dos tipos de normalización de datos que se encuentran con frecuencia en las referencias bibliografías y en especial en Audrain, Lesquoy-de-Turckeim, Miller y Tomassone, 1992, Pág. 179-180. El primero consiste en restar para cada una de las variables su media, y dividir por la raíz cuadrada de la suma de las desviaciones a su media.

$$y_i^{[1]} = \frac{y_i - \bar{y}}{\sqrt{\sum_{i\to 1}^{n}(y_i - \bar{y})^2}} \quad (i = 1,2,....,n)$$

$$x_{i,j}^{[1]} = \frac{x_{i,j} - \bar{x}_j}{\sqrt{\sum_{i\to 1}^{n}(x_{i,j} - \bar{x}_j)^2}} \quad (i = 1,2,...,n)\,(j = 1,2,..,p)$$

El segundo consiste en restar para cada una de las variables su media y, dividir por la raíz cuadrada de la suma de cuadrados de las desviaciones a su media por (n-1).

$$y_i^{[2]} = \frac{y_i - \bar{y}}{\sqrt{\frac{\sum_{i\to 1}^{n}(y_i - \bar{y})^2}{n-1}}} \quad (i = 1,2,....,n)$$

$$x_{i,j}{}^{2} = \frac{x_{i,j} - \bar{x}_j}{\sqrt{\frac{\sum_{i\to 1}^{n}(x_{i,j} - \bar{x}_j)^2}{n-1}}} \quad (i = 1,2,...,n)\,(j = 1,2,.....,p)$$

3. Cálculo de coeficientes de regresión para ambas normalizaciones

Aunque las operaciones intermedias que hay que realizar para llegar a los coeficientes de regresión difieran del tipo de normalización de los datos, los coeficientes de regresión asociados a las variables

$$x_1^{[1]}, x_2^{[1]}, ..., x_p^{[1]}$$

Como resultado de la primera normalización y a las variables.

$$x_1^{[2]}, x_2^{[2]}, ..., x_p^{[2]}$$

Como resultado de la segunda normalización, son los mismos.

- En la primera normalización de los datos las operaciones son las siguientes:

$$\sum_{i=1}^{n}\left(y_i^{[1]}\right)^2 = 1$$

$$\sum_{i=1}^{n}\left(x_{i,j}^{[1]}\right)^2 = 1 \qquad j = 1,2,....,p$$

$$\sum_{i=1}^{n}\left(y_i^{[1]} x_{i,j}^{[1]}\right) = r_{y,x_j} \qquad j = 1,2,....,p$$

$$\sum_{i=1}^{n}\left(x_{i,j}^{[1]} x_{i,j'}^{[1]}\right) = r_{x_j,x_{i,j'}} \quad (j = 1,2,....,p; j' \neq j)$$

Donde r_{y,x_j} representa la correlación muestral entre x_j e y.

- En la segunda normalización de los datos las operaciones son las siguientes:

$$\sum_{i=1}^{n}\left(y_i^{[2]}\right)^2 = n-1$$

$$\sum_{i=1}^{n}\left(x_{i,j}^{[2]}\right)^2 = n-1 \qquad j = 1,2,....,p$$

$$\sum_{i=1}^{n}\left(y_i^{[2]} x_{i,j}^{[2]}\right) = (n-1) r_{y,x_j} \qquad j = 1,2,....,p$$

$$\sum_{i=1}^{n}\left(x_{i,j}^{[2]} x_{i,j'}{}^{2}\right) = (n-1) r_{x_j,x_{i,j'}} \quad (j = 1,2,....,p; j' \neq j)$$

Tanto en la primera como en la segunda normalización de los datos los coeficientes de regresión afectados, tanto por unas como por otras variables, son los mismos.

Partimos de la expresión que nos permite calcular los coeficientes de regresión.

Para la primera normalización:

$$\hat{\beta}^{[1]} = (X^{[1]\prime} X^{[1]})^{-1} X^{[1]\prime} y^{[1]}$$

Dado que $X^{[1]\prime} X^{[1]} = R$, matraz de correlación de las variables explicativas.

$$X^{[1]\prime} Y^{[1]} = \begin{pmatrix} r_{y,x_1} \\ r_{y,x_2} \\ . \\ . \\ r_{y,x_p} \end{pmatrix}$$

La fórmula para l cálculo de los coeficientes de regresión es la que a continuación mostramos.

$$\hat{\beta}^{[1]} = R^{-1} \begin{pmatrix} r_{y,x_1} \\ r_{y,x_2} \\ . \\ . \\ r_{y,x_p} \end{pmatrix}$$

Para la segunda normalización:

$$\hat{\beta}^{[2]} = (X^{[2]\prime} X^{[2]})^{-1} X^{[2]\prime} y^{[2]}$$

Dado que,

$$X^{[2]\prime} X^{[2]} = (n-1)R \qquad \text{y}$$

$$X^{[2]\prime} Y^{[2]} = (n-1) \begin{pmatrix} r_{y,x_1} \\ r_{y,x_2} \\ . \\ . \\ r_{y,x_p} \end{pmatrix}$$

La fórmula para el cálculo de los coeficientes de regresión es la que a continuación mostramos.

$$\hat{\beta}^{[2]} = R^{-1} \begin{pmatrix} r_{y,x_1} \\ r_{y,x_2} \\ . \\ . \\ r_{y,x_p} \end{pmatrix}$$

De lo que concluimos que los coeficientes de regresión son invariantes con respecto al tipo de normalización.

$$\hat{\beta}^{[1]} = \hat{\beta}^{[2]}$$

4. Ecuaciones de predicción lineal

La ecuación de predicción lineal en función de las variables normalizadas por primer caso:

$$y^{[1]^*} = \sum_{j=1}^{p} \hat{\beta}_j^{[1]^*} x_j^{[1]}$$

Deshaciendo el cambio, llegamos a la ecuación de predicción en función de las variables originales.

$$\frac{y^* - \bar{y}}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \sum_{j=1}^{p} \hat{\beta}_j^{[1]^*} \left(\frac{(x_{ij} - \bar{x}_j)}{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \right) \Rightarrow$$

$$y^* = \left(\bar{y} - \sum_{j=1}^{p} \sqrt{\frac{SCD_y}{SCD_{x_j}}}\, \hat{\beta}_j^{[1]^*} \bar{x}_j \right) + \sum_{j=1}^{p} \sqrt{\frac{SCD_y}{SCD_{x_j}}}\, \hat{\beta}_j^{[1]^*} x_j$$

donde

$$SCD_y = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \text{y}$$

$$SCD_{x_j} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

La ecuación de predicción lineal en función de las variables normalizadas por el segundo caso:

$$y^{[1]^*} = \sum_{j=1}^{p} \hat{\beta}_j^{[1]^*} x_j^{[1]}$$

Deshaciendo el cambio, llegamos a la ecuación de predicción en función de las variables originales.

$$y^* = \left(\bar{y} - \sum_{j=1}^{p} \sqrt{\frac{SCD_y}{SCD_{x_j}}}\, \hat{\beta}_j^{[2]^*} \bar{x}_j \right) + \sum_{j=1}^{p} \sqrt{\frac{SCD_y}{SCD_{x_j}}}\, \hat{\beta}_j^{[2]^*} x_j$$

Dado que: $\hat{\beta}^{[1]^*} = \hat{\beta}^{[2]^*} \quad (j = 1,2,\ldots\ldots, p)$ se concluye que el tipo de normalización no influye en la ecuación de predicción lineal.

5. Bibliografía

Es bastante difícil encontrar bibliografía sobre este tema. De hecho no existe ninguna publicación en español; generalmente aparecen resultados dentro del campo de la química, por ser precisos en el campo de la cromatografía y espectrofotometría.

Incluyo los más fundamentales al ser realizados por los creadores del método.

[1] *Bry, X* (1996): Analices factorielles multiples, Paris, Ed.Economia.

En este libro no solo se ve la manera didáctica de la regresión PLS, sino también la relación que existe entre la regresión PLS y el análisis de componentes principales.

[2] *Hoskuldsson, A.* (1988): PLS regresión methods. Journal of chemometrics.

En este artículo se desarrolla la estructura matemática y estadistica de la regresión PLS. Es algo difícil de leer.

[3] *Tenenhaus, M.* (1998): La regresión PLS. Theorie et pratique, Paris, Editions Techip.

Este libro contiene una introducción a las técnicas que proporcionaron la regresión PLS como el análisis canónico utilizado por la regresión PLS, tanto para datos cuantitativos (PLS1 y PLS2) como para datos cualitativos.

[4] *R. Martínez Arias*: El análisis Multivariante en la investigación científica.

[5] *J. Etxeberria*: Regresión múltiple.

"Tengo mis resultados hace tiempo, pero no sé cómo llegar a ellos"

C. F. Gauss

Índice

Encuesta de Intención de Voto Docente Estudiantil para las Elecciones de Rector y Vicerrector de la Universidad Mayor de San Andrés, gestión 2010

Proyecto de Investigación IETA

En Asamblea Docente – Estudiantil de la Carrera de Estadística reunida en fecha 12 de febrero de 2010, se decidió levantar una encuesta por muestreo sobre la Percepción de Voto de Estudiantes y Docentes de la Universidad Mayor de San Andrés para las elecciones de Rector y Vicerrector, gestión 2010 – 2013, como anticipo a las elecciones para Rector y Vicerrector de la UMSA. Este trabajo fue encargado al Instituto de Estadística Teórica y Aplicada (IETA), dependiente de la Carrera.

El Instituto realizó la planificación, ejecución y análisis de la encuesta logrando resultados precisos y confiables del porcentaje de votación, semana y media antes de las elecciones previstas.

Los resultados de la elección, confirmaron que la encuesta proporcionó buenos resultados, demostrando que la investigación fue realizada con responsabilidad profesional a cargo de docentes y estudiantes de la Carrera de Estadística y la participación de estudiantes de la Carrera de Trabajo Social de la UMSA.

Las elecciones para Rector y Vicerrector se llevaron a cabo la primera semana de abril del presente año, con la presentación de cuatro frentes.

Objetivo

El objetivo de la investigación fue dar resultados porcentuales de la preferencia docente y estudiantil, sobre alguno de los frentes participantes, y generar información de solicitud de necesidades y requerimientos de docentes y estudiantes hacia las futuras autoridades de nuestra principal casa de estudios.

El trabajo realizado por docentes y estudiantes tenía carácter investigativo sin la intención de perjudicar a los frentes participantes en las elecciones.

Sin embargo, la información de la encuesta permitió a los frentes, mejorar sus acciones y propuestas, analizar las necesidades que tiene el estamento docente – estudiantil y reflexionar en sus decisiones como autoridades en bien de la Universidad.

Metodología

La población objeto de estudio, fueron docentes y estudiantes de las 13 facultades de la Universidad Mayor de San Andrés además los docentes del CIDES[1].

El tipo de muestreo empleado fue estratificado por niveles de estudio en estudiantes, es decir, se ha considerado en la muestra estudiantes de distintos años o semestres. El Marco Muestral fue construido a partir de la información de registros de estudiantes matriculados por carreras y número de docentes por Facultad de la última gestión 2009 del Documento Datos Estadísticos de la Población Universitaria 1995-2009, la UMSA en cifras.

[1] Centro de Postgrado en Ciencias del Desarrollo

Asimismo se recopiló información de aulas, paralelos y horarios de clases ajustados a los días que se levantó la encuesta, de todas las carreras de la Universidad. De esta manera las aulas con estudiantes en clases, en distintos horarios, fueron seleccionadas aleatoriamente, con el fin de cuidar la representatividad de la muestra.

El tamaño de la muestra de estudiantes y docentes fue de 4.000 y 400, respectivamente. Sin embargo en el trabajo de campo se llegó a encuestar efectivamente a 3.874 estudiantes y 308 docentes, la disminución de la muestra se debe a incidencias de rechazo a la encuesta.

En la recopilación de información participaron 13 docentes encargados de cada Facultad, 111 estudiantes encuestadores, 29 estudiantes en crítica y codificación y 26 estudiantes en transcripción de datos.

Culminada la encuesta a estudiantes y docentes de las distintas facultades, se realizó la crítica, codificación y transcripción, posteriormente fueron remitidas para la validación y análisis de datos.

Resultados de la Encuesta de Percepción Docente–Estudiantil vs. Resultados de las Elecciones

Los cuadros y gráficos siguientes describen los principales resultados de la encuesta a nivel global, efectuada los días miércoles 17 y jueves 18 de marzo, comparados con los resultados oficiales de las elecciones realizadas a principios de abril.

El cuadro N° 1 muestra los resultados de la encuesta, las elecciones oficiales y la diferencia de porcentaje observado. Tanto en la encuesta como en los resultados oficiales, se encuentra en primer lugar el frente *VIA-U* y en segundo lugar el frente *Revolución*. Lo interesante de los resultados es que se mantiene la misma estructura del voto, independientemente de los resultados porcentuales en ambos estamentos y voto ponderado, excepto en los votos en blanco o nulo, esto debido a la indecisión o desconocimiento de las elecciones o frentes participantes.

Algunos docentes y estudiantes no quisieron dar información, argumentando que el voto es secreto pese a que la encuesta era anónima.

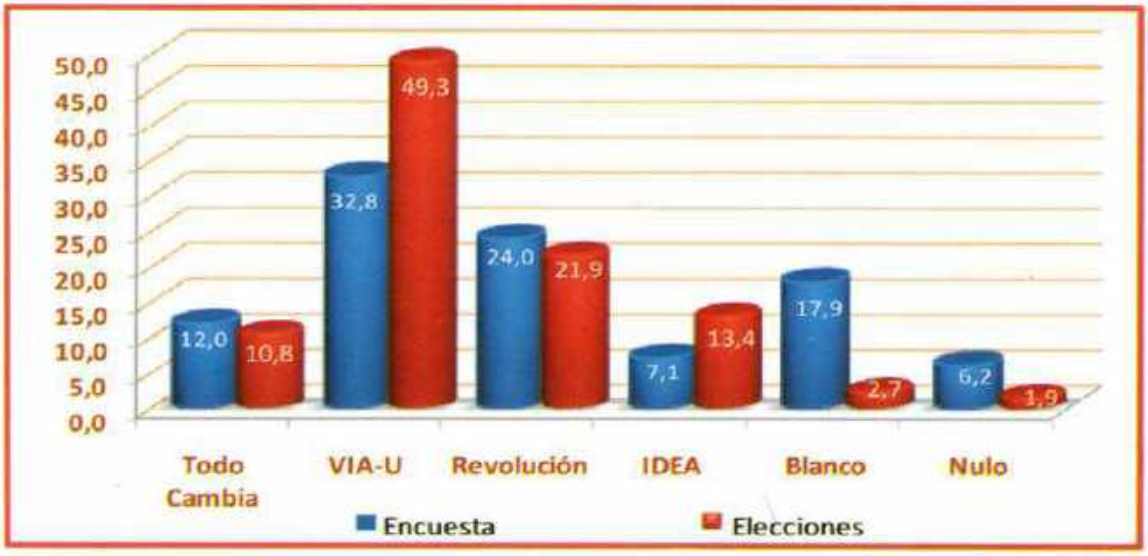
Cuadro Nº 1
Encuesta vs. Elecciones

FRENTE	DOCENTES			ESTUDIANTES			PONDERADO		
	Encuesta	Elecciones	Diferencia	Encuesta	Elecciones	Diferencia	Encuesta	Elecciones	Diferencia
Todo Cambia	12,0	10,8	1,2	15,6	20,1	-4,5	13,8	16,2	-2,4
VIA-U	32,8	49,3	-16,5	20,2	24,9	-4,7	26,5	38,3	-11,8
Revolución	24,0	21,9	2,1	22,0	33,5	-11,5	23,0	29,0	-6,0
IDEA	7,1	13,4	-6,3	8,0	14,1	-6,1	7,6	14,3	-6,7
Blanco	17,9	2,7	15,2	20,8	1,6	19,2	19,3	2,2	17,1
Nulo	6,2	1,9	4,3	13,4	5,8	7,6	9,8	0,0	9,8
Total	100,0	100,0		100,0	100,0		100,0	100,0	

Fuente: Carrera de Estadística UMSA – Instituto de EstadísticaTeórica y Aplicada IETA

La Gráfica Nº 1 presenta los resultados de la Encuesta de Percepción de Voto de Docentes, comparativo a los resultados oficiales de las elecciones. Los porcentajes para cada frente, son parecidos excepto al de *VIA-U* (32.8% contra 49.3%) y los votos en blanco y nulo. Nuevamente la estructura del voto de los cuatro candidatos, es similar en la encuesta y los datos oficiales.

Gráfica Nº 1
Encuesta vs. Elecciones
DOCENTES



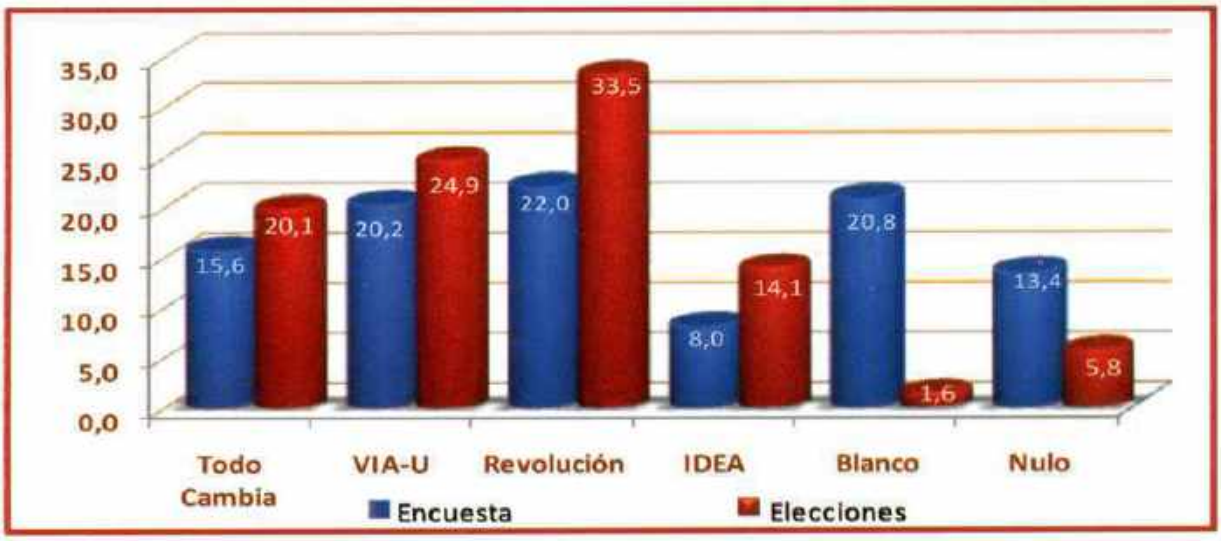
Fuente: Carrera de Estadística UMSA – Instituto de EstadísticaTeórica y Aplicada IETA

La Gráfica Nº 2, presenta los resultados comparativos en estudiantes. El frente ganador según encuesta es *Revolución* con 22,0 % y según los datos oficiales también es el frente Revolución con el 33.5%. El frente *VIA-U* alcanzó un

segundo lugar con un 20,2% en la encuesta y 24,9% en las elecciones oficiales. El porcentaje de voto en "blanco" alcanza a un 20,8% y el voto "nulo" de 13,4%. La suma de ambos porcentajes es de 34,2% con diferencia de 7.4% de los datos oficiales.

Los resultados en estudiantes presentan una estructura mucho más parecida a los de docentes entre encuesta y elecciones, sin contar los votos blanco y nulo.

Gráfica N° 2
Encuesta vs. Elecciones



Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

La Gráfica N° 3 muestra el comportamiento porcentual de la votación ponderada entre ambos estamentos, donde nos permite observar una gran similitud estructural de los datos oficiales con la encuesta realizada.

El voto en blanco y nulo difieren puesto bastante y esto es debido a que muchos de los frentes en el periodo de encuesta, todavía no eran conocidos y los electores estaban indecisos de su voto.

En conclusión, de acuerdo a la encuesta y los datos oficiales, la Carrera de Estadística en base a este estudio, se anticipó a los resultados adecuadamente.

Gráfica N° 3
Encuesta vs. Elecciones

Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

Encargados de la Investigación

- Lic. Raúl Delgado Alvarez **DIRECTOR DE CARRERA**
- Lic. Fernando Rivero Suguiura **DIRECTOR DEL INSTITUTO DE ESTADISTICA TEORICA Y APLICADA (IETA)**
- Lic. Jaime Pinto Ajuacho **DIRECTOR ACADEMICO CARRERA DE ESTADISTICA**
- Lic. Amilcar Miranda G. **DOCENTE INVESTIGADOR**
- Lic. Ivan Marquez C. **DOCENTE INVESTIGADOR**
- Univ. Consuelo Barrios G. **ESTUDIANTE INVESTIGADOR**
- Univ. Mercedes Heredia C. **ESTUDIANTE INVESTIGADOR**
- Univ. Beatriz Vallejos M. **ESTUDIANTE INVESTIGADOR**
- Univ. José Manuel Calderón **ESTUDIANTE INVESTIGADOR**
- Egr. Zulema Vargas C. **ASISTENTE DE INFORMATICA**
- Docentes y Estudiantes de la Carrera de Estadística–Estudiantes de la Carrera de Trabajo Social

"En estadística, lo que desaparece detrás de los números es la muerte."

Günter Grass

Índice

Encuesta Sobre las Causas de la Migración en las Ciudades de La Paz y El Alto gestión 2009

Proyecto de Investigación IETA

La Carrera de Estadística mediante el Instituto de Estadística Teórica y Aplicada (IETA) puso en marcha la Encuesta sobre Causas de la Migración en las Ciudades de La Paz y El Alto, con la finalidad de identificar las diferentes problemáticas, causas y consecuencias que trae consigo la migración dentro y fuera de nuestro país, donde actualmente Bolivia es uno de los países con una tasa migratoria alta en Latinoamérica.

Objetivo

El objetivo de la investigación es el de determinar las causas por las cuales las personas de las ciudades de La Paz y El Alto migran. Conocer si el cambio de residencia de un lugar a otro, interno y externo del país, mejoran o empeoran las condiciones de vida de las personas y los miembros del hogar.

Metodología

Se ha aplicado una encuesta por muestreo en las ciudades de La Paz y El Alto en una muestra aleatoria de 1113 viviendas u hogares (552 hogares en la ciudad de La Paz y 561 en la ciudad de El Alto).

Han participado como supervisores y encuestadores en el trabajo de campo aproximadamente 120 estudiantes de las Carreras de Estadística y Trabajo Social de la UMSA y 9 docentes de la Carrera de Estadística. El proyecto se realizó desde el mes de junio a diciembre del año 2009.

En la ciudad de La Paz

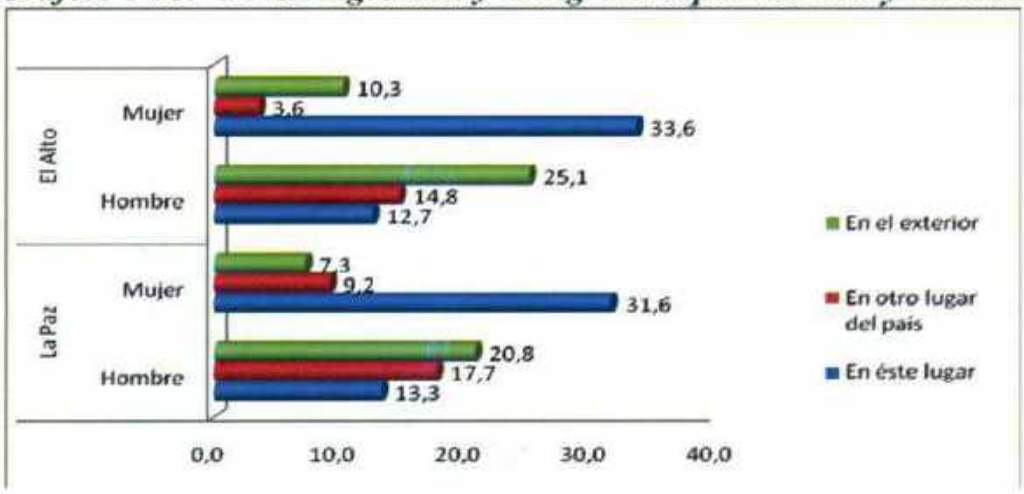
De cada 3 personas, uno se va de su ciudad, y por cada 4 personas que viven fuera de esta ciudad, uno llega a la ciudad, es decir, de cada 100 personas que se van o emigran de la ciudad, 75 personas llegan de otros lugares.

En la ciudad de El Alto

1 de cada 4 personas es emigrante, y 1 de cada 3 personas es inmigrante, de otra forma, 115 personas que provienen de diferentes lugares, llegan a la ciudad de El Alto, de 100 personas que emigran.

Resultados

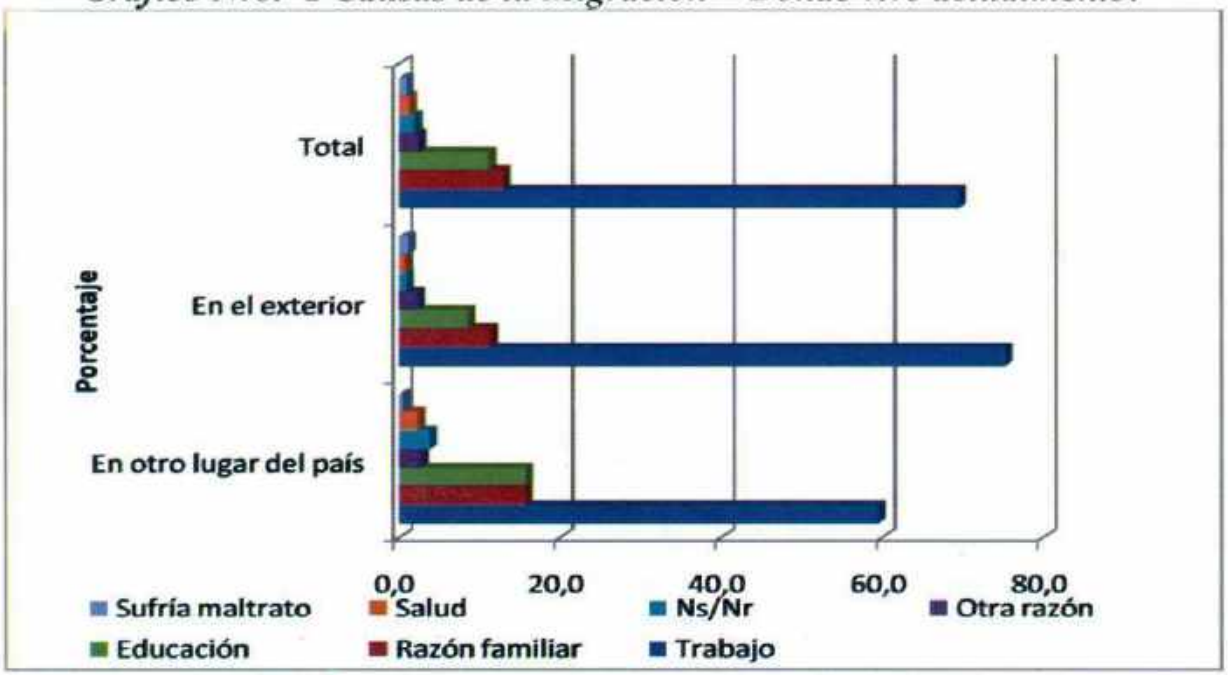
Gráfico Nro. 1 No Migrantes y Emigrantes por Género y Ciudad



Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

De acuerdo al gráfico N°1, las mujeres son las que menos migran en relación a los hombres, tanto al exterior como a otro lugar del país. Es de esperar ya que los hombres se ven obligados a migrar para mantener a sus familias y la mayoría de las mujeres se quedan al cuidado de sus hijos.

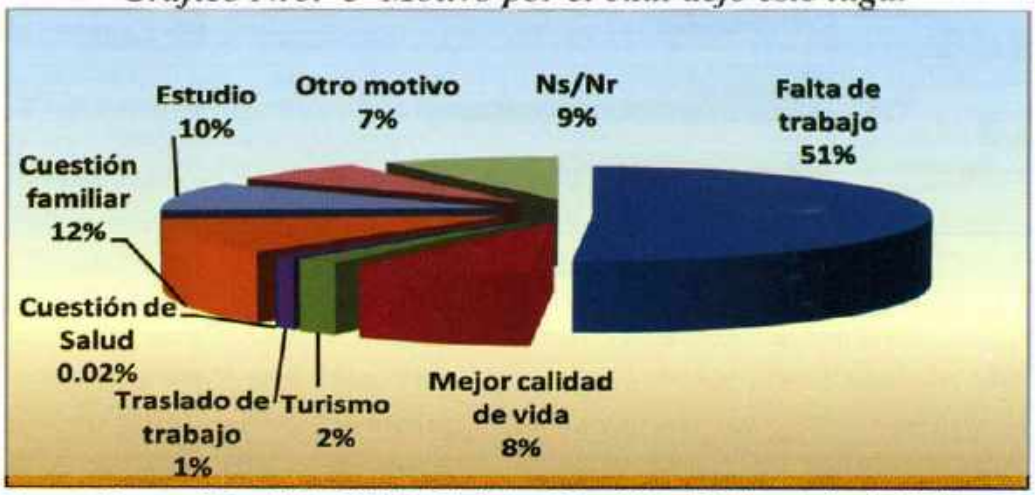
Gráfico Nro. 2 Causas de la Migración – Dónde vive actualmente?



Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

El gráfico N°2 muestra las causas por las cuales la población de La Paz y El Alto migran. La mayoría lo hace por trabajo, más al exterior (75%) que a otro lugar del país (58%). Las causas de la migración por razón familiar y por educación, son también importantes y están por encima del 10%. Las otras causas no son significativas (menor del 5%).

Gráfico Nro. 3 Motivo por el cual dejó este lugar



Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

De acuerdo al gráfico N°3, la mayoría de las personas de ambas ciudades migran por falta de trabajo (51 %), luego por cuestión familiar (12%), por estudio el 10% y mejor calidad de vida que seguramente tiene que ver con el trabajo también, en 8%.

Gráfico Nro. 4 Hogares de Personas Emigrantes Niños y Adolescentes separados de sus Padres, entre 0 y 17 años

Fuente: Carrera de Estadística UMSA – Instituto de Estadística Teórica y Aplicada IETA

El Gráfico Nro. 4, muestra que el 15.9% de los hogares de emigrantes, presentan niños y adolescentes entre 0 y 17 años separados de sus padres tanto en otro lugar del país como en el exterior.

Calidad de Vida de los Migrantes

Cuadro Nro. 2 ¿El nivel de vida en su hogar ha cambiado? (Ciudad de La Paz) y (Ciudad de El Alto)

Ciudad	Opción	%
LA PAZ	Si	60.4
	No	39.6
Total		**100.0**

UMSA – Instituto de EstadísticaTeórica y Aplicada IETA

Ciudad	Opción	%
EL ALTO	Si	63.3
	No	36.7
Total		**100.0**

El Cuadro Nro. 2, de la izquierda muestra a los hogares migrantes en la ciudad de La Paz, el 60.4% afirman que cambió el nivel de vida en su hogar y el 39.6% afirman que nada cambió desde que al menos uno de sus miembros ha migrado.

En el cuadro de la derecha se muestra a los hogares migrantes en la ciudad de El Alto, el 63.3% afirman que existen cambios y el 36.7% indican que no existe ningún cambio.

Conclusiones

Las causas de la migración, son:

Culturales: La base cultural de nuestra población determinada es un factor muy importante a la hora de decidir a qué país o lugar va a emigrar (influye el idioma, la forma de vida, etc.).

Socioeconómicas: Existe una relación directa entre desarrollo socioeconómico y la migración, por ende, entre subdesarrollo y emigración. La mayor parte de los migrantes bolivianos lo hacen por motivos económicos, buscando un mejor nivel de vida.

Familiares: Los vínculos familiares también resultan ser un factor importante en la decisión de migrar. En su mayoría, el lazo familiar se rompe a consecuencia de la migración. Lo preocupante es que a consecuencia de la migración se desunen las familias y dejan a los hijos separados de sus padres.

- **En el caso de la emigración**:
  Constituyen consecuencias *positivas*, el alivio de algunos problemas de superpoblación como es el caso de la ciudad de La Paz. La disminución de la presión sobre los recursos, la inversión de las remesas de dinero que envían los emigrantes, la disminución del desempleo, el aumento de la productividad, etc.

  Constituyen consecuencias *negativas*: la desvitalización; el envejecimiento de la población (los que emigran suelen ser jóvenes), la población que queda se hace más tradicionalista y más reacia al cambio. Suelen irse las personas más productivas y con mayor afán de superación.

- **En el caso de la inmigración**:

  Constituyen consecuencias *positivas*: el rejuvenecimiento de la población, la población está más dispuesta a los cambios, aportes de capital y de mano de obra, aportes de nuevas técnicas, llegan personas capacitadas y/o preparadas sin que el país haya invertido en su preparación.

  Constituyen consecuencias *negativas*: aparecen desequilibrios en cuanto a la estructura por edad y sexo, se introducen una mayor diversidad cultural, política, lingüística, religiosa, llegando a formar grupos completamente segregados y marginales.

Encargados de la Investigación

- Lic. Raúl Delgado Alvarez
  DIRECTOR DE CARRERA
- Lic. Fernando Rivero Suguiura
  DIRECTOR DEL INSTITUTO DE ESTADISTICA TEORICA Y APLICADA (IETA)
- Lic. Jaime Pinto Ajuacho
  DOCENTE INVESTIGADOR
- Lic. Lucy Cuarita
  DOCENTE INVESTIGADORA
- Egr. Zulema Vargas C.
  ASISTENTE DE INFORMATICA
- Univ. Ivan Aliaga C.
  ESTUDIANTE INVESTIGADOR
- Univ. Consuelo Barrios G.
  ESTUDIANTE INVESTIGADORA
- Univ. Beatriz Vallejos M.
  ESTUDIANTE INVESTIGADORA
- Univ. Graciela Heredia C.
  ESTUDIANTE INVESTIGADORA
- Docentes y Estudiantes de la Carrera de Estadística–Estudiantes de la Carrera de Trabajo Social.

Índice

Cuando las estadísticas nos dicen que la familia mexicana tiene en promedio cuatro hijos y medio, nos explicamos por qué siempre hay uno chaparrito."

Marco Aurelio Almazán

Sistema de Indicadores Sociales y Económicos "S.I.S.E."

Proyecto de Investigación IETA

El proyecto denominado Sistema de Indicadores Sociales y Económicos (SISE) ha sido creado en una reunión del 10 de abril de 2010, por parte de un grupo de docentes de la Carrera de Estadística, con la inquietud de promover el conocimiento práctico y científico en el manejo de indicadores sociales y económicos en favor de los estudiantes de la Carrera.

El objetivo fue conformar un taller de discusión entre docentes y estudiantes en la comprensión y utilización de las diferentes metodologías disponibles para el cálculo de indicadores en estos rubros, de tal manera que se puedan desarrollar sesiones durante la semana, para la preparación de la información estadística y la interpretación metodológica de un conjunto de indicadores, que permita al grupo participante, calcular e interpretar dichos indicadores, comprender el sentido que tienen y llegar, así, a la conclusión de cómo se manejan las estadísticas en nuestro país.

Paralelamente, utilizar un software estadístico que permita sistematizar estos indicadores y que en el avance del proyecto, se convierta en un importante sistema de indicadores puestos en servicio de la sociedad boliviana.

El Instituto de Estadística Teórica y Aplicada (IETA) dependiente de la Carrera de Estadística de la UMSA, ha invitado a estudiantes de la Carrera de últimos cursos a participar en el proyecto. Aproximadamente 25 estudiantes se inscribieron y a partir del 1ro. de mayo el SISE comenzó a funcionar.

Los temas de inició, fueron referidos a la parte económica, se han estudiado las Cuentas Nacionales, el Índice de Precios (IPC), Exportaciones e Importaciones de productos y la Inflación.

Complementariamente, se ha explicado la metodología de las estadísticas sociales y se incursionó brevemente en Indicadores de Empleo.

El proyecto tiene la finalidad de continuar investigando diferentes áreas del campo social y económico (salud, educación, demografía, agropecuaria, pobreza, empleo, vivienda, etc.) de tal manera que los estudiantes complementen sus estudios académicos con la parte práctica, y así cuando lleguen a ser profesionales, puedan defenderse en el campo laboral.

Índice

Historia de la Estadística

Los primeros instantes...

Hasta el comienzo del siglo XIX no se utilizaba la palabra Estadística en el sentido que le damos hoy día. En su origen, el término Estadística aparecía ligado a la actividad gubernamental, y el de estadístico con el de estadista o político. Ello se debía principalmente a que el primer y fundamental uso de la Estadística era empleado por los gobernantes que deseaban conocer la extensión de sus dominios, la población residente en ellos y la cantidad de impuestos que se podían obtener de la población.

Hasta 1850 la palabra "Estadística" se usaba en un sentido diferente del actual, significando básicamente información sobre estados políticos. Tal información tendía a crecer rápidamente en cantidad y en extensión teniendo que adoptarse formas de recoger la información en tablas, cuadros, gráficos y matrices. Se llegó así de manera natural a que la palabra Estadística significase todo el material surgido de la observación del mundo, aceptándose este uso a partir del siglo XIX.

En la actualidad el término Estadística tiene un sentido más amplio, y significa ciencia que estudia el comportamiento de los fenómenos de masas, no ocupándose de comportamientos individuales.

El Diccionario de la Lengua Española, considera que el término Estadística procede de estadista, voz de Estado y Estado a la palabra latina "Status", o situación en que una persona o cosa está sujeta a cambios que influyen en su condición. Se llega a dos acepciones:

- Como censo o recuento de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de un estado, provincia, pueblo o conjunto.

- Estudio de los hechos morales o físicos del mundo que se prestan a numeración o recuento y a comparación de las cifras referentes a ellos.

Aritmética de la Política ?

Cuando se construyen estadísticas para comparar cifras, aparece de forma natural la existencia del valor medio, concepto que facilitó un campo de investigación a los aritméticos políticos como John Graunt, Petty o Süssmilch.


JOHN GRAUNT

Los precursores de la Estadística la entendían como una aritmética de la política. Sus trabajos se referían principalmente a lo que hoy denominamos estadísticas demográficas.

Así Graunt establece las primeras tasas de mortalidad, con intervalos de edades

definidos, utilizando para ello los registros de defunciones de Londres. Las tablas de defunciones de Halley que son la iniciación de los trabajos actuariales actuales.

Desarrollo evolutivo...

1ª Etapa. Las estadísticas son tan antiguas como la propia sociedad humana, remontándose a las primeras civilizaciones, en donde se hacen los primeros intentos estadísticos con fines fiscales o militares. Se hacían recuentos de poblaciones o censos con fines demográficos, así como inventarios. El cargo de censor era de los más importante de la administración romana.

2ª Etapa. Se produce, entre los siglos XVIII y XIX y se caracteriza por el esfuerzo en describir y analizar los conjuntos estadísticos, sea cual sea su naturaleza. En 1790 se levanta el primer censo en Estados Unidos, realizándolo, a partir de ese momento, decenalmente.

3ª Etapa. Se aplican a grandes masas de datos, técnicas de análisis matemático y probabilístico entre otros.

4ª Etapa. Es la etapa más moderna y en ella estadística y probabilidad, van ligadas a las grandes herramientas de las matemáticas, álgebra, cálculo y análisis.

Con independencia de todo lo anterior, aparece y se desarrolla el cálculo de probabilidades y la inferencia estadística.

Aporte de los Personajes Célebres a la Estadística...

Aparece una aportación básica de algunos matemáticos al ámbito de la estadística, fundamentalmente hacia el siglo XVII. El cálculo de probabilidades que se fusionó con las investigaciones de los aritméticos políticos para dar lugar a la estadística matemática.

En 1654 el caballero de Méré planteó a Pascal, cuestiones relativas a fenómenos de azar, como: en ocho lanzamientos sucesivos de un dado, intenta un jugador obtener un uno, pero el juego se interrumpe después de tres intentos fallidos ¿en qué proporción debe ser compensado el jugador?. Pascal consultó a Fermat y este estudio dio origen al cálculo de probabilidades, aunque el primer trabajo publicado formalmente al respecto se debe a Huygens, el cual presentó en 1657 un breve tratado "Sobre los Razonamientos Relativos al Juego de Dados", basado en los trabajos de Pascal y Fermat.

BLAISE PASCAL

Pascal (1623 - 1662), relacionó el estudio de probabilidades con el triángulo aritmético que lleva su nombre (Triángulo de Pascal), del cual descubrió muy destacadas propiedades y la manera de representar coeficientes binomiales con dicho triángulo. Posteriormente Pascal formuló la llamada "Apuesta de Pascal", una reflexión filosófica sobre la creencia en Dios, basada en consideraciones

probabilísticas. La famosa apuesta analiza la creencia en Dios en términos de apuestas sobre la existencia, pues si el hombre cree y finalmente Dios no existe, nada se pierde en realidad.

Pascal murió a la edad de 39 años, después de sufrir un dolor intenso debido al crecimiento de un tumor maligno en el estómago, que luego se le propagó en el cerebro.

Experimentos tipo Bernoulli...

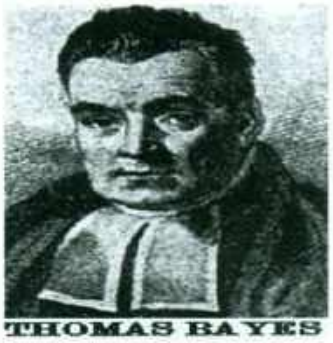
Dentro de la primera etapa del cálculo de probabilidades, se debe destacar la obra de Jacob Bernoulli (1654 – 1705), y en especial su famoso libro "Ars Conjectandi" (El Arte de la Conjetura), en él se establecen los principios fundamentales de dos ramas de la estadística, como son el Cálculo de Probabilidades y la Teoría Combinatoria.


JACOB BERNOULLI

En este trabajo se presenta una definición de probabilidad que coincide con la que más adelante se considerará regla de Laplace. También se desarrolla la Ley de los Grandes Números, o Teorema de Bernoulli.

El Teorema de Bayes...

Thomas Bayes (1702 - 1761), fue un gran matemático. El teorema que lleva su nombre, trata sobre la probabilidad de un suceso condicionado por la ocurrencia de otro suceso. Bayes fue uno de los primeros en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística.


THOMAS BAYES

Actualmente, con base a su obra, se ha desarrollado una poderosa teoría que ha conseguido notables aplicaciones en las más diversas áreas del conocimiento. En el campo sanitario, el enfoque de la inferencia bayesiana experimenta un desarrollo sostenido, especialmente en lo que concierne al análisis de ensayos clínicos, donde dicho enfoque ha venido interesando de manera creciente a las agencias reguladoras de los medicamentos.

Bayes fue ordenado como ministro disidente, y en 1731 se convirtió en reverendo de la iglesia presbiteriana en Tunbridge Wells; aparentemente trató de retirarse en 1749, pero continuó ejerciendo hasta 1752, y permaneció en este lugar hasta su muerte.

Es sentido común la Teoría de las Probabilidades ?

Pierre Laplace (1749 – 1827) desde 1774 publicó muchos trabajos sobre la teoría de probabilidades, y se le considera el verdadero creador de la Teoría de Probabilidades. Posteriormente publicó el

libro clásico "Theorie Analytique des Probabilites" donde Laplace sostenía que la teoría de probabilidades era sólo sentido común expresado en números. En su libro se presenta en detalle la forma de potenciar el cálculo de probabilidades con el cálculo analítico.

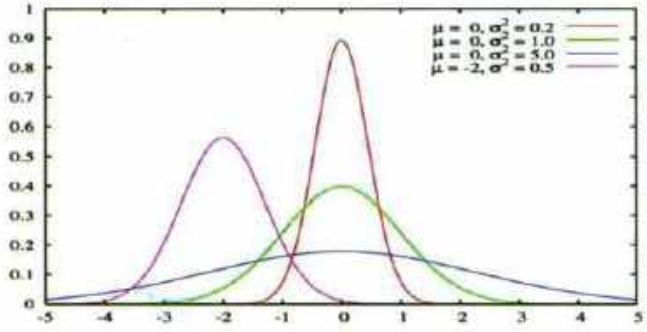
LAPLACE

Laplace utilizó las funciones "Gamma" y "Beta" de Euler y estudió las distribuciones asociadas con detalle, aunque no con la concepción con la que hoy en día se pueden tratar.

La famosa Campana de Gauss...

Otro de los precursores de la Estadística fue Johann Gauss (1777 – 1855), considerado el príncipe de las matemáticas debido a que desde muy pequeño mostró su talento para con los números, siendo un niño prodigio, de quién existen muchas anécdotas acerca de su asombrosa precosidad.

En 1823 publicó la obra denominada "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae", dedicada a la estadística, concretamente a la Distribución Normal cuya curva característica es denominada como "Campana de Gauss", muy usada en distintas disciplinas de la ciencia.



No cabe duda que los trabajos de Gauss han dado lugar al modelo más usado en la estadística moderna. No se podría hablar de técnicas de control de calidad, por ejemplo, sin estos trabajos de Gauss.

Ambos autores, Gauss y Laplace, son los creadores del conocido método de Mínimos Cuadrados, que hacen nacer en el amplio campo de la experimentación de las ciencias físico - naturales, por la necesidad que tienen los científicos de funcionalizar sus resultados experimentales. Estos trabajos se constituyen en el punto de partida de la teoría de la regresión y por ende de los más modernos métodos econométricos.

Johann Gauss

La distribución de Poisson ...

Otro personaje importante en la historia de la estadística es Siméon Denis Poisson (1781 – 1840), fue astrónomo, físico y matemático francés al que se le conoce

por sus diferentes trabajos en el campo de la electricidad, la geometría diferencial y la teoría de probabilidades.

Siméon Poisson

Su ocupación fue estudiar la teoría de la probabilidad y el análisis complejo. Su contribución al estudio de la teoría de probabilidades se fundamenta en los resultados de Laplace. En 1837, publicó en Rerecherchés sur la "Probabilite des Jugements", el desarrollo de una fórmula para el cálculo de la probabilidad de ocurrencia en sucesos cuando ésta es muy pequeña, que tiene gran aplicación práctica. A partir de esta fórmula obtuvo una distribución que lleva su nombre, y que, más tarde se demostraría como un caso aproximado de la distribución Binomial.

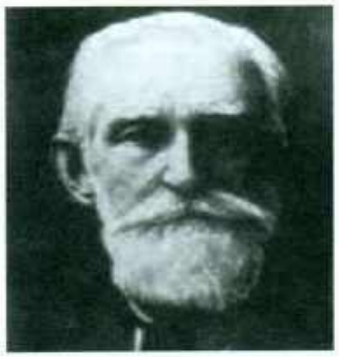
Poisson enseñaba en la escuela Politécnica desde el año 1802 hasta 1808, en que llegó a ser un astrónomo del Bureau des Longitudes. En el campo de la astronomía estuvo fundamentalmente interesado en el movimiento de la luna. Durante toda su vida publicó entre 300 y 400 trabajos matemáticos incluyendo aplicaciones a la electricidad, el magnetismo y la astronomía.

Poisson muere en 1840, siendo miembro de la Academia de Ciencias de París.

La unificación de la estadística y la probabilidad fue realizada por Quetelet y por los rusos Chebyshev (1821-1894) y Markov (1856 – 1922), además de los trabajos del francés Poincaré, que publicó en 1896 un trabajo de síntesis que se denominó "Calcule des Probabilities".

La Desigualdad de Chebyshev ...

Pafnuti Chebyshev matemático ruso, su nombre se translitera también como Tchebychev, Tchebycheff, Tschebyscheff o Čebišëv.



Pafnuti Chebyshev

En 1846 defendió su tesis "Un Intento de Análisis Elemental de la Teoría Probabilística" y en 1847 ingreso como profesor a la Universidad de San Petersburgo.

Es conocido por su trabajo en el área de la probabilidad y estadística. La desigualdad de Chebyshev dice que la probabilidad de que una variable aleatoria esté distanciada de su media en más de a veces la desviación típica es menor o igual que $1/a^2$.

Las Cadenas de Márkov ...

Andréi A. Márkov (1856 – 1922), fue matemático, el año 1878 comenzó su carrera como profesor. Enfocó sus trabajos en análisis y la teoría del número, fracciones continuas, limite de

integrales, teoría de la aproximación y la serie de convergencias.

Andréi Márkov

Desde el principio mostró cierto talento a las matemáticas. En la Universidad fue discípulo de Chebyshev, Márkov impartió clases y, cuando el propio Chebyshev dejó la Universidad tres años después, fue Márkov quien le sustituyó en los cursos de teoría de la probabilidad y la secuencia de las variables mutuamente dependientes, esperando con ello establecer, de manera general, las leyes limitantes de las probabilidades. Probó el Teorema del Límite Central bajo supuestos generales.

Su trabajo teórico en el campo de los procesos en los que están involucrados componentes aleatorios (Procesos Estocásticos) darían fruto en un instrumento matemático que actualmente se conoce como las Cadenas de Márkov, éstas, hoy día, se consideran una herramienta esencial en disciplinas como la economía, la ingeniería, la investigación de operaciones y muchas otras.

El Coeficiente de Correlación de Pearson...

Karl Pearson (1857 – 1936), fue un prominente científico, matemático, historiador y pensador británico, que estableció la disciplina de la estadística matemática. Desarrolló una intensa investigación sobre la aplicación de los métodos estadísticos en la biología y fue fundador de la bioestadística.

Fue un positivista radical y socialista. "Carl" se convirtió en "Karl" inadvetidamente cuando la Universidad de Heidelberg cambió la manera de escribir su nombre cuando se matriculó en 1879, aunque usó ambas variantes de su nombre hasta 1884 cuando finalmente adoptó "Karl" – presuntamente por Karl Marx, a quién conoció en vida – y eventualmente llegó a ser conocido universalmente como "KP".

Karl Pearson

En 1900 Pearson popularizó el criterio de la "Chi-Cuadrado" y el Coeficiente de Correlación que lleva su nombre. En 1911 fundó el primer departamento de Estadística en la Universidad de Londres, donde fue profesor y donde dirigió el Laboratory of National Eugenics creado por Sir Francis Galton. Fundó en 1902 la revista Biometrika, desde entonces una de las más importantes en el campo de la estadística.

El Coeficiente de Spearman...

Otro personaje importante en el aporte a la Estadística es Charles Spearman (1863 - 1945), psicólogo de profesión, estudio estadística y logro desarrollar notables

aplicaciones de la estadística en el campo de la psicología. Formuló la teoría del análisis factorial y dentro de ello demostró que la inteligencia se compone de un factor general y otros específicos.

Charles Spearman

Creó y desarrolló la metodología de los llamados Experimentos Factoriales, que son aquellos en los que se estudia simultáneamente dos o más factores, y donde los tratamientos se forman por la combinación de los diferentes niveles de cada uno de los factores. Los experimentos factoriales se emplean en todos los campos de la investigación, son muy útiles en investigaciones exploratorias en las que poco se sabe acerca de muchos factores. Spearman logró con el estudio de la psicología completar su estudio de la Estadística y viceversa, para él una se completaba con la otra. Por todo esto es que Charles Spearman es considerado uno de los grandes estadísticos de todos los tiempos.

De dónde el nombre de la distribución T-Student?

Quizás uno de los más misteriosos personajes de la Estadística fue William Gosset (1876 – 1937).

Estadístico, mejor conocido por su sobrenombre literario Student. Gosset fue amigo tanto de Pearson como de Fisher. Gosset trabajaba con Guinness, negociante agroquímico dueño de una destilería. Gosset aplicaba sus conocimientos estadísticos en dicho negocio, para seleccionar las mejores variedades de cebada.

Otro investigador de Guinness había publicado anteriormente un artículo que contenía secretos industriales de la destilería. Para evitar futuras exposiciones de información confidencial, Guinness prohibió a sus empleados la publicación de artículos, independien-temente de la información que contuviesen. Esto significó que Gosset no podía publicar su trabajo usando su propio nombre, de ahí el uso de su pseudónimo Student en sus publicaciones, para evitar que su empleador lo detectara. Por tanto, su logro más famoso se conoce ahora como la Distribución t de Student, que de otra manera hubiera sido la Distribución t de Gosset.

Primero la distribución F-Fisher o la distribución T-Student?

Fue Fisher quién apreció la importancia de los trabajos de Gosset sobre muestras pequeñas, tras recibir las famosas Tablas de Student. Fisher creyó que Gosset había efectuado una "revolución lógica". Irónicamente el estadístico t por la que Gosset es famoso, fue realmente la creación de Fisher.

William Gosset

Fisher introdujo la forma t debido a que se ajustaba a su teoría de grados de libertad. Fisher es responsable también de la aplicación de la distribución t a la regresión.

Ronald Fisher (1890 – 1962), científico, matemático, estadístico, biólogo evolutivo y genetista inglés. Fisher realizó muchos avances en la Estadística, siendo una de sus más importantes contribuciones, la Inferencia Estadística creada por él en 1920.

Ronald Fisher

Publicó varios artículos sobre biometría, incluyendo el célebre "The Correlation Between Relatives on the Supposition of Mendelian Inheritance", que inauguró la fundación de la llamada genética biométrica e introdujo la metodología del análisis de varianza, considerablemente superior a la de la correlación.

En 1919 Fisher empezó a trabajar en la Rothamsted Experimental Station (Harpenden, Hertfordshire, Inglaterra). Allí comenzó el estudio de una extensa colección de datos, cuyos resultados fueron publicados bajo el título general de Studies in Crop Variation. Durante los siguientes siete años, se dedicó al estudio pionero de los principios del Diseño de Experimentos (The Design of Experiments, 1935), elaboró sus trabajos sobre el análisis de varianza y comenzó a prestar una atención especial a las ventajas metodológicas de la computación de datos (Statistical Methods for Research Workers, 1925).

Otros aportes de estadísticos son:

Sir Francis Galton (1822-1911) estudió los fenómenos de regresión en mayor profundidad. En Rusia, Kolmogorov contribuyó al progreso del estudio de los fenómenos de Márkov y en general al de los procesos estocásticos.

A principios del siglo XX matemáticos como Yule, Box, Piercé y otros de la escuela anglosajona desarrollaron, entre otros aspectos, la teoría de procesos estocásticos, fundamentando el campo de las series temporales que tendría gran influencia en múltiples campos como la meteorología o la economía.

Kendall, Jenkins, Fisher, Von Mises, etc, configuran el cuerpo teórico de gran amplitud que da lugar a lo que hoy en día conocemos como estadística matemática.

Edición y recopilación de información:

Consuelo Beatriz Barrios G.
Graciela Mercedes Heredia C.
Beatriz Vilma Vallejos M.
Mauricio Jorge Gamarra U.
Fernando Oday Rivero S.

Índice

El Uso de la Estadística en el Voleibol

(Oscar Luis Villamea-Profesor de Educación Física)

Realmente es muy difícil saber quién fue el primero en aplicar la estadística individual en el voleibol. Sí sabemos que Julio Velasco fue y es el gran promotor sobre este tema. Es quien dio impulso a la creación de un software aplicado al voleibol.

Este soft se llama Data Volley y no solamente permite obtener la estadística individual sino la estadística del ataque rival en los dos sistemas (de cambio de saque y de punto) y la dirección de la misma.

Julio Velasco, dirigiendo la selección italiana, se hizo famoso en los Mundiales y Juegos Olímpicos en donde se lo veía junto a los auxiliares con un auricular y un micrófono, con el cual se comunicaban con otro auxiliar ubicado en la tribuna y llevando la estadística en la computadora (notebook).

Hoy en día la estadística individual es tan importante que los jugadores en los altos niveles de competencia, se cotizan acorde a sus porcentajes y ellos mismos terminado el encuentro o durante la semana tiene la necesidad de averiguar su estadística con el fin de saber en qué aspecto deben mejorar. La estadística individual se utiliza en todos los torneos locales e internacionales. La ***Federación Internacional de Voleibol (FIVB)*** en su página Web, tiene datos estadísticos de los jugadores que participan de torneos internacionales.

ESTADISTICA

¿Qué es?

La estadística es una forma de expresar las vicisitudes del juego en números. Estos números pueden manifestarse en indicadores de porcentajes de rendimiento, eficacia, eficiencia, error, etc., en forma individual o grupal, por sistemas o en forma global.

Todas estas formas de medir la estadística son válidas solamente cuando sea en función de mejorar el rendimiento del grupo o el individual y no que sea un fin en sí mismo, o sea que es una herramienta en función de la mejora del entrenamiento y del rendimiento individual.

¿Para qué sirve?

La estadística individual tiene como función poder darle al entrenador una clara visión de cómo se desempeña cada jugador dentro del campo de juego en cada acción que realiza.

¿Cuántas opciones de estadísticas hay?

Cada entrenador puede realizar el seguimiento de datos que le interese,

durante o post partido, del equipo propio o del rival.

Considerar que para decidir qué estadística llevar durante el partido, es fundamental la elección de una información que sirva en ese momento, como por ejemplo el ataque rival, dirección, distribución, etc., y la estadística individual dejarla para el análisis después del partido con cada jugador.

Particularmente, luego de varios años de experiencia, la estadística individual es utilizada mediante el vídeo (post-partido) y la que me da información del equipo oponente, se capta durante el partido.

Estadística Individual

La estadística individual se basa en recabar datos de cada jugador en cada acción que realiza y colocarle una notación acorde al resultado de la misma. Las acciones del jugador que se pueden evaluar son:

- Saque
- Recepción
- Armado
- Ataque
- Bloqueo
- Defensa

Pero, para poder ordenar estas acciones es importante dividirlas acordes a la situación del partido:

- Sistema de punto
- Sistema de cambio de saque

En el *sistema de punto,* se toman en cuenta todas aquellas acciones donde el propio equipo tiene la posesión del saque. Y en el *sistema de cambio de saque* se toman en cuenta las acciones en que el rival tiene la posesión del saque y estaría con posibilidad de sumar en el tablero.

Considerando estas dos opciones podemos dividir la estadística individual en:

Sistema de Punto	**Sistema de Cambio de Saque**
• **Saque**	• **Recepción**
• **Bloqueo**	• **Armado**
• **Defensa**	• **Ataque**
• **Armado**	• **Bloqueo**
• **Ataque**	• **Defensa**

Profundizando aún más, se puede agregar la cobertura del bloqueo, sin embargo es suficiente iniciar una base de datos estadísticas con estos ítems.

A cada uno, de las opciones en cada sistema, para poder evaluarlos se los divide en tres, cuatro o cinco items de puntaje (obviamente acorde a como se realiza la acción). Estas pueden llamarse como cada uno crea conveniente, ejemplo:

Acción: **Saque**

- Saque Positivo - Saque Neutro - Saque Negativo (opción de tres ítems)
- Saque doble Positivo - Saque Positivo - Saque Neutro - Saque Negativo - Saque doble Negativo (opción de cinco ítems)

Para comenzar a llevar estadística de los jugadores, primero es importante poder hacerlo con tres items debido a que es más simple apreciar las altas y bajas de cada jugador y hacer una clara evaluación del mismo. Al dividir en más ítems hay

que tener muy presente qué; importancia le damos a todas las acciones que no son evaluadas como culminación de una acción de juego, ya que pueden representar el mayor porcentaje de ejecuciones de una misma acción.

Es por eso que cada acción la dividiremos en tres items de evaluación:

- *Positivo*: Siempre que la acción de juego culmine a favor del propio equipo.
- *Neutro*: Siempre que la acción de juego permita la continuidad de la misma.
- *Negativa*: Siempre que la acción de juego culmine en contra del propio equipo.

¿Porcentajes o números verdaderos?

Después de haber explicado al jugador cómo va a ser la forma en que se lo va a evaluar en los partidos y de haberla puesto en práctica en algunos partidos preparativos, es importante darle al jugador la cantidad exacta de acciones que realizó durante el encuentro, si es posible lo más cercano a la culminación del partido donde tiene su memoria fresca, y dejar los porcentajes de los mismos a la posterior evaluación del cuerpo técnico. Estos porcentajes permitirán que el entrenador pueda medir la evolución de cada jugador y poder conversar individualmente con cada uno de ellos para analizar las falencias y progresos sobre cada acción de juego evaluada.

Eficacia, Eficiencia y Error: para qué sirven y qué son?

La forma de poder tener un parámetro para medir la evolución de cada jugador sobre cómo actúa en cada partido y poder llevar una media de ella, son los porcentajes de eficacia, eficiencia y error.

Al calcular los porcentajes permite tener una medida común con respecto a cada encuentro deportivo que no dependerá del número de acciones que realice un jugador. Para poder ejemplificarlo, podemos decir que si un jugador en un partido realiza 10 saques con ataque en total, y en el partido siguiente en la misma situación realiza 30 saques de ataque, cuál sería la forma de medir la evolución del jugador en esta acción, sino fuera por los porcentajes que dan la eficacia, eficiencia y error.

Como se dijo anteriormente, se le asigna tres valores a cada acción de juego, acorde a los datos que se recaba:

- ***Eficacia***: Es el porcentaje con que el jugador realizó todas las acciones positivas.
- ***Error***: Es el porcentaje con que el jugador realizó todas las acciones negativas.

- ***Eficiencia***: Es el porcentaje de cuan provechoso fue el trabajo del jugador sobre la acción evaluada.

Para poder calcular estos porcentajes se cuenta con los siguientes datos: Por ejemplo, siendo la acción evaluada el **Saque** (obviamente dentro del sistema de punto).

- Total de pelotas con las que realizó puntos de saque (saques positivos).
- Total de pelotas con las que el rival mantuvo en juego el balón (saques neutros).
- Total de pelotas en que se erró el saque (saques negativos).
- Total de saques (sumatoria de todos los saques).

Entonces, para realizar los cálculos se tienen los siguientes indicadores:

$$Eficacia = \frac{Total\ de\ saques\ positivos}{Total\ de\ saques\ realizados} 100$$

$$Error = \frac{Total\ de\ saques\ negativos}{Total\ de\ saques\ realizados} 100$$

$$Eficiencia = \frac{Total\ de\ saques\ positivos - Total\ de\ saques\ negativos}{Total\ de\ saques\ realizados} 100$$

Base de Datos

Una vez obtenido los primeros resultados, es importante crear una base de datos de cada jugador donde se puede ir observando su evolución en el transcurso del año deportivo, y poder decir que tal jugador tiene una media de tanto en recepción, saque, bloqueo, etc.

También sirve para poder mostrarle al jugador su desarrollo en cuanto a los resultados prácticos de su accionar y que este pueda saber en dónde debe mejorar y en qué aspecto está realizando las cosas bien.

Posibles parámetros con los cuales evaluar a cada jugador

Los siguientes pueden ser una medida como para evaluar a cada jugador en cada acción de juego:

Saque

- *Saque Positivo*: Todos los saques en que se realiza punto directo.
- *Saque Negativo*: Todos los saques que son errados.
- *Saque Neutro*: Todos los saques que permita seguir jugando al equipo rival.

Bloqueo

- *Bloqueo Positivo*: Todas las acciones de bloqueo que después de realizarla la pelota pique en el campo rival u ocasionalmente rebote contra algún adversario.

- *Bloqueo Negativo*: Todas las acciones de bloqueo que después de realizarla, el balón pique en campo propio o pegue en manos propias y salga del campo de juego.
- *Bloqueo Neutro*: Todas las acciones de bloqueo que después de realizarla, la pelota pueda seguir en juego por cualquiera de los dos equipos.

Ataque

- *Ataque Positivo*: Todas las acciones de ataque que luego de realizarla la pelota pique en el campo contrario o golpee contra el bloqueo y el rival no pueda seguir jugando el balón.
- *Ataque Negativo*: Todas las acciones de ataque que luego de realizarla la pelota pique fuera del campo de juego, se quede en la red o permita una acción positiva del bloqueo.
- *Ataque Neutro*: Todas las acciones de ataque que después de realizarla, la pelota pueda seguir en juego por cualquiera de los dos equipos.

Recepción

- *Recepción Positivo*: Todas las acciones de recepción en la cual son colocadas perfectas las pelotas al armador, acorde a la zona de armado.
- *Recepción Negativa*: Todas las acciones de recepción en la cual el saque es positivo ya que se le hizo punto directo al receptor.
- *Recepción Neutra*: Todas las acciones de recepción en la cual son colocadas las pelotas al armador no tan perfectas con respecto a la zona de armado, o sea que puede llegar el armador en forma exigida, golpeando de abajo, etc.

Defensa

- *Defensa Positiva*: Todas las acciones de defensa que ante un ataque potente o una colocada o un desvío en el bloqueo se recupere un balón y permita rearmar el ataque.

- *Defensa Negativa*: Todas las acciones de defensa, estando en el lugar correcto, no permitan continuar la acción de juego.
- *Defensa Neutra*: Todas las acciones de defensa que permitan continuar la acción de juego sin armar el ataque y no se comentan errores posicionales.

Armado

En el armado se evalúa fundamentalmente la *precisión* ya que la parte táctica dependerá exclusivamente de cómo este preparado el partido.

En todos los casos, es importante tener bien claro los parámetros, que pueden o no ser estos, para que la medición sea válida.

Particularmente se toma en cuenta la decisión del árbitro y no la observación propia.

Conclusiones

Todas las posibilidades de estadísticas y parámetros que se enumeraron no son únicos ni exclusivos, sino que se pueden mejorar. También se puede realizar otro tipo de estadísticas u otro tipo de parámetros, pero se debe saber que antes de empezar, hay que tener bien claro qué es lo que se busca desarrollar, para saber cómo iniciar esta búsqueda y no perder tiempo cambiando y no definiendo nunca qué es lo que se quiere.

El fin de este artículo es poder dar una guía o un orden a los que recién se inician en el tema de la estadística en el voleibol.

Índice

El Mérito Estadístico del Pulpo Paul

(Escrita en http://alt1040.com el 9 de Julio de 2010 por Pepe Flores)

Y el **Pulpo Paul** habló. Su nuevo candidato, para beneplácito de muchos lectores, es **España**. El simpático cefalópodo se ha convertido en el fenómeno pop del Mundial. Yo me he preguntado si lo del pulpo es pura suerte, no sólo en un sentido mágico-cómico-musical, sino desde una perspectiva estadística. Tengo la buenaventura de compartir casa con un actuario, así que la tarde de ayer lo asedié con preguntas sobre cómo lo hizo un desgraciado molusco para atinarle a todos los pronósticos.

Partamos del supuesto de que el pulpo goza de aleatoriedad máxima; es decir, que no hay ninguna manipulación por parte de los dueños del acuario. Su verdadero mérito radicó en atinarle a los tres primeros partidos de Alemania. Verán, la probabilidad de que eligiera correctamente una victoria era aproximadamente de **40%** en cada uno de los casos (considerando que la posibilidad de un empate es de 20%). La chance de que concatenara los tres resultados correctos era de **6.4%**. A partir de ahí, comienza la verdadera magia matemática de Paul.

A partir de la siguiente ronda, la predicción se tornaba un escaso más fácil para nuestro cefalópodo, puesto que la posibilidad era **50/50** (no hay empates en esta fase). El pulpo dijo **Alemania**, y los teutones se la llevaron ante **Inglaterra** y ante **Argentina**. En esta ocasión, la chance de Paul de predecir cuatro juegos correctos seguidos era de **3.7%**; y de juntar cinco, del **1.6%**. ¿Ya es demasiada suerte, no?. Bueno, pues en la siguiente el pulpo se la jugó con **España**, y otra vez, acertó, rompiendo una probabilidad de **0.8%** para juntar seis partidos invicto.

Acá ya hay algo raro. ¿Podrá irse invicto Paul?. La posibilidad de que atine al séptimo juego (el de la final) era de un **0.4%**. Y un octavo (Alemania Uruguay), sólo de **0.2%**. Como me comentó mi amigo actuario, acá sólo hay una explicación: algo sesga las decisiones del pulpo. Algunos dirán que tiene poderes, otros dirán que los que ponen la comida deciden qué elegirá. Sin embargo, otra hipótesis fuerte es que **la elección del pulpo afecta el efecto debido al furor que causa**. Por esta razón, cuando Paul habla, las casas de apuesta modifican sus números.

La explicación apela al modelo estadístico. Cuando la probabilidad es menor del **1%**, forzosamente se entiende que hay otros factores que afectan; o de lo contrario, estaríamos ante un caso que se da una de cada 250 veces (ó 500, si le atina también al de tercer lugar).

Acá hay una explicación muy lógica: cuando el pulpo predijo la semifinal, los alemanes estaban confiados en que *su* pulpo les daría la victoria.

La prensa española, por el contrario, esperaba que Paul los diera por derrotados. Al darse lo contrario, ocurre un efecto de proporciones impredecibles. **Alemania** recibe un golpe anímico. España crece, los diarios hablan, la noticia corre como reguero, la *tuitósfera* ebulle, las casas de apuesta ajustan, y al final, el entorno cambia. Conclusión: lo que diga Paul sí influye.

¿Eso significa que **España** será el campeón?. Bueno, en esta ocasión el pulpo no tendrá tanto protagonismo. Por una parte, los dueños del bar se han curado en salud diciendo que **Paul** es especialista en **Alemania**, lo que le da un margen de yerro en la final. Por la otra parte, había expectativa del lado español, contrario a la aparente indiferencia holandesa, que el pulpo dé ganadora a la *Furia Roja*, no afecta tanto el guión.

Curioso hubiera sido ver que se decantara por **Holanda**, por el efecto anímico (para ambas partes) habría puesto más interesantes las cosas. Ese es el mérito estadístico de Paul, una magia regida por la forma en que los números afectan nuestra vida cotidiana.

Estadísticas Sorprendentes Sobre las Causas del Divorcio

(Extraído de http://actualidad.rt.com)

Índice

La periodista y escritora estadounidense Anneli Rufus realizó una investigación tratando de determinar qué factores y cómo influyen en la probabilidad de divorcio de una pareja. Rufus publicó los resultados de su estudio en *The Daily Beasts*. En la lista compuesta por 15 puntos figuran diferentes factores, incluyendo la edad a la que los esposos contrajeron el matrimonio, el género de los hijos y hasta el nivel de testosterona.

Así, la probabilidad de divorcio durante los 10 primeros años del matrimonio para las mujeres que se casaron antes de cumplir 18 años es de un 48.0%. Si el deseo de la mujer de tener un hijo es mayor que el del hombre, la posibilidad de que se divorcien crece. Los esposos que tienen dos hijas son más propensos al divorcio, que los que tienen dos hijos (el 43.1% contra el 36.9%, respectivamente). Los hombres que tienen un nivel alto de testosterona son un 43.0% más propensos al divorcio que sus congéneres con un nivel normal de hormonas en la sangre.

El estado de salud también importa a la hora de divorciarse, señala Rufus. Así, si a la mujer se le diagnostica cáncer o esclerosis múltiple, la probabilidad de que su

matrimonio se disuelva, es seis veces mayor que si fuera el esposo el que se enferma.

La profesión también influye en el divorcio, conforme el estudio de la

periodista. El oficio más peligroso en este sentido es el de bailarín, que tienen el 43.0% de probabilidad de rupturas. mientras que los optometristas y los físicos nucleares corren el menor riesgo de divorciarse (con el 4.0% y el 7.3%, respectivamente).

De acuerdo con la información publicada, la raza también es un factor importante en los asuntos familiares. Así, el primer matrimonio de las mujeres de color se disuelve durante los primeros 10 años en el 47.0% de los casos; las hispanas se divorcian en el 34.0%; las caucásicas en el 32.0%, mientras que las asiáticas en tan sólo 20.0%. Entre las premisas del divorcio, figura también un factor curioso. Según el estudio, las personas que no salen con una sonrisa en sus fotos de la niñez tienen pocas oportunidades de crear un matrimonio sólido.

La lista hecha por Rufus mira desde otro ángulo las relaciones entre los esposos y las causas que los llevan al divorcio. Según el sondeo, una mujer tiene en promedio 7 amantes y los hombres 13. El director de la página web, Mark Pierson, aconseja empezar las relaciones de forma honesta. "Si empiezan las relaciones con mentiras, lo más probable es que las terminen con mentiras aun más grandes", dice Pierson.

Estadísticas Amorosas

(Extraído de http://actualidad.rt.com)

Aproximadamente una de cada tres mujeres no dice la verdad al responder a la pregunta de cuántos amantes tuvo antes, informa el Diario Británico "The Daily Mail" basándose en un sondeo realizado por la página web *www.MyVoucher-Codes.co.uk.*

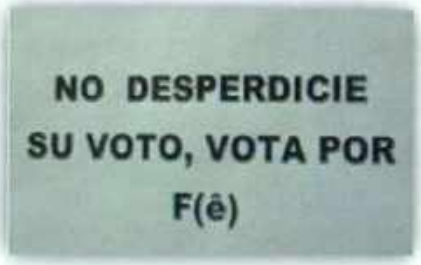
El 64% de las mujeres encuestadas tiende a reducir el número de amantes anteriores (la mitad por vergüenza, y un 19% para no parecer poco cuidadosa en sus relaciones).

Cada una de cinco mujeres oculta la verdad en el supuesto de que su novio tenga menos relaciones amorosas que ella.

Entre los hombres, 43% están predispuestos a mentir y 60% exagera la cantidad de sus conquistas amorosas.

Índice

Comentario Sobre las Elecciones Para Director de la Carrera de Estadística



En algo más de un mes, se han realizado dos elecciones para elegir al nuevo Director de Carrera de Estadística para la gestión 2010 – 2013 (entre septiembre y octubre).

Ambos escrutinios no han aportado a la Carrera. En la primera y segunda elección, se presentaron distintos candidatos, y en ambos casos se ha impuesto el voto en blanco, más en el estamento docente que en el de estudiantes.

Hoy no corresponde analizar los resultados, sino más bien hacer una reflexión en torno a ellos, preguntándonos hacia dónde vamos. Se supone que se requiere contar con un Director(a) que lleve adelante nuestra Carrera y nos represente en cada una de las instancias universitarias.

Quién está en desventaja?, claramente la institución (la institución que es la Carrera). Es que no pensamos y no velamos por la institución?, más bien, profesamos nuestros intereses?, sabiendo que, ante todo, deberíamos cuidar nuestra casa.

Qué realmente queremos de la Carrera?, qué perfil debería tener el candidato(a) para ser elegido(a)?, al parecer, no estamos de acuerdo con nadie. Será posible que alguien se anime a presentarse para ser elegido(a) Director(a), en la situación en la que nos encontramos?... son tantas preguntas que seguramente no tienen respuesta ahora.

Entonces, que debemos hacer?. Posiblemente y lo más ideal, es buscar un consenso entre estudiantes y docentes, reunirnos para definir qué deseamos para esta Carrera a la cual nos debemos.

En épocas pasadas no se necesitaba conformar frentes en pugna, para elegir a nuestras autoridades docentes y estudiantiles, los elegíamos por consenso siguiendo las normas de la Universidad y poniéndonos de acuerdo en un sólo candidato que sea de agrado para todos.

Una Carrera tan pequeña en relación a otras, se constituía en una gran familia conformada por los tres estamentos, hoy eso se desvanece. Ojala reflexionemos sobre el tema antes de la próxima elección.

Fernando Rivero S.

Índice

Estadística, Estadistas y Mentirosos

"La estadística no miente; el estadista, sí"

Esta mañana, al salir de clase, he llegado a tomar un bus en forma tranquila, pensando si podía presentarme al examen de matemática de mañana, cuando de pronto me ha sorprendido una conversación de tres chicas que iban a mi lado. No es que vaya escuchando las conversaciones de la gente (normalmente, paso), pero cuando la gente habla a voz en grito tienes dos opciones: oírles o escucharles.

Una de ellas hablaba sobre su padre y la mala relación de éste con las estadísticas que publican en los periódicos: según parece, creía que la estadística es toda una gran mentira, y que la gente se inventa los datos y los porcentajes con tal de que den lo que a ellos les conviene.

Ésta reflexión no es la primera ni la última vez que la oigo: después de todo, incluso mi madre alega que la Estadística en su totalidad es una enorme patraña, y que no sabía por qué estudiaba yo eso.

La Estadística, por definición, es la ciencia encargada de recoger, organizar e interpretar los datos. Resumiendo, es la ciencia de los datos, y en un mundo como el actual, donde la afluencia de datos es evidente, se hace necesario tener un conocimiento básico sobre cómo tratarlos. Ése conjunto de conocimientos básicos (y no tan básicos) es la Estadística.

No es un conocimiento trivial: si nos engañan con las estadísticas, es porque no tenemos ni idea de cómo manejarlas. Después de todo, abundan por ahí ejemplos de estadísticas (publicadas en periódicos), donde la suma de todos los porcentajes puestos en juego da más del 100% o menos incluso.

En cuanto a mi experiencia personal, para mí la estadística es, además de ciencia, una herramienta. Una herramienta, por cierto, poderosísima, como casi toda la matemática que estoy aprendiendo en la carrera en estos momentos: ciencias brillantes, herramientas potentísimas. Cuando pienso que esto es lo básico, y me fascino al mismo tiempo.

Como decía, la Estadística está presente en casi todo momento: raro es el día que, al hacer una memoria de laboratorio, no tengo que hacer algún ajuste por mínimos cuadrados (lineal, si hay suerte). Y, aunque no toque hacer el ajuste, trabajo continuamente con medias, medianas, desviaciones,.... mi pan de cada día en el laboratorio: el mío, y el de todos los que pasan o han pasado por ahí.

Todo el trabajo en el laboratorio no serviría para nada (salvo para entretenerse un rato), si no fuese gracias a la

Estadística que, en mi caso, me permite afirmar que ciertos datos siguen cierta progresión, o cumplen cierta relación, o que falsean o se ajustan a cierta hipótesis que he realizado.

Pero, por supuesto, la Estadística tiene su "lado oscuro": como toda herramienta, es susceptible de ser usada bien... o mal. Después de todo, hoy en día resulta más fácil obtener unos datos, inventarse los valores que tienen que dar y ponerle a todo la etiqueta de "demostrado estadísticamente" que, pasando de etiquetas y demás cosas, hacer un tratamiento estadístico riguroso de los datos, exponiéndose a que devuelva valores inesperados (lo cual, aunque parezca un fracaso, muchas veces termina siendo una gran bendición). Al fin y al cabo:

"Si quieres demostrar algo absurdo toma un montón de datos, tortúralos hasta que digan lo que quieres demostrar, y a la confesión así obtenida llámale "estadística". (Darrel Huff, "How to lie with statistics")"

En resumen: la Estadística no miente: quién miente es el que hace las estadísticas. Siempre habrá casos de mal uso de la Estadística: personas que abusarán de la enorme cantidad de herramientas de que dispone el estadista, y que retorcerán y torturarán a los números hasta que éstos digan lo que ellos quieren, y nada más. Ellos son los culpables de que la Estadística aparezca tan desprestigiada hoy en día. Y es que, después de todo, gracias a ellos da la impresión de que:

"Si se reúnen suficientes datos, se puede demostrar cualquier cosa con ayuda de la Estadística. (Ley de Williams y Holland)"

"Eso es lo que quiero decir. Vea si puede encontrar algunas estadísticas para probarlas"

Dedicado a todos los que abusan de la Estadística...

Un saludo de un estudiante de la Carrera

Índice

"La falacia del cuadro estadístico estriba en que es unilateral, en la medida en que representa sólo el aspecto promedio de la realidad y excluye el cuadro total. La concepción estadística del mundo es una mera abstracción, y es incluso falaz, en particular cuando atañe a la psicología del hombre."

Carl Jung

Fraternidad "Potolos de Estadística"

"Estadística no sólo es ciencia con base matemática, es también difundir cultura, aportamos a nuestra riqueza cultural"

La Fraternidad

La fraternidad **"Potolos de Estadística"** fue fundada el 22 de mayo de 2002, por la docente Lic. Nilda Flores S. y los estudiantes de la Carrera incentivados en esta actividad.

El objetivo es participar de la Entrada Universitaria y así promover la cultura, la tradición, la belleza de nuestro folklore y recuperar una de las danzas menos conocida y vista en nuestro país.

No se puede precisar con exactitud el origen de la danza de los Potolos, pero se afirma que nace en la comunidad Potolo, la cual deriva de la cultura de la comunidad de los *Jalq'as*.

La fraternidad reactiva una fiesta tradicional que se desarrolla en las localidades de Potolo y Potobamba del departamento de Chuquisaca y Potosí, respectivamente, danza boliviana que brinda un homenaje a nuestras riquezas culturales y ancestrales, que es reconocida y apreciada por todos los bolivianos.

La comunidad Potolo está ubicada en la provincia Oropeza, localidad Sacapaya, cantón Potolo. Está situada en el área marginal entre las colinas y límites con los departamentos de Potosí y Chuquisaca, a una altura de 3.080 mts. sobre el nivel del mar.

Significado de la danza y el tipo de traje utilizado

Según la tradición, se dice que los Potolos bailaban cuando se encontraban tristes, sin embargo, el baile pregona la supervivencia de su clase, el consuelo de su tristeza como algo propio nacido de su suelo y de su ser. La danza del Potolo es atrevida e insinuante al mover las caderas y hombros acompañado del movimiento de cabeza como acto de coqueterio. Se baila en pareja ya que se muestra el enamoramiento del hombre a la mujer mostrando así el carisma y el fervor de su danza.

El origen del nombre Potolo se debe a la vestimenta que llevaron los primeros pobladores llamados *"Phututus"*.

Traje del varón

Calzona: es un pantalón color blanco de bayeta de la tierra.
Allmilla: Camisa de color blanca de bayeta de la tierra.

Chumpi: Faja tejida de colores vistosos que sostiene la allmilla al cuerpo, contiene dibujos de viscachas, cóndores murciélagos y lechuzas, dedicada al tata Santiago.
Aguayo: Sirve para cargar su coca y merienda.
Chalina: Tejida de hilos de colores vistosos se lleva en la parte del hombro y cuello.
Sombrero o Toco: Fabricado de lana de oveja, es lo más llamativo y simpático de la danza, aparenta un cotillón pequeño.
Ojotas: Conocidas como abarcas, fabricadas de goma de neumáticos en deshecho.
Chuspa: Bolsa donde ponen la coca sagrada, la lijtta o lejia, la chuspa hasta no hace mucho tiempo llevaba dibujos de animales, ahora en su gran mayoría se teje con listas de colores blanco, café y negro.

Traje de la mujer

Allmilla: Vestido con manga tres cuartos, la falda del vestido es campana y el cuello puede ser redondo, en "v" o cuadrado, existe allmillas en forma de carniceros, ambos están hechos en tela de bayeta o tocuyo, esta tela es conocida como cañarita por los campesinos. El vestido tiene adornos con figuras incaicas sencillas especie de soles o rayos solares y flores, el cuello cerrado con ribetes de cinco centímetros de altura bordados con lanas de color.
Chumpi: Es una faja tejida de varios colores, para dar dos vueltas alrededor de la cintura, sirve para sostener la allmilla al cuerpo. Tiene dibujos similares al del hombre.
Lijlla o Rebozo: Manta prendida con alfileres que sirve de protección al frio.
Ajsu: Es un vestido que le sirve de saya, se ponen como sotana cubriendo de la cintura a los pies.
Aguayo o Mantilla: Sirve para envolver y se usa para cargar al niño en la espalda.
Sombrero o Toco: Sombrero de paño, el cual identifica a los Jalq'as, es de copa pequeña de color blanco, más grande que de los varones.
Las mujeres casadas llevan en el sombrero una cinta gruesa, en cambio las solteras dos cintas de color.
Ojotas: Son similares al de los hombres.
Chuspa: Bolsa con características similares al de los hombres.

Instrumentos y tipo de música

La música que acompaña esta danza es originaria, se manifiesta a través de charangos, flautas, erkes, bombo, zampoña, guitarrilla, caja, cornetines y pequeños tambores, componiendo hermosos versos de acuerdo a sus ocasiones y fiestas en el cual se expresa el respeto sentimental.

Ocho años de participación

Son ocho años de participación de la Carrera con la danza **"Potolos de Estadística"** en la entrada universitaria, cuenta con al menos 140 fraternos, de los cuales 56 son integrantes entre estudiantes y docentes de la Carrera de Estadística.

Desde el inicio de su participación hasta la fecha, son tres trofeos los que se han ganado como premiación, el año 2003 en segundo lugar en la Entrada Universitaria, el año 2006 se obtuvo el primer lugar en la categoría, y el año 2009 se logró el tercer lugar en la categoría "Autóctona con Banda".

La primera ñusta el año 2002 recargó en la representación de la Univ. Paola Aramayo. Las ñustas de la fraternidad, lograron estar entre las diez finalistas del certamen Ñusta Universidad Mayor de San Andrés quienes fueron, la Univ. Claudia Fernández D. el año 2003, y el año 2010 la Univ. Carla Choque S.

Conclusión

Es un orgullo para la Carrera de Estadística ser protagonista de transmitir y rescatar un patrimonio de nuestro país con una danza no muy conocida en nuestro ámbito, como es el de los Potolos.

Es una satisfacción el cumplir con nuestra sociedad, promoviendo la valoración de nuestras raíces y ancestros culturales con el fin de enriquecer el patrimonio cultural de nuestro país.

Índice

"La cultura es el producto del pensamiento reflexivo, cultura es identidad, todos somos parte de la cultura, valoremos lo nuestro y defendamos nuestra identidad"

XII Congreso Nacional de Estudiantes de Estadística

"La Estadística como herramienta de cambio frente a un mundo globalizado"

Los estudiantes de Estadística de la Universidad Nacional de Piura Perú, recibieron con los brazos abiertos a docentes y estudiantes de diferentes lugares. Doce delegaciones de universidades de Perú y dos delegaciones de Bolivia participaron del "XII Congreso Nacional de Estudiantes de Estadística" en Piura Perú, del 13 al 18 de Septiembre del presente año.

La delegación boliviana estaba conformada por seis estudiantes y un docente de la Universidad Mayor de San Andrés, diez estudiantes y dos docentes de la Universidad Autónoma Tomás Frías de Potosí.

Se destacó la presencia del Director del Instituto Nacional de Estadística e Informática (INEI) Msc. Renán Quispe Llanos, quién fue el invitado de honor del XII CONEEST.

Participaron un total de veintidós expositores de diferentes países; quienes brindaron ponencias magistrales, y algunos de ellos dictaron minicursos. El detalle de los temas y ponentes fueron los siguientes:

- Construcción de Pruebas Psicológicas (Toshiya Tanaka de la Universidad de Kansai de Japón) ,
- Métodos Estadísticos Para Análisis de Vida de Supervivencia (Luis Alberto Escobar de EE.UU),
- El Modelo Jerárquico Bayesiano Robusto y sus Aplicaciones a Meta-Análisis, Ensayos Clínicos y Predicción (Luis Pericchi Guerra de Puerto Rico),
- Estadística Industrial (Enrique Raúl Villa Diharce de México),
- Análisis de Supervivencia con EPI INFO (Diego Coria Villca de Bolivia),
- Análisis de Datos Faltantes (Luis Mauricio Castro Cepero de Chile),
- Modelos Discretos y Continuos con y sin Regresores (Heleno Bolfarine de Brasil),

- Técnicas de Data Mining con SAS Enterprise Miner (Luis Cajachahua Espinoza de la UNI),
- Introducción al Análisis de Datos Longitudinales (Juan de Jesús Sandoval de Colombia),
- Modelos de Series de Tiempo de Memoria Larga ARFIMA con aplicaciones en R (Mario Piscoya de Brasil),
- Muestreo Complejo con SPSS (Fernando Rivero Suguiura de Bolivia),
- Manejo de G-Stat (Víctor Sánchez Cáceres de la UNC).

Además de las conferencias y los mini-cursos también se efectuaron visitas técnicas a diversas empresas de la región, orientadas al control estadístico, como ser: Caña Brava, Alicorp, Eco-acuícola, AJE PER, entre otras.

Otra actividad digna de destacar, fue el concurso de conocimientos sobre Estadística realizado entre los estudiantes de las universidades presentes; así como la realización de diferentes actividades deportivas, lo cual logró generar un ambiente cordial y amistoso entre las delegaciones presentes.

El próximo año, se realizará el XIII Congreso en la ciudad de Cuzco Perú, al cual se espera una nueva participación boliviana.

Índice

IX Congreso Latinoamericano de Sociedades de Estadística - CLATSE

Luego de la participación en Piura Perú, una delegación de ocho estudiantes de la Carrera de Estadística de la UMSA, partió hacia Valparaiso Chile para participar del "IX Congreso Latinoamericano de Sociedades de Estadística - CLATSE" realizado entre el 19 al 22 de octubre del año en curso.

Tras un largo viaje hasta esa ciudad chilena, nuestros estudiantes pudieron compartir, durante el Congreso, con un grupo de docentes de diferentes especialidades de la Estadística y estudiantes de distintos países de Latinoamérica.

Fueron invitados a ser partícipes de este evento internacional, sociedades, instituciones y universidades vinculadas a la Estadística, y varios profesores expertos de Brasil, México, Paraguay, Chile y Venezuela, entre otros, en diversas áreas de la Estadística, los cuales demostraron calidad y experiencia al momento de exponer los temas de interés.

Concurrieron al evento, la Sociedad Argentina de Estadística-SAE, la Sociedad Chilena de Estadística-SOCHE y la Sociedad Uruguaya de Estadística-SUE; además de las universidades de Valparaíso y Pontificia Universidad Católica de Valparaíso.

El CLATSE ha sido una experiencia importantísima para todas aquellas personas presentes, que comparten el amplio mundo de la Estadística en todas sus áreas de estudio; pues se trataron temas en el área de análisis multivariante, series de tiempo, estadística bayesiana (de concurrida participación), muestreo, modelos lineales y no lineales, estadística educativa (con nuevas técnicas de enseñanza y visión); así como las exposiciones de nuevas investigaciones, realizadas tanto por las Sociedades como por las Universidades participantes.

Felicitaciones a nuestros estudiantes y docentes participantes de ambos eventos internacionales, ojala, este sea un paso inicial para nuevas participaciones, que permiten ampliar los criterios y conocimientos de la Estadística, y porque no, dar a conocer nuestra Carrera en el ámbito internacional.

Que este tipo de eventos donde participan estudiantes y docentes de la Carrera, se intensifique el próximo año, con el propósito de abrir caminos a un mejor futuro profesional.

Índice

"La estadística ha demostrado que la mortalidad de los militares aumenta perceptiblemente durante tiempos de guerra."

Alphonse Allais

Aniversario de la Carrera de Estadística

Del 11 al 15 de Octubre del presente año, la Carrera de Estadística de la Universidad Mayor de San Andrés, celebró su 26 aniversario de existencia.

Se realizaron diferentes actos en su conmemoración. Se inició el día lunes 11 con la inauguración y la presentación de proyectos de investigación e interacción social de los docentes de la Carrera y el Instituto de Estadística Teórica y Aplicada (I.E.T.A.), por la tarde y durante la semana los docentes y estudiantes participaron en competencias deportivas.

Los días siguientes se efectuaron ponencias del que hacer estadístico en diferentes áreas. Estuvieron como invitados de honor el M.Sc. Ernesto Cupé C. y el Lic. Diego Coria V.; docentes de la Carrera de Matemática de la Universidad Mayor de San Andrés y de la Carrera de Estadística de la Universidad Autónoma Tomás Frías de Potosí, respectivamente.

El día martes 12 se realizó la elección de Miss Estadística, siendo elegida la señorita Univ. Vanesa Morales; de igual manera se realizó la elección de Mister Estadística, siendo electo el Univ. Danny Aguilar y también se efectuó la coronación Bufa.

Al día siguiente se realizó el bautizo a los estudiantes nuevos de la Carrera; y culminaron los festejos con la realización de una reunión social de confraternización.

Varianza, Covarianza.....

Estadística es la Esperanza!!!

FELICIDADES CARRERA....!!

"La esencia de la vida es la improbabilidad estadística a escala colosal."

Richard Dawkins

Índice

Día Mundial de la Estadística

En su 41ª sesión, celebrada del 23 al 26 de febrero de 2010, la Comisión de Estadística de las Naciones Unidas, convino en fijar el 20 de octubre como fecha conmemorativa del primer Día Mundial de la Estadística, haciendo suyos el tema general de celebración de los numerosos logros de las estadísticas oficiales y los valores básicos de servicio, integridad y profesionalidad. Con ello se pretende contribuir a incrementar la visibilidad y la confianza de la sociedad en la estadística oficial.

Las estadísticas son cruciales para el desarrollo económico y social, contribuyen al progreso de nuestra sociedad, siendo indispensables para la investigación, en el planteamiento de políticas públicas sociales y económicas, para el desarrollo de la sociedad civil.

De hecho, las estadísticas son de utilidad para todos los integrantes de nuestra sociedad. El desarrollo tecnológico ha hecho crecer el colectivo de usuarios que pueden acceder a la información estadística al tiempo que la globalización hace que las necesidades de información no queden exclusivamente restringidas al ámbito gubernamental o político. En nuestros días, las estadísticas oficiales influyen en las decisiones de todos los integrantes de nuestra sociedad. Dichas decisiones están basadas en la evidencia disponible (las estadísticas), pudiendo verse afectadas, por tanto, si medimos mal o si no existe plena confianza en las estadísticas. De ahí la importancia de esta última y de la credibilidad de las estadísticas oficiales.

Las estadísticas oficiales deben servir para atender los nuevos retos que se plantea, aspecto de especial relevancia en este momento en el que se ha abierto un cierto debate sobre cómo debe abordarse la medición del progreso de las sociedades y acerca del papel que juega la Estadística a la hora de medir el bienestar. De ahí la oportunidad de la celebración de este primer Día Mundial de la Estadística, dedicado a la estadística oficial, que debe servir para renovar y ahondar el compromiso de servicio, integridad y profesionalidad que el desarrollo de esta actividad pública requiere.

Por su parte, la Carrera de Estadística no quedó al margen de ésta celebración, puesto que la Carrera aporta a la Universidad y a la sociedad con conocimiento, teórico y práctico y con el desarrollo de la investigación y las actividades de interacción social.

A iniciativa del Instituto Nacional de Estadística (INE) se llevó a cabo una feria en el atrio de la Universidad Mayor de San Andrés, en la cual participaron instituciones gubernamentales como el

INE, Ministerio de Planificación del Desarrollo, Ministerio de Educación, Ministerio de Salud y Deportes, Ministerio de Economía y Finanzas Públicas, Banco Central de Bolivia, UDAPE y las Carreras de Economía y Estadística de la UMSA.

La Carrera de Estadística y el Instituto de Estadística Teórica y Aplicada (I.E.T.A.) expusieron específicamente los objetivos de la Carrera, su plan de estudios, el perfil del profesional estadístico, las diferentes ediciones de la Revista Varianza, los proyectos de investigación del Instituto y algunas tesis de aplicación estadística.

"Hagamos de este histórico Día Mundial de la Estadística como un éxito por el reconocimiento y la celebración de la función de las estadísticas en el desarrollo económico y social de nuestras sociedades y por dedicar más esfuerzos y recursos para fortalecer la capacidad estadística nacional"

Ban Ki-Moon (Secretario General de las Naciones Unidas)

Índice

Saben por qué Dios jamás recibió una cátedra en una Universidad ?

- *Sólo tiene una publicación importante*
- *Está escrita en hebreo*
- *No tiene referencias*
- *Además, hay quienes dudan que el fuese el autor*
- *Si, es posible que haya creado el universo, pero no ha publicado los resultados*
- *Los científicos han tenido problemas para demostrar experimentalmente su creación*
- *Y lo más difícil, resulta complicado trabajar con él*