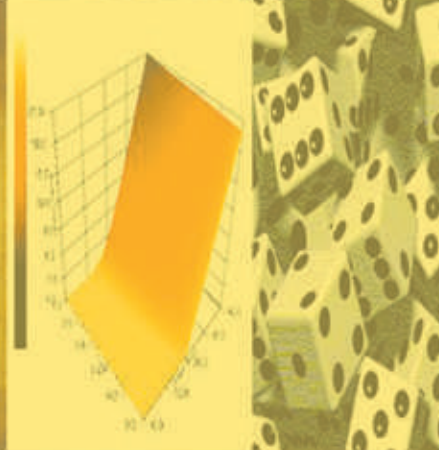




Universidad Mayor  
de San Andrés

# Varianza

*Revista de la Carrera de Estadística*





# Varianza

## Revista de la Carrera de Estadística

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

*Número 8  
Noviembre de 2011*

Revista Varianza  
N° 8 - Noviembre, 2011

Director del I.E.T.A.  
Fernando O. Rivero Sugiura

Diagramación y Diseño  
Enrique J. Villagra Arancibia

Redacción y Edición  
Consuelo Barrios  
Patricia Loza  
Beatriz Mamani  
Enrique Villagra  
Fernando Rivero

Colaboradores  
Juan Carlos Flores López  
Nicolás Chávez Quisbert  
Raul Delgado Alvarez  
Veronica Cuenca Ramallo  
Jaime Pinto Ajhuacho  
Deyvis Nina Canaviri

## Presentación

La Carrera de Estadística dependiente de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés (U.M.S.A.), realiza investigaciones anuales con la participación activa de sus docentes, el Instituto de Estadística Teórica y Aplicada (I.E.T.A.) y nuestros estudiantes.

Como se sabe, toda investigación realizada para que sea productiva necesariamente debe ser publicada para el conocimiento de nuestros estudiantes y el reconocimiento de nuestra sociedad, con la mejor intención de que sea un aporte de las distintas metodologías que tiene la estadística. Es así que nuestra carrera presenta los resultados en la revista VARIANZA, cada gestión. Esta publicación nos permite presentar los aportes teórico práctico y los resultados de las investigaciones realizadas en la gestión 2011 de la Carrera de Estadística. Esperando que sea de utilidad para la comunidad universitaria y el conjunto de nuestra sociedad, damos a conocer los distintos temas de investigación desarrollados. La dirección a mi cargo agradece y reconoce la colaboración de nuestros docentes, estudiantes por el aporte de sus investigaciones así también al personal de la carrera, sin cuya participación no habría sido posible la realización de la presente publicación.

Lic. Juan Carlos Flores López  
DIRECTOR 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 - 2612816

*Dedicada a:*

*Los estudiantes universitarios de la Carrera de Estadística, por su perseverancia en pos del conocimiento y aceptar los retos y problemas que se tienen que enfrentar en el largo camino de la profesión.*



## *Contenido*

<b>Investigación</b> .....	<b>1</b>
Modelación Estocástica de Series de Tiempo Univariantes con Influencia de Tendencia y Estacionalidad. M.Sc. Nicolás Chávez Quisbert.....	1
Identificación de Coordenadas de las Ciudades Capitales de Bolivia utilizando el Escalamiento Multidimensional. Lic. Fernando O. Rivero Suguiura .....	4
Regresión Logística. Lic. Verónica Cuenca Ramallo .....	9
Soluciones Alternativas. Lic. Raúl Delgado Álvarez .....	12
Pautas de una Encuesta por Muestreo. Lic. Jaime Pinto Ajhuacho .....	15
Modelos de Programacion Lineal Y No Lineal con Multiobjetivos. M.Sc. Nicolas Chavez Quisbert .....	19
Análisis de Medidas Repetidas mediante Métodos de Máxima Verosimilitud. Deyvis Nina Canaviri .....	22
El nuevo I.E.T.A.....	26
Estadística Clásica e Inferencia .....	28
<b>Historia</b> .....	<b>38</b>
Historia de la Estadística .....	38
<b>Curiosidades</b> .....	<b>47</b>
Estadísticas del Titanic .....	47
Paradoja de los Cumpleaños .....	52
Paradoja de la Placa de Automóvil.....	55
Infalible Probabilidad Bayesianas en las Series de Aventuras de la Televisión.....	56

Mezclando los naipes 7 veces .....	58
Sumando las caras ocultas de los dados .....	58
La Martingala .....	59
Estadísticas del 11-S a 10 años de los atentados .....	60
<b>Opinión.....</b>	<b>69</b>
El futuro Censo de Población y Vivienda en nuestro país.....	69
<b>Actividades.....</b>	<b>71</b>
XXIV versión de la Entrada Universitaria .....	71
Elecciones para Director de la Carrera de Estadística.....	74
Homenaje a un gran maestro de la Estadística. Luis Zapata Escobar .....	75
Dia Mundial de la Estadística.....	78
XXVII Aniversario de la Carrera de Estadística .....	80
<b>Anexos .....</b>	<b>85</b>

## Modelación Estocástica de Series de Tiempo Univariantes con Influencia de Tendencia y Estacionalidad

**Autor: M.Sc. Nicolás Chávez Quisbert**

El modelo más apropiado basado en los procesos estocásticos para modelar series de tiempo con dichas características es:

$$\text{ARIMA}(p,d,p) * \text{SARIMA}(P,D,Q)$$

### 1. Modelo Multiplicativo

$$\text{ARIMA}(p, d, p) * \text{SARIMA}(P, D, Q)$$

El modelo multiplicativo sugerido por Box y Jenkins para explicar a series con ambos efectos, parte del siguiente hecho.

Sea el modelo estacional:

$$\phi_p(B^s) \nabla_s^D (Z_t - \mu) = \theta_q(B^s) \alpha_t \quad (1)$$

Donde  $\alpha_t$  es un proceso de no ruido blanco, sino más bien generado por un proceso ARIMA (p, d, p) de la forma:

$$\phi_p(B) \nabla^d \alpha_t = \theta_q(B) a_t \quad (2)$$

Con  $a_t$  proceso de ruido blanco.

De la expresión (2) despejando  $\alpha_t$  se tiene que:

$$\alpha_t = \frac{\theta_q(B) a_t}{\phi_p(B) \nabla^d} \quad (3)$$

Reemplazando la expresión (3) en (1) se obtiene el modelo multiplicativo.

$$\phi_p(B) \phi_p(B^s) \nabla^d \nabla_s^D (Z_t - \mu) = \theta_q(B) \theta_q(B^s) a_t \quad (4)$$

### 2. Identificación de un Modelo Multiplicativo

Esta etapa, es una de las más complicadas e

importantes, ya que aquí se decidirla el orden de los procesos autorregresivos y de medias móviles tanto en la parte regular como estacional, para un modelo estacional multiplicativo.

Para identificación de los modelos, es bueno hacer uso de las funciones de autocovarianza, y autocorrelación y autocorrelación parcial, los cuales permitirán determinar el orden de los procesos ya sea en la parte regular, y estacional.

La Función de Autocorrelación da los órdenes de la parte MA ya sea en la parte regular o estacional, es decir los valores de q y Q, y la Función de Autocorrelación Parcial da los órdenes de la parte AR ya sea en la parte regular o estacional, es decir los valores de p y P.

#### 2.1. Función Generadora de Autocovarianzas

Esta función permite obtener funciones de autocovarianzas para modelos multiplicativos estacionales.

Dada la sucesión de autocovarianzas  $\{Y_t\}$ , la función generadora de autocovarianzas viene a ser (GUERRERO: 2003)

$$Y(B) = \sum_{k=-\infty}^{\infty} Y_k B^k \quad (5)$$

En donde la autocovarianza de orden k se obtiene como coeficiente tanto de  $B^k$  como  $B^{-k}$

Sea el modelo:

$$\phi(B) \phi_s(B^s) \nabla^d \nabla_s^D (T(Z_t) - \mu_t) = \theta(B) \theta_s(B^s) a_t \quad (6)$$

Originado por el producto de los siguientes procesos:

$$\phi_s(B^s) \nabla_s^D (T(Z_t) - \mu_T) = \theta_s(B^s) \alpha_t \quad (7)$$





Donde  $\mu_T = E(T(Z_t))$ , con función de autocovarianza  $Y_1(B^s)$  y el proceso

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t \quad (8)$$

Con función de autocovarianza  $Y_2(B)$  y  $a_t$  proceso de ruido blanco.

Entonces el modelo multiplicativo estacional definido en (6) tiene como función de autocovarianzas a:

$$Y(B) = Y_1(B^s)Y_2(B) \quad (9)$$

Por ser independientes los procesos.

Reemplazando  $Y_1(B^s)$  y  $Y_2(B)$  como funciones de autocovarianzas similares a (3.18) se tiene:

$$Y(B) = \left( \sum_{k=-\infty}^{\infty} Y_{1,k} B^k \right) \left( \sum_{k=-\infty}^{\infty} Y_{2,k} B^k \right) \quad (10)$$

De (10) se puede obtener la varianza y las autocovarianzas del proceso multiplicativo definido dado en (6).

$$Y_0 = Y_{1,0} Y_{2,0} + 2 \sum_{i=1}^{\infty} Y_{1,i} Y_{2,i}$$

$$Y_k = Y_{1,0} Y_{2,k} + \sum_{i=1}^{\infty} Y_{1,i} (Y_{2,i-k} + Y_{2,i+k}) \quad K \neq 0, s, 2s, \dots$$

$$Y_s = \sum_{i=0}^{\infty} (Y_{1,(i+1)s} Y_{2,i} + Y_{1,i} Y_{2,(i+1)s})$$

$$Y_{2s} = Y_{1,s} Y_{2,s} + \sum_{i=0}^{\infty} (Y_{1,(i+2)s} Y_{2,i} + Y_{1,i} Y_{2,(i+2)s})$$

## 2.2. Verificación de Supuestos

Para la verificación de una serie de supuestos que deben cumplir los residuales, en el ajuste de un modelo multiplicativo estacional se deben calcular los siguientes estadísticos, los cuales permitirán determinar si alguna de las hipótesis se está violando.

### Hipótesis Primera

Los  $a_t$  tienen media cero y varianza constante:

Para probar determinar:

$$m(\hat{a}) = \sum_{t=t'}^N (\hat{a}_t / (N - d - p - (D + P)s)) \quad (11)$$

$$\hat{\sigma}_a^2 = \sum_{t=t'}^N (\hat{a}_t - m(\hat{a}))^2 / (N - d - p - q - (D + P + Q)s)$$

Donde  $t' = d + p + (D + P)s + 1$

La media de los residuales se dirá que es significativamente diferente de cero con un 95% de confianza si:

$$\frac{\{N - d - p - (D + P)s\}^{1/2} m(\hat{a})}{\hat{\sigma}_a} \geq 2$$

Y para verificar que los  $a_t$  tienen varianza constante, se debe graficar la serie de los residuales en función al tiempo.

### Hipótesis Segunda

Los  $a_t$  son mutuamente independientes. Para probar determinar: la función de autocorrelación de los residuos.

$$r_k(\hat{a}) = \frac{\sum_{t=t'}^{N-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=t'}^{N-k} \hat{a}_t^2} \quad \text{Con } k=1, 2, 3, \dots$$

$$\text{Var}(r_k(\hat{a})) = 1 / (N - d - p - (D + P)s) \quad (12)$$

Con un 95% de confianza se puede afirmar que la autocorrelación para un rezago de k-ésimo es significativamente diferente de cero sí:

$$|r_k(\hat{a})| \geq 2 \left\{ \text{var}(r_k(\hat{a})) \right\}^{1/2}$$

El test simultáneo de Pormenteau, para modelos estacionales multiplicativos lo presenta Jung y Box (BOX – JENKINS: 1970).

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$$

$$H_1 : \rho_1 \neq \rho_2 \neq \dots \neq \rho_m \neq 0$$

$$Q' = (N - d - p - (D + P)s + 2)^*$$

$$\sum r_k^2(\hat{a}) / (N - d - p - (D + P)s + k) \quad (13)$$

El cual se distribuye asintóticamente según una distribución  $J_i$  cuadrado

$$\chi^2(m - p - q - P - Q) \quad (14)$$

Si  $Q'$  calculado supera el teórico de tablas dado, entonces podrá rechazarse la hipótesis nula.

### 3. Bibliografía

[1] Box, G. E. P. Jenkins, G. M. “Time Series Analysis, Forecasting and Control”, Holden – Day; San Francisco. 1970

[2] Fernández – Corugedo “Curso Modelos Macroeconómicos para la Política Monetaria” Banco Central de la República de Argentina, Mimeo – Argentina, 2003

[3] Guerrero Víctor M. “Análisis Estadístico De Series De Tiempo Económicas”; Universidad Autónoma Metropolitana; México. 1991

[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

*“El muestreo no es la simple sustitución de una cobertura parcial para una cobertura total. Muestreo es la ciencia y el arte de controlar y medir la fiabilidad de la información estadística útil a través de la teoría de la probabilidad”*

*Deming (1950).*

## Identificación de Coordenadas de las Ciudades Capitales de Bolivia utilizando el Escalamiento Multidimensional

Autor: Lic. Fernando O. Rivero Suguira

### 1. Introducción

El “escalamiento multidimensional” es una técnica multivariante que tiene el objetivo de identificar la representación gráfica perceptual de un conjunto de objetos, cuando se conocen las similitudes entre dichos objetos.

Shepard demuestra empíricamente, que si se conoce una ordenación de las distancias entre objetos, podría encontrarse las coordenadas de estos objetos en un espacio euclidiano, cuyas interdistancias reproducen prácticamente la ordenación original. Las ideas de Shepard fueron después mejoradas por Kruskal en los años 60 y desarrolladas por otros autores como Guttman y Lingoes donde su aplicación se inicia en la Psicología y después a muchas áreas del que hacer científico.

El presente artículo presenta una aplicación de este método como parte del Análisis Multivariante de manera práctica y didáctica, en un ejemplo de la medición de distancias euclídeas entre las nueve ciudades capitales de nuestro país (Sucre, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Trinidad y Cobija), con el objetivo de identificar, luego las coordenadas de las ciudades nombradas sobre un plano, de tal manera que se pueda comprender mejor lo que hace metodológicamente el análisis de Escalamiento Multidimensional.

Sea la matriz  $D$  de distancias euclídeas de las nueve ciudades capitales de Bolivia, medido en centímetros, obtenidas de un mapa del país a escala 1:2'500'000, representado en menor escala en el gráfico N° 1. La matriz de distancias, es:

Ciudades	Sucre	La Paz	Cochabamba	Oruro	Potosí	Tarija	Santa Cruz	Trinidad	Cobija
Sucre	-	28,5	14,2	15,9	5,4	19	18,4	32,1	64,5
La Paz		-	16	13,2	29	45,1	37,7	28	40,3
Cochabamba			-	8,2	16,6	32,8	22,3	22,1	50,6
Oruro				-	15,7	32,1	29,2	29	52,3
Potosí					-	16,6	23,3	36,4	66,5
Tarija						-	30,5	50,4	83
Santa Cruz							-	25,3	64,6
Trinidad								-	40,7
Cobija									-

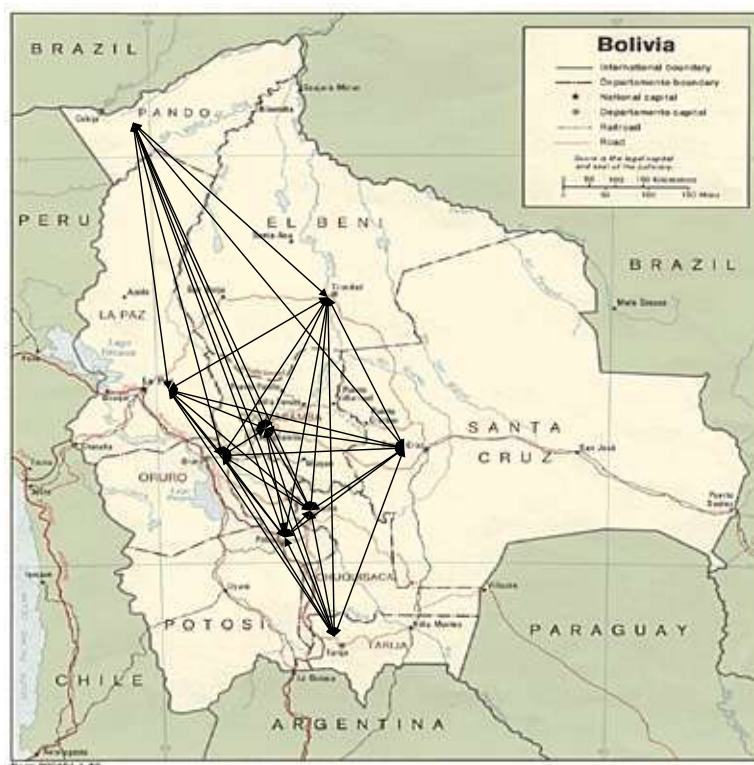
Fuente: Propia. Medición en centímetros de ciudad a ciudad de un mapa de Bolivia a escala 1:2'500'000.

Por ejemplo, 28.5 cm. corresponde a la distancia en línea recta entre las ciudades de Sucre y La Paz. 36.4

cm. es la distancia entre las ciudades de Potosí y Trinidad, y así sucesivamente.

Gráfico N° 1

Bolivia: Trazado de distancias entre las capitales de Departamento y medidas en centímetros (Escala inferior a 1:2'500'000).



## 2. Metodología

El método consiste en determinar la matriz de distancias euclidianas  $D$  a partir de una matriz de proximidades  $\Delta$ .

Sea la matriz  $A = -\frac{1}{2}D^{(2)}$  definida a partir de la matriz  $D$  de distancias euclídeas al cuadrado. Sea  $B = H'AH$  de dimensión  $9 \times 9$  denominada “matriz de productos escalares”, con  $H$  “matriz de centrado” donde  $H = I - n^{-1}11'$ .  $I$  matriz identidad,  $n$  número de objetos en la muestra, y  $1$  vector columna de “unos”. La matriz de centrado es simétrica e idempotente de rango  $n - 1$ , que al aplicarla sobre un vector de datos lo transforma en sus desviaciones respecto a la media de sus componentes. Es decir, cumple con las siguientes propiedades:

- i)  $H' = H$
- ii)  $H^2 = H$
- iii)  $H1 = 1'H = 0$
- iv) los valores propios de  $H$  están entre 0 y 1

Entonces, la matriz  $B$  queda como:

$$\begin{aligned} B &= H'AH \\ &= (I - n^{-1}11')'A(I - n^{-1}11') \\ &= A - n^{-1}A11' - n^{-1}11'A + n^{-2}11'A11' \end{aligned}$$

Es semidefinida positiva, puesto que:

$$B = HAH = -\frac{1}{2}HD^{(2)}H = XX' \geq 0$$

Donde  $X$  es matriz de dimensión  $n * p$  de “coordenadas” en el Análisis de Escalado Multidimensional, siendo  $p$  la dimensión de las coordenadas.

Por la descomposición espectral de una matriz, se tiene que:

$$B = P\Delta P' = XX' \geq 0$$

$\Delta$  es la matriz de valores propios normalizados de  $B$ . Con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > \lambda_{p+1} = 0$ , y la matriz  $P$  de dimensión  $n * p$ , es la que contiene los vectores propios asociados a los valores propios correspondientes. Es decir:

$$B = P\Delta P' = (P\Delta^{1/2})(P\Delta^{1/2})'$$

Con  $X = P\Delta^{1/2}$ . Las coordenadas principales resultantes de los objetos en estudio, tienen dimensión  $q < p$ .

Si  $q = 2$ , las dos primeras coordenadas de  $X$  proporcionan la representación gráfica en el plano. El gráfico N° 2, muestra la salida en SPSS de las coordenadas en un plano, resultante del ejemplo, siendo los primeros dos valores y vectores propios representativos, de la matriz  $B$ , los siguientes:

$$\lambda_1 = 4.50$$

$$v_1 = (-0.21 \quad 0.17 \quad -0.00 \quad -0.02 \\ -0.24 \quad -0.49 \quad -0.17 \quad 0.20 \quad 0.75)$$

$$\lambda_2 = 0.83$$

$$v_2 = (0.01 \quad 0.42 \quad 0.05 \quad 0.32 \quad 0.17 \\ 0.13 \quad -0.62 \quad -0.54 \quad 0.06)$$

En resumen el procedimiento consiste en transformar:

$\Delta$  (Proximidades)  $\rightarrow D$  (Distancias)  $\rightarrow B$  (Productos escalares)  $\rightarrow X$  (coordenadas)

Una medida que nos informa de la bondad del Modelo de Escalamiento Multidimensional en un plano, es el estadístico *Stress* que Kruskal definió como:

$$Stress = \sqrt{\frac{\sum_{i,j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j} d_{ij}^2}}$$

Mientras mayor sea la diferencia entre las disparidades y las distancias, mayor será el *Stress* y por tanto peor será el modelo. Kruskal (1964) sugiere las siguientes interpretaciones del *Stress*:

0.200	→ Pobre
- 0.100	→ Regular
- 0.050	→ Bueno
- 0.025	→ Aceptable
- 0.010	→ Excelente

El resultado del ejemplo, da como estadístico de *Stress* igual a:

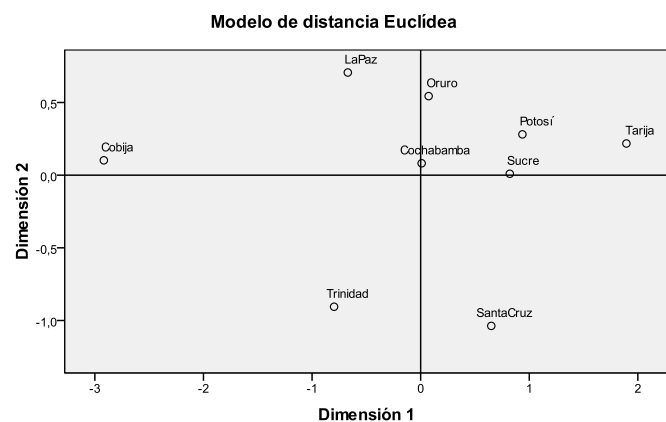
For matrix  
Stress = ,00170 RSQ = ,99999

De acuerdo a la clasificación anterior, está entre aceptable y excelente.

Gráfico N° 2

Diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz  $B$ , (salida en SPSS).

Configuración de estímulos derivada

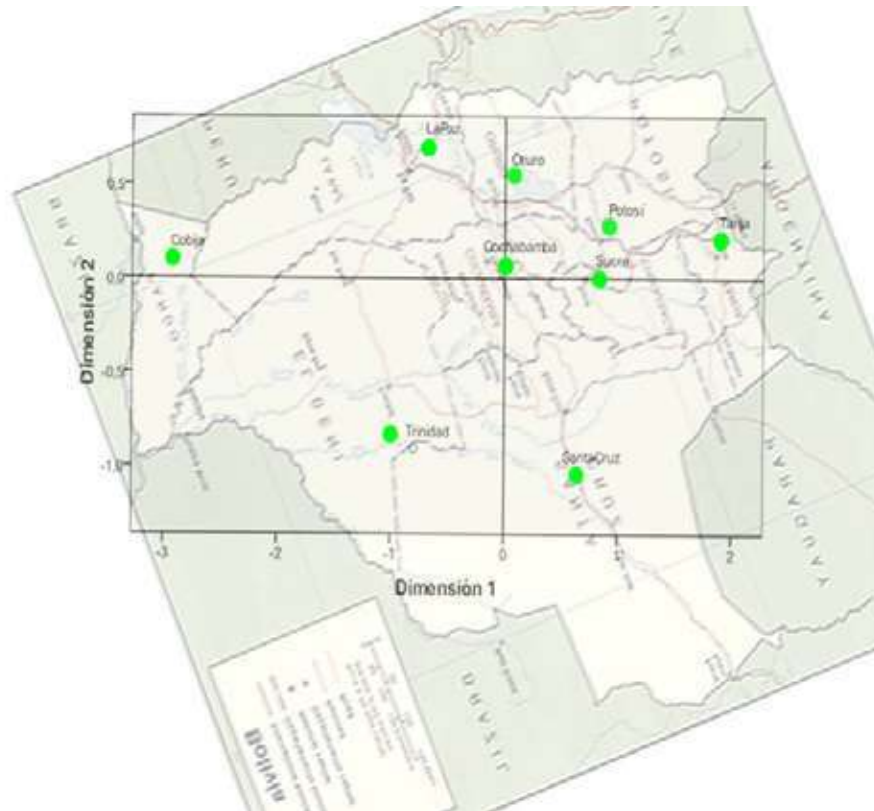


Gráficamente, este es el resultado de identificación de coordenadas de las ciudades capitales de Bolivia en un plano, según el método de Escalamiento Multidimensional.

Para comprender mejor su ubicación, se puede observar el Gráfico N° 3, donde se traslada y rota el mapa de Bolivia, haciendo coincidir con el Gráfico N° 2.

Gráfico N° 3

Diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz **B**, sobreponiendo el mapa de Bolivia ajustado al resultado.



Como se puede notar en el Gráfico N° 3, el punto de la ciudad de La Paz, aparece en la parte superior del cuadrante 2. Cerca al punto La Paz, aparece Oruro y luego el punto Cochabamba, Sucre, Potosí, Tarija, y así sucesivamente. Se puede observar cierta coincidencia y precisión en la ubicación de los puntos.

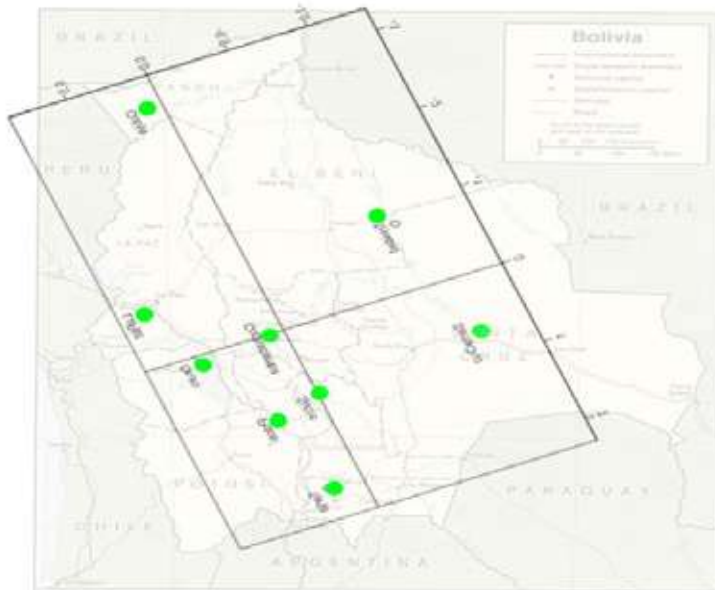
El Gráfico N° 4, muestra la rotación y traslación del dispersograma, hasta ubicar de forma cercana y precisa, los puntos de las ciudades capitales en el mapa de Bolivia.

*“No existe la suerte. Sólo hay preparación adecuada o inadecuada para hacer frente a una estadística”.*

*Robert Heinlein*

Gráfico N° 4

Mapa de Bolivia sobreponiendo el diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz **B** ajustado al resultado.



### 3. Conclusiones y Recomendaciones

El artículo es una demostración sencilla de un ejemplo didáctico, que da a comprender lo que realmente hace el Escalamiento Multidimensional con las distancias observadas de ciudad a ciudad y con el resultado de la identificación final de las posiciones o coordenadas de cada una de las nueve ciudades con alta aproximación. Para ello se utilizó un mapa de Bolivia a escala 1:2'500'000 representada a escala menor en el Gráfico N° 1.

A partir de este ejemplo, el analista se da cuenta de lo importante y aplicado que es este método para otro tipo de investigación. Por ejemplo, reagrupar gráficamente en un mapa perceptual, municipios de Bolivia que presentan niveles de pobreza altos, medios y bajos. Ayudaría bastante a comprender el agrupamiento que se da de estos municipios por una característica como es el nivel de pobreza.

El método requiere que la medición de las similitudes entre objetos o elementos, cumplan la condición de distancia euclidiana. Si no fuese así, se puede cometer errores en la aplicación del Escalamiento Multidimensional, y es que la matriz **B** de productos escalares puede no ser semidefinida positiva, y

complicar el análisis. Especialmente, las similitudes pueden no cumplir la propiedad de la desigualdad triangular  $d_{ij} \leq d_{ij'} + d_{j'j}$ . Sin embargo, existe un método denominado “Estimación de la Constante Aditiva” debido a Torgerson que soluciona el problema.

En la próxima versión de la Revista Varianza, se puede realizar un artículo aplicativo a un campo de investigación como el mencionado.

### 4. Referencias

- [1] S. JAMES PRESS (1948): “Applied Multivariate Analysis”, University of Chicago. New York.
- [2] K. V. MARDIA (1979): “Multivariate Analysis”, Academic Press. Inc. London.
- [3] DONALD F. MORRISON (1967): “Multivariate Statistical Methods”, Mc Graw Hill Book Company. New York.
- [4] CUADRAS, CARLES M. (2006): “Métodos de Análisis Multivariante”. Barcelona, España.
- [5] EZEQUIEL URIEL – JOAQUIN ALDAS (2005): “Análisis Multivariante Aplicado”. Thomson. España.

## Regresión Logística

*Autor: Lic. Verónica Cuenca Ramallo*

Un método conocido en un modelo de regresión es la regresión logística, este método está destinado a utilizarlo cuando se presente datos binomiales (dicotómicas) o multinomiales (multivariadas, politómicas).

En la actualidad existe muchos estudios sobre lo que es la regresión logística y sus aplicaciones en diferentes áreas, en este artículo se encargará esta regresión.

### 1. Introducción

En estadística, la regresión logística es un modelo de regresión para variables dependientes, es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. Es un modelo lineal generalizado que usa como función de enlace la función logit.

La regresión logística es usada extensamente en las ciencias médicas, sociales y otras, esto significa áreas del conocimiento humano donde las variables que se analizan son en su mayoría cualitativas.

Otros nombres para la regresión logística usados en varias áreas de aplicación incluyen:

#### 1.1 Modelo logístico, modelo logit, y clasificador de máxima entropía.

En los modelos de regresión se quiere conocer la relación entre:

- ❖ Una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial) o con más de dos valores (regresión logística multinomial).
- ❖ Una o más variables explicativas independientes, o covariables, ya sean cualitativas o cuantitativas.

Siendo la ecuación inicial del modelo de tipo exponencial, si bien su transformación logarítmica (logit) permite su uso como una función lineal.

Como vemos, las covariables pueden ser cuantitativas o cualitativas. Las covariables cualitativas deben ser dicotómicas, tomando valores 0 para su ausencia y 1 para su presencia (esta codificación es importante, ya que cualquier otra codificación provocaría modificaciones en la interpretación del modelo).

Pero si la covariable cualitativa tuviera más de dos categorías, para su inclusión en el modelo deberíamos realizar una transformación de la misma en varias covariables cualitativas dicotómicas ficticias o de diseño (las llamadas variables dummy), de forma que una de las categorías se tomaría como categoría de referencia. Con ello cada categoría entraría en el modelo de forma individual.

En general, si la covariable cualitativa posee  $n$  categorías, habrá que realizar  $n-1$  covariables ficticias.

En resumen el modelo de regresión logística se utiliza cuando estamos interesados en pronosticar la probabilidad de que ocurra o no un suceso determinado.

Por ejemplo, a la vista de un conjunto de pruebas médicas, que una persona tenga una determinada enfermedad, o bien que un cliente devuelva un crédito bancario, el comportamiento electoral, color de ojos etc.

### 2. El modelo de la regresión logística

El modelo de regresión en el caso multivariante se considera una agrupación de  $p$  variables independientes que está denotado a voluntad por el vector  $x' = (x_1, x_2, \dots, x_k)$ .

Se asume que cada uno de estas variables es una escala mínima de intervalo, permite la probabilidad condicional que la respuesta se presenta con la notación siguiente:

$$P(Y=1 | x) = \pi(x)$$



El modelo de regresión logística (logit) múltiple esta dada por la siguiente ecuación.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + \mu$$

Donde:

$$\pi(x) = \frac{e^Y}{1+e^Y}$$

La regresión logística analiza datos distribuidos binomialmente de la forma:

$$Y \sim B(p_i, n_i), i = 1, 2, \dots, m$$

$$Y \sim B(p_i, n_i), i = 1, 2, \dots, m$$

Donde los números de ensayos Bernoulli  $n_i$  son conocidos y las probabilidades de éxito  $p_i$  son desconocidas.

El modelo es entonces obtenido en base a lo que cada ensayo (valor de  $i$ ) y el conjunto de variables explicativas/independientes puedan informar acerca de la probabilidad final. Estas variables explicativas pueden pensarse como un vector  $X_i$   $k$ -dimensional y el modelo toma entonces la forma

$$p_i = E\left(\frac{Y_i}{n_i} | X_i\right)$$

Los logits de las probabilidades binomiales desconocidas es decir los logaritmos de los odds o ratios de probabilidad que indica cuanto se modifican las probabilidades por unidad de cambio en las variables  $x$ , son modeladas como una función lineal de los  $x_i$ .

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_{k,i}$$

Es importante notar que el elemento particular  $x_i$  puede ser ajustado a 1 para todo  $i$  obteniéndose un intercepto en el modelo. Los parámetros desconocidos  $\beta_j$  son usualmente estimados a través de máximo verosimilitud.

La interpretación de los estimados del parámetro  $\beta_j$  es como los efectos aditivos en el log odds ratio para una unidad de cambio en la  $j$ -ésima variable explicativa.

En el caso de una variable explicativa dicotómica, por ejemplo género,  $e^\beta$  es la estimación del odds ratio de tener el resultado para, poder decir algo, hombres comparados con mujeres.

El modelo tiene una formulación equivalente dada por:

$$p_i = \frac{1}{1 + e^{-\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}$$

Esta forma funcional es comúnmente identificada como una red neuronal de una sola capa, calcula una salida continua en lugar de una función por partes. La derivada de  $p_i$  con respecto a  $x = x_1, \dots, x_k$  es calculada de la forma general:

$$y = \frac{1}{1+e^{-f(x)}}$$

Donde  $f(X)$  es una función analítica en  $X$ .

3. Algunos aspectos a tener en cuenta para el uso de la regresión logística.

- ❖ Tamaño de la muestra y número de variables independientes.
- ❖ Una de las ventajas de la regresión logística es que permite el uso de múltiples variables con pocos casos esto es relativo, sin embargo hay que tener en cuenta algunas precauciones.
- ❖ Sugerencia el número de sujetos debe ser superior a  $10(k+1)$  donde  $K$  es el número de variables explicativas.

- ❖ Si se introducen variables Dummy, el número de elementos en la muestra debe aumentar.
  - ❖ Si una de las variables es dicotómica, además de respuesta esta no tiene al menos 10 casos en cada uno de sus dos respuestas posibles ocurre que sus estimaciones no son confiables.
  - ❖ Cuando las variables independientes se presenta en gran número, puede indicar que no se ha reflexionado suficientemente sobre el problema.
  - ❖ Se debe tener en cuenta el efecto sobre el riesgo de que ocurra el evento, de los cambios de las variables explicativas cuando son cuantitativas (continuas).
  - ❖ Cuando algunas de las variables independientes analizadas están altamente correlacionadas, los resultados que se obtienen pueden no ser satisfactorios.
  - ❖ Para no tener problemas en la aplicación primero se debe realizar un análisis previo univariado entre las distintas variables explicativas.
  - ❖ Para que la regresión logística tenga un sentido claro, debe existir una relación monótona entre las variables explicativas y la de respuesta, esto significa que el aumento de las unas se acompañe del aumento a lo disminución aproximadamente constante de la otra, para todo el rango de valores estudiados.
- ❖ El exponencial de los  $\beta_1$  se corresponde con el riesgo relativo, o sea, es una medida de la influencia de la variable  $x_1$  sobre el riesgo de que ocurra ese hecho y suponiendo que el resto de las variables del modelo permanezcan constantes.
  - ❖ Una vez estimados los valores de  $\beta_0$  y  $\beta_1$ , podemos determinar la probabilidad del suceso para distintos valores de los  $X_i$ .
  - ❖ Para las variables explicativas (categóricas) ya sean nominales u ordinales de más de 2 categorías (politómicas), para incluirlas en el modelo hay que darles un tratamiento especial.
  - ❖ Si estamos en presencia de una variable nominal con C categorías, debemos incluirla en el modelo de regresión logística como variable categórica, de manera que a partir de ella se crean C-1 variables dicotómicas llamadas *dummy* se debe precisar con cuál de las categorías de la variable original interesa comparar el resto y esa será la llamada categoría de referencia.
  - ❖ En el caso de las variables ordinales se puede asumir que la escala funciona aproximadamente a un nivel cuantitativo, se pueden manejarse como variables *dummy*.

#### 4. Algunas observaciones de tipo práctico

- ❖ Para una mejor interpretación de los coeficientes cualesquiera  $\beta_1$  es necesario referirnos al concepto riesgo relativo.

#### 5. Bibliografía.

[1] David W.Hosmer (1989) “APPLIED LOGISTIC REGRESSION “

[2] David W.Hosmer (2000) “LOGISTIC REGRESSION”

NOTA.-El ejemplo aplicativo se realizará en la siguiente revista N° 9 se tratará para casos dicotómicos y casos multivariados, también se explicará la terminología.

## Soluciones Alternativas

Autor: Lic. Raúl Delgado Álvarez

Cuando nos preguntamos quienes son los lectores de la revista de la carrera, con seguridad son los estudiantes de Estadística entre otros lectores, es innovador que una actividad deseable en los estudiantes debe ser la de buscar soluciones distintas a las mostradas en clases que por años puede ser la misma, este aspecto psicopedagógico del aprendizaje permite comprender, hacer suyo el conocimiento, por lo tanto, mostrar que existen soluciones alternativas a problemas lo cual afianza los conocimientos, permite sostener la comprensión de la teoría en problemas prácticos, convencido de este proceso es importante para la investigación, me permito plantear un problema y los argumentos necesarios para su solución de dos maneras.

### PROBLEMA

Demostrar que el coeficiente de curtosis (de Fisher) para la Distribución Binomial de parámetros  $n$  y  $p$  es:

$$g_2 = 3 + \frac{1 - 6p + 6p^2}{np(1-p)}$$

Primera solución

La función denominada de los Cumulantes definida como:

$$C_x(t) = \ln [M_x(t)]$$

goza de las propiedades:

- (1)  $C_x'(0) = E[x]C_x'(0) = E[x]$
- (2)  $C_x''(0) = V[x]C_x''(0) = V[x]$
- (3)  $C_x'''(0) = \mu_3 C_x'''(0) = \mu_3$
- (4)  $C_x^{iv}(0) = \mu_4 - 3(\mu_2)^2$   
 $C_x^{iv}(0) = \mu_4 - 3(\mu_2)^2$

Dada la f.g.m. de la Distribución Binomial

$$M_x(t) = [1 - p + pe^t]^n M_x(t) = [1 - p + pe^t]^n,$$

Haciendo  $q = 1 - p$   $q = 1 - p$

$$M_x(t) = (q + pe^t)^n$$

aplicando logaritmos

$$\ln[M_x(t)] = n \ln(q + pe^t)$$

derivando:

$$C_x'(t) = \frac{n}{q+pe^t} pe^t = npe^t(q + pe^t)^{-1} \quad (1)$$

$$C_x''(t) = np \{e^t(q + pe^t)^{-1} + e^t(-1)(q + pe^t)^{-2}pe^t\}$$

$$C_x''(t) = npqe^t(q + pe^t)^{-2} \quad (2)$$

$$C_x'''(t) = npq \{e^t(q + pe^t)^{-2} + e^t(-2)(q + pe^t)^{-3}pe^t\}$$

$$C_x'''(t) = npq(qe^t + pe^{2t})(q + pe^t)^{-3} \quad (3)$$

$$C_x^{iv}(t) = npq \{(qe^t - 2pe^{2t})(q + pe^t)^{-3} + (qe^t - pe^{2t})(-3)(q + pe^t)^{-4}pe^t\}$$

$$C_x^{iv}(t) = npq[(qe^t - 2pe^{2t})$$

$$(q + pe^t)^{-3} - 3pe^t(qe^t - pe^{2t})(q + pe^t)^{-4} \quad (4)$$

Evaluando (2) y (4) para  $t=0$  se tiene:

$$C_x''(0) = npq = np(1-p) = V[x] \text{ ó } \mu_2 = np(1-p)$$

$$C_x^{iv}(0) = npq\{(q-2p) - 3(q-p)p\}$$

volviendo a  $q = 1-p$

$$C_x^{iv}(0) = np(1-p)(1-6p+6p^2)$$

$$\rightarrow \mu_4 = C_x^{iv}(0) + 3(\mu_2)^2$$

$$\mu_4 = C_x^{iv}(0) + 3(\mu_2)^2$$

$$\mu_4 = np(1-p)(1-6p+6p^2) + 3[np(1-p)]^2$$

$$\mu_4 = 3n^2p^2(1-p)^2 + np(1-p)(1-6p+6p^2)$$

El coeficiente de Curtosis de Fisher es:

$$g_2 = \frac{\mu_4}{(\mu_2)^2} \rightarrow g_2 = \frac{3n^2p^2(1-p)^2 + np(1-p)(1-6p+6p^2)}{n^2p^2(1-p)^2}$$

$$g_2 = 3 + \frac{1-6p+6p^2}{np(1-p)}$$

### Segunda Solución

Una solución alternativa para encontrar el coeficiente de curtosis para la Distribución Binomial de parámetros  $n$  y  $p$  es aplicar la fórmula de recurrencia para los momentos centrales

$$\mu_{r+1} = p(1-p)[nr\mu_{r-1} + \frac{d\mu_r}{dp}]$$

cuya demostración se basa en la derivada del momento central de orden  $r$ .

### Demostración

$$\begin{aligned} \mu_r &= E[x = E(x)]^r \\ &= \sum_{x=0}^n (x - np)^r \binom{n}{x} p^x (1-p)^{n-x} \end{aligned}$$

Derivando:

$$\begin{aligned} \frac{d\mu_r}{dp} &= \sum_{x=0}^n \binom{n}{x} \frac{d}{dp} \{(x - np)^r p^x (1-p)^{n-x}\} \\ &= \sum_{x=0}^n \binom{n}{x} \{r(x - np)^{r-1}(-n)p^x(1-p)^{n-x} + (x - np)^r x p^{x-1}(1-p)^{n-x} + (x - np)^r p^x (n-x)(-1)(1-p)^{n-x-1}\} \\ &= \sum_{x=0}^n \binom{n}{x} \{-nr(x - np)^{r-1}p^x(1-p)^{n-x} + (x - np)^r [(x - np) + np] p^{x-1}(1-p)^{n-x} - (n-x)(x - np)^r p^x(1-p)^{n-x-1}\} \end{aligned}$$

$$\begin{aligned} &= \sum_{x=0}^n \binom{n}{x} \{-nr(x - np)^{r-1}p^x(1-p)^{n-x} + (x - np)^{r+1}p^{x-1}(1-p)^{n-x} + np^{x-1}(1-p)^{n-x}(x - np)^r - n(x - np)^r p^x(1-p)^{n-x-1} + [(x - np) + np](x - np)^r p^x(1-p)^{n-x-1}\} \end{aligned}$$

Introduciendo el símbolo sumatorio se tiene:

$$= -nr \sum_{x=0}^n \binom{n}{x} \{(x - np)^{r-1}p^x(1-p)^{n-x} + \frac{1}{p} \sum_{x=0}^n \binom{n}{x} (x - np)^{r+1}p^x(1-p)^{n-x} - \sum_{x=0}^n \binom{n}{x} (x - np)^r p^x(1-p)^{n-x-1}\}$$

$$+n \sum_{x=0}^n \binom{n}{x} (x-np)^r p^x (1-p)^{n-x} - \frac{n}{1-p} \sum_{x=0}^n \binom{n}{x} (x-np)^r p^x (1-p)^{n-x} + \frac{1}{1-p} \sum_{x=0}^n \binom{n}{x} (x-np)^{r+1} p^x (1-p)^{n-x} - \frac{np}{1-p} \sum_{x=0}^n \binom{n}{x} (x-np)^r p^x (1-p)^{n-x}$$

Identificando los momentos centrales:

$$\frac{d\mu_r}{dp} = -nr\mu_{r-1} + \frac{1}{p}\mu_{r+1} + n\mu_r - \frac{n}{1-p}\mu_r + \frac{1}{1-p}\mu_{r+1} + \frac{np}{1-p}\mu_r$$

$$= -nr\mu_{r-1} + \frac{1-p+p}{p(1-p)}\mu_{r+1}$$

$$+n\mu_r - \left(\frac{n-np-n+np}{1-p}\right)\mu_r$$

$$\frac{d\mu_r}{dp} = -nr\mu_{r-1} + \frac{\mu_{r+1}}{p(1-p)} \rightarrow \mu_{r+1} = p(1-p)\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]$$

Aplicando la formula de recurrencia

$$\mu_{r+1} = p(1-p)\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]$$

$$\mu_0 = E[x - E(x)]^0 = 1$$

$$\mu_0 = 1 \quad \frac{d(1)}{dp} = 0$$

$$\frac{d\mu_0}{dp} = 0;$$

$$\mu_1 = p(1-p)\left[n(0)\mu_0 + \frac{d\mu_0}{dp}\right]$$

$$= p(1-p)[0+0] = 0$$

$$\mu_2 = p(1-p)[n(1)(\mu_0) + 0]$$

$$= p(1-p)n$$

$$= np(1-p)$$

$$\mu_3 = p(1-p)\left[n(2)(\mu_1) + \frac{d\mu_2}{dp}\right]$$

$$= p(1-p)[0+n(1-2p)]$$

$$\mu_3 = p(1-p)n(1-2p)$$

$$= n[(p-p^2)(1-2p)]$$

$$\mu_3 = n(p-p^2-2p^2+2p^3)$$

$$= n(p-3p^2+2p^3)$$

$$\frac{d\mu_3}{dp} = n(1-6p+6p^2)$$

$$\mu_4 = p(1-p)\left[n(3)\mu_3 + \frac{d\mu_3}{dp}\right]$$

$$= p(1-p)[3n^2p(1-p)$$

$$+n(1-6p+6p^2)]$$

$$\mu_4 = np(1-p)[3np(1-p)$$

$$+(1-6p+6p^2)]$$

Aplicando los momentos centrales, en el coeficiente de curtosis:

$$g_2 = \frac{\mu_4}{(\mu_2)^2}$$

$$= \frac{np(1-p)[3np(1-p) + (1-6p+6p^2)]}{n^2p^2(1-p)^2}$$

$$g_2 = \frac{[3np(1-p) + (1-6p+6p^2)]}{np(1-p)}$$

$$g_2 = 3 + \frac{1-6p+6p^2}{np(1-p)}$$

## *Pautas de una Encuesta por Muestreo*

*Autor: Lic. Jaime Pinto Ajhuacho*

En la investigación social se desea conocer características de una Población, puede llevarse a cabo sobre el total de la población, labor generalmente muy compleja y costosa o también puede realizarse sobre una parte de esa población, a la que se le denomina “muestra”.

La teoría del muestreo, proporciona toda una serie de técnicas que nos permiten conocer los parámetros poblacionales, es mas ofrecen metodologías de captura de datos; su objetivo del muestreo es hacer inferencias acerca de una población con base en la información contenida en una muestra.

### 1. Etapas de una encuesta por muestreo

Los pasos involucrados en la planeación y ejecución de una encuesta por muestreo pueden variar de acuerdo a su complejidad.

Los principales pasos de una encuesta por muestreo son los siguientes:

- Consideraciones generales:  
Objetivos de la encuesta - ambitos (Poblacional, Geográfico y Temporal).
- Metodología:  
Contenido temático – Variables - Conceptos Teóricos - Definiciones - Periodos de Referencia - Métodos de recolección de datos - Descripción de la boleta -Organización del material de encuesta.
- Diseño de la muestra:  
Marco muestral - Tipo de muestreo - Tipo de muestreo - Tamaño muestral - Período muestral Selección de la muestra - Método de estimación – Error de Muestreo.

- Operativo de campo:  
Reclutamiento del personal - Capacitación - Organización del operativo de campo - Relevamiento de la información.
- Revisión y procesamiento:  
Proceso de crítica y codificación - Organización y control de boletas - Vaciado de los datos – Validación y Consistencia.
- Análisis de los datos:  
Vaciado de Plan de Tabulados - Cálculo de estimaciones - Proceso de armado e informe de encuesta.
- Publicación de datos.  
Ampliando los conceptos de los pasos en la encuesta por muestreo, tendremos:

### 2. Planificación de la encuesta

La encuesta inicialmente se reduce a una etapa pensante en la que un equipo de personas diseña las características y procedimientos de lo que es el tema a investigar y la forma de recoger la información, analizando al mínimo detalle, para este propósito se destinaran un tiempo, espacio y recursos a utilizar.

### 3. Objetivos de la investigación

Una vez asociadas las ideas en torno al tema, la misma se ve expresada en lo que se conoce como objetivos de la investigación, que expresara lo que se requiere para conocer características del universo de estudio.

### 4. Objetivos de la encuesta

Una exposición clara de los objetivos es lo más útil de lo contrario, es fácil olvidarlos en una encuesta compleja al preocuparse por los detalles de la

planeación y por lo tanto tomar decisiones que se aparten de los objetivos.

## 5. Población bajo muestreo

Se debe definir cuidadosamente la población que va a ser muestreada; por ejemplo si son los jóvenes los que van a ser muestreados, entonces se debe definir qué se entiende por joven (todos aquellos entre 18 y 25 años de edad, por ejemplo).

## 6. Ámbitos

a).- **Ámbito geográfico.**- Este Ámbito de la encuesta corresponde al área física, relación cartográfica del universo de estudio (ejemplo el área urbana de la ciudad de La Paz ).

b).- **Ámbito poblacional.**- La Población o universo representa el total de unidades objeto de estudio, las que pueden estar constituidas por todas las personas de un país, una zona geográfica o un grupo étnico, dependiendo de la cobertura de la muestra.( por ejemplo, el ámbito poblacional para una encuesta a hogares en la ciudad de La Paz, está identificado como todos los hogares residentes en viviendas particulares).

c).- **Ámbito temporal.**- Se refiere al tiempo que durara la encuesta, es decir se refiere al carácter puntual de la Encuesta, (por ejemplo el operativo de campo de la encuesta puede durar aproximadamente un mes dependiendo del tamaño de la muestra para la ciudad de La Paz).

## 7. Contenido temático

Es en esta etapa que se considera el tema o temas de la investigación, que son el conjunto de las variables de estudio. En todo proceso de relevamiento de información por el método de la entrevista directa, como son las encuestas se tiene como instrumento fundamental a los formularios o boletas de encuesta.

## 8. Instrumentos de investigación

Luego de determinar la temática a ser analizado por la Encuesta, se procede al diseño de los instrumentos de investigación: el cuestionario de la encuesta, manuales e instructivos sobre el llenado y operativización de la boleta.

## 9. Diseño de la boleta

La boleta toma en cuenta los conceptos y definiciones de la Encuesta, a fin de reflejar en las preguntas el objetivo que se persigue. Esta es una de las etapas más complejas dentro de la preparación de la encuesta.

La Boleta o cuestionario tiene diferentes tipos de preguntas como ser Preguntas Directas, Preguntas de sondeo, Preguntas pre-codificadas, Preguntas Parcialmente pre-codificadas, Preguntas abiertas.

## 10. Diseño muestral

Comprende el planteamiento estadístico, es decir diseñar el trabajo muestral, la cual podemos resumir en 4 partes:

- **Elección del marco muestral**

Elegir el marco (o marcos) de tal manera que la lista de unidades de muestreo y la población objetivo concuerde lo más posible. Se debe tener en cuenta que marcos múltiples pueden hacer el muestreo más eficiente (por ejemplo, los residentes de una ciudad pueden ser muestreados de una lista de manzanas de la ciudad juntamente con una lista de residentes dentro de las manzanas).

- **Determinación del tamaño de la muestra**

Existe gran variedad de diseños para seleccionar la muestra; por cada diseño, se pueden hacer cálculos del tamaño de muestra, partiendo del nivel de precisión deseado. Los costos relativos

y el tiempo empleado para cada diseño también deben ser comparados antes de tomar una decisión. Dependiendo la técnica utilizada, se puede plantear el algoritmo con el cual se tendrá el tamaño de la muestra.

- Selección de las unidades de la muestra

Determinada la muestra, se debe seleccionar las unidades de la muestra recurriendo a alguna técnica que la teoría del muestreo sugiera, es así que tenemos la selección al azar de las unidades, la selección sistemática, y otras técnicas.

- Estimadores

Es aplicación del proceso de la estimación, que consiste en la utilización de de Estimadores, los cuales son expresiones matemáticas y/o estadísticas, que nos permiten conocer los parámetros poblacionales, conociéndose también a todo este proceso con el nombre de inferencia estadística.

## 11. Encuesta piloto

Es de gran utilidad probar el cuestionario y los métodos de campo en pequeña escala. Esto casi siempre da por resultado mejoras del cuestionario y puede evitar otros problemas que serían serios a mayor escala, por ejemplo, que el costo fuera mucho mayor que el esperado.

## 12. Trabajo de campo

Se debe considerar la convocatoria del personal que debe cumplir algunas condiciones, su capacitación y su posterior trabajo que será el llenado del formulario de encuesta. Es así que debemos tomar en cuenta lo siguiente:

- ✓ Diseño del Programa de Capacitación

El diseño del programa de capacitación está íntimamente relacionado con el Manual

del encuestador y del Supervisor de grupo, estableciéndose el orden en que se trataran los diversos temas.

- ✓ Estrategia del operativo de campo

En base al tamaño de la muestra y al presupuesto para la captura de información, se diseña la estrategia de operativo de campo, que incluye aspectos referentes al número de personal que será capacitado en cada uno de los ámbitos de investigación, número de encuestadores y supervisores de grupo necesarios para la recopilación de la información, la carga diaria de trabajo que tendrá cada encuestador y el número de días de trabajo.

- ✓ Operativo de campo

El operativo de campo abarca dos etapas de vital importancia: una referida a la capacitación del personal de campo y otra al proceso mismo de captación de información.

- ✓ Organizar el trabajo de campo

En encuestas extensas se encuentra muchos problemas de orden administrativo; el personal debe recibir un entrenamiento sobre el propósito de la encuesta y los métodos de medición que se emplearán.

- ✓ Captación de los datos

Es conveniente cerciorarse que todos los datos sean pertinentes a la encuesta y que no se omiten datos esenciales, suele existir en algunos casos la tendencia a hacer un número excesivo de preguntas que no se analizan posteriormente.

## 13. Revisión (crítica y codificación de los cuestionarios)

Se realiza la revisión de los formularios empleados y con información, a causa del operativo de campo. En esta etapa se diferencian dos procesos importantes y complementarios entre sí:



La *depuración de la información*, es la verificación de que las respuestas estén anotadas en forma legible, completa y en el lugar correcto, eliminando toda aquella información innecesaria; controlando por otra parte, la coherencia y consistencia de la información.

La *codificación*, es la asignación de códigos a aquella información eminentemente cualitativa; mediante este proceso se posibilita que la información literal pueda ser introducida a la computadora.

Los criterios para la ejecución de la crítica y codificación se exponen en un instructivo, denominado Manual de Crítica y Codificación.

#### 14. Tratamiento de la información

Aunque la etapa de procesamiento de datos constituye una de las fases finales de la encuesta, gran parte de su planificación se realiza con mucha anticipación.

#### 15. Transcripción

La transcripción se realiza paralelamente a la crítica y codificación, en la cual se introduce la información a la computadora (generando la Base de Datos), mediante un programa inteligente de entrada de

datos, previamente diseñando, que impide el ingreso de información errónea o contradictoria entre sí.

#### 16. Resumir y analizar los datos

Una vez realizada la encuesta es necesaria la edición de los cuestionarios obtenidos, con el fin de corregir errores o desechar los datos que obviamente están equivocados; habrá necesidad de tomar ciertas decisiones respecto al procedimiento de cálculo en los casos de falta de respuesta o de ausencia de datos en el proceso de edición; luego se realizarán los cálculos que conduzcan a las estimaciones.

#### 17. Proporcionar información para encuestas futuras

Cuanta más información de una población se tenga inicialmente, más fácil será el diseño de una muestra que proporcione estimaciones exactas. Siempre la muestra obtenida es un antecedente de futuros muestreos debido a los datos que revela sobre las medias, las desviaciones estándar y la naturaleza de la variabilidad de las medidas principales, así como sobre los costos de obtención de datos.

El objeto de describir los pasos de una encuesta por muestreo es recalcar que el muestreo es un asunto práctico y exige muchas y diversas habilidades.

*“La vida sería intolerable si los fenómenos ocurrieran al azar de una forma completamente impredecible y carecería de interés si, en el otro extremo, todo fuera determinista y completamente predecible“*

*Radhakrishna Rao*

## Modelos de Programación Lineal y No Lineal con Multiobjetivos

*Autor: M.Sc. Nicolas Chavez Quisbert*

### 1. Introducción

La programación por metas (Goal Programming) fue inicialmente introducida por Charnes y Cooper en los años 50 desarrollada en los años 70 por Ljiri, Lee, Ignizio y Romero, es actualmente uno de los enfoques multicriterio que más se utilizan.

En principio fue dirigida a resolver problemas industriales, sin embargo posteriormente se ha extendido a muchos otros campos como la economía, agricultura, recursos ambientales, recursos pesqueros, etc.

En este artículo lo que quiere es mostrar la mejor toma de decisiones desde un punto de vista Gerencial, y de cómo lograr tomar estas mejores decisiones, utilizando Modelos de Programación Lineal y No Lineal con Multiobjetivos, todo con el fin de maximizar beneficios minimizando costos y recursos.

### 2. Aspectos básicos de la programación por objetivos

Entre los aspectos básicos de la programación con multiobjetivos se tiene que:

La Programación por Objetivos proporciona una manera racional de intentar alcanzar varios objetivos de manera simultánea, jerarquizando los mismos o asociándole una ponderación a cada uno.

El enfoque básico de la Programación por Objetivos es establecer un objetivo numérico específico para cada uno de los objetivos, formular una función objetivo para cada uno y después buscar una solución que minimice la suma ponderada de las desviaciones de estas funciones objetivo de sus metas respectivas.

Tres tipos de metas:

Meta unilateral inferior: establece un límite inferior por abajo del cual no se quiere ir (pero se aceptan desvíos a la meta que deberá minimizarse)

$$\text{Ej. : } a_{11} x_1 + a_{12} x_2 \geq \text{Meta1}$$

Meta unilateral superior: establece un límite superior que no se quiere exceder (pero se aceptan desvíos a la meta que deberá minimizarse)

$$\text{Ej. : } a_{21} x_1 + a_{22} x_2 \leq \text{Meta2}$$

Meta bilateral: establece un “blanco” específico que no se quiere perder hacia ningún lado.

$$\text{Ej. : } a_{31} x_1 + a_{32} x_2 \leq \text{Meta3}$$

### 3. Formulación de una ecuación objetivo para cada meta

En el modelo de Programación por Objetivos existen dos tipos de restricciones funcionales: las restricciones ordinarias de Programación Lineal (restricciones “duras” o estrictas) y las ecuaciones objetivo (“blandas” o flexibles). Las restricciones “duras” requieren ser cumplidas de manera estricta. Las restricciones “blandas” pueden admitir desvíos a la meta establecida, pero estos desvíos estarán asociados a una penalización que se reflejará en un parámetro en la Función Objetivo.

Valor objetivo de la Meta

El Valor de la Meta se descompone en dos elementos: (1°) el valor correspondiente al nivel de la Meta alcanzado

efectivamente  $\sum_j a_j x_j$  y (2°) el desvío o diferencia entre el valor meta y el nivel alcanzado ( $d_j$ ):

$$\sum_j a_j x_j = M_i d_i$$

Variables de desvío

Para formalizar los desvíos aceptados a cada una de las metas se emplea las variables auxiliares  $d_j$ ; las que por definición pueden obtener valores positivos o negativos.

Para poder hacer operativo el modelo de Programación Lineal cada  $d_i$  se sustituirá por la diferencia de dos variables no-negativas:

$$d_i = d_i^+ - d_i^- \quad \text{donde } d_i^+, d_i^- \geq 0$$

Con la siguiente definición:

$$d_i^+ = \begin{cases} d_i & \text{si } d_i \geq 0 \\ 0 & \text{otra manera} \end{cases}$$

$$d_i^- = \begin{cases} |d_i| & \text{si } d_i \leq 0 \\ 0 & \text{otra manera} \end{cases}$$



## 4. La función objetivo

Depende del procedimiento, si se consideran los diferentes objetivos de manera simultánea (Objetivos sin prioridades) o sí por el contrario se adopta un procedimiento secuencial (Objetivos con prioridades). En el primer caso se tratará de Minimizar una función ponderada de las variables de desvío.

En el segundo caso se identificarán “m” modelos de PL a ser optimizados de manera secuencial y de acuerdo a los “m” niveles de jerarquía asignados a los objetivos. La secuencia se resuelve por nivel de jerarquía, desde el nivel de mayor prioridad al de menor.

## 5. Modelo de programación meta lineal

Las suposiciones básicas que caracterizan el modelo de programación lineal se aplican igualmente al modelo de programación meta. La diferencia principal en la estructura es que la programación meta no intenta minimizar o maximizar la función objetivo como lo hace el modelo de programación lineal, en vez busca minimizar las desviaciones entre las metas deseadas y los resultados reales de acuerdo a las prioridades asignadas. El objetivo o función de preferencia de un modelo de programación meta es expresado en términos de las desviaciones de las metas a que se apunta. Esto es, las variables de holgura o sobrantes de las restricciones se colocan en la función objetivo y deben minimizarse. El modelo general de la programación meta se define como:

$$\text{Minimizar } Z = \sum_{i=1}^m w_i (d_i^+ + d_i^-) \quad (1)$$

Sujeto a

$$\sum_{j=1}^n a_{ij} x_j + d_i^- - d_i^+ = b_i \quad \text{para todo } i \quad (2)$$

$$x_j, d_i^-, d_i^+ \geq 0 \quad \text{para todo } j \quad (3)$$

La variable  $x_j$  representa una variable de decisión,  $w_i$  representan los pesos de ponderación (ordinal o cardinal) asignados a cada una de las meta, y  $d_i^+$  y  $d_i^-$  presenta el grado de sobrelogro y sublogro de la meta, respectivamente. Puesto que al mismo tiempo no podemos tener logro por encima y por debajo de la meta, o una o ambas de estas variables debe ser igual a cero.

## 6. Modelo de programación meta cuadrática

En toda la teoría desarrollada hasta ahora, hemos supuesto que la función objetivo de la programación meta es lineal. Por tanto, el incremento en cualquier desviación,  $d_i$ , siempre adiciona una cantidad igual de desutilidad, independiente del nivel de todas las otras desviaciones meta. En esta parte supondremos que la función objetivo del modelo de programación meta es cuadrática y sujeta a restricciones lineales. Esto no presenta dificultades de cómputo, puesto que los algoritmos de programación cuadrática estándar pueden utilizarse para resolver dichos problemas. Parte de la discusión actual en programación cuadrática meta se ha tomado de Shim y Siegel (1975).

El problema de programación cuadrática meta se define como:

$$\text{Minimizar } Z = \sum_{i=1}^l \sum_{j=1}^m w_i (d_i^- d_j^- + d_i^+ d_j^+ + d_i^- d_j^+) \quad (4)$$

Sujeto a

$$\sum_{j=1}^n a_{ij} x_j + d_i^- - d_i^+ = b_i \quad \text{para todo } i \quad (5)$$

$$x_j, d_i^-, d_i^+ \geq 0 \quad \text{para todo } i, j. \quad (6)$$

Como antes, deseamos encontrar variables  $x_j, d_i^-$  y  $d_i^+$  para minimizar la función objetivo anterior, que está compuesta de los términos cuadráticos,  $d_1^{-2}$ , y una iteración de los términos del producto cruzado  $d_1^- d_2^+$ . La función objetivo, por consiguiente, debe ser una función convexa para garantizar una función global óptima para este problema de minimización utilizando un algoritmo de programación cuadrática existente.

Existen dos metodologías clásicas de solución para modelos con multiobjetivos, Taha (2004), los cuales son, el Método de Factores de Ponderación y el Método de Jerarquías.

## 7. Bibliografía

Cohen Daniel (1993), Sistemas De Información Para La Toma De Decisiones, Mcgraw Hill, México

Heizer Jay (1997). Dirección De La Producción Decisiones Técnicas, Prentice Hall, España

Hiller F. Liberman G. (1991), Introducción A La Investigación De Operaciones, Ediciones Mcgpaw – Hill, México

Hodson William, (1996) Manual Del Ingeniero Industrial, Mcgraw Hill, México

Mathur K. Solow D. (1996), Investigación De Operaciones, Prentice All, México

Moskowitz H. Wright G. (1982), Investigación De Operaciones,

Prentice Hall, México

Taha Hamdy, (2004) Investigación De Operaciones Prentice All, México

*El pensamiento estadístico es una filosofía dentro de la cual el aprendizaje y la acción se basan en tres principios, cualquier trabajo ocurre en procesos interconectados, dentro de éstos existen variaciones y para poder tener éxito al aplicar este pensamiento hay que entender los procesos y reducir las variaciones.*

**LA ESTADÍSTICA ES LA BASE PARA LA PLANIFICACIÓN Y LA PLANIFICACIÓN ES EL INSTRUMENTO DEL DESARROLLO**

## Análisis de Medidas Repetidas mediante Métodos de Máxima Verosimilitud

Autor: Univ. Deyvis Nina Canaviri

### 1 Introducción

En el análisis de experimentos con medidas repetidas ocurre con cierta frecuencia que la matriz de covarianzas no se ajusta al supuesto restrictivo de esfericidad en el análisis de varianza, o de igualdad de varianzas para cualquier par de diferencias entre tratamientos (Huyhn&Feldt, 1970). Una condición suficiente para el cumplimiento de este supuesto se da cuando las variables aleatorias están igualmente correlacionadas y tienen varianzas iguales.

Situaciones experimentales comunes provocan sin embargo que las medidas repetidas con mayor contigüidad temporal o espacial presenten correlaciones más altas. Estas situaciones generan matrices de covarianzas que incumplen el supuesto de esfericidad y pruebas  $F$  sesgadas que inflan la probabilidad de rechazar la hipótesis nula cuando es cierta por encima del criterio fijado por el investigador.

Tradicionalmente se han propuesto dos soluciones que pueden ser familiares al lector: la utilización de una prueba  $F$  con grados de libertad ajustados (Greenhouse&Geisser, 1959), y el análisis de varianza multivariado (O'Brien & Kaiser, 1985). Esta última estrategia es típicamente recomendada cuando existe una grave desviación del supuesto de esfericidad y la muestra sea grande (Jensen, 1982, 1987; Marcucci, 1986). El modelado de la matriz de covarianza mediante procedimientos de estimación de máxima

verosimilitud nos lleva a una tercera estrategia que incorpora en el análisis información paramétrica más ajustada al estudio del comportamiento de interés (Wolfinger, 1993). Esta última opción ha mostrado tener más potencia estadística que la prueba  $F$  con grados de libertad ajustados en el caso de matrices de covarianza autorregresivas (Albohali, 1983;

Milliken& Johnson, 1994), así como resultados favorables más parsimoniosos que el análisis multivariado de varianza (Elston&Grizzle, 1962). El objetivo de la presente comunicación es presentar dos estudios de casos que ilustren las oportunidades y problemas de esta tercera vía en el análisis de medidas repetidas, así como describir algunas características de los programas informáticos disponibles.

### 2 Un Caso

El estudio consiste en evaluar el efecto sobre la resistencia muscular de tres programas de entrenamiento físico (Littell, Freund&Spector, 1991). En el primero de ellos, se aumentó semana a semana el número de repeticiones de un ejercicio de levantamiento de pesas. En el segundo programa, se aumentó el peso a levantar según los sujetos ganaban en resistencia. En la tercera condición de control, los participantes no hicieron el ejercicio. Las medidas de resistencia se realizaron en días alternos durante un período de dos semanas.

El modelo lineal para un *análisis de varianza univariado* puede ser expresado como:

$$y_{ik} = \mu_{ik} + s_j + e_{ijk} \quad (1)$$

Para  $i = 1, 2, 3$   $i = 1, 2, 3$  programas de entrenamiento y  $k = 1, 2, \dots, 7$   $k = 1, 2, \dots, 7$  días alternos y  $j = 1, 2, \dots, n_i$   $j = 1, 2, \dots, n_i$  sujetos por programas.

En donde  $s_j \approx N(0, \sigma_s^2)$ ,  $e_{ijk} \approx N(0, \sigma_e^2)$   $s_j \approx N(0, \sigma_s^2)$ ,  $e_{ijk} \approx N(0, \sigma_e^2)$  y ambos son independientes. La matriz de covarianzas puede expresarse como:

$$\begin{cases} \sigma_s^2 + \sigma_e^2 & \text{si } i = i', j = j' \text{ y } k = k' \\ \sigma_s^2 & \text{si } i = i', j = j' \text{ y } k \neq k' \\ 0 & \text{En cualquier otro caso} \end{cases}$$

o equivalentemente:

$$\begin{cases} \sigma_y^2 & \text{si } i = i', j = j' \text{ y } k = k' \\ \sigma_y^2 \rho & \text{si } i = i', j = j' \text{ y } k \neq k' \\ 0 & \text{En cualquier otro caso} \end{cases}$$

Contando en ambos casos con dos parámetros. Podemos formular las hipótesis estadísticas de efectos principales de Programa ( $H_{01}$ ), y Tiempo ( $H_{02}$ ), así como la interacción entre Programa y Tiempo ( $H_{03}$ ) del modo siguiente:

$$H_{01}: \mu_{1..} = \mu_{2..} = \mu_{3..}$$

$$H_{02}: \mu_{..k} = \mu_{..k'} \text{ para cualquier par de tiempos } k, k'$$

$$H_{03}: \mu_{i.k} - \mu_{i.k'} = \mu_{i'.k} - \mu_{i'.k'} \text{ para cualquier par de programas } i, i' \text{ y de tiempos } k, k'$$

Sus resultados pueden apreciarse en la Tabla 1, resultando significativa la interacción entre programa y tiempo, pero no así el efecto principal de programa.

Tabla 1			
<i>Resultados de análisis de varianza univariado</i>			
Efecto	g.l	F	p
Programa	2	3.07	0.0548
Sujeto (Programa)	54		
Tiempo	6	7.43	0.0001
Programa x Tiempo	12	2.99	0.0005
Tiempo x Sujeto (Programa)	32		

Un diagnóstico de la adecuación del modelo mediante el análisis de residuales revela la siguiente matriz de correlaciones, con valores decrecientes en función de la distancia temporal entre observaciones.

$$R = \begin{bmatrix} 1 & 0.96 & 0.92 & 0.87 & 0.84 & 0.81 & 0.80 \\ & 1 & 0.94 & 0.88 & 0.86 & 0.83 & 0.79 \\ & & 1 & 0.96 & 0.94 & 0.91 & 0.89 \\ & & & 1 & 0.96 & 0.91 & 0.89 \\ & & & & 1 & 0.95 & 0.92 \\ & & & & & 1 & 0.95 \\ & & & & & & 1 \end{bmatrix}$$

La inspección visual nos sugiere un mal ajuste al criterio de esfericidad. Esto es apoyado por los resultados de la prueba de Mauchly, así como por el valor e de Geisser-Greenhouse estimado en 0.42, que se aleja del valor 1 requerido por la prueba F insesgada para los contrastes intrasujetos. A la vista del tamaño de la muestra ( $N = 57$ ), y de la desviación considerable respecto de la condición de esfericidad, podemos considerar la alternativa del *análisis multivariado*.

El modelo lineal en este caso puede expresarse como:

$$Y = X\beta + E \quad (2)$$

en donde Y es la matriz de observaciones (Sujetos x Tiempos), X es la matriz del diseño (Sujetos x Programas),  $\beta$  es el vector de parámetros de efectos fijos o medias (Programas x Tiempos), E es la matriz de diferencias individuales (Sujetos x Tiempos) y  $E \approx N, [0, \Sigma]$

La matriz de covarianzas en este caso no tiene restricciones, por lo que cuenta con 28 parámetros distintos, 7 varianzas y 21 covarianzas.

$$\begin{cases} \sigma_k^2 & \text{si } i = i', j = j' \text{ y } k = k' \\ \sigma_{kk'} & \text{si } i = i', j = j' \text{ y } k \neq k' \\ 0 & \text{En cualquier otro caso} \end{cases}$$

El contraste de las hipótesis para los efectos intrasujetos equivalentes al modelo univariado ha de realizarse mediante contrastes multivariados. La hipótesis lineal general multivariada viene dada por:

$$H_0: H\beta M' = 0$$

en donde H es la matriz de contrastes experimentales de interés para los tratamientos entre sujetos, y M es la matriz de combinaciones lineales intrasujetos o entre tiempos de interés. Podemos expresar así las hipótesis estadísticas de efecto principal de tiempo ( $H_{02}$ ), e interacción entre tiempo y tratamiento ( $H_{03}$ ) como:

$$H_{02}: H_{INT}\beta M_T' = 0 \quad y$$

$$H_{03}: H_{PROG}\beta M_T' = 0$$

siendo  $H_{INT}$  un vector 1 x 3 de unos,  $H_{PROG}$  la matriz 2 x 3 de hipótesis de efecto principal de programa expresada de forma matricial, y  $M_T$  la matriz 6 x 7 de hipótesis de efecto principal de tiempo. Estos contrastes son programados por defecto a través de los comandos de medidas repetidas en paquetes estadísticos usuales tales como *Statistical Package for the Social Sciences* (SPSS) o *Statistical Analysis System* (SAS). En los resultados en la Tabla 2 no se aprecian cambios considerables en cuanto al efecto principal de tiempo, mientras que la interacción entre programa y tiempo ha dejado de ser significativa.

Una segunda alternativa al modelo univariado clásico consiste en utilizar un *modelo univariado autorregresivo*. Aunque podemos expresar el modelo lineal como en la ecuación (1), la diferencia radica en la matriz de covarianzas. Esta cuenta también con dos parámetros, pero indica una correlación decreciente en función de la distancia temporal de las observaciones ( $d$ ):

$$\begin{cases} \sigma_y^2 & \text{si } i = i', j = j', k = k' \\ \sigma_y^2 \rho^d & \text{si } i = i', j = j', k \neq k' \text{ para } 0 \leq d \leq 6 \\ 0 & \text{En cualquier otro caso} \end{cases}$$

La programación en SAS del procedimiento de análisis mediante métodos de máxima verosimilitud puede consultarse en el resultado en la Tabla 3. La fundamentación teórica de los estadísticos utilizados en el contraste de estos efectos intrasujeto es expuesta en Littell, Milliken, Stroup & Wolfinger (1996).

Los métodos de máxima verosimilitud nos permiten también contrastar el ajuste de las tres matrices de covarianza utilizadas en los modelos anteriores. Como puede verse en la Tabla 4, los criterios de Akaike y Schwarz son más cercanos a cero en el caso de la matriz autorregresiva, lo que suele considerarse como un índice de mejor ajuste (Littell, Milliken, Stroup & Wolfinger, 1996). Este modelo es además más parsimonioso en cuanto que parece conseguir mejores resultados con un ahorro de 26 parámetros respecto de la matriz de covarianzas multivariada sin restricciones. Aunque todavía faltan más trabajos de investigación en este sentido, los resultados citados anteriormente (Elston & Grizzle, 1962; Albohali, 1986) apuntan al valor de una precisa especificación de la matriz de covarianzas y por lo tanto favorecerían también al último modelo.

### Resultados del análisis mediante contrastes multivariados

Efecto	g.l.	$\lambda$ de Wilks	p
Tiempo	6	6.44	0.0001
programa x tiempo	12	1.38	0.1880

### Razones F aproximadas obtenidas tras procedimientos de estimación por máxima verosimilitud

Efecto	g.l.	F Appr.	p
Tiempo	6	4.30	0.0003
programa x tiempo	12	1.17	0.3007

### Medidas de ajuste de las matrices de covarianza utilizadas en los modelos univariados, multivariados y autorregresivo

Efecto	SC	UN	AR(1)
Tiempo	-712,12	-645,45	-635,4
programa x tiempo	-716,35	-700,54	-639,34

Una deficiencia de la metodología tradicional del análisis de varianza en la investigación longitudinal es que no modela explícitamente las diferencias individuales en patrones de crecimiento, relegándolas al término de error. Asume también la existencia de un diseño completo con el mismo número de observaciones por persona y espaciado temporal, considerando al factor de medidas repetidas como cruzado al factor sujeto. Cuando existen datos faltantes puede ocurrir que algunas razones F en el modelo univariado no tengan distribuciones exactas, y sólo sean aproximaciones que empeoran conforme aumenta el número de valores perdidos (Milliken & Johnson, 1994).

En el análisis de varianza multivariado, los paquetes estadísticos comunes tales como SAS o SPSS eliminan cualquier registro incompleto antes de ejecutar los algoritmos de cálculo. Los procedimientos de máxima verosimilitud se presentan como alternativas en este sentido cuando existen muestras

grandes. Permiten además considerar las medidas repetidas como anidadas al factor sujeto, con un número desigual de observaciones por persona y distinto espaciado temporal. Incorporan también parámetros de variabilidad interindividual en crecimiento, cuya estimación tiene en cuenta las diferentes precisiones derivadas del distinto número de medidas hechas en cada participante.

### 3 Bibliografía

[1] Searle, S. R., Casella, G. & McCulloch, C. W. (1992). *Variance components*. New York: John Wiley.

[2] Smith, D. W. & Murray, L. W. (1984). An alternative to Eisenhart's model II and mixed model in the case of negative variance estimates. *Journal of the American Statistical Association*, 79, 145-151.

[3] Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in statistics: Simulation*, 22 (4), 1.079-1.106.

*“Si los estadísticos teóricos hacen caso omiso al reto de enfrentar las encuestas multipropósito, entonces el vacío existente entre ellos y los estadísticos prácticos se hará cada vez más grande. El diseño y análisis de encuestas multivariantes debe ser una de las próximas áreas de mayor investigación.”*

*J. M. F. Smith (1976)*



## El nuevo IETA

El Instituto de Estadística Teórica y Aplicada (IETA), ha entrado a una dinámica de trabajo y aplicación como jamás lo había hecho. En esta etapa, ha logrado ser parte de un convenio rectoral entre el Gobierno Autónomo Departamental y la UMSA, en el cual la universidad se transforma en el brazo derecho de apoyo científico y tecnológico para el Departamento de La Paz, mediante la Gobernación.

Existen diferentes proyectos que se vienen realizando entre ambas instituciones, y otros que se planifican actualmente. Desde esa perspectiva, el IETA se ha convertido en la entidad generadora de información estadística para el departamento de La Paz como insumo fundamental para la toma de decisiones de futuros proyectos del Departamento.

En ello, el IETA asume la responsabilidad del proyecto para la generación de información estadística mediante una investigación por muestreo denominada Encuesta Socio-demográfica a nivel de los municipios del departamento de La Paz, después de una lucha plausible entre unidades académicas interesadas de realizarla.



Es el IETA que logra el proyecto, gracias a ciertas personas de nuestra entidad y de la gobernación que creen en nuestra capacidad y compromiso para hacerlo, tales personas son la Sra. Rectora de la Universidad Dra. Teresa Rescala, el Sr. Director de Planificación y Coordinación de la UMSA Msc.

Julio Alvarado, el Sr. ex Director de Información de la Secretaría de Planificación de la Gobernación del Departamento de La Paz Lic. Ivan Toro T. y otras personas que sin dudarlo, apoyan el proyecto en favor del IETA y la Carrera de Estadística.



Hoy el IETA con la publicación de sus dos primeros libros de información estadística referidos al Altiplano Sur y Norte del departamento de La Paz, demuestra a autoridades universitarias, personas de estado, autoridades originarias y locales de los municipios y a la sociedad boliviana en sí que cuando se quiere hacer las cosas con compromiso y responsabilidad, se puede, pese a las limitaciones que se han encontrado en el camino y que seguirán hasta el final del proyecto.

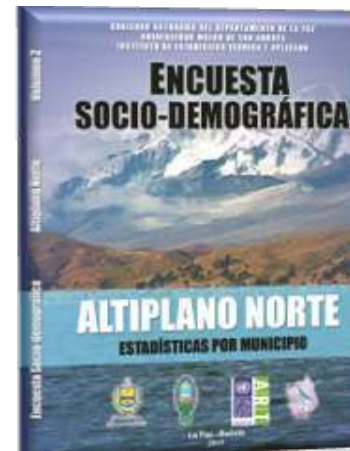
En este principio, tenemos que agradecer a todas aquellas personas que le han puesto el “lomo” a este trabajo, entre ellos están los docentes, administrativos y estudiantes de la Carrera de Estadística, al igual que muchos estudiantes de la UMSA que han hecho de encuestadores y supervisores, dándoles la oportunidad de conocer e interrelacionar con la gente de las comunidades campesinas de los municipios que hemos visitado. No sólo son ellos, están profesores rurales de las comunidades que han aceptado el reto de la toma de datos de la encuesta en sus comunidades con el sólo compromiso de apoyo al desarrollo del departamento de La Paz.



En un periodo de un año de proyecto, hemos realizado la encuesta en cuatro regiones del departamento de La Paz (Altiplano Sur y Norte, Amazonía y Yungas). Para este fin se han comprometido el apoyo de 30 becarios estudiantes de diferentes niveles de la Carrera de Estadística para el trabajo de supervisión, crítica y codificación, transcripción y análisis de los datos. De igual forma se han contratado los servicios de dos técnicos para la labor de coordinación, administración y procesamiento de información estadística. Se ha trabajado con dos profesionales analistas conocidos en nuestro país en materia de análisis. Se han contratado los servicios de un diseñador gráfico para la estética y diseño de dichos documentos con el apoyo de personal adicional en la comunicación social y la corrección de estilos.



Sin duda esta ha sido una labor de todos los que pensamos y trabajamos como institución y no por intereses personales. En esa perspectiva reafirmamos nuestro compromiso de continuar la siguiente gestión con mayor fuerza después de la experiencia vivida, siempre con la profesionalidad demostrada de la gente que es parte del IETA y la Carrera de Estadística de la UMSA.



*“Las cosas complejas y estadísticamente improbables, son por naturaleza más difíciles de explicar que las cosas simples y estadísticamente probables.*

*Richard Dawkins*

## Estadística Clásica e Inferencia

### 1.- Introducción

En la actualidad, la inferencia estadística clásica, desarrollada en el primer tercio del siglo XX, da respuesta suficiente para tratar de forma adecuada la problemática suscitada.

Tomar el valor de una medida (población), pasa por realizar unas concretas determinaciones (muestra) y a partir de ella, dar el resultado final de la medida y su incertidumbre.

La incertidumbre es un parámetro asociado al resultado de una medición que caracteriza la dispersión de los valores que pueden atribuirse a la medición.

De esta forma, la muestra y las técnicas de muestreo se convierten en el fundamento de la inferencia estadística clásica.

### 2.- Estadística Clásica

La estadística tiene muchos significados, siendo una técnica que debe establecer una guía para determinar situaciones de incertidumbre.

La información y la incertidumbre participan de forma decisiva. Es, por lo tanto, la estadística una ligazón entre pasado y futuro, de especial trascendencia en la mayor parte de los procesos.

Todo procedimiento estadístico que emplea información para obtener conocimiento, está en el ámbito de la inferencia y cuando hay que determinar acciones, se aborda la denominada decisión estadística.

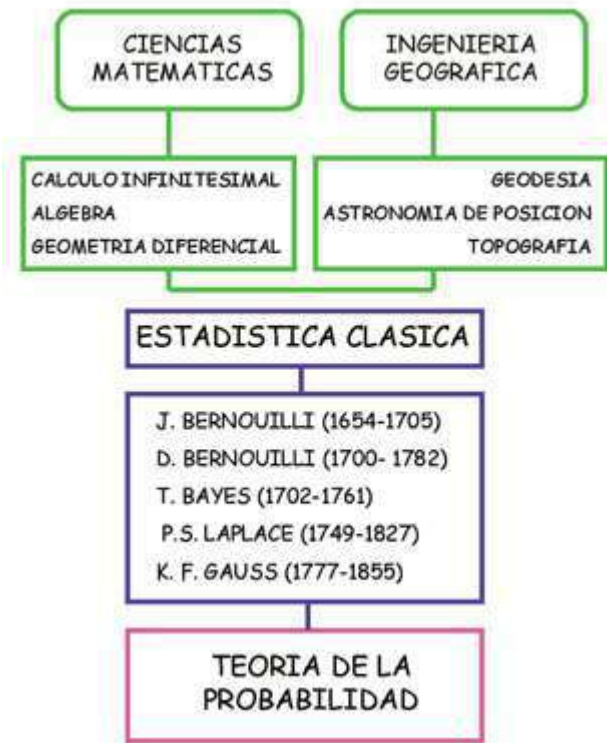


Figura 1.- Marco de la estadística clásica

Los primeros nombres ligados a esta etapa significativa son los Bernoulli. Originarios de los Países Bajos, huyeron a Suiza por la persecución que en el reinado de Felipe II hicieron los españoles a los no católicos.

### 3.- La Teoría de la Probabilidad

En el primer tercio del siglo XX irrumpen con fuerza en el ámbito científico-matemático con corrientes que tratan de configurar un marco propio: la teoría de la probabilidad.

Crear una axiomática propia y establecer las bases para difundir una nueva estadística, ocupó un amplio período de tiempo donde científicos de gran renombre fueron plasmando sus teorías.

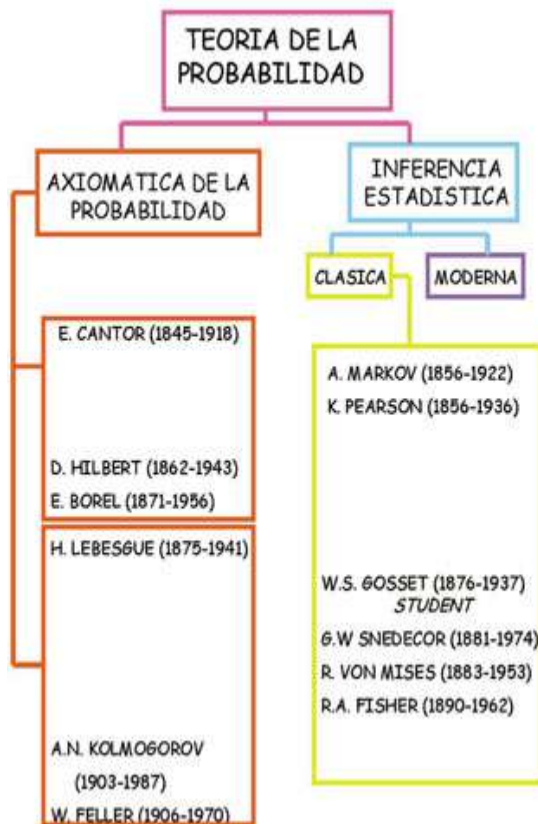


Figura N 2.- Flujo significativo

Es necesario destacar los tres grandes bloques de tratamiento en esta etapa tan interesante que trata de dar contenidos a una nueva forma en el ámbito del conocimiento matemático.

Preparación de una base topológica y de las estructuras asociadas necesarias para establecer una nueva axiomática (Cantor, Hilbert, Borel).

Inicio y consolidación de la axiomática de la teoría de la probabilidad (Lebesgue, Kolmogorov, Feller).

Desarrollo de la inferencia estadística clásica y su aplicación a Ingeniería, Industria (Markov, Pearson, Gosset, Snedecor, Von Mises, Fisher).

Los tres bloques fueron desarrollándose a la vez, consiguiendo en tan solo cincuenta años desgajar unos contenidos que han formado una estructura propia en el ámbito de las ciencias matemáticas.

## 4.- Inferencia Estadística Moderna

A partir de mediados del siglo XX, tras la implantación de la denominada **inferencia clásica**, surgen investigadores que intentan reforzar nuevas formas, tratando de implementar aplicaciones dirigidas hacia las ciencias sociales y ámbito económico.

La **moderna inferencia** tiene dos pilares fundamentales: la teoría de la decisión y los métodos bayesianos.

A mediados de 1.800 ya se distinguen aspectos como utilidad marginal y utilidad total, renovándose el primitivo concepto de Bernoulli.

Con posterioridad Von Neumann y Oscar Morgenstern, hacia 1.945, desarrollan la moderna teoría probabilística de la utilidad, definiendo una axiomática común para la probabilidad y para la utilidad.

En 1.954 aparece la obra de Savage, **iniciador de la línea bayesiana**, en el marco de la teoría de la utilidad.

En los últimos años el empleo de la teoría ha sido predictivo, aplicado en muchas ocasiones a las ciencias sociales y económicas, destacando la utilidad colectiva de Arrow (problema de bienestar social) y hace poco tiempo Black y Rethenberg.

Recientemente trabajos de McCrimmon (1968) han contrastado la axiomática de Savage, empleando experimentación adecuada, obteniendo interesantes resultados.

Fundamentalmente la **inferencia estadística moderna** sigue dos sentidos que, aunque convergentes, tienen enfoque diferenciado:

**Teoría bayesiana.** Admite conocimientos a priori, así como observaciones nuevas. Las a priori son modificadas por las verosimilitudes.

Teoría de la decisión. Punto de partida en los trabajos de Wald (1.950). Pretende establecer una teoría, unas reglas de acción en situación de incertidumbre o riesgo. El valor de cualquier decisión es medido por su pérdida.

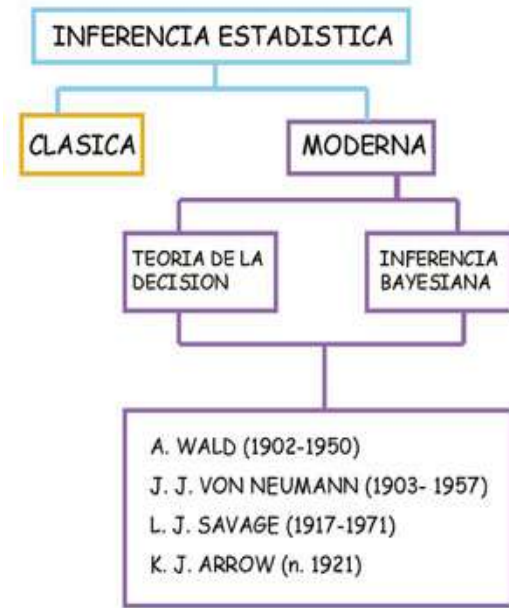


Figura 3.- Inferencia estadística moderna

## 5.- Inferencia Estadística Robusta

La inferencia estadística clásica establece mecanismos de actuación ampliamente desarrollados como la estimación genérica y los contrastes de hipótesis.

Ambos constituyen una parte esencial del trabajo habitual en la estadística actual. En ellos impera el dogma de la normalidad, quedando invalidadas las observaciones discordantes con esta norma.

Etapas de transición han intentado establecer mecanismos fundados en incluir los datos anómalos y diseñar funciones de distribución mixtas que modelen ambos comportamientos.

Los métodos robustos son complementarios de los métodos estadísticos clásicos y configuran una teoría más afinada en aspectos tan interesantes como la parte de estimación y de contraste.

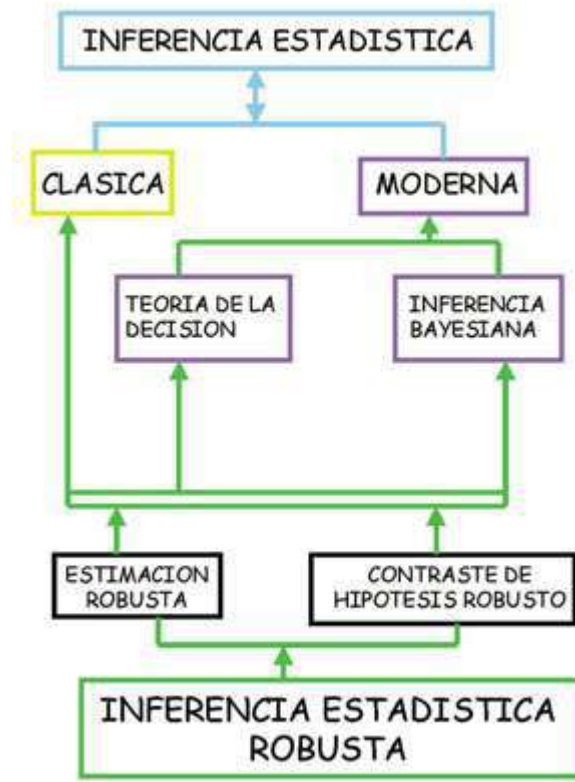


Figura 4.- Ambito de la inferencia estadística robusta

El término robusto fue establecido por Box (1.953), tratando de forzar hacia una modelización más completa que la establecida. Trabajos posteriores fueron reforzando la teoría, destacando:

Quenouille (1.956), con técnicas que permiten reducir el sesgo y establecer nuevos entornos para la estimación.

Anscombel (1.960), estimulando la investigación teórica y experimental en datos extremos.

Huber (1.964), diseñando los procedimientos estadísticos robustos.

Hampel (1.968), introduciendo el uso de curvas de influencia para analizar la sensibilidad de los estimadores.

Jaekel (1.971) y Berger (1.976), sobre aspectos de admisibilidad de los estimadores.

En la actualidad, y una vez establecidos los elementos teóricos necesarios, es preciso aplicar el potencial de la nueva herramienta a las múltiples utilidades que diferentes áreas tienen establecidas.

## **BAYES, Thomas (1702 – 1761)**

Teólogo y matemático británico. Hijo de uno de los primeros ministros presbiterianos ordenados en Inglaterra. También fue ministro y asistió a su padre.

Fue el primero en utilizar la probabilidad inductivamente, estableciendo un fundamento matemático para la inferencia probabilística.

En un ensayo póstumo quedó patente el intento de establecer las probabilidades tras el conocimiento de causas. Laplace, a los once años de su muerte, dio a conocer definitivamente sus revolucionarias teorías.

La estadística bayesiana le debe su nombre al trabajo pionero del reverendo Thomas Bayes titulado *“An Essay towards solving a Problem in the Doctrine of Chances”* ( *Una composición para solucionar un problema en la doctrina de las oportunidades* ), publicado póstumamente en 1764 en la *“Philosophical Transactions of the Royal Society of London”*.

El artículo fue enviado a la Real Sociedad de Londres por Richard Price, amigo de Bayes, en 1763, quien escribió:

*“Yo ahora le mando un ensayo que he encontrado entre los papeles de nuestro fallecido amigo Thomas Bayes, y el cual, en mi opinión, tiene un gran mérito, y bien merece ser preservado... En una introducción que él ha escrito para este ensayo, él dice, que su objetivo en un principio fue, descubrir un método por el cual se pueda juzgar la probabilidad de que un evento tenga que ocurrir bajo circunstancias dadas, y bajo la suposición de que nada es conocido sobre dicho evento, salvo que, bajo las mismas circunstancias, éste ha ocurrido un cierto número de veces y fallado otro tanto... Cualquiera persona juiciosa verá que el problema aquí mencionado no es de ninguna manera una*

*simple especulación producto de la curiosidad, sino un problema que se necesita resolver para contar con un fundamento seguro para todos nuestros razonamientos concernientes a hechos pasados y a lo que probablemente ocurra de ahí en adelante... El propósito a mí me parece es, mostrar qué razones nosotros tenemos para creer que en la constitución de las cosas existen leyes fijas de acuerdo con las cuales las cosas pasan, y que, por lo tanto, el funcionamiento del mundo debe ser el efecto de la sabiduría y el poder de una causa inteligente, y así, confirmar el argumento tomado desde las causas finales para la existencia de la deidad.”*

Aunque la obra de Thomas Bayes data ya de hace más de dos siglos, la estadística bayesiana es relativamente nueva, y actualmente ostenta un gran desarrollo aunque no ajeno a también grandes controversias.

Los métodos bayesianos, con una interpretación diferente del concepto de probabilidad, constituyen una alternativa a la estadística tradicional centrada en el contraste de hipótesis, denominada por contraposición estadística frecuentista, y están siendo motivo actual de debate.

En esencia se diferencian en que incorporan información externa al estudio para con ella y los propios datos observados estimar una distribución de probabilidad para la magnitud -efecto- que se está investigando.

Se denomina método bayesiano por basarse originalmente en el teorema de Bayes, publicación póstuma de Thomas Bayes en 1764, que en esencia nos permite, si conocemos la probabilidad de que ocurra un suceso, modificar su valor cuando disponemos de nueva información.

## **6.- Teorema de Bayes**

Recordando la terminología y el teorema de Bayes. Vamos a llamar  $P(A)$  a la probabilidad de que ocurra el suceso  $A$ .

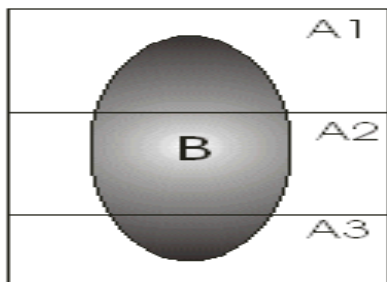
$P(AB)$  a la probabilidad de que ocurran los sucesos  $A$  y  $B$  (ambos).

$P(A / B)$  a la probabilidad de que ocurra A cuando sabemos que ha ocurrido B (se denomina **probabilidad condicionada**).

La probabilidad de que ocurra A y B es igual a la probabilidad de B multiplicada por la probabilidad de A condicionada a que haya ocurrido B.

$$P(AB) = P(B) \times P(A / B) = P(A) \times P(B / A)$$

Por simetría es obvio que se cumple la tercera igualdad.



Si tenemos un conjunto de posibles sucesos  $A_i$  ( $A_1 \dots A_n$ ), mutuamente excluyentes (no puede ocurrir dos de ellos a la vez) y que constituyen todas las posibles situaciones (o lo que es lo mismo  $P(A_1) + P(A_2) + \dots + P(A_n) = 1$ , el que ocurra alguno de los sucesos A tiene probabilidad 1, suceso seguro). Lo representamos gráficamente en la figura. El cuadrado corresponde a todas las situaciones posibles, que en este caso pueden dividirse en tres:  $A_1, A_2, A_3$ . El suceso B se puede producir en cualquiera de las tres situaciones.

Si reescribimos ahora la anterior ecuación por ejemplo para  $A_1$  tenemos

$$P(A_1 \cdot B) = P(A_1 / B) \times P(B) = P(B / A_1) \times P(A_1)$$

Con un poco de álgebra tenemos

$$P(A_1 / B) = \frac{P(B / A_1) \cdot P(A_1)}{P(B)}$$

$$P(B) = P(B \cdot A_1) + P(B \cdot A_2) + P(B \cdot A_3) = \sum P(B \cdot A_i) = \sum P(B / A_i) \cdot P(A_i)$$

$$P(A_1 / B) = \frac{P(B / A_1) \cdot P(A_1)}{\sum P(B / A_i) \cdot P(A_i)}$$

que constituye el famoso **teorema de Bayes**. Para cualquiera de las otras situaciones ( $A_2, A_3$ ) la fórmula es similar.

## 7.- Metodología bayesiana

En la metodología estadística clásica -frecuentista- se calcula la probabilidad de observar un resultado suponiendo que la realidad sea de una manera determinada (hipótesis nula), sin embargo en la práctica necesitamos los conocimientos para tomar decisiones, y lo que realmente nos interesa es conocer la probabilidad de que las cosas sean de una manera determinada dados los datos (condicionado a..) que hemos observado.

Esta es la diferencia que radica en el enfoque bayesiano. En el caso de las pruebas diagnósticas lo que nos interesa en la práctica es el valor predictivo, positivo o negativo, de la prueba no la sensibilidad o especificidad de éstas.

Aunque estamos habituados a la presentación de los métodos bayesianos con sucesos binarios o dicotómicos (enfermo o sano), también son aplicables cuando los resultados son continuos (por ejemplo proporción de pacientes que sobreviven).

En el análisis estadístico clásico para evaluar por ejemplo la eficacia de un nuevo tratamiento frente al tratamiento anterior se utiliza exclusivamente la información obtenida en el estudio.

El teorema de Bayes es bastante importante, pues relaciona las probabilidades “**a priori**”  $P(A_i)$  con las probabilidades “**aposteriori**”  $P(A_i / B)$  (Probabilidad de  $A_i$  después que ha ocurrido el evento B).

Supongamos que se está comparando la tasa de mortalidad. Para utilizar la terminología habitual vamos a llamar a esa magnitud que interesa calcular  $\theta$ .

Lo primero que hay que determinar es la distribución de probabilidad de esa magnitud con la información externa de la que se dispone, es lo que se denomina **probabilidad a priori** y vamos a representar como  $P(\theta)$ . Seguidamente se cuantifica

la información que aportan los datos observados en nuestro estudio mediante lo que se denomina **función de verosimilitud** (likelihood), que denotaremos como  $P(\text{Datos} / \theta)$ . La verosimilitud representa la probabilidad de los datos observados para cualquier valor del parámetro  $\theta$ . Podemos ahora utilizar el teorema de Bayes para actualizar el valor a priori  $P(\theta)$  a la luz de los datos obtenidos y calcular  $P(\theta / \text{Datos})$ , la denominada función de **probabilidad a posteriori**, es decir cómo de probables son los diferentes valores posibles de  $\theta$  una vez obtenidos nuestros datos. Según el teorema de Bayes tenemos que

$$P(\theta / \text{Datos}) \propto P(\theta) \cdot P(\text{Datos} / \theta)$$

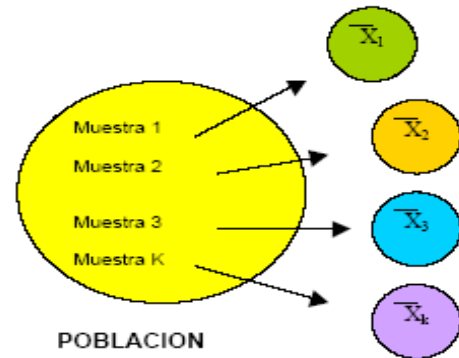
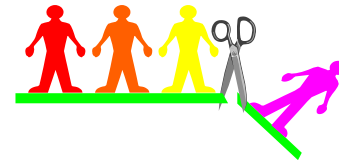
donde el símbolo  $\propto$  indica que el lado de la izquierda es proporcional al lado de la derecha, es decir que son iguales salvo por un término constante (el denominador del teorema de Bayes) que no depende del parámetro de interés  $\theta$ .

Así pues los resultados se expresan como una función de la probabilidad a posteriori de los diferentes valores de  $\theta$ .

## 8.- Inferencia Estadística

Inferencia Estadística es la ciencia que permite hacer conclusiones acerca de una población a partir de la información de una muestra. Esta definición deja algunas preguntas por responder tales como:

- ¿Qué es una población?
- ¿Qué es una muestra?
- ¿Cómo se debe extraer la muestra?
- ¿Cómo se relacionan la muestra y la población?

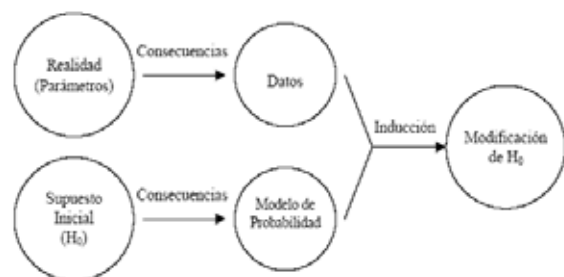


Una respuesta a esta última pregunta es “a través de probabilidades” (condicionales). En pruebas de hipótesis clásicas se definen dos tipos de error cuyas probabilidades están dadas por:

$\alpha$  = Probabilidad de rechazar  $H_0 / H_0$  es verdadera.

$\beta$  = Probabilidad de aceptar  $H_0 / H_0$  es falsa.

El proceso de prueba de hipótesis consiste en evaluar cuán probable es la muestra, dada una población desconocida pero definida a partir de una hipótesis nula. Este proceso se puede esquematizar de la siguiente manera:



1. Se tiene una “realidad” desconocida la cual se desea estudiar (parámetros).
2. La realidad desconocida es no observable pero sus consecuencias sí lo son (datos).
3. Aunque la realidad es desconocida, se tiene un supuesto inicial acerca de ella (hipótesis nula).
4. Esta hipótesis inicial conduce, por un proceso de deducción, a ciertas consecuencias necesarias



(modelo de probabilidad) que pueden ser comparadas con los datos.

5. Si las consecuencias de la hipótesis nula y los datos no concuerdan, la discrepancia puede conducir, por un proceso de inducción, al rechazo de la hipótesis nula.

La pregunta ahora es, cómo evaluar esa discrepancia. La respuesta es el “valor de probabilidad”.

El valor de probabilidad ( $p$ ) mide la probabilidad de obtener datos al menos tan discordantes como los obtenidos en la muestra bajo el modelo de probabilidad supuesto.

Si el valor de probabilidad es alto, los datos obtenidos son muy probables bajo el modelo de probabilidad supuesto (poco discordantes con las consecuencias de  $H_0$ ) y por lo tanto se concluye en que no hay suficiente evidencia para rechazar  $H_0$ .

Si el valor de probabilidad es bajo, los datos obtenidos son poco probables bajo el modelo de probabilidad supuesto (muy discordantes con las consecuencias de  $H_0$ ) y por lo tanto se concluye en que sí hay suficiente evidencia para rechazar  $H_0$ .

## 9.- Inferencia Bayesiana

Un punto importante en la definición clásica de inferencia es que el parámetro  $\theta$ , el cual es desconocido, es tratado como constante en vez que como variable. Esta es la idea fundamental de la teoría clásica, pero conduce a ciertos problemas de interpretación.

Si  $[0.2, 0.3]$  es un intervalo del 95% de confianza para  $\theta$ , sería cómodo decir que hay un 95% de probabilidad de que  $\theta$  esté en el intervalo. Sin embargo, esto es incongruente con la idea de que  $\theta$  no es aleatorio. Dado que  $\theta$  es fijo, solo existen dos opciones:  $\theta$  está dentro o está fuera del intervalo. El único elemento aleatorio en este modelo es  $y$ , por lo que la interpretación correcta del intervalo consistiría en decir que, si se repite el procedimiento muchas veces, entonces a la larga, el 95% de los intervalos contendrían a  $\theta$ .

Toda inferencia basada en la estadística clásica es forzada a tener este tipo de interpretación

frecuencial, aunque sin embargo, nosotros solo contamos con un intervalo para interpretar.

El marco teórico en el cual se desarrolla la inferencia bayesiana es idéntico al de la teoría clásica.

Se tiene un parámetro poblacional  $\theta$  sobre el cual se desea hacer inferencias y se tiene un modelo de probabilidad  $p(y|\theta)$  el cual determina la probabilidad de los datos observados  $y$  bajo diferentes valores de  $\theta$ . La diferencia fundamental entre la teoría clásica y la bayesiana está en que  $\theta$  es tratado como una cantidad aleatoria. Así, la inferencia bayesiana se basa en  $p(\theta|y)$  en vez de  $p(y|\theta)$ , esto es, en la distribución de probabilidades del parámetro dados los datos.

La inferencia bayesiana, se puede resumir como el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir los resultados mediante una distribución de probabilidades para los parámetros del modelo y para cantidades desconocidas pero observables tales como predicciones para nuevas observaciones.

La característica esencial de los métodos bayesianos está en su uso explícito de probabilidades para cuantificar la incertidumbre en inferencias basadas en el análisis estadístico de los datos. Esto permite un manejo mucho más natural e intuitivo de la inferencia, salvando por ejemplo el problema de la interpretación frecuencial de los resultados. Sin embargo, para hacer uso de un enfoque bayesiano, es necesario especificar una distribución de probabilidades a priori  $p(\theta)$ , la cual representa el conocimiento que se tiene sobre la distribución de  $\theta$  previo a la obtención de los datos.

Esta noción de una distribución a priori para el parámetro constituye el centro del pensamiento bayesiano y, dependiendo de si se es un defensor o un opositor a esta metodología, su principal ventaja sobre la teoría clásica o su mayor vulnerabilidad.

## 10.- Características de la Aproximación Bayesiana

De acuerdo con O'Hagan (1994), se pueden identificar cuatro aspectos fundamentales que caracterizan la aproximación bayesiana a la inferencia estadística:

- **Información a Priori.** Todos los problemas son únicos y tienen su propio contexto. De tal contexto se deriva información a priori, y es la formulación y uso de esta información a priori la que diferencia la inferencia bayesiana de la estadística clásica.

- **Probabilidad Subjetiva.** La estadística bayesiana formaliza la noción de que todas las probabilidades son subjetivas, dependiendo de las creencias individuales y la información disponible. Así, el análisis bayesiano resulta personal, único de acuerdo con las creencias individuales de cada uno.

- **Auto consistente.** Al tratar al parámetro  $\theta$  como aleatorio, la inferencia bayesiana se basa completamente en la teoría de la probabilidad. Esto tiene muchas ventajas y significa que toda inferencia puede ser tratada en términos de declaraciones probabilísticas para  $\theta$ .

- **No “adhockery”.** Debido a que la inferencia clásica no puede hacer declaraciones probabilísticas acerca de  $\theta$ , varios criterios son desarrollados para juzgar si un estimador particular es en algún sentido “bueno”. Esto ha conducido a una proliferación de procedimientos, frecuentemente en conflicto unos con otros. La inferencia bayesiana deja de lado esta tendencia a inventar criterios ad hoc para juzgar y comparar estimadores al basarse exclusivamente en la distribución posterior para expresar en términos exclusivamente probabilísticos toda inferencia referente al parámetro.

## 11.- Objeciones a la Inferencia Bayesiana.

La principal objeción a la inferencia bayesiana, es que las conclusiones dependen de la selección específica de la distribución a priori. Aunque para otros esto es lo interesante de la aproximación bayesiana, este es un debate aún no cerrado. Sin embargo, antes de dejar esta característica, se debe señalar que inclusive en inferencia clásica, y además en investigación científica en general, estos conocimientos a priori son utilizados implícitamente. Así por ejemplo, el conocimiento a priori es utilizado para formular un modelo de verosimilitud apropiado. En pruebas de hipótesis, las creencias a priori acerca de la plausibilidad de una hipótesis son frecuentemente utilizadas para ajustar el nivel de significancia de la prueba. Así,

si se cree que los datos pueden conducir al rechazo de la hipótesis, esto se puede ajustar escogiendo un nivel de significancia bastante alto. En este sentido entonces, la inferencia bayesiana formaliza la incorporación de la información a priori, la cual es incorporada frecuentemente “debajo de la mesa” en el análisis clásico.

## 12.- Ampliación de Inferencia bayesiana: distribuciones a priori y a posteriori

Por lo comentado se conocen técnicas y procedimientos de la denominada “inferencia estadística clásica” donde, en los casos en que se exponían métodos basados en parámetros poblacionales, se suponía que estos eran desconocidos pero no aleatorios.

La diferencia fundamental que plantea el “**enfoque inferencial bayesiano**” radica en que considera el parámetro desconocido  $\theta$  como variable aleatoria,

a la que se asigna, por tanto, una determinada distribución de probabilidad, que recibe el nombre de **distribución a priori** y se representa por  $g(\theta)$ .

La información suministrada por una muestra aleatoria simple  $\mathbf{X}$  bajo la óptica bayesiana, puede cambiar la idea del comportamiento probabilístico que se tuviera sobre el parámetro  $\theta$  y, en este

sentido, es posible aceptar que exista una **distribución a posteriori** del parámetro  $\theta$ ,

representa por  $g(\theta/X)$ , donde se recoge la

modificación de la distribución de probabilidad de dicho parámetro  $\theta$  cuando se dispone de la información muestral  $\mathbf{X}$ .

Si la distribución a priori del parámetro  $\theta$  se postula

de tipo discreto, aplicando el teorema de Bayes pueden determinarse las correspondientes probabilidades a posteriori para cada uno de los valores que tome el parámetro  $\theta$ . En efecto

$$P(\theta = \theta_i / X) = \frac{P(X / \theta = \theta_i)P(\theta = \theta_i)}{\sum_{i=1}^K P(X / \theta = \theta_i)P(\theta = \theta_i)}$$

Si la distribución a priori  $g(\theta)$  es continua, y la población de donde se extrae la muestra también, puede obtenerse la distribución a posteriori del parámetro  $\theta$  teniendo en cuenta que:

$f(X; \theta)$  representa la función de densidad conjunta del vector aleatorio muestral  $\mathbf{X}$  y del parámetro  $\theta$ .

$f(X / \theta)$  es la función de densidad de la muestra  $\mathbf{X}$  para un  $\theta$  dado.

$f(X)$  es la distribución marginal de la muestra  $\mathbf{X}$

Como se sabe

$$f(X / \theta) = \frac{f(X; \theta)}{g(\theta)}$$

$$g(\theta / X) = \frac{f(X; \theta)}{f(X)}$$

donde

$$\begin{aligned} f(X; \theta) &= f(X / \theta)g(\theta) \\ f(X; \theta) &= g(\theta / X)f(X) \end{aligned}$$

por lo que  $f(X / \theta)g(\theta) = g(\theta / X)f(X)$

y la distribución a posteriori del parámetro  $\theta$  resulta

$$g(\theta / X) = \frac{f(X / \theta)g(\theta)}{f(X)}$$

obteniéndose la densidad marginal  $f(X)$  como

$$f(X) = \int_{\Theta} f(X; \theta)d\theta = \int_{\Theta} f(X / \theta)g(\theta)d\theta$$

llegándose, finalmente, a que

$$g(\theta / X) = \frac{f(X / \theta)g(\theta)}{\int_{\Theta} f(X / \theta)g(\theta)d\theta}$$

versión continua del Teorema de Bayes.

### 13.- Estimacion de Bayes.

El **estimador de Bayes** es la función de decisión, establecida sobre el espacio paramétrico  $\Theta$ , que produzca el menor riesgo esperado.

El estimador será  $\hat{\theta} = d^*(X)$  tal que minimice la expresión

$$B(d) = \int_{\Theta} \int_X l[\theta; d(X)]f(X / \theta)g(\theta)d\theta d\mathbf{X}$$

Existe un procedimiento operativo para determinar la decisión de Bayes sin tener que describir todo el espacio muestral (y, por tanto, trabajar en él).

En el epígrafe anterior establecimos la distribución a posteriori del parámetro  $\theta$ , calculada como

$$g(\theta / X) = \frac{f(X / \theta)g(\theta)}{f(X)}$$

por consiguiente, el riesgo de Bayes queda.

$$\begin{aligned} B(d) &= \int_{\Theta} \int_X l[\theta; d(X)]f(X / \theta)g(\theta)d\theta d\mathbf{X} = \\ &= \int_{\Theta} \int_X l[\theta; d(X)]g(\theta / X)f(X)d\theta d\mathbf{X} = \\ &= \int_X f(X) \left[ \int_{\Theta} l[\theta; d(X)]g(\theta / X)d\theta \right] d\mathbf{X} = \\ &= \int_X f(X) E_{\theta / X} [l(\theta; d(X))] d\mathbf{X} \end{aligned}$$

siendo  $E_{\theta / X} \{ l[\theta; d(X)] \}$ , riesgo de Bayes a posteriori, es decir,

$$E_{\theta / X} \{ l[\theta; d(X)] \} = \int_{\Theta} l[\theta; d(X)]g(\theta / X)d(\theta)$$

Dado que la función de decisión  $d(X)$  que minimiza

$B(d)$  es la misma que hace mínima la expresión (1),

pues  $d(X)$  no depende de  $f(X)$ , puede utilizarse

esta última para determinar el estimador de Bayes, de manera que dicho estimador será aquel que verifique

$$\frac{\partial E_{\Theta/X} [l(\theta; d(X))] }{\partial d(X)} =$$

$$= \Psi[\theta; d(X)] = 0$$

$$\left[ \frac{\partial^2 E_{\Theta/X} [l(\theta; d(X))] }{\partial d^2(X)} \right]_{d(X)=d^*(X)} > 0$$

## Referencias Bibliográficas

YAÑEZ DE DIEGO, I.: *Teoría de muestras*. UNED. Madrid, 1.991.

INFANTE MACIAS, R.: *Teoría de la decisión*. UNED. Madrid, 1.992.

CASTILLO RON, E.: *Introducción a la estadística aplicada*. Ed. el autor. Santander, 1.993.

FELLER, W.: *Introducción a la Teoría de probabilidades y sus aplicaciones*. Volumen 1. Editorial Limusa. México, 1.996.

GARCIA PEREZ, A.: *Inferencia estadística robusta*. UNED. Primer Curso sobre métodos de la teoría de la probabilidad y la inferencia estadística. Avila, Otoño 1.997.

GOMEZ VILLEGAS, M.A.: *?R.A. Fisher: el inicio del análisis multivariante?. 100cias@uned*. N?3. 2.000. Pág.: 51-55.

Diversas consultas en Internet y en Enciclopedias.

*Un barco no debería navegar con una sola ancla,  
ni la vida con una sola esperanza.*

*Epicteto*

## Historia de la Estadística

Esta historia de la estadística es continuación a la historia presentada en la Revista Varianza N° 7. El presente testimonio, muestra a personajes del quehacer estadístico a partir del siglo XIX.

se requiere predecir los resultados futuros y administrarlo económicamente. Durante los 90's, el ingenio de Shewhart fue redescubierto por una tercera generación de directivos industriales como el enfoque "Seis Sigma".

WALTER A. SHEWHART



JOHN WISHART



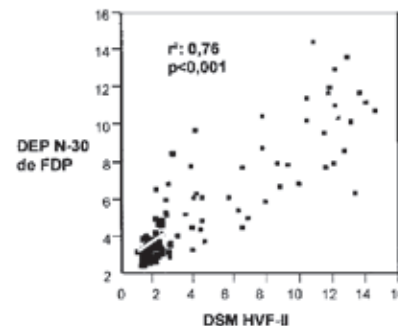
Walter A. Shewhart (1891 - 1967) físico, ingeniero y estadístico estadounidense, es conocido como el padre del **Control Estadístico de la Calidad**, nació en New Canton, Illinois, U.S.A., acudió a la Universidad de Illinois, y se doctoró en física en la Universidad de California en Berkeley en 1917. Shewhart sirvió al Departamento de Guerra de U.S.A. de las Naciones Unidas y del gobierno de la India, de igual forma, trabajó en el Consejo Nacional de Investigación y en el Instituto Internacional de Estadística.

John Wishart (1898 - 1956) nació en Montrose Escocia, asistió a la Academia de Perth y luego, en 1916 ingresó en la Universidad de Edimburgo. Allí fue profesor de matemáticas de ET Whittaker.

La Primera Guerra Mundial hizo que la carrera universitaria de Wishart fuese interrumpida. Completó sus estudios universitarios en 1922, donde se graduó

Fue miembro honorario de la Real Sociedad de Estadística en Inglaterra, y de la Asociación de Estadística de Calcuta. Fue director del Instituto de Estadísticas Matemáticas de la Asociación Americana para el Avance de la Ciencia y formó parte de la Asociación Americana de Estadística.

En 1924, Shewhart delimitó el problema de la calidad en términos de variación de causas asignables y ocasionales e introdujo la gráfica de control como una herramienta para distinguir entre las dos. Shewhart hizo hincapié en que para llevar un proceso de producción a un estado de control estadístico, donde sólo existan variaciones de causas ocasionales, y mantenerlo en control,



en matemáticas y física. Había tomado un curso de formación del profesorado en la Casa de Moray, como parte de su carrera y, después de graduarse, se trasladó a Leeds tras aceptar un puesto como

## GERTRUDE MARY COX

profesor de matemáticas en la preparatoria West Leeds.

En 1924, después de una recomendación de Whittaker, le ofrecieron el puesto de asistente de Pearson en el University College de Londres. Wishart aprendió mucho de la estadística durante sus tres años con Pearson.

Wishart creó un laboratorio en Cambridge para sus alumnos de postgrado, donde Williams Cochran fue su alumno. Wishart tenía mucho talento para la estadística matemática y un instinto para realizar aplicaciones prácticas en diseño experimental.

Posteriormente, Wishart se convirtió en el Jefe del Laboratorio de Estadística de la Facultad de Matemáticas en Cambridge.

Algunas de las publicaciones más importantes de Wishart se realizaron en el período 1928-1932 antes de que se involucrara en la docencia en Cambridge.

En 1928 generalizó la distribución Chi-cuadrado que se denomina la Distribución de Wishart. También estudió las propiedades de la distribución del coeficiente de correlación múltiple que Fisher había considerado antes. También escribió numerosos artículos sobre las aplicaciones agrícolas de las estadísticas.

Wishart estuvo muy involucrado con el trabajo de la Royal Statistical Society. Él fue uno de los becarios que formaron Comité Organizador de la Sección de Investigación de la Agricultura en 1933. En 1945 se convirtió en presidente de la Royal Statistical Society de la sección de investigaciones.

Wishart murió en un accidente en Acapulco, México cuando se encontraba de visita en calidad de representante de la Organización de las Naciones Unidas para la Agricultura y organizando la creación de un centro de investigación para aplicar técnicas estadísticas en la investigación agrícola.



Gertrude Mary Cox (1900 - 1978) nació en Dayton Iowa USA. En 1918 se graduó de la Escuela Secundaria de Perry en Iowa. Tras convertirse en diaconisa de la Iglesia Metodista en 1925 continuó sus estudios en la Universidad de Iowa (Iowa State College), donde estudió matemática y estadística graduándose en 1929 y posteriormente concluyó su maestría en estadística el año 1931.

En el periodo 1931 a 1933, efectuó estudios de postgrado en estadística en la UCB (University of California, Berkeley), luego, retornó a la Universidad Estatal de Iowa, a trabajar como asistente en un laboratorio de Estadística, donde empleó metodologías alternativas en Diseño de Experimentos. En 1939 fue nombrada profesora asistente en el Departamento de Estadística de la Universidad. En 1940 fue nombrada profesora de estadística en la Universidad de Carolina del Norte en Raleigh.

En 1945 se convirtió en directora de la División de Investigación en Estadística de la Universidad Estatal de Carolina del Norte (North Carolina State College) que fue dirigida por Cochran. En 1947 se convirtió en miembro fundador de la Sociedad de Biometría.

En 1950 en colaboración con William Cochran publicó el libro *Diseño Experimental*, que se convirtió en un clásico sobre el tema. En 1960 realizó su último trabajo como Directora de Estadísticas de Research Triangle Institute en Durham, Carolina del Norte. Permaneció en este cargo hasta su jubilación en 1964.

En 1949 se convirtió en la primera mujer elegida por el Instituto Internacional de Estadística (ISI - Instituto Internacional de Estadística) por su eficiente labor a favor de la profesión. En 1956 fue elegida presidente de la American Statistical Association (ASA - American Statistical Association) y en 1975 fue tomada como miembro de la Academia Nacional de Ciencias.

## FRANK YATES



Frank Yates (1902–1994) nació en Manchester, Inglaterra. Fue educado en Wadham House, una escuela privada. Obtuvo una beca para Clifton College en 1916. Cuatro años más tarde se le concedió una beca para estudiar en la Universidad de San Juan en Cambridge. Se graduó en la “Clase Matrícula de Honor” (con honores de primera clase) en 1924.

Después de dos años como docente en matemáticas, decidió abandonar la enseñanza y se unió a la compañía Gold Coast Survey como asesor matemático. Debido a su estado de salud, decidió trabajar en Inglaterra y después de hacer una solicitud a RA Fisher, fue nombrado asistente estadístico en la Estación Experimental de Rothamsted en 1931. En 1933 se convirtió en jefe de Estadísticas de la mencionada institución, ocupó este cargo hasta su jubilación en 1968.

Yates trabajó bastante con el Diseño Experimental, colaborando a Fisher en varios proyectos de investigación. Desarrolló los métodos computacionales para el análisis de experimentos a través de mínimos cuadrados y creó los bloques incompletos modernos como los lattices.

Durante la Segunda Guerra Mundial, estudió el suministro de alimentos y la aplicación de fertilizantes para aumentar los rendimientos de la tierra. Aplicó sus técnicas estadísticas de diseño de experimentos en problemas tales como el control de plagas. Después de 1945 se continuó aplicando sus técnicas estadísticas a los problemas de la nutrición humana.

Frank fue nombrado parte de la Comisión Estadística de Muestreo de las Naciones Unidas y publicó los métodos de muestreo para encuestas y censos en 1949.

Yates era un entusiasta usuario de la computadora y comentaba: *“Para ser un buen estadístico teórico también se debe utilizar el equipo y por lo tanto debe tener los mejores recursos de computación”*.

Fue una de las personas que influyeron en la creación de la British Computer Society (Sociedad Británica de Computación) y fue su presidente en 1960 y 1961. Después de su retiro se convirtió en investigador senior en el Imperial College de Londres.

En 1966 La Royal Society de Londres le dio la Medalla Real por sus contribuciones de largo alcance profundo en los métodos de la biología experimental.

## JOHN WILDER TUKEY



John Wilder Tukey (1915-2000) estadístico nacido en New Bedford, Massachusetts. Tukey obtuvo el título de Bachiller en Artes en 1936 y una maestría

en Ciencias en 1937, ambas en química, en la Universidad de Brown, luego obtuvo su doctorado en Matemáticas en la Universidad de Princeton.

Durante la Segunda Guerra Mundial, Tukey trabajó en la Oficina de la Investigación de Control de Fuego de Artillería y colaboró con Samuel Wilks y William Cochran. Después de la guerra, regresó a Princeton dividiendo su tiempo entre la universidad y los laboratorios AT&T Bell.

Tukey ha sido uno de los grandes talentos estadísticos del siglo XX, con importantes contribuciones a la Topología, Visualización de Información y en especial a la estadística, incluyendo la filosofía.

Tukey introdujo las modernas técnicas para estimar el espectro de las Series Temporales. En 1965, en un artículo conjunto con J. W. Cooley, publicado en la revista *Mathematics of Computation*, introdujo el algoritmo de la Transformada Rápida de Fourier (FFT), fundamental para crear el procesamiento digital de datos.

En 1970, contribuyó significativamente en lo que ahora conocemos como la Estimación Jackknife, también denominada la Quenouille-Tukey Jackknife.

Fundó el Análisis Exploratorio de Datos o EDA (Exploratory Data Analysis), una nueva aproximación a la estadística que usa fuertemente un conjunto de técnicas basadas en el uso de gráficos. Introdujo los diagramas de caja (Box Plot) en su libro de 1977, *Análisis Exploratorio de Datos*. Tukey está detrás de muchas nuevas y sencillas formas de tener en cuenta las magnitudes estadísticas. Entre ellas destacan los gráficos “Box and Whisker Plot” (diagrama de caja y bigotes), el “Stem-and-Leaf Diagram” (diagrama de tallo y hojas), los “radigramas” (rootograms) y los diagramas de ajuste.

## GEORGE EDWARD PELHAM BOX



George Edward Pelham Box (1919) químico, matemático, estadístico inglés, en el área de la estadística realizó importantes contribuciones en las áreas de Control de Calidad, Análisis de Series Temporales, Diseño de Experimentos y la Inferencia Bayesiana.

Box nació en Gravesend, Kent, Inglaterra y se formó como químico. Durante la Segunda Guerra Mundial, trabajó en experimentos bioquímicos sobre el efecto de gases tóxicos en los animales pequeños para el ejército británico. Necesitaba asesoramiento estadístico para analizar los resultados de sus experimentos, pero no pudo encontrar un estadístico que podría darle orientación, por lo que él mismo estudio la estadística de los textos disponibles. Después de la guerra, se matriculó en la Universidad College de Londres y obtuvo una licenciatura en matemáticas y estadística. Recibió un doctorado de la Universidad de Londres en 1953.

Box ha escrito artículos de investigación y libros publicados, estos incluyen *Análisis de Series de Tiempo: Previsión y Control* (1979, con Gwilym Jenkins) e *Inferencia Bayesiana en el Análisis Estadístico* (1973, con George C. Tiao). Hoy, su nombre está asociado con resultados importantes en las estadísticas como el Modelo de Box-Jenkins, transformaciones de Box-Cox, diseños de experimentos y muchos otros.

George Box se casó con Joan Fisher, la segunda hija de Ronald Fisher.



El estadístico George Box acuñó, en 1953, el término robustez, para designar los métodos estadísticos que procuran asegurar resultados aceptables, cuando no se cumplen los supuestos estándares en que se basan los métodos estadísticos regulares. Ya desde fines del siglo pasado hubo científicos que se preocuparon del tema. Se dieron cuenta de los peligros de hacer inferencias, cuando los datos aparecen contaminados con valores extraños, y llegaron a proponer modelos y estimadores robustos como alternativas para estos casos. Sin embargo, no es sino a partir de la década de 1960, que este tópico es reconocido como un tema de investigación en estadística.

Box se desempeñó como presidente de la Asociación Americana de Estadística en 1978 y del Instituto de Estadística Matemática en 1979. Recibió la medalla Shewhart de la Sociedad Americana de Control de Calidad en 1968, el Premio en memoria a Wilks de la Asociación Americana de Estadística en 1972, el Conferenciante RA Fisher en 1974, y la Medalla de Guy de Oro de la Royal Statistical Society en 1993. Box fue elegido miembro de la Academia Americana de Artes y Ciencias en 1974 y miembro de la Royal Society en 1979.

## HAROLD HOTELLING



Harold Hotelling (1895 – 1973) norteamericano, nació en Fulda Minnesota, fue un estadístico matemático y un economista teórico influyente.

Demostrando que no todo depende de Fisher, trabajó con la prueba multivariada (T de Hotelling) que generaliza la prueba de Student a varias dimensiones.

Su nombre es conocido por todos los estadísticos debido a la Distribución de T-Hotelling y su uso en pruebas de hipótesis y de confianza.

También presentó el Análisis de Correlación Canónica, y es el epónimo de la Ley de Hotelling, el Lema de Hotelling y la Regla de Hotelling en la economía.

Hotelling ocupa un lugar crucial en la genealogía de la teoría económica moderna, varias áreas de investigación activa se vieron influidos por sus artículos de economía.

Además de su enseñanza, escritura y actividades profesionales, Hotelling trabajó en los consejos editoriales de la revista *Annals of Mathematical Statistics* y la *American Journal of Economics and Sociology*. Fue elegido presidente del Instituto de Estadística Matemática y la Sociedad Econométrica. También fue elegido miembro de la Asociación Americana de Estadística en 1937.

## JERZY NEYMAN



Jerzy Neyman (1894-1981) matemático y estadístico nacido en Moldavia Polonia.

En 1934, Jerzy Neyman publicó en la Royal Statistical Society, de Londres, lo que puede ser considerado como el primer trabajo científico sobre el Muestreo de Poblaciones Finitas. Estableció, sin lugar a dudas, que la selección aleatoria es la base de una teoría científica que permite predecir la validez de las estimaciones muestrales.

También dejó establecido toda una filosofía sobre la eficiencia de la estrategia muestral. Neyman y Egon Pearson, hijo de Karl Pearson, presentaron en 1936 una teoría sobre la forma de probar hipótesis estadísticas, en base a datos. Esta presentación promovió mucho interés, estimuló una considerable cantidad de investigación, y muchos de los resultados hasta hoy aún se usan.

Ellos resolvieron dificultades fundamentales para la comprensión de las pruebas de hipótesis, introduciendo las nociones de hipótesis alternativa, y los dos tipos de error, el de rechazar una hipótesis que es verdadera, y el de no rechazar una hipótesis que es falsa. De ello, surge un resultado fundamental, el Lema de Neyman - Pearson, y se crea una larga controversia con R.A. Fisher, que visualizaba la prueba de hipótesis como un procedimiento mediante el cual el investigador podía formarse una opinión sobre alguna característica de la población, o parámetro. Neyman y Pearson vieron la prueba de hipótesis como un medio para que el investigador tomara una decisión sobre un parámetro de la población. El Lema de Neyman - Pearson es considerado como la base de la moderna teoría de la decisión.

Neyman introdujo, en 1934, la teoría de los intervalos de confianza. Es una forma de estimar un parámetro, contrapuesta a la estimación puntual, que determina un intervalo que contiene al parámetro, y un coeficiente de confianza, que representa la probabilidad que el intervalo efectivamente contenga al parámetro. Los intervalos de confianza y las pruebas de hipótesis son dos elementos de la inferencia estadística fundamentales.

En las décadas de los años 30 y 40 se centra en el desarrollo de la técnica del Muestreo Estratificado en la población. Jerzy Neyman fue un arduo defensor del método de muestreo estratificado, y su trabajo en este tema, abrió nuevas áreas de investigación. También desarrollo el método de muestreo por conglomerados. Durante el periodo de 1930 y 1940 se establecen las condiciones bajo las cuales estos métodos, el muestreo estratificado y el muestreo por conglomerados, resultan más eficientes.

## WILLIAM GEMMELL COCHRAN



William Gemmell Cochran (1909 – 1980) fue matemático y un destacado estadístico escocés.

Estudió matemáticas en las universidades de Glasgow y Cambridge y posteriormente trabajó en la “Estación Experimental de Rothamsted” entre los años 1934 a 1939, fecha en que emigró a EE.UU. Allí colaboró en la creación de varios departamentos de estadística y trabajó en la Universidad de Harvard desde 1957 hasta la fecha de su jubilación en 1976.

En el área de la estadística, realizó contribuciones al Diseño de Experimentos y a la Teoría del Muestreo.

Escribió muchos artículos y libros, algunos de los cuales se han convertido en referencias fundamentales y de los que podemos mencionar:

- Cálculo e Interpretación del Análisis de Varianza y Covarianza, que publicó en 1934 - Experimental Designs (con Gertrude Mary Cox), 1950.

- Statistical Methods Applied to Experiments in Agriculture and Biology, de George W. Snedecor, pero al que Cochran contribuyó a partir de la quinta edición.

## FRANK WILCOXON



Frank Wilcoxon (1892-1965) químico y estadístico nacido en Estados Unidos, entre sus contribuciones en la estadística se tienen las Pruebas de Rango, las Comparaciones Múltiples, Rangos Secuenciales y Experimentos Factoriales.

A lo largo de su investigación, encontró que los métodos estadísticos eran numéricamente simples, fáciles de comprender y de aplicar.

Wilcoxon introdujo sus dos pruebas de rango, la prueba de la suma de rangos para dos muestras y la prueba del rango con signo para muestras apareadas, en un artículo el año 1945. Con este artículo y el de su contemporáneo Mann Whitney, comenzó el amplio desarrollo de la Estadística No Paramétrica. No hay duda que este artículo fue la contribución más importante de Wilcoxon ya que se convirtió en la mayor inspiración para el desarrollo de los métodos no paramétricos, además la metodología introducida ha tenido un amplio impacto en la estadística aplicada, particularmente por sus aplicaciones a las ciencias sociales, convirtiéndose en la herramienta estadística más popular.

En 1963 estudió las propiedades de la prueba de la suma de rangos donde se permite la dependencia dentro de las muestras. Wilcoxon estaba interesado

en la generalización de su procedimiento básico de la prueba del rango a nuevas situaciones. Frank fue el iniciador de la investigación en métodos secuenciales no-paramétricos. Debido a que las técnicas no paramétricas fueron tan exitosas en situaciones con muestras de tamaño fijo, pensó que sus buenas propiedades se podrían ampliar en forma natural a las situaciones secuenciales. Esta idea condujo a un gran número de procedimientos de rangos secuenciales desarrollados por Wilcoxon.

## ANDRÉI NIKOLÁYEVICH KOLMOGÓROV



Andréi Nikoláyevich Kolmogórov (1903 - 1987) matemático ruso que hizo progresos importantes en los campos de la Teoría de la Probabilidad y de la Topología. En particular, desarrolló una base axiomática que supone el pilar básico de la Teoría de la Probabilidad a partir de la Teoría de Conjuntos. Al principio de su carrera trabajó en lógica constructivista y en las Series de Fourier.

También trabajó en la mecánica clásica. Asimismo, fue el fundador de la Teoría de la Complejidad Algorítmica. Obtuvo su doctorado en la Universidad Estatal de Moscú.

En 1922 Kolmogórov publicó sus primeros resultados en teoría de conjuntos y un año más tarde, construyó una Serie de Fourier, obteniendo un notable reconocimiento internacional.

Alrededor de este tiempo, decidió dedicar su vida a la matemática y publicó ocho trabajos sobre la

Teoría de la Integración, Análisis de Fourier y sobre la Teoría de Probabilidad.

Kolmogórov y el matemático Sydney Chapman desarrollaron el conjunto de ecuaciones fundamentales en el campo, las ecuaciones de Chapman-Kolmogórov.

## ABRAHAM WALD



Abraham Wald (1902-1950) matemático austríaco, nacionalizado estadounidense. Exiliado a EE.UU. en 1938, se especializó en estadística y aportó a esta ciencia con un elevado rigor matemático. Fue el fundador del Análisis Secuencial.

Desde 1927 curso estudios en la Universidad de Viena, doctorándose en 1931. Formó parte del Institute for Business Cycle Research. Debido a la invasión nazi, se va a los Estados Unidos en 1938, donde es parte del equipo de la Cowles Commission Research y alumno de Harold Hotelling en Columbia University (New York).

Después fue profesor, caracterizándose por explicar complejos conceptos de manera muy sencilla. También desarrolló proyectos militares para el Statistics Research Group y desarrolló la teoría del Análisis Secuencial, aplicada a la II Guerra Mundial para mejorar los estudios de control

de la calidad. Destacó en programación Lineal, geometría y estadística, propuso varias aplicaciones matemáticas a la economía de cierta relevancia. Se interesó en la econometría, proponiendo modelos para suavizar la estacionalidad de las series temporales, y creando la “Prueba de Wald”, que data de 1939.

## CALYAMPUDI RADHAKRISHNA RAO



Calyampudi Radhakrishna Rao nació en Hadagali, Estado de Karnataka India, el 10 de septiembre de 1920. Actualmente es profesor emérito de la Universidad Estatal de Pensilvania.

Trabajó en el Instituto Estadístico de la India y en el Museo Antropológico de Cambridge antes de conseguir el doctorado en el King's College de la Universidad de Cambridge. El tutor de tesis doctoral, fue R.A. Fisher.

Sus mayores descubrimientos fueron la Cota de Cramér-Rao, y el Teorema de Rao-Blackwell, ambos relacionados con la calidad de los estimadores. También trabajó en Análisis Multivariante y Geometría Diferencial.

Rao obtuvo el premio de la American Statistical Association, el Wilks Memorial Award en 1988. También fue premiado con la Medalla Mahalanobis. En junio de 2002, Rao fue premiado con la Medalla Nacional de Ciencias de Estados Unidos.

GEORGE CASELLA



Distinguido profesor en el Departamento de Estadísticas y miembro del Instituto de Genética para la Universidad de Florida. Investigador activo en muchos aspectos de la estadística, contribuyendo en la Teoría Estadística en las áreas de Teoría de la Decisión y Estadística confidencial para la Estadística medioambiental. Su más recientemente esfuerzo, está concentrado en la Estadística Genoma y Metodología en Ciencias Políticas. También mantiene activa investigación en la Teoría y Aplicación de Monte Carlo y otros métodos intensivos computacionales.

Es miembro de la Asociación Americana de Estadística. El instituto de Estadística Matemática, y un electo miembro del Instituto Internacional de Estadística (IIE) donde se habla de él como un muy buen investigador y recientemente fue electo miembro extranjero de la Academia de Ciencia Española Royal.

Casella se desenvuelve como editor de teoría y métodos de la revista de la Asociación de Estadística Americana, actualmente es editor conjunto de la revista de la Sociedad Estadístico Royal.

Casella es autor de 8 libros:

- Variance Components.
- Theory of Point Estimation.
- Statistical Inference.
- Monte Carlo Statistical Methods.
- Statistical Analysis of Quantitative Traits.
- Statistical Design.
- Introduction to Monte Carlo Methods with R.

ROGER BERGER



Recibido en matemáticas de la Universidad de Kansas, tiene una maestría y doctorado en Estadística de la Universidad de Purdue, finalizó sus estudios de PhD en 1977. Brindó sus servicios como asistente y profesor para la Universidad del Estado de Florida. En 1982, paso a la facultad del departamento de Estadística en la Universidad del Estado de Carolina del Norte como docente destacado.

Su interés profesional incluye la teoría estadística y la educación. Recientes investigaciones incluyen métodos exactos para datos categóricos y biomédicos incluyendo los test de equivalencia. Fue uno de los primeros en desarrollar la teoría del test de unión e intersección. Es co-autor con George Casella del libro, Inferencia Estadística.

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

*Richard Dawkins*

## Estadísticas del Titanic

Eran las 13:30 horas del jueves 11 de abril de 1912. Los aproximadamente 1.500 pasajeros del Titanic flameaban alegremente sus pañuelos. Tripulado por 765 hombres al mando del Capitán Edward J. Smith, el buque más seguro y lujoso conocido, se separaba lentamente de los muelles de Queenstown (Irlanda) en su viaje inaugural a Nueva York.

Considerado insumergible, era el mayor (52.310 toneladas y la altura de un edificio de 11 plantas) ingenio móvil jamás construido por el hombre. Tres días más tarde la fastuosa mansión flotante avanzaba a 23 nudos sobre las gélidas aguas del Atlántico Norte. A las 23:40 del domingo 14 de abril, el vigía de la campana de cofa avistó un iceberg, a unos 450 mts. al frente.

Sólo 37 segundos después, el hielo abría una brecha de 75 mts. de longitud en su costado de estribor por debajo de su línea de flotación. Dos horas y cuarenta minutos más tarde, el buque desaparecía bajo las heladas aguas llevándose con él a más de 1.500 personas.

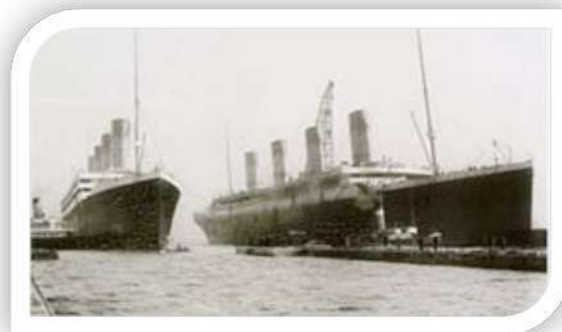
La historia del Titanic ha fascinado a millones de personas. Su tragedia fue debida a un exceso de arrogancia y prepotencia del hombre, que marcó aquella época llamada “La Belle Époque”

Su plan consistía en construir 3 grandes transatlánticos. Cada uno debería ser el 50% más largo que el Lusitania que desplazaba 30.000 toneladas. Un año después, el dique seco acogía al primero de ellos, El Olympic. Tres meses después, el 31 de marzo de 1909, el dique seco acogía al Titanic.



El Titanic junto a su gemelo el Olympic, ambos en fase de construcción (foto de la época).

El 2 de abril de 1912, el Titanic estaba terminado y listo para entrar en servicio. El Titanic tardó 3 años y varios millones de horas de trabajo en su construcción.



Con sus ventanas de vidrios tintados, maderas nobles, y adornos chapados en oro, el Titanic era en realidad una verdadera obra de arte.

En 1907, J. Bruce Ismay, Director de la White Star Line, junto con Lord James Pirrie, de la Harland & Wolff, acordaron construir un barco, el mas grande de todos los construidos para la White Star Line.



J. Bruce Ismay.  
(Foto de la Epoca)



Camarote y gran escalera de Primera Clase. (foto de la época)

Estadísticas del total de pasajeros a bordo del buque, total de pasajeros que sobrevivieron y de los que murieron, se muestra en el cuadro siguiente:

### Cuadro

Estadísticas del Titanic de pasajeros a bordo, sobrevivientes y fallecidos

Clase	Contingencia	Tipo de persona							
		Hombres		Mujeres		Niños		Total	
		Conteo	(%)	Conteo	(%)	Conteo	(%)	Conteo	(%)
1ra	A bordo	176	21,8	143	34,6	7	6,1	382	24,9
	Sobrevivientes	58	7,2	139	33,7	6	5,2	244	15,9
	Fallecidos	118	14,6	4	1,0	1	0,9	139	9,0
2da	A bordo	168	20,8	92	22,3	25	21,7	328	21,4
	Sobrevivientes	14	1,7	80	19,4	25	21,7	140	9,1
	Fallecidos	154	19,1	12	2,9	0	0,0	188	12,3
3ra	A bordo	462	57,3	178	43,1	83	72,2	823	53,7
	Sobrevivientes	75	9,3	91	22,0	30	26,1	227	14,8
	Fallecidos	387	48,0	87	21,1	53	46,1	596	38,9
Total	A bordo	806	100,0	413	100,0	115	100,0	1534	100,0
	Sobrevivientes	147	18,2	310	75,1	61	53,0	611	39,8
	Fallecidos	659	81,8	103	24,9	54	47,0	923	60,2

Fuente: Elaboración propia

Como se puede ver, del 100% de los pasajeros a bordo, el 24,9% eran de primera clase, 21,4% de segunda clase, y 53,7% de tercera clase. De ellos, el 15,9% sobrevivieron en primera clase, 9,1% en segunda clase, y 14,8% sobrevivieron en tercera clase.

Lo más gravitante de los resultados es que el mayor porcentaje de personas fallecidas se da en tercera clase con 38,9%. Como era de esperar, en segunda clase los fallecidos alcanzan en un 12,3%, y sólo el

9,0% de los pasajeros fallecidos son de primera clase.

Del total de hombres, el 81,8% fallecieron. De 413 mujeres a bordo del Titanic, 103 fallecieron, esto significa un 24,9%, y de 115 niños a bordo el 47,0% fallecieron. Estos datos diferenciados por clase, cambian fuertemente. El 48,0% de los hombres de tercera clase fallecieron, mientras que sólo el 19,1% y 14,6% de segunda y primera clase en hombres, murieron. En el caso de las mujeres, el 21,1% de

tercera clase fallecieron, mientras que en segunda clase, sólo un 2,9% murieron y 1,0% en primera clase.

En el caso de los niños, casi todos los fallecidos fueron de tercera clase, que alcanzan a un total de 53.

Los sobrevivientes mayores del Titanic, fueron mujeres en 75,1% de ellas. En el caso de los hombres, sólo un 18,2% sobrevivieron, y en los niños, un 53,0% salvaron la vida.

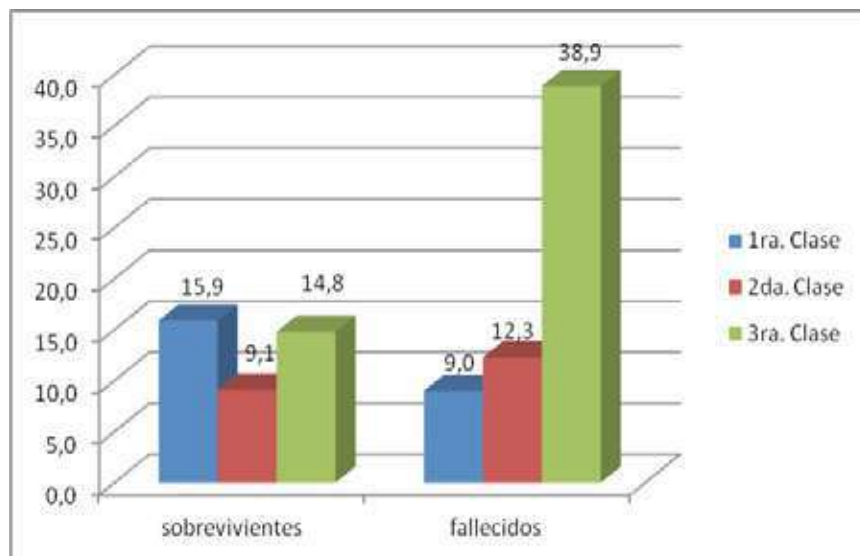
Se tenían a disposición 16 botes salvavidas y 4 plegables, admitibles para 1.100 personas de las 2.300 que se encontraban en el barco. De las 1.100 plazas, sólo se llegaron a ocupar 750 ya que la

evacuación se realizó de forma apresurada y muchos botes se arriaron ocupados por debajo de su capacidad.

De las 2.300 personas nombradas, 1.534 eran pasajeros, a ello se suman las personas de la tripulación con un total de 766.

Del grupo de 766 personas que pertenecían a la tripulación, fallecieron 600 personas aproximadamente. Haciendo un total de 1.523 almas desaparecidas. Finalmente sobrevivieron un total de 777 personas entre pasajeros y tripulantes haciendo un porcentaje de 33,8%. Las pérdidas humanas en porcentaje, es de 66,2%. Se pudieron rescatar un total de 306 cuerpos del agua.

Gráfico de fallecidos y sobrevivientes



Fuente: Elaboración propia

En conclusión, los datos demuestran la discriminación entre personas que viajaban en primera, segunda y tercera clase del barco más importante de la época. La prioridad de incluir a la tripulación y a pasajeros de primera clase en botes salvavidas es evidente, diferenciando incluso, a mujeres y niños de tercera clase. De igual manera, los hombres no tenían mucha posibilidad de salvarse de la tragedia y mucho más aún aquellos que eran de tercera clase.

En el lado opuesto de la balanza, ser hombre y pasajero del Titanic era casi una sentencia de muerte ya que sólo se salvaron el 18,2% sin gran diferenciación de clases. Es cierto que en proporción se salvaron más hombres de primera clase, pero la tragedia se cebó con los hombres que viajaban en tercera, sobreviviendo solamente un 9,3%.

Entre los miembros de la tripulación se repitió el mismo esquema; sólo sobrevivieron aproximadamente



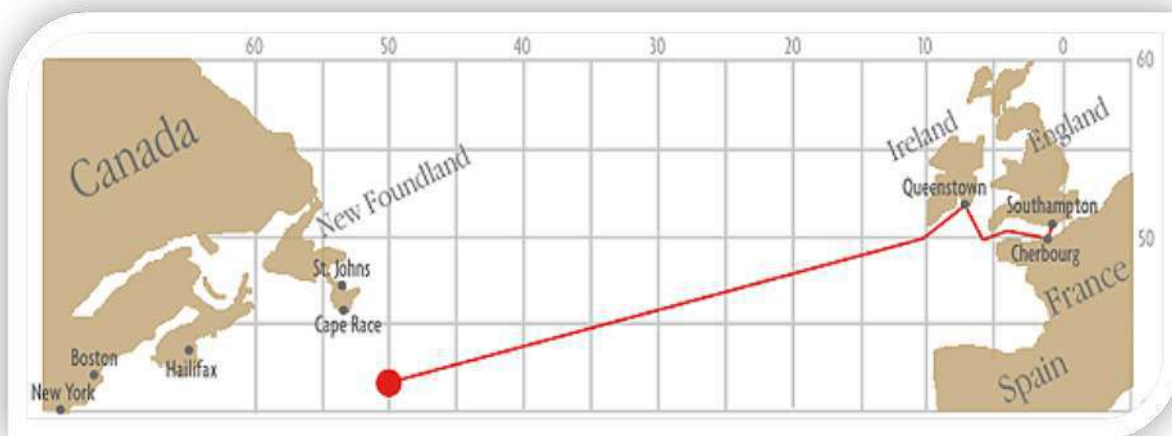
el 20,0% de los o hombres mientras que sólo se perdieron 3 de las 20 mujeres que la integraban.

A partir de las 2:00 de la mañana las posibilidades de sobrevivir se redujeron drásticamente. El último bote salvavidas, el N° 4, fue fletado a las 2:10 y en ese momento sólo quedaban 4 botes plegables para casi 1600 personas.

A partir de las 2:00 la mejor opción para sobrevivir era lanzarse al agua y tratar de alcanzar uno de los 4 botes plegables o al bote N° 4. Algunos pocos pasajeros consiguieron nadar hasta el que fue el primero en alejarse y se sabe con seguridad

que un mínimo de entre 44 y 48 personas y un máximo de 79, lograron sobrevivir mediante este método. El dato exacto no se conoce porque al ser interrogados en la investigación que hubo después, muchos mintieron y dijeron que habían saltado al agua, por la vergüenza de no haber cedido su plaza en el bote a alguna mujer o niño. De las más de 1.500 personas que se quedaron en el barco o que no saltaron, sólo sobrevivieron 4.

Según los registros, el Titanic se hundió a eso de las 2:20 de la mañana del 15 de abril de 1912. Poco después de la colisión, Thomas Andrews se reunió con el Capitán Smith. Fue durante esa reunión cuando Andrews supo que el barco se hundía.



Mapa del lugar donde ocurrió la tragedia

Sabiendo que el barco estaba condenado, el Capitán Smith ordenó el inmediato abandono del Titanic. Su mayor problema era que el Titanic no disponía de los suficientes botes salvavidas para evacuar a las 2.200 personas que había a bordo.

Esto fue debido a que la regulación Inglesa requería que cualquier barco que superase las 10.000 toneladas llevase 16 botes salvavidas, y el Titanic desplazaba 52.310.

Eran las 2:20 de la mañana, y habían transcurrido 3 horas desde que chocara con el Iceberg. Poco tiempo después del hundimiento del Titanic, los lamentos fueron apagándose, todo lo que quedó fueron los botes salvavidas y la oscuridad.



Uno de los botes salvavidas del Titanic. (foto de la época)

Hasta el 1 de Septiembre de 1985, siempre se había pensado que los restos del Titanic permanecerían en muy buen estado, dada la profundidad de las aguas donde reposaba, ya que al ser una profundidad muy elevada (12.000 pies o 4.000 metros), se pensó que la acción de los microorganismos en los restos del barco habría sido mínima, incluso se llegó a pensar

en poder sacar a la superficie al Titanic.

Pero después de la expedición y de encontrar los restos, la realidad era otra. Lo que parecían formaciones orgánicas por todo el barco, no era sino herrumbre producida por la corrosión del acero, el barco ‘sangraba herrumbre’.



Sección de Proa y de Popa del Titanic, tal y como se descubrió.

*Para triunfar en la vida, no es importante llegar primero. Para triunfar simplemente hay que llegar, levantándose cada vez que se cae en el camino.*

## Paradoja de los Cumpleaños

<http://www.estadisticaparatodos.es/taller/cumpleanos/cumpleanos.html>

La paradoja del cumpleaños establece que si hay 23 personas reunidas existe una probabilidad del 50,7% de que al menos dos personas de ellas cumplan años el mismo día. Para 60 o más personas la probabilidad es mayor del 99%. Obviamente es casi del 100% para 366 personas (teniendo en cuenta los años bisiestos).

En sentido estricto esto no es una paradoja ya que no es una contradicción lógica; es una paradoja en el sentido que es una verdad matemática que contradice la común intuición. Cuando se propone este problema por primera vez y se pide una estimación sobre el tamaño mínimo que debería tener un grupo para que sea más probable que improbable que dos personas compartan el día del cumpleaños, la mayoría de las personas se equivoca por completo. La respuesta intuitiva que se da a menudo es 183, es decir 365 dividido entre dos. La cantidad correcta no es algo a lo que la gente pueda llegar fácilmente y, ciertamente, no por intuición. Es bastante extraño que las primeras estimaciones sean inferiores a 40. Y sin embargo la respuesta es 23.

La clave para entender estas “sorprendentes” recurrencias es pensar que hay muchas posibilidades de encontrar parejas que cumplan años el mismo día.

Un análisis superficial asume que 23 días (cumpleaños de las 23 personas) es una fracción demasiado pequeña del posible número de días distintos (365) para esperar repeticiones. Y así sería si esperáramos la repetición de un día dado. Pero las repeticiones, en el caso supuesto, pueden darse entre dos días cualesquiera, con lo que éstas pueden combinarse entre sí de un número de formas que aumenta rápidamente con el número de elementos a considerar. Así:

- Entre dos personas C1 y C2 sólo cabe una posibilidad de repetición de cumpleaños:  $C1=C2$ .

- Con tres ya hay tres posibilidades ( $C1=C2$ ;  $C1=C3$ ;  $C2=C3$ ).
- Con cuatro ya habría seis,  $(4 \times 3)/2=6$ .
- Con un grupo de 10 personas,  $(10 \times 9)/2=45$  posibilidades.
- Con 23 personas, hay  $(23 \times 22)/2 = 253$  parejas distintas, cada uno de ellas es una candidata potencial para cumplir la paradoja.
- Y así sucesivamente, en uno de 40, ya son 780 las parejas, y 1770 si juntamos 60 personas.

No hay que mal interpretar lo que nos dice esta paradoja: si entramos en una habitación con 22 personas, la probabilidad de que cualquiera cumpla años el mismo día que usted, no es del 50%, es mucho más baja, sólo hay un 6% de probabilidades. Esto es debido a que ahora sólo hay 22 parejas posibles y se necesitan 253 personas para que haya más de un 50% de probabilidades de que esto ocurra.

El problema real de la paradoja del cumpleaños consiste en preguntar si el cumpleaños de cualquiera de las 23 personas coincide con el cumpleaños de alguna de las otras personas.

Ejemplos de coincidencias

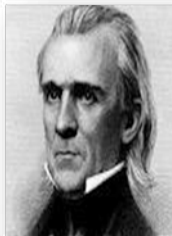
- En los jugadores del Osasuna (liga 2005/06) hay coincidencias de cumpleaños.



- De un total de sólo 19 monarcas españoles desde los reyes Católicos, coinciden Carlos II con Carlos IV (11 de noviembre) y José I con Juan Carlos I (5 de enero).



- De los 40 presidentes de USA hasta Reagan: Polk y Harding nacieron un 2 de noviembre.



Puedes creer que esto puede ser casualidad o si eres mal pensado que esta preparado, por eso nada mejor que hacer un estudio probabilístico riguroso de esta paradoja.

Estimación de la probabilidad

¿Cuál es la probabilidad de que en un grupo de  $n$  personas, por lo menos dos de ellas coincidan en su cumpleaños?, desechamos los años bisiestos y los gemelos, y asumimos que existen 365 cumpleaños que tienen la misma probabilidad.

Solución utilizando la regla de Laplace y la Combinatoria

El truco es calcular primero la probabilidad de  $n$  personas tengan cumpleaños diferentes

Sea el suceso:

$A = \{\text{"al menos dos personas celebran su cumpleaños a la vez"}\}$  y su complementario

$A^c = \{\text{"no hay dos personas que celebren su cumpleaños a la vez"}\}$

Caso particular:  $n=5$

El número de casos posibles de celebración de cumpleaños, suponiendo el año de 365 días, es:  
 $365^5 = 6,478 \times 10^{12}$

El número de casos favorables: como la primera de las personas puede haber nacido uno de los 365 días del año, la siguiente uno de los 364 días restantes y así sucesivamente, resultan  $365 \times 364 \times 363 \times 362 \times 361 = 6,303 \times 10^{12}$  casos de que no existan dos personas que hayan nacido el mismo día.

Aplicando la regla de Laplace

$P(A^c) = \text{casos favorables/casos posibles} = 6,303 / 6,478 = 0,973$

$p(A) = 1 - p(A^c) = 1 - 0,973 = 0,027$

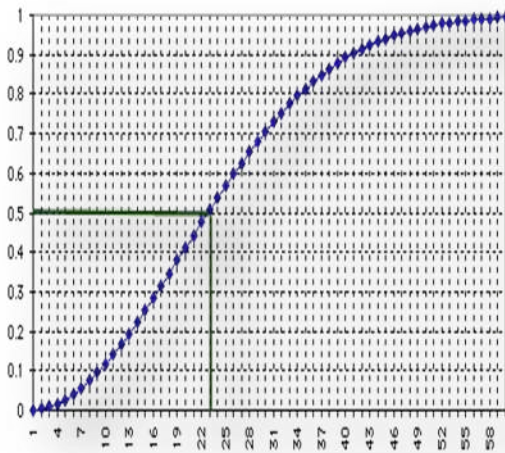
El problema puede generalizarse para una reunión de  $n$  personas. La probabilidad de que al menos dos de ellas cumplan años el mismo día es:

$$1 - \frac{365 \times 364 \times 363 \times \dots \times (365 - n + 1)}{365^n}$$

Para  $n = 23, 30$  y  $50$  la probabilidad mencionada es: 0.51, 0.71 y 0.97 respectivamente. Como se ve, para  $n = 23$  existe, aproximadamente, una probabilidad a la par que por lo menos coincidan dos cumpleaños, y cuando  $n = 50$ , tenemos casi la certeza de que ocurrirá la coincidencia

Numérica y gráficamente, las posibilidades de que distintos grupos de personas compartan cumpleaños son las siguientes:

n	prob	n	prob
5	0.027	30	0.706
10	0.117	35	0.814
15	0.253	40	0.891
18	0.347	50	0.970
20	0.411	60	0.995
23	0.507	70	0.999
25	0.569	80	0.999
27	0.627	90	0.999



Solución utilizando la regla de Laplace y la regla del producto

Otra manera de plantear el problema para que la solución no parezca tan sorprendente es que si se juntan 23 personas en una habitación, calculamos las opciones de que no compartan cumpleaños.

- Supongamos que está solo en una habitación. Las opciones de que todos los de la habitación tengan un cumpleaños diferente son obviamente del 100% o, en el lenguaje de probabilidad, 1.
- Ahora entra otra persona. Las probabilidades de que tenga un cumpleaños distinto del nuestro son  $364/365$  (Vamos a ignorar los años bisieptos en este cálculo), o 0,9973, que es lo mismo que el 99,73%.
- Entra un tercero. La probabilidad de que esa persona tenga un cumpleaños distinto del de usted y del segundo es  $363/365$ . La probabilidad de que los tres tengan cumpleaños diferentes es  $364/365$  veces  $363/365$ , o 0,9918.
- Así que las opciones de que 23 personas tengan distintos cumpleaños son  $364/365 * 363/365 * 362/365 * 361/365 * \dots * 343/365$ , que da como resultado 0,493. Esto significa que hay un 49,3% de posibilidades de que todos los de la habitación tengan cumpleaños diferentes y, a la inversa, un 50,7% de posibilidades de que al menos dos

$$1 - \left(\frac{364}{365}\right)^n$$

compartan cumpleaños.

En contraste, la probabilidad que cualquiera en una habitación de n personas tengan el mismo día de cumpleaños que usted está dada por

que para  $n = 23$  sólo da alrededor de 0,061, y se necesitaría al menos una n de 253 para dar un valor de 0,5.

*Para empezar un gran proyecto, hace falta valentía. Para terminar un gran proyecto, hace falta perseverancia.*

## Paradoja de la Placa de Automóvil



Esta paradoja reviste múltiples formas. Con la siguiente puedes ganar varias apuestas: Esperar si se repiten las dos últimas cifras de la matrícula en quince automóviles anotados al azar. La probabilidad es ahora de 0,67: ganaré dos de cada tres veces, pero ganaría, como antes, en cinco de cada seis tomando diecinueve matrículas en vez de quince.

La explicación es análoga a la anterior.

El procedimiento más cómodo para calcular la probabilidad de repetición es considerar la contraria. La probabilidad de no repetición entre dos matrículas, según el teorema de las probabilidades compuestas, será:

$$p_2 = 99/100$$

Ya que la segunda matrícula puede “optar” entre 99 casos favorables sobre 100 para no repetirse. En el caso de tres matrículas:

$$p_3 = \frac{99}{100} \cdot \frac{98}{100} = 0,9703$$

Para las 15 matrículas :

$$p_{15} = \frac{99}{100} \cdot \frac{98}{100} \cdot \frac{97}{100} \cdot \frac{96}{100} \dots \dots \frac{86}{100} = 0.3313$$

Por lo que la probabilidad de que haya al menos una repetición:

$$q_{15} = 1 - 0,3313 = 0,6687$$

Para un número mayor de repeticiones, la probabilidad aumenta rápidamente. Así:

$$q_{20} = 0,87 ; \quad q_{25} = 0,96 ; \quad q_{30} = 0,99 ;$$

Es decir, que un número de matrículas igual al cuarto de las posibles garantiza prácticamente la repetición.

Curiosidad: al estudiar estas dos versiones de esta paradoja hemos podido comprobar que: Cuanto mayor es el número de casos posibles considerados, menor es en términos relativos el número de elementos que hay que comparar entre sí para garantizar repeticiones

- El número de matrículas que habría que ir cotejando para que la probabilidad de alguna repetición fuera al menos 0,50 (¡el punto donde la apuesta sería equitativa!), sería 13, lo que supone un 13 % del número de matrículas posibles
- Sin embargo, para el caso del cumpleaños, ya vimos que esta probabilidad se alcanzaba para 23 personas, lo que, respecto a las 365 fechas posibles.

## Infalible Probabilidad Bayesiana en las Series de Aventuras de la Televisión

Gracias a la gran magia de Hollywood, ahora en la televisión abundan nuevos héroes que usan la Probabilidad Bayesiana de forma infalible, cual si fueron unos dioses que todo lo saben y todo lo ven, llegando mucho más allá de los resultados

que conseguía Sherlock Holmes en la Londres victoriana, como se puede ver actualmente en series de la televisión como X Files, CSI, Bones, Dr. House, John Doe, The Mentalist, etc.



Los nuevos héroes que basan su poder en la Probabilidad Bayesiana son detectives, policías, agentes del FBI, investigadores forenses, médicos, antropólogos, psíquicos, adivinos, etc.



Algunos de ellos son muy excéntricos, otros muy humildes, otros tienen una personalidad intolerable, pero en el fondo la gran mayoría son petulantes debido a que se saben poseedores de unas facultades que no tienen los demás pobres mortales.

Así, algunos de ellos son unos «sabelotodo» con una gran memoria, como lo era Sherlock Holmes. Otros casi tienen el donde la «omnisciencia», porque son apoyados por equipos de especialistas que son expertos en todos los campos del saber: la química, la biología, la medicina, la balística, la psicología, la computación, la ingeniería, la estadística, la geología, etc. Por consiguiente, no sólo saben formular en cada caso la hipótesis bayesiana correcta (H), sino que además son semejantes al ojo en el cielo del Gran Hermano que todo lo ve y lo oye, ya que también disponen de toda la última tecnología para descubrir hasta los datos (D) más imperceptibles que pasan desapercibidos para el común de las gentes, y por eso ellos con luces mágicas encuentran rastros de sangre que antes no se veían a simple vista, identifican el código genético del malhechor a partir de una simple partícula de caspa que dejó caer en el lugar del crimen, reconocen al sospechoso a partir del

tratamiento digital de las imágenes que fueron grabadas por una de las tantas videocámaras de la ciudad que permanentemente vigilan a toda la población, descubren al mentiroso por sus tic nerviosos o por su lenguaje no verbal ante las cámaras o por su agitación emocional ante el polígrafo, reconstruyen la forma cómo ocurrieron los hechos mediante hologramas tridimensionales, etc. Y si todo esto no es suficiente o no funciona, entonces acuden a las facultades paranormales de mentalistas, adivinos y psíquicos que descubren lo que parece insondable.

Por supuesto, estos nuevos héroes, gracias a la mágica fantasía de Hollywood, también a menudo hacen trampas para dejar con la boca abierta a los televidentes, burlándose así de su capacidad

de análisis. Por ejemplo, en algún capítulo del Dr. House la Probabilidad Bayesiana indica que el síntoma X en un 92% de los casos puede ser causado por la enfermedad W y sólo en un 8% de los casos puede ser causado por la enfermedad Z, y sin embargo, contra toda la lógica matemática y contra todos los fundamentos del Teorema de Bayes, el héroe de la trama de forma arrogante concluye que la real enfermedad es Z, y sorprendentemente acierta en su diagnóstico llegando a un desenlace feliz del episodio, basado en ese miserable 8% de probabilidad.

Es conveniente aprender a distinguir los alcances verdaderos de la Probabilidad Bayesiana frente a los alcances fantásticos, ficticios o exagerados que abundan en las actuales series televisivas.



*Si quieres triunfar, no te quedes mirando la escalera. Empieza a subir, escalón por escalón, hasta que llegues arriba.*



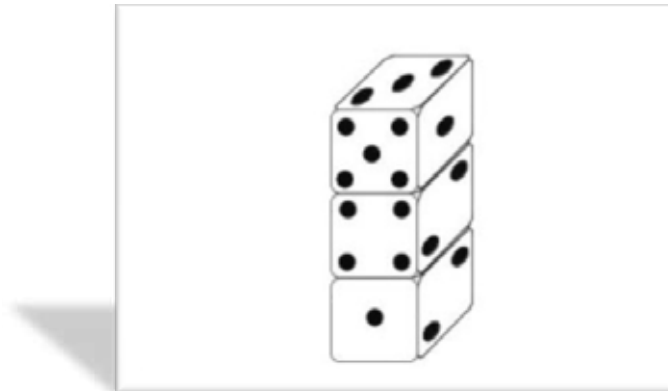
## Mezclando los naipes 7 veces



En una partida de naipes es frecuente que el jugador que ha tenido una mala mano acuse a quién barajó de no haber mezclado bien las cartas. También podemos observar que quien pierde más tiempo barajando no es otro que el que está teniendo peor suerte en la partida e intenta que ésta cambie mezclando a conciencia las cartas.

En 1991 los matemáticos estadounidenses Persi Diaconis y David Bayer recurrieron a la computadora para estudiar este problema y comprobaron que basta mezclar las cartas 7 veces para que su distribución sea aleatoria dentro de una baraja de 52 naipes. Esto quiere decir que cualquier carta tiene la misma probabilidad de encontrarse en cualquier posición. Mezclar las cartas más de 7 veces es innecesario y menos de 7 insuficiente.

## Sumando las caras ocultas de los dados



Este es un pequeño juego o truco con el que puedes demostrar a tus amigos que eres capaz de sumar las caras ocultas de una torre de tres dados o más. Tendrás que pedirle a uno de los presentes que junte los dados sin que tú le veas y que te avise cuando acabe.

Habrás que restarle a 21 (El 21 sale de que las caras opuestas de un dado siempre suman 7, y puesto que tienes 3 dados, multiplicas 7 por 3), el número que marque el dado de la cima de la torre y esa será la suma de las caras ocultas (en el caso de la figura 18). Puedes pedir que te lo pongan más difícil juntando 4 dados, y esta vez para acertar la suma tendrás que restarle a 28 la cara superior.

## La Martingala



La Martingala es un método para apostar en juegos de azar que nació en Francia en el siglo XVIII. La primera aplicación del método fue diseñada para jugar al cara o cruz. El método consiste en multiplicar sucesivamente la apuesta inicial en caso de pérdida hasta ganar una vez (gana el doble de lo apostado). En el momento en el que se gana se obtiene un beneficio igual a la apuesta inicial. Entonces, se vuelve a hacer de nuevo la apuesta inicial.

En el juego de la ruleta, la martingala consiste en apostar una cantidad, diez bolivianos por ejemplo, a un color, en este caso al rojo. Si se pierde, se duplica la última apuesta: veinte bolivianos al rojo. En caso de volver a perder, se vuelve a duplicar la última apuesta: cuarenta bolivianos al rojo, y así sucesivamente.... En el momento en el que se gane una vez, se logra el beneficio inicial de la primera apuesta, es decir, diez bolivianos.

Apostar 10 Bs. al rojo -> Sale negro, se pierde y se duplica la apuesta.

Apostar 20 Bs. al rojo -> Sale negro, se pierde y se duplica la apuesta.

Apostar 40 Bs. al rojo -> Sale rojo, ganó!! recibe 80 Bs.

Apostó en total 70 Bs. Y recibió 80 Bs. Va en progresión aritmética.

Este método está muy extendido y no son pocos los que creen que con él pueden derrotar a la banca. A primera vista es engañoso y por ello es utilizado por muchos casinos para incitar a jugar a incautos. Sin embargo, existe un tope de apuestas que llegado a él, habría que detener el método y asumir las pérdidas. No se puede duplicar la apuesta aunque se disponga de dinero.

*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*

## Estadísticas del 11 – S a 10 años de los atentados

Los atentados terroristas suicidas cometidos el 11 de septiembre de 2001 (11 – S) en los Estados Unidos por miembros de la red Yihadista Al-Qaeda mediante el secuestro de aviones de línea para ser impactados contra varios objetivos, trae consigo la destrucción de uno de los complejos comerciales más grandes del mundo, el World Trade Center en Nueva York y graves daños al Pentágono del Estado de Virginia, siendo el episodio que procedería a la guerra de Afganistán y a la adopción por el gobierno estadounidense y aliados de la política denominada Guerra contra el Terrorismo.

La presente nota, tiene el objetivo de presentar estadísticas de los hechos importantes acaecidos en esa jornada dramática, después de 10 años.



Un día normal. Vista de las Torres Gemelas, antes del 11 de septiembre de 2001

### ❖ Personas afectadas, muertas y desaparecidas

Circunstancia	Total
Número de personas fallecidas dentro del Complejo World Trade Center	2.602
Número de personas desaparecidas dentro del Complejo World Trade Center	24
Tripulación del Vuelo 11 de American Airlines*	93
Tripulación del Vuelo 175 de United Airlines*	64
Tripulación del Vuelo 77 de American Airlines*	64
Tripulación del Vuelo 93 de United Airlines**	44
Número de personas fallecidas dentro del Pentágono	125
<b>Total</b>	<b>3.016</b>

\* Incluye 5 tripulantes secuestradores

\*\* Incluye 4 tripulantes secuestradores

Según la comisión de investigación del 11-S, aproximadamente 16.000 personas se encontraban en las zonas de impacto del complejo del World Trade Center en el momento de los ataques. La gran mayoría de ellos sobrevivió, gracias a las labores de evacuación antes del derrumbe de las Torres.

Los atentados causaron más de 6.000 heridos, la muerte de aproximadamente 3.000 personas y la desaparición de otras, resultando muertos igualmente 19 terroristas. Fallecieron 246 personas en los cuatro aparatos estrellados (ninguno de los pasajeros de los aviones secuestrados sobrevivió – no se incluye en la cifra los 19 terroristas). Entre las víctimas del edificio del Pentágono, se contaron a 343 bomberos del departamento de bomberos de Nueva York, 23 policías del departamento de policía y 37 policías de la autoridad portuaria de Nueva York y Nueva Jersey. El cuadro anterior, resume las estadísticas de las personas fallecidas y desaparecidas en diferentes circunstancias.

De las 3.016 personas, tan sólo 1.379 fueron muertes confirmadas gracias a las pruebas de ADN. 1.350 personas restantes han sido reconocidas como muertas por las autoridades, después de que sus familiares reclamasen sus certificados de defunción. La suma de ambas cifras es 2.729, muy por debajo de los 3.016. Los 287 restantes fueron un grupo impreciso de desaparecidos, cuyo paradero en el momento de los atentados se desconocía. Según el diario The New York Times, informaba que el departamento de policía que registraba a las víctimas en el apartado de homicidios, el número de muertes podía ser menor, de aproximadamente 2.800.



Las Torres Gemelas, el día del atentado

Un avance del estudio del Instituto Nacional de Normas y Tecnología, difundido en julio de 2003, señaló que la mayoría de las muertes se produjo por encima de los pisos donde impactó cada uno de los aviones. En la Torre Norte murieron 1.356 personas a la altura o por encima del piso 93 (considerando que las Torres tenían 110 pisos), atrapados porque las tres escaleras disponibles quedaron destruidas, mientras que 618 personas murieron en la Torre Sur a la altura o por encima del piso 77, aunque la escalera de emergencia quedó intacta, no pudieron acceder a ella por la nube de humo existente. Sólo 6 personas fallecieron por debajo de la zona afectada. Muchos investigadores indican que al menos 200 es el número de personas que optaron por saltar al vacío.

Entre los desaparecidos se incluyeron a 17 personas que formaban parte del cuerpo de bomberos y de compañías con sede en las Torres. También figuraron algunas que no habían fallecido, tratándose, en la mayoría de los casos de individuos sin lazos familiares que solían merodear diariamente por el lugar del siniestro. Varios de ellos aparecieron en centros psiquiátricos.

Por ejemplo, George Sims comenzó a despertar de su amnesia en un hospital de Nueva York, donde era tratado por esquizofrenia. Su madre temía que hubiese fallecido bajo las Torres. Un vagabundo llamado Albert Vaughan, un habitual del metro del WTC (World Trade Center), fue borrado de la

lista de desaparecidos una vez que su familia fue notificada que Albert seguía con vida. “No sabía que todo el mundo pensaba que estaba muerto”, manifestó Vaughan, cuya memoria no le alcanzó a recordar dónde se encontraba el 11 de septiembre de 2001. Una mujer barcelonesa llamada Alicia Esteve, se hizo pasar por sobreviviente del atentado, adoptó una identidad falsa (Tania Head); e incluso llegó a ser presidenta de la Red de Sobrevivientes de la catástrofe del World Trade Center. Gracias a The New York Times se descubrió el fraude; y se reveló su verdadera identidad. Se han dado varios casos como los nombrados, manifestó Ellen Borakove, portavoz de los servicios forenses.

La mayor dificultad estribó en la veintena de desaparecidos de origen extranjero, en su mayoría trabajadores ilegales, con números de identificación falsos, y hasta el nombre, en algunos casos. El registro de víctimas extranjeras fue lo más difícil de todas. Los denunciantes de la desaparición se encontraban fuera del país y la policía no lograba dar con ellos. Se cree que 500 extranjeros perecieron, en su mayoría residentes en el área de Nueva York.

La identificación de mujeres también fue complicada, ya que varias de ellas habían sido registradas por familiares con sus apellidos de soltera y sus nombres estaban duplicados. También se dieron 25 casos de personas declaradas muertas por algún pariente con fines de cobro de seguros, indemnización o simplemente evasión de deudas. Familias que, desesperadas por encontrar a un ser querido, lo incluyeron entre los desaparecidos.



Impacto a la Torre Norte



El segundo avión a punto de estrellarse en la Torre Sur

Según las cifras presentadas por el Departamento de Salud en enero de 2002, 247 latinos estuvieron entre los muertos del atentado terrorista de Al-Qaeda contra las Torres Gemelas, representando un 8.1% del total. De estos, 25 eran oriundos de la República Dominicana, 18 de Colombia, 13 de Ecuador, 6 de Cuba, 4 de Argentina y 1 de Chile. En otros sitios, se habla de 15 muertos de México, así como otros de El Salvador, Honduras, Jamaica, Perú, Venezuela, Paraguay y Guyana. Se sabe de una persona de nacionalidad boliviana que trabajaba en el World Trade Center, y justamente llegó tarde ese día a su trabajo, sorprendiéndose del suceso en puertas de los rascacielos, puesto que el primer impacto se dio a las 8:46 de la mañana.

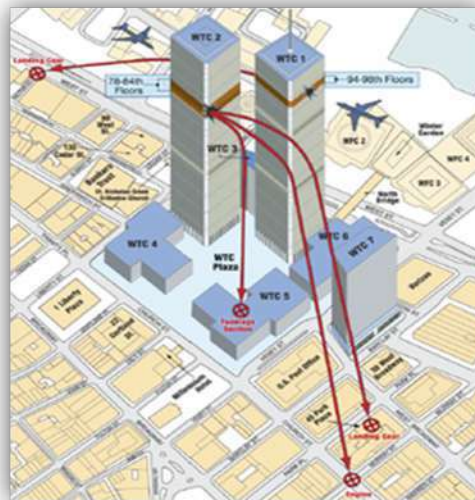
Los atentados supusieron el ataque terrorista de mayor importancia contra los Estados Unidos de América, superando al atentado de Oklahoma City cometido por los terroristas de ultraderecha Timothy McVeigh y Terry Nichols que causó 168 muertos y los ataques llevados a cabo por células de Al-Qaeda en 1998 contra embajadas de EEUU en los países africanos de Kenia y Tanzania.

## ❖ Destino de los aviones secuestrados

Los atentados fueron cometidos por 19 miembros de la red de Al-Qaeda, divididos en 4 grupos de secuestradores, cada uno de ellos con un terrorista

piloto que se encargaría de pilotar el avión una vez ya reducida la tripulación de la cabina. Los aviones de los vuelos 11 de American Airlines y 175 de United Airlines fueron los primeros en ser secuestrados siendo ambos estrellados contra las dos Torres Gemelas del World Trade Center, el primero contra la Torre Norte, el segundo poco después contra la Sur, provocando que ambos rascacielos se derrumbaran en las dos horas siguientes.

El tercer avión secuestrado pertenecía al vuelo 77 de American Airlines y fue empleado para ser impactado contra una de las fachadas del Pentágono en el Condado de Arlington cerca de Washington DC, en Virginia. El cuarto avión, perteneciente al vuelo 93 de United Airlines, no alcanzó ningún objetivo al resultar estrellado en campo abierto, cerca de Shanksville, en Pensilvania, tras perder el control en cabina como consecuencia del enfrentamiento entre los pasajeros y tripulantes con el comando terrorista.



Los 4 aviones de pasajeros secuestrados tenían como destino el estado de California, desde el Aeropuerto Internacional de Boston, el Aeropuerto Internacional Washington Dulles y el Aeropuerto Internacional Libertad de Newark, por lo que sus depósitos de combustible iban llenos con unos 91.000 litros y unos 65.455 kg de peso.

Según revelaron los testimonios desde los propios aviones, los secuestradores habían tomado el control de éstos usando simples navajas con las que mataron a azafatas de vuelo y al menos a un piloto o pasajero. Según las investigaciones de la Comisión de Investigación del 11-S, se tiene también constancia de que fue usado algún *spray* para retener a los pasajeros en la cabina de primera clase. Asimismo se amenazó con la presencia de una bomba en 3 de los aviones; no fue así en el American Airlines 77. Según las conclusiones de la comisión, se piensa que los avisos de bomba eran probablemente falsos.

El gráfico muestra los impactos en las Torres Gemelas de ambos aviones, y la trayectoria (y punto de impacto en la calle) que siguieron los trenes de aterrizaje, y uno de los motores.

En el cuarto avión, la caja negra reveló que los pasajeros, después de enterarse de que el resto de aviones habían sido estrellados deliberadamente, trataron de retomar el control de los aparatos, a lo que los secuestradores reaccionaron moviendo el avión en un fallido intento para someter a los pasajeros. Poco después, el avión se estrelló a las 10:03, 11 a.m. hora local. Posteriormente el líder de Al-Qaeda capturado Khalid Shaikh Mohammed dijo que el vuelo 93 tenía como objetivo el Congreso de los Estados Unidos.

El sector aéreo fue uno de los más afectados, el espacio aéreo estadounidense permaneció cerrado durante varios días por primera vez en su historia. Se estima que el negocio perdió un 20% de su total. La demanda de vuelos comerciales se redujo drásticamente, debido sobre todo al temor de que se repitieran las acciones terroristas, y la resistencia del público a someterse a medidas rigurosas de seguridad en los aeropuertos. En un intento por aliviar esta situación, el Congreso aprobó un paquete financiero de 15.000 millones de dólares para el sector, en tanto que el Gobierno de Bush adelantó un recorte adicional de los impuestos para revitalizar el consumo; esta medida tuvo efectos negativos en el presupuesto, ya de por sí mermado por los gastos de la guerra.

## ❖ Cronología de los Atentados

El horario está establecido según la hora local de verano en el Este de los Estados Unidos.

- 08.00: el vuelo 11 de American Airlines, un Boeing 767 con 93 personas a bordo, despegó del Aeropuerto Internacional Logan de Boston hacia Los Ángeles.
- 08.14: el vuelo 175 de United Airlines, un Boeing 767 con 64 personas a bordo, despegó del Aeropuerto Internacional Logan de Boston hacia Los Ángeles.
- 08.19: La azafata Betty Ong del vuelo 11 de American Airlines se comunica con la oficina de reservaciones de la compañía para informar que el avión parece estar siendo secuestrado.
- 08.21: el vuelo 77 de American Airlines, un Boeing 757 con 64 personas a bordo, despegó del Aeropuerto Internacional Dulles de Washington DC hacia Los Ángeles.
- 08.37: la FAA (La Administración Federal Aeronáutica) notifica al NORAD (*North American Aerospace Defense Command*, Mando de Defensa Aérea de América del Norte) que existe una sospecha de secuestro del vuelo 11 de American.
- 08.41: el vuelo 93 de United Airlines, un Boeing 757 con 44 personas a bordo, despegó del Aeropuerto Internacional de Newark hacia San Francisco.
- 08.47: el vuelo 11 de American se incrusta por completo en la Torre Norte del World Trade Center (WTC).
- 09.03: el vuelo 175 de United embiste la Torre Sur del World Trade Center. El suceso es cubierto en directo por diversas cámaras de televisión que enfocaban a las Torres Gemelas a causa de la densa humareda que surgía de la Torre Norte.





Explosión del Boeing 767 contra la Torre Sur

- 09.03: la FAA notifica al NORAD que existe una sospecha de secuestro del vuelo 175 de United.
- 09.07: el Jefe de Empleados de la Casa Blanca avisa al presidente George W. Bush que embistieron la segunda Torre y que Estados Unidos se encuentra bajo ataque. El presidente se encontraba reunido con unos niños de una escuela primaria de Sarasota (Florida).
- 09.08: la FAA prohíbe todos los despegues que se dirijan hacia o pasen sobre el espacio de Nueva York.
- 09.26: la FAA prohíbe el despegue de todos los aviones civiles.
- 09.31: el presidente George W. Bush pronuncia un discurso en la escuela primaria en la que se encuentra, informando que se trata aparentemente de un acto terrorista.
- 09.34: la FAA notifica al NORAD que existe una sospecha de secuestro del vuelo 77 de American.
- 09.38: el vuelo 77 de American se estrella contra el Pentágono.
- 09.45: la FAA, (con órdenes del Presidente) ordena a todos los aviones en vuelo aterrizar inmediatamente en el aeropuerto más cercano.
- 09.48: el Congreso y la Casa Blanca son evacuados.
- 09.57: el presidente Bush abandona Florida.
- 09.59: se derrumba la Torre Sur.
- 10.03: el vuelo 93 de United cae en un campo abierto en Shanksville, Pensilvania.
- 10.06: la FAA notifica al NORAD que existe una sospecha de secuestro del vuelo 93 de United.
- 10.28: cae la Torre Norte.
- 11.04: el edificio de las Naciones Unidas en Nueva York es evacuado.
- 11.45: el presidente Bush aterriza en la base aérea Barksdale en Luisiana.
- 13.04: el presidente Bush declara “alerta máxima” en todo el mundo y se dirige a la nación a través de los medios de comunicación desde la base aérea Barksdale.
- 13.37: el Presidente Bush deja en Air Force One, la base Barksdale.
- 14.51: la Marina envía destructores misilísticos a Nueva York.
- 15.07: el Presidente Bush aterriza en la base aérea Offutt en Nebraska.
- 16.36: el presidente Bush deja en Air Force One, la base Offutt y ordena ser llevado a la Casa Blanca.
- 17.25: se derrumba el edificio 7 de 47 pisos del World Trade Center.
- 19.00: el Presidente Bush llega a la Casa Blanca.

- 20.30: el presidente Bush habla a la nación desde el Despacho Oval, en la Casa Blanca.
- ❖ Daños al Complejo Comercial del World Trade Center y al Pentágono

La Torre Sur cayó a las 9:59 (hora local en Nueva York), tras estar en llamas durante 56 minutos en un fuego causado por el impacto del vuelo 175 de United Airlines a las 9:03. La Torre Norte cayó a las 10:28, tras estar en llamas aproximadamente 102 minutos en un fuego causado por el impacto del vuelo 11 de American Airlines a las 8:46. Un tercer edificio, el World Trade Center 7, se derrumbó a las 17:20, al parecer tras haber sido seriamente dañado por los escombros de las Torres Gemelas al caer, junto con una serie de incendios. El hotel Marriott, 4 estaciones del metro de Nueva York y la iglesia cristiana ortodoxa de San Nicolás afectados entre otros.



Una de las Torres a punto de colapsar

En total, en Manhattan, 25 edificios sufrieron daños y 7 edificios del complejo de negocios del World Trade Center fueron arrasados. Más tarde, el *Deutsche Bank Building* situado en la calle *Liberty street* y *Borough of Manhattan Community College's Fiterman Hall* en el *30 de West Broadway* tuvieron que ser demolidos debido al estado en que quedaron, que los hacía inhabitables. Varios equipos de comunicaciones también sufrieron daños. Sin ir más lejos, las antenas de telecomunicaciones de la Torre Norte cayeron con su derrumbe, mientras que otras antenas de radio de Torres colindantes resultaron también gravemente dañadas.

Los incendios en las Torres Gemelas, debilitaron las cerchas que sostenían los pisos, e hicieron que los pisos se combaran. A su vez, los pisos al combarse, tiraron de las columnas de acero exteriores hasta el punto que las columnas exteriores se inclinaron hacia el interior. Con los daños a las columnas principales, las columnas exteriores torcidas no pudieron soportar el peso de los edificios, produciéndose el derrumbe.



Derrumbe del Complejo World Trade Center 1





Derrumbe del Complejo World Trade Center 2

En el condado de Arlington, una porción del Pentágono fue severamente dañada por el fuego y el impacto del avión. Al cabo de un rato, una sección entera del edificio se derrumbó.

## ❖ Detalle de los Secuestradores

19 hombres árabes embarcaron en los 4 aviones, 5 en cada uno, excepto el Vuelo 93 de United Airlines, que tuvo 4 secuestradores. De los atacantes, 15 eran de Arabia Saudita, 2 de los Emiratos Árabes Unidos, 1 de Egipto, y 1 del Líbano. En general, era gente con estudios y de familias acomodadas, fallecidos en apenas 72 horas. La lista completa es:

En el Vuelo 11 de American Airlines:

- Mohammed Atta (egipcio y supuesto piloto)
- Waleed al-Shehri (saudí)
- Wail al-Shehri (saudí)
- Abdulaziz al-Omari (saudí)
- Satam al-Suqami (saudí)

En el Vuelo 175 de United Airlines:

- Marwan al-Shehhi (emiratí y supuesto piloto)
- Fayez Banihammad (emiratí)
- Mohand al-Shehri (saudí)
- Hamza al-Ghamdi (saudí)
- Ahmed al-Ghamdi (saudí)

En el Vuelo 77 de American Airlines:

- Hani Hanjour (saudí y supuesto piloto)
- Khalid al-Mihdhar (saudí)
- Majed Moqed (saudí)
- Nawaf al-Hazmi (saudí)
- Salem al-Hazmi (saudí)

En el Vuelo 93 de United Airlines:

- Ziad Jarrah (libanés)
- Ahmed al-Haznawi (saudí)
- Ahmed al-Nami (saudí)
- Saeed al-Ghamdi (saudí)

27 miembros de Al-Qaeda trataron de ingresar en los Estados Unidos para tomar parte en el atentado. Sólo 19 participaron, los otros 8 son llamados “el vigésimo secuestrador” de los cuales, fueron denegadas las visas y algunos detenidos posteriormente.

Pocos habían tratado de ocultar sus nombres o identificación, eran casi los únicos pasajeros de origen árabe en los vuelos. Así, el FBI pudo determinar sus nombres y en muchos casos detalles, como la fecha de nacimiento, las residencias, el estado del visado, y la identidad específica de los sospechosos pilotos.



Las pesquisas del gobierno de los Estados Unidos, incluyeron la operación del FBI, la mayor de la historia con más de 7.000 agentes involucrados. Los resultados de ésta, determinaron que al-Qaeda y Osama Bin Laden tenían la responsabilidad de los atentados. A idéntica conclusión llegaron los estudios encargados por el gobierno británico. La declaración de una Guerra Santa contra los Estados Unidos, y otros pronunciamientos de Bin Laden llamando a matar a civiles estadounidenses en 1998, son consideradas por muchos como evidencia de su motivación para cometer estos actos. El 11 de septiembre de 2001, Bin Laden emite un comunicado en el que dice: *“Califico de héroes a los pilotos de los aviones”*.

Finalmente, los conspiradores del 11-S erogaron un gasto entre 400.000 y 500.000 dólares americanos para planificar y conducir su ataque. Los orígenes específicos del dinero utilizado permanece desconocido

## ❖ Guerra contra el Terrorismo

El primer paso dado por EEUU en la Guerra contra el Terrorismo fue la invasión de Afganistán el 7 de octubre de 2001 por fuerzas de la OTAN y la Alianza del Norte con apoyo de las Naciones Unidas, ante la negativa del gobernante régimen talibán de entregar a Osama Bin Laden, que supuestamente se había refugiado en ese país.



Un grupo de soldados estadounidenses se dispone a entrar en uno de los palacios de Saddam Hussein en Bagdad el 9 de abril de 2003.

El 13 de noviembre de 2001, la capital Kabul fue tomada por la Alianza del Norte y el gobierno quedó en manos de EEUU/OTAN y la Alianza del Norte. Desde entonces Al-Qaeda y los talibán se unieron y se reorganizaron como guerrilla insurgente. El 2 de mayo de 2011, Bin Laden fue abatido por tropas de élite estadounidenses en Pakistán.



Transporte de un helicóptero americano a soldados que asisten a la guerra contra Irak.

El segundo paso de la Guerra contra el Terrorismo de EEUU fue la invasión de Irak el 20 de marzo de 2003. Esta acción militar fue realizada por Estados Unidos y Gran Bretaña sin autorización de las Naciones Unidas. Además España, Italia y

otros países, se aliaron con EEUU en esta acción y enviaron ayuda humanitaria a la zona. Estados Unidos sostuvo que la invasión era indispensable debido a que Iraq poseía armas de destrucción masiva ocultas. La invasión desencadenó una guerra, con cientos de muertos, y causó el derrocamiento del gobierno encabezado por Saddam Hussein el 9 de abril de 2003. Estados Unidos sostuvo entonces que la razón de la invasión se debía a que existían informaciones de los servicios de inteligencia que permitían suponer que Saddam Hussein mantenía relaciones secretas con Al-Qaeda. Recientes informes indican que nunca hubo una relación de Hussein con Al-Qaeda, y el presidente Bush trató de relacionar a Iraq con la Guerra contra el Terrorismo.

Desde entonces varios grupos iraquíes opositores a la invasión han organizado un movimiento de resistencia que se ha mostrado muy activo en la realización de ataques contra objetivos militares.

Alrededor de 1.200 extranjeros han sido arrestados

y encarcelados en secreto en relación con la investigación de los ataques del 11 de septiembre, aunque el gobierno no ha divulgado el número exacto. Los métodos utilizados por el Estado para investigar y detener sospechosos han sido severamente criticados por organizaciones de derechos humanos como Human Rights Watch y jefes de gobierno como la canciller alemana Angela Merkel.

Al día de hoy, las consecuencias continúan al haberse detonado una guerra civil sectaria “no declarada”, que tiene como consecuencia la muerte de más de 34.000 civiles (solamente en el 2006, según la ONU) y según cifras de Acnur, hay 1,7 millones de iraquíes desplazados internamente y otros 2 millones que han huido a países vecinos. Además, a junio del 2007 las bajas del ejército de los Estados Unidos ascienden a más de 4.000 caídos, superando por mucho las muertes producto de los atentados del 11 de septiembre del 2001.

*Los hombres y pueblos en decadencia viven acordándose de dónde vienen; los hombres geniales y pueblos fuertes sólo necesitan saber a dónde van*

## *El futuro Censo de Población y Vivienda en nuestro país*

La realización del futuro Censo de Población y Vivienda en nuestro país, está previsto, según el gobierno, para el año 2012, cuando han transcurrido ya 10 años del último censo.

Una de las preocupaciones, es el tiempo que queda para el día del empadronamiento programado para el año siguiente, que según cálculos y el trabajo que se debe realizar, no será posible. Otro aspecto a tomar en cuenta, es la manera cómo percibe la población boliviana del censo, debido a algunas decisiones erradas que se tomaron el año 2001. En este documento de opinión, se comentan estos aspectos.



El censo programado, requiere de un proceso de planificación adecuado en el que se inicia con la actualización cartográfica (actividad más importante antes del censo), donde se incluye la incorporación y actualización de áreas geográficas amanzanadas de las ciudades y centros poblados en crecimiento, en relación al Censo 2001. Además, se necesita efectuar un recuento de viviendas y población por ciudad, centro poblado y comunidad, lo más cercano posible, de tal forma que permita la delimitación y división de sectores y segmentos censales de acuerdo al número de viviendas que se pueda empadronar en un día. Los datos obtenidos de la actualización, también sirven para la preparación de la logística del trabajo de campo, y otros aspectos técnicos propios del censo. Por lo visto, todavía no se ha empezado esta importante tarea que lleva por lo menos un año de trabajo, considerando la actualización de áreas amanzanadas y dispersas del país.

Lo demás, son actividades paralelas, como: definir y consolidar la boleta censal, concientizar a la población sobre la realización del censo, planificar la prueba piloto, etc., tareas que pueden desarrollarse en un periodo de tiempo prudente.

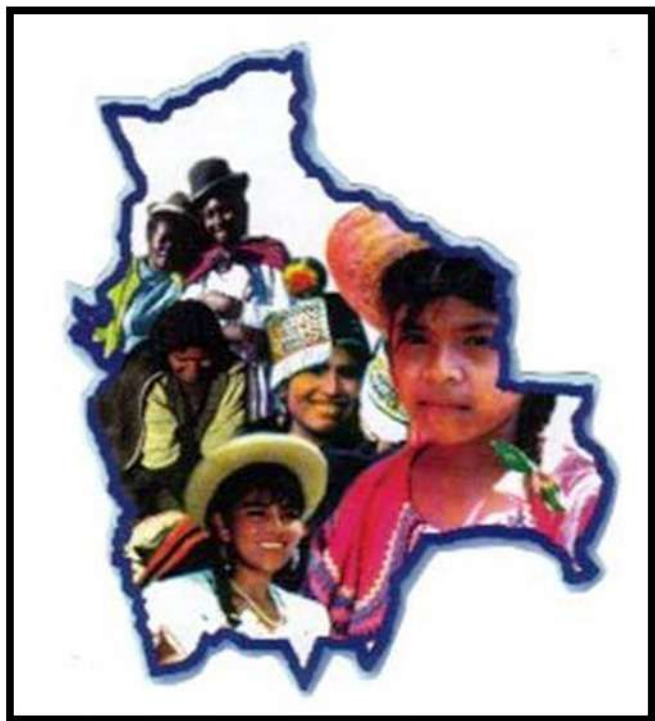
Una deficiente actualización, acortando el tiempo previsto de trabajo, trae como consecuencia una desfavorable delimitación y división de sectores y segmentos censales, una inadecuada distribución de empadronadores y equipos de trabajo, una mala asignación de boletas censales por segmento, etc. Puede ser, que el día del censo, sobren o falten empadronadores y boletas censales por sector y segmento censal por motivo de una incompleta actualización.



No es el único problema, otro y de fondo, es que las demandas sociales piden la realización de un censo para que, de acuerdo a los resultados de recuento de población censal, se destinen mayores recursos económicos a su región, concretamente a su municipio. Uno se pregunta, ¿de dónde viene esta demanda social actual para el futuro censo?; la respuesta es evidente, se trata de una política desacertada de los asesores del Censo 2001, afirmando que:

*“Los pobladores que sean censados en sus lugares de origen, y que según el recuento de población, recibirán recursos económicos de la participación popular, que vayan en beneficio de sus comunidades y de su municipio”.*

Esta determinación hizo que mucha gente se dirigiera a sus lugares de origen, unos días antes del censo; fueran censados allí, y luego al día siguiente, migrar nuevamente a las ciudades u otros lugares del país, reflejando la verdadera estadía de los pobladores.



En ese sentido, los datos de población y vivienda del censo a nivel municipal, sufren cambios que no refleja la realidad boliviana.

¿En qué queda haber realizado un trabajo de actualización cartográfico, de recuento de viviendas previa al censo, cuando se aplica una política de migración poblacional momentánea como la nombrada?.

En el aspecto técnico, uno de los insumos más importantes del censo, es la construcción de un marco muestral para las futuras encuestas a hogares que el Instituto Nacional de Estadística realiza. Un marco muestral, es una lista exhaustiva de municipios, comunidades, áreas geográficas bien

delimitadas amanzanadas y dispersas acompañadas del total de viviendas, personas y otro tipo de información, de todo el territorio nacional; incluye la cartografía bien diseñada. Este listado, que viene de la información censal, sirve de insumo para el cálculo, distribución y selección de la muestra para las futuras encuestas en los diez años siguientes.

La pregunta es, ¿si este listado con información tan importante, más los mapas y planos que se cuenten, no son confiables y precisos, por el problema migratorio censal comentado; cómo se determinarán las muestras futuras a partir de un marco muestral afectado por la política migratoria nombrada?. Indudablemente que mal, reflejando cantidades de población y viviendas ocupadas fuera de la verdad a niveles como el municipio.



Al advenimiento de un nuevo censo en nuestro país, esperar que se consideren estos aspectos con la seriedad que se merece.

Fernando Rivero Suguiura

## XXIV versión de la Entrada Universitaria

La Entrada Universitaria se inicia un 5 de marzo de 1988 a la cabeza del señor Luis Sempértegui quién es el fundador de esta actividad cultural.

La primera exhibición de la Entrada Universitaria contó con la participación de nueve fraternidades que interpretaron los bailes más típicos de nuestra región. Ese año nuestra entrada sólo tuvo una duración de aproximadamente dos hora.

El año siguiente (1989), participaron más de 40 grupos, entre ellos Caporales, Antawaras, Diablada, etc.. El año 1990 fue considerado como el de la consolidación de la Entrada Folklórica Universitaria, se habilitaron libros de inscripciones, convocando a la tercera versión de la entrada. En esta ocasión el número de participantes superaba a los 70 grupos y se inició cerca al medio día, concluyendo pasada las 9 de la noche.



Actualmente la Entrada Universitaria es la mayor representación juvenil de colorido y folklore del país con la participación de más de 70 agrupaciones folklóricas, además a incitado a que distintas universidades del interior del país realicen las mismas actividades folklóricas en sus respectivas ciudades.

## Elección de la Ñusta Universitaria

Como se ha hecho habitual, este año 2011, como antesala a la XXIV Entrada Universitaria cada fraternidad eligió a su representante o Ñusta.

La gran elección de la Ñusta Universitaria, se realizó el 20 de Julio del presente año, entre 53 candidatas. El título de Ñusta Universitaria “Patrimonio de Bolivia” 2011 recayó sobre la señorita Mariana Deheza Ramos de 21 años, estudiante de la carrera de Administración de Empresas y representante de la fraternidad Tinkus Wistus. El segundo lugar le perteneció a la señorita Wara Sotomayor Marquez de 19 años, representante de la fraternidad Pujllay de Ciencias de la Educación, obteniendo el título de Ñusta “Autonomía Universitaria”. Por último, el tercer lugar fue para la señorita Mary Cruz Guzmán Vegazo de 22 años, estudiante de Tecnología Médica, en representación de la fraternidad Tinkus Huayna Lisos de Electromecánica con el título de Ñusta “Ciencia y Cultura”.



Como no podía ser de otra manera, la fraternidad Potosinos de Estadística eligió a su Ñusta en un evento que se realizó el día 14 de julio, donde la ganadora fue la señorita Jhenny Mercado quién cautivó a los presentes con sus alegres movimientos de la danza de los Potosinos. El jurado estaba compuesto por los docentes de la carrera Juan Carlos Flores L. y Nicolás Chávez Q.; y en representación de la parte estudiantil, la señorita Ana Rosa Aliendre.



*Jhenny Mercado*

Nuestra soberana, tuvo la difícil labor de representar a la fraternidad Potosinos en la gran elección de la Ñusta Universitaria, entre 53 candidatas.

## *Presentacion de Monografías*

La exposición de monografías se llevó a cabo el 21 de julio del presente año en la plaza del Bicentenario y el atrio de la UMSA. Los representantes de cada una de las fraternidades realizaron un trabajo de investigación determinando las características y vestimenta de sus propias danzas que se ejecutan año tras año en esta gran fiesta. Los organizadores informaron que esta actividad contó con el apoyo del Museo de Etnografía y Folklore (MUSEF) que se encargó de calificar las presentaciones de los universitarios.



La presentación de la monografía fue uno de los requisitos fundamentales para participar en el acontecimiento folklórico, éste año se inscribieron 85 fraternidades de distintas carreras y facultades, de las cuales sólo 73 cumplieron los requisitos.

## *Pre – Entrada Universitaria*

Más de 60 fraternidades de estudiantes de la Universidad Mayor de San Andrés efectuaron el 24 de julio la pre-entrada actividad que se inició a las 9:30 horas. La concentración fue en la plaza Triangular continuando por lo largo de la Av. Busch, y concluyendo en la Plaza Villaruel de nuestra ciudad.

## Entrada Universitaria

En fecha 30 de julio se realizó la XXIV Entrada Folklórica de la Universidad Mayor de San Andrés, con 73 grupos de fraternidades, acontecimiento denominado con el título de patrimonio cultural, vivo e intangible de Bolivia mediante la promulgación de la Ley N°155.

El recorrido de la entrada por las principales calles y avenidas del centro paceño fue desde la avenida Montes, pasando por la avenida Mariscal Santa Cruz, 16 de julio, y calle Bueno, Camacho hasta la avenida Simón Bolívar donde culminó.



*Baile de los Potos*

## Ganadores de la Entrada Folklórica

La fraternidad Morenos Achachis Trabajadores San Andrés se impuso en la Entrada Universitaria 2011, en segundo lugar se ubicó el Taller de Proyección Cultural de la carrera de Arquitectura con la “Danza de Mímulas” y en tercer lugar se colocó la fraternidad “Reyes Morenos de Administración” de la carrera de Administración de Empresas.

La fraternidad Potos de Estadística obtuvo el tercer lugar en la categoría “B” y la decima posición en la calificación general de la Entrada Folklórica de la Universidad.



## Recorrido Entrada Universitaria 2011

Como era de esperar, la fraternidad Potos de Estadística tuvo un despliegue particular empezando su participación a horas 13:30 y concluyendo a las 17:00. Contó con la participación de más de 120 estudiantes ordenados en seis bloques. Se destaca que los estudiantes que bailan en esta fraternidad, no sólo son universitarios de la carrera de Estadística sino también de otras carreras. Además se observa en estos últimos años mayor participación de personas del sexo femenino.

Se adjunta en anexo la tabla de posiciones de las fraternidades participantes en la Entrada Universitaria 2011.





## Elecciones para Director de la Carrera de Estadística

Después de un periodo largo de cuatro directores interinos, el 16 de junio del presente año, se llevó a cabo los comicios electorales para elegir al Director de la Carrera de Estadística, gestión 2011 – 2014.

Con la inscripción de dos frentes liderizados por la licenciada Nilda Gloria Flores Salinas con el frente *Nivel* y el licenciado Juan Carlos Flores López con el frente *Fuerza Estadística*, llegaron a la votación con grandes expectativas.



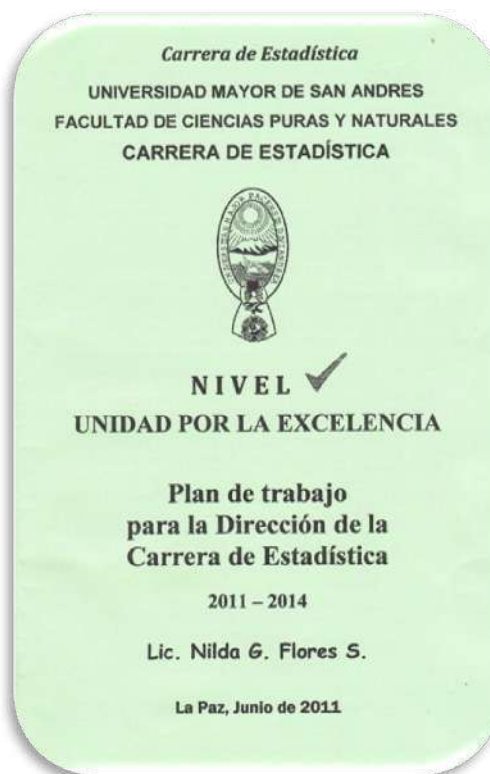
El desarrollo de las elecciones se realizó de forma normal con participación total de los docentes de la Carrera habilitados para el voto, que no alcanzan a más de 12 personas. La cobertura estudiantil votante, fue de 86%, un porcentaje importante de participación comparada con anteriores elecciones.

El ganador de la contienda eleccionaria fue el frente *Fuerza Estadística* y su representante tomó posesión del cargo el 26 de julio del año en curso.

El ganador oficial es el Lic. Juan Carlos Flores, y aún más, el gran ganador de la elección es el grupo de docentes y estudiantes de la Carrera que creen

en este tipo proceso eleccionario como alternativa democrática, y manifiestan de manera consciente, sus intereses al servicio de la profesión. Lo fundamental es que hemos dejado atrás momentos de incertidumbre con el voto en blanco y la consecuencia de la falta de autoridades en nuestra Carrera.

Desde el IETA y todos los que componemos esta institución dependiente de la Carrera de Estadística, le hacemos llegar nuestra felicitación al nuevo director y que emprenda mejores ideas para mejorar la parte académica, investigación e interacción social de nuestra querida Carrera.



*Cada fracaso supone un capítulo más en la historia de nuestra vida y una lección que nos ayuda a crecer. No te dejes desanimar por los fracasos. Aprende de ellos, y sigue adelante.*

## Homenaje a un gran maestro de la Estadística

### Luis Zapata Escobar

Estudió Ingeniería Eléctrica en la Universidad Mayor de San Andrés entre los años 1963 y 1968, obtuvo una maestría en Estadística Matemática en el Centro Interamericano de Enseñanza de Estadística (CIENES) de la Universidad de Santiago de Chile en 1972.



Durante sus primeros años, fue jefe de trabajos prácticos del Departamento de Matemática del Instituto Superior de Ciencias Básicas de la UMSA. En 1973 inicio la labor de docente en la Carrera de Estadística de la UMSA, luego fue docente de la Escuela Militar de Ingeniería “Mariscal Antonio José de Sucre” (EMI), y posteriormente en 1978 fue director del Departamento de Matemática. Los años 1982 a 1984 alcanzó a ser Jefe de la Carrera de Estadística y, a continuación llegó a ser Decano de la Facultad de Ciencias Puras y Naturales donde cumplió una labor encomiable a favor de nuestra facultad.

Fue docente de la maestría en Administración y Dirección de Empresas de la Escuela Militar de Ingeniería “Mariscal Antonio José de Sucre”, que a su vez, fue docente de la EMI en más de una década. Profesor de la maestría en Estadística Experimental en Psicología de la Universidad Católica Boliviana “San Pablo”, y también fue docente en pregrado de dicha universidad por varias gestiones. Instructor

de la maestría del Centro de Formación Docente del CEPIES de la UMSA en la asignatura de Métodos Cuantitativos en el año 2002.



Al mismo tiempo escribió varios textos que benefician a los estudiantes de la UMSA, y a la sociedad en la labor académica y laboral de nuestro país.

En diferentes situaciones, Luis Zapata trabajó como consultor y también ocupó cargos importantes en instituciones públicas y privadas de nuestro país. El año 1987 llegó a ser Jefe del Departamento de Muestreo y Encuestas del Instituto Nacional de Estadística (INE), durante un año. Posteriormente trabajó como consultor en la Encuesta Nacional de Población y Vivienda (ENPV-88).

En su labor de consultor, escribió metodologías que fueron base para las encuestas que realizó el INE en esa temporada, algunos de los mencionados, son: “Precisión de los resultados de la Encuesta Permanente de Hogares 1987”, “Documento

Metodológico del Diseño de la Muestra en la Encuesta Nacional de Población y Vivienda y sus errores de Muestreo”. De igual manera, escribió los siguientes libros: “La situación educativa en Bolivia” y “La situación del empleo en Bolivia”. En noviembre de 1992 publicó un documento denominado “El subempleo medido a través de la Encuesta Integrada de Hogares” INE, del Departamento de Estadísticas Sociales.

Otras investigaciones donde estuvo apoyando como consultor, fueron: “Encuesta de cobertura de saneamiento básico de Oruro” proyecto ejecutado por el Ministerio de Asuntos Urbanos en 1988. ”Encuesta de línea de base en los distritos de salud de Ayo Ayo, Samaipata y Carrasco”. Preparó el “Diseño de la muestra para la encuesta de medición de impacto y evaluación de las acciones de PROSALUD en la ciudad de El Alto” ejecutado en los meses de octubre a diciembre de 1992, siendo estas sólo algunas de las producciones logradas en su carrera profesional.

Como se puede notar, este es simplemente un pequeño resumen de las muchas cosas que don luchito (denominación cariñosa que muchas personas le decían cuando se referían a él) ha hecho en la Carrera.

Hoy docentes, profesionales en Estadística y estudiantes que hemos sido orgullosamente sus estudiantes, no podemos olvidar las caminatas que hemos realizado con él. Nos ha dejado un importante mensaje de amistad entre todos. Las caminatas que hicimos, perdurarán en el recuerdo como parte de esparcimiento entre los que hemos vivido esas lindas experiencias junto a Luis Zapata.



Desde este sitio de comunicación de la revista Varianza, hemos querido mandar un mensaje de homenaje, reconocimiento, admiración y respeto a uno de los grandes profesores que hemos tenido en la Carrera el Licenciado Luis Zapata Escobar.





## Día Mundial de la Estadística

El Día Mundial de la Estadística, ha sido fijado por la Comisión Estadística de las Naciones Unidas para el 20 de octubre de 2010 (como primer aniversario). La celebración de este día permite reconocer el servicio prestado por organismos internacionales, nacionales, regionales y subregionales para promover y facilitar el avance de los sistemas estadísticos nacionales.

La resolución de las Naciones Unidas, reafirma que es fundamental que los países cuenten con capacidad estadística nacional para producir estadísticas e indicadores fiables y oportunos que sirvan de base para tomar decisiones informadas, basadas en los valores básicos de servicio, integridad y profesionalidad.

El ente oficial de las estadísticas en nuestro país, es el Instituto Nacional de Estadística de Bolivia (INE), y como tal, debemos nosotros los estadísticos reafirmar nuestro compromiso de aportar con dicha institución en la función técnica y científica de la generación información estadística en beneficio de nuestra sociedad.

La UMSA, no queda al margen de las múltiples necesidades de información estadística requerida, no sólo para la administración y planificación de sus labores internas, sino más bien para el apoyo a organismos nacionales e internacionales, y la sociedad en sí, que actúan en nuestro país.

Es así que hoy nuestras autoridades universitarias, recurren a nuestro servicio profesional de estadística (docentes y estudiantes) para aportar en este fin. Nosotros como profesionales en el campo de la estadística, no podemos quedar al margen, es por eso que hoy en el día mundial de la estadística reafirmamos nuestro compromiso de continuar trabajando.

Tras un año de celebrarse por primera vez el Día Mundial de la Estadística, la carrera de Estadística no estuvo indiferente a tal celebración.



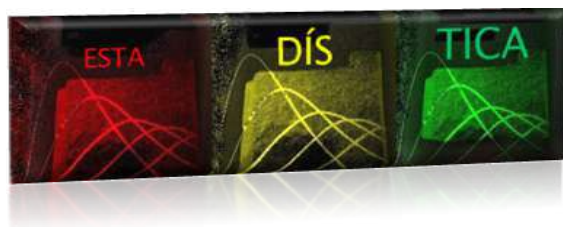
En un pequeño y significativo acto, el director de la Carrera de Estadística Lic. Juan Carlos Flores L., realizó este acontecimiento mundial. De igual forma, nuestros docentes más antiguos Rubén Belmonte C. y Luis Zapata E. fueron invitados para dar un discurso de la historia e importancia de la estadística para nuestra sociedad, al igual que los universitarios representantes del Centro de Estudiantes de Estadística no quedaron al margen en esta celebración. Lo más significativo de los discursos fue: la aportación de la estadística en la historia, la necesidad de las estadísticas oficiales al bienestar de la sociedad, donde se resaltaron los servicios que presta el Sistema Nacional de Estadísticas Públicas. Se destacó también la profesionalidad e integridad de los pocos estadísticos bolivianos existentes promoviendo la confianza del público en la generación de nuevas metodologías, procesamiento y análisis de estadísticas oficiales.



*Profesor emérito de la Carrera de Estadística  
M.Sc. Ruben Belmonte Coloma*

Se manifestó la importancia de servicio técnico y científico de nuestros profesionales formados en nuestra universidad, para bien de la sociedad en temas teóricos y prácticos del quehacer de la estadística.

Desde la Universidad Mayor de San Andrés, los estadísticos, docentes y estudiantes, debemos comprometernos a trabajar en beneficio de nuestro país que confía en nuestra capacidad y servicio profesional.



**20**  
**Octubre**

**Día Mundial  
de la  
Estadística**



*La Asamblea General de las Naciones Unidas designó que cada 20 de Octubre se conmemore el Día Mundial de la Estadística para elevar la conciencia de los muchos logros de las estadísticas oficiales basadas en los valores fundamentales de servicio, profesionalismo e integridad.*

## XXVII Aniversario de la Carrera de Estadística

La Carrera de Estadística de la Universidad Mayor de San Andrés, cumple un año más de vida. Son 27 años de servicio en la formación de profesionales en el área de la estadística.



La Carrera de Estadística está instalada en el edificio antiguo del Monoblock Central de la UMSA, Av. Villazon. Allí se desarrolla la mayoría de las actividades académicas donde se encuentra la dirección de la Carrera. En la calle 27 de Cota Cota, Campus Universitario funciona el Instituto de Estadística Teórica y Aplicada (IETA) con actividades de investigación e interacción social.



Allá por el año 1972 se crea la especialidad de Estadística dentro del Departamento de Matemática con un plan de estudios apenas diferenciado, junto con la especialidad de Informática.

Esta especialidad se inicia con pocos estudiantes y docentes, con etapas dificultosas en cuanto a dotación de los medios necesarios para su funcionamiento que paulatinamente se van solucionando hasta su consolidación. Se estructura un plan de de estudios acorde a las necesidades de la licenciatura. En 1984 por resolución del VII Congreso de Universidades, la Carrera de Estadística se separa de la tutela de la Carrera de Matemática y a partir de ese año, funciona en forma independiente.



En 1985, se crea el Instituto de Estadística Teórica y Aplicada (IETA), para el desarrollo de la investigación e interacción social en el campo de la estadística, siendo dependiente de la carrera con objetivos definidos.

Actualmente, con el pasar del tiempo, la carrera ha dado un número reducido de profesionales en estadística y la demanda de ellos es grande, cubierta por instituciones públicas y privadas. Otros profesionales se incorporan a la docencia en la Carrera y en otras facultades.



Del perfil profesional, se puede decir que los graduados en estadística tienen la capacidad de formular, planificar y ejecutar proyectos de investigación teóricos y aplicados, fruto de la información con ética y actitud creativa.

Se proponen soluciones a problemas de optimización en la industria, transporte y otros sistemas. Participan en equipos profesionales multidisciplinarios de asesoramiento en la industria, banca, agropecuaria, estudios socioeconómicos, etc.

Como se puede ver, en tan poco tiempo, la carrera ha dado grandes pasos.

Respecto a las actividades en conmemoración de la creación de la carrera, se inició el 24 de octubre del presente año con la inauguración de la semana aniversario a cargo del nuevo director Lic. Juan Carlos Flores L.

Durante la semana se realizaron diferentes actividades entre conferencias por parte de invitados profesionales y docentes de la carrera. Por la tarde, se realizaron actividades deportivas y culturales.

La noche de talentos se efectuó en el auditorium de Cota Cota con la participación de estudiantes y docentes. En esa oportunidad se realizó la elección del Mister de Estadística donde fue elegido el universitario Bladimir Morales T., estudiante de los cursos básicos de la carrera.

El día jueves 27 de octubre, se realizó el bautizo a los nuevos estudiantes, dándoles de esa manera la bienvenida a esta importante carrera.



Se cerró con broche de oro la semana aniversario de Estadística el día viernes 28 de octubre con una recepción social en el salón de eventos "Baltor". Una de las actividades de la fiesta, fue elegir a la "Miss Estadística" donde fue elegida la señorita María Eugenia Almeida G. digna representante.

Desde la Revista Varianza, deseamos todo lo mejor para el futuro de nuestra carrera, siempre con el compromiso de servirla.

Felicidades Carrera de Estadística!!!!





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

**La Paz - Bolivia**