LA REVISTA DE LA CARRERA DE ESTADISTICA

VARIANZA

Universidad Mayor de San Andrés
Facultad de Ciencias Puras y Naturales

AÑO 2
N° 2

REVISTA VARIANZA

Número 2, Año 2
Noviembre del 2002

Dirección:
José Aníbal Angulo A.
Director IETA

Colaboradores
Luis Zapata
Nicolás Chávez
Raúl Delgado
Jaime Pinto A.
Jorge Troche
Juan Carlos Flores
Augusto Solís
Jaime Chumacero
David Barrera
Fernando Rivero S.
Aníbal Angulo

Impresión
¡DEAGRAF
Cel. 719-46681

Diseño Gráfico
Américo Gonzales O.
Telf. 2230455

PRESENTACIÓN

En reconocimiento de la importancia que los conceptos y técnicas Estadísticas tienen en las diferentes disciplinas, motivó a la anterior Dirección del I.E.T. A. ya los docentes de la Carrera de Estadística a publicar el 1er numero de la revista VARIANZA.

Con la publicación de este 2do número, se pretende por una parte poner a consideración del lector los temas que han motivado a los autores a realizar dichos artículos y por otra se motiva a los lectores en diferentes temas para que puedan tener suficiente interés en profundizar los fundamentos teóricos de esta importante disciplina.

Si bien los temas presentados en esta oportunidad son diversos, el orden y el enfoque de los mismos han sido introducidos de manera de ser accesibles a la comprensión de la comunidad Universitaria, ofreciendo en muchos casos herramientas efectivas para estructurar, analizar y sintetizar una gran variedad de problemas referidos al manejo de información estadística.

El esfuerzo de la Dirección del I.E.T.A. es un ejemplo motivador y conservar el entusiasmo por el valor y la importancia de la Estadística es una tarea en la cual toda la comunidad de la Carrera esta comprometida.

Apreciaremos mucho a los lectores que nos hagan conocer sus opiniones y sugerencias para mejorar.

Lic. Nilda Flores Salinas
JEFE CARRERA DE ESTADÍSTICA

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

La Paz - Bolivia
Edificio Antiguo Planta Baja pb05
Tel: 2442100

VARIANZA

REVISTA DE LA CARRERA DE ESTADÍSTICA

EL PENSAMIENTO ESTADÍSTICO

H. G. Wells

“El pensamiento estadístico será un día tan necesario para el ciudadano eficiente como la capacidad de leer y escribir. “

Existen dos formas principales de pensamiento lógico, la deducción y la inducción. La primera se debe principalmente a los griegos, que fueron los primeros en ver claramente la gran potencia de proponer axiomas o hipótesis generales y deducir de ellos. El pensamiento inductivo, no comenzó a constituir una herramienta sistemática del hombre hasta la última parte del siglo XVIII. La inducción procede en la dirección opuesta a la deducción. Partiendo de hechos experimentales, nos conduce a inferir conclusiones generales.

Francis Bacon fue el primero que subrayó adecuadamente los métodos inductivos como base del procedimiento científico, pero no fue sino en 1763 cuando el clérigo Thomas Bayes construyó la primera base matemática de esta rama de la lógica. A fin de formarnos una idea de lo que hizo Bayes consideremos un ejemplo que es, por supuesto, artificioso.

Supongamos que tenemos una caja cerrada que contiene un gran número de bolas blancas y negras, pero tenemos alguna razón para pensar que existe la probabilidad 2/3 de que hay tantas bolas blancas como negras en la caja. Entonces tomamos la caja sacamos una muestra de bolos y encontramos que 3/4 de las muestras son negras ahora bien, antes de sacar esta muestra nos inclinábamos fuertemente a pensar que la mezcla desconocida era mitad de blancas y mitad de negras. Después de haber tomado la muestra, claramente deberíamos cambiar nuestra idea y comenzar a inclinarnos hacia la opinión de que las bolas negras superen en número a las blancas que hay en la caja. Bayes elaboró un teorema que indica exactamente cómo deberían ser modificadas las opiniones mantenidas, antecedentemente a los experimento, por razón de la evidencia del muestreo. Aunque la utilidad de este teorema mismo ha resultado ser muy limitada, éste fue el comienzo de toda una teoría matemática del razonamiento inductivo.

¿Cómo es esto?, podría decirse, ¿Es la estadística algo tan general y profundo? ¿No es la estadística meramente el nombre de una información numérica con la cual los publicistas tratan de convencernos y a veces nos confunden?.

La palabra estadística tiene dos significados algo diferentes. En el uso más familiar en cierto que la Estadística significa simplemente información numérica, ordenada en tablas o en graficas. En este sentido decimos que el Almanaque Mundial contiene una gran cantidad de estadísticas útiles; pero más ampliamente, y más técnicamente, la estadística es el nombre de la ciencia y del arte que trata de la inferencia incierta, la cual usa los números para obtener algún conocimiento acerca de la naturaleza y de la experiencia.

Lo importante del razonamiento inductivo se basa en el hecho de que, dejando a un lado excepciones triviales, los sucesos y los fenómenos de la naturaleza son demasiado multiformes, demasiado numerosos, demasiado extensos o demasiado inaccesibles para permitir una observación completa. Como hizo observar el autor de Eclesiastés, “no hay hombre alguno que no pueda conocer la obra que Dios hace desde el principio hasta el fin”. No podemos medir los rayos cósmicos en todas partes y en cada instante. No podemos ensayar un nuevo medicamento en todas las personas. No podemos comprobar cada una de las granadas y las bombas que fabricamos, entre otras razones porque no quedaría ninguna por usar. Así hemos de contestarnos con muestras. Las medidas obtenidas en cada experimento científico constituyen una muestra del conjunto ilimitado de mediciones que resultarían si uno realizase el mismo experimento una y otra vez indefinidamente. Este conjunto total de mediciones potenciales se suele denominar la Población .Casi siempre se interesa uno en la muestra solamente en cuanto es capaz de revelar algo acerca de la población de la cual procede.

Las cuatro cuestiones principales que uno se pregunta acerca de las muestras son las siguientes:

1. ¿Cómo se puede describir la muestra de forma útil y clara?

2. A partir del conocimiento de esta muestra, ¿cómo se deben inferir de la mejor manera posibles conclusiones que se refieran a la población total?.

3. ¿Hasta qué punto son de fiar estas conclusiones?.

4. ¿Cómo se deberían tomar las muestras a fin de que puedan ser tan iluminadoras y tan de garantía como sea posible?.

Es evidente que el estadístico nunca puede afirmar con certeza (al 100%) cómo es la población original, mediante un mero muestreo, porque las demás variarán. Si, por ejemplo, vamos sacando muestras de una mezcla que contenga 70% de bolas blancas y 40% de bolas negras, en modo alguno obtendremos esta relación de 70 a 40 de bolas blancas a negras en cada una de las muestras que tenemos. Sin embargo, para una cierta población y con métodos adecuados de muestreo es posible elaborar teóricamente el esquema de variación de las muestras da al estadístico un firme apoyo. Le permite considerar las muestras y obtener conclusiones acerca de la población original.

Debemos recordar ahora que la Estadística trata de conclusiones inciertas. No podemos esperar que el estadístico llegue a una conclusión absolutamente firme. Lo que podemos esperar es que nos proporcionen una respuesta doble a nuestra cuestión. Una parte de su respuesta puede ser: “Mi estimación mejor es .....” . La otra parte inevitable de su respuesta es: “El grado de confianza que usted está justificando en dar mi estimación es .....“

Autor Warren Weaver
Matemáticas en el Mundo Moderno
Selecciones de Scientific American

Índice

AUTOESTIMA ¿INNATA O INDUCIDA?

Luis Zapata Escobar

Introducción

Como resultado de la asesoría de un postulante al título de Licenciatura en Psicología, el presente trabajo es una aplicación del muestreo estratificado para transformar una investigación cuasi experimental en otra de carácter experimental. El estudio se realizó en estudiantes de segundo medio de un colegio de la ciudad que atiende en dos turnos, de la mañana y de la tarde, divididos en áreas de especialidad, Exactas y Sociales. Los jóvenes fluctúan entre los 15 y 16 años, tanto hombres como mujeres.

El estudio esta referido a medir el nivel de autoestima y la forma en que, mediante charlas de motivación, talleres y dinámicas de grupo se podría incrementar estos niveles, teniendo en cuenta que para esta edad, los conflictos internos propios del cambio de niño a adolescente son considerados cruciales a la luz de la inquietud del sicólogo investigador del colegio de referencia.

Hipótesis de investigación: Es posible mejorar el nivel de autoestima ofreciendo temas de motivación para incrementar dicho nivel a través de talleres, conferencias y dinámicas de grupo,

Objetivos específicos:

a. Determinar si los niveles de autoestima se diferencian según los factores: curso, turno, y especialidad.

b. Analizar la medida en que un cambio en los niveles de autoestima pueden ser afectados por motivaciones de un programa dirigido.

c. Analizar si hay o no diferencias entre factores de curso, turno, especialidad y sexo.

d. Analizar si en media la diferencia antes del programa dirigido y, después de él, produce un cambio significativo.

Población objetivo. Alumnos de segundo medio con edades entre los 15 y 16 años. Que cursan estudios secundarios en 8 cursos paralelos divididos en 5 paralelos en el turno de la mañana y 4 cursos paralelos en el turno de la tarde, separados en dos áreas de estudio.

Marco muestral. Listado en orden alfabético de los 8 cursos divididos de la siguiente manera:

DISTRIBUCIÓN DE ALUMNOS SEGUN AREA POR TURNO, CURSOS Y NUMERO				
	TURNO MAÑANA		TURNO TARDE	
AREA	Nº DE CURSOS	Nº DE ALUMNOS	Nº DE CURSOS	Nº DE ALUMNOS
Exactas	2	104	2	90
Sociales	3	183	2	130
ToTal	5	287	4	220

Diseño de la investigación. Todos los alumnos en ambos turnos son atendidos por el mismo equipo de docentes en cada una de las asignaturas según el área. Por esta razón se consideran grupos ya formados, en su origen no son asignados al azar para ser considerados en una investigación experimental. A todos los estudiantes, se aplicó un test inicial (Pre test) y luego se los invitó a participar en conferencias, talleres y dinámicas de grupo que demandaron aproximadamente 3 meses. Las actividades incluyeron temas de motivación de los que se esperaba influyen en un aumento del nivel de autoestima. Transcurrido ese tiempo se aplico un segundo test (Post test). Cada test fue calificado en escala de 1 al 100. Los formularios contenían siete ítems, de ellos el más importante, motivo del presente análisis es el test estándar de autoestima F35B.

Estratificación. Los estratos están definidos por las áreas de estudio y el turno de asistencia, en consecuencia son 4 estratos.

Estrato 1. Alumnos*de Exactas turno de la mañana. Cursos E11 y E12

Estrato 2. Alumnos de Sociales turno de la mañana, Cursos E21, E22, E23

Estrato 3. Alumnos de Exactas turno de la tarde. Cursos E31, E32

Estrato 4 Alumnos de Sociales turno de la tarde.. Cursos E41, E42

Tamaño de muestra. De los resultados del formulario aplicado al principio de la gestión escolar, Pre test se obtuvo la desviación estándar de 11, 65 puntos para los 507 jóvenes que se presentaron. (Desviación estándar poblacional) Para un 95% de confiabilidad y un error de 3 puntos el tamaño de la muestra resultante es:

$$n_0 = \frac{Z^2_{1-\alpha/2}\sigma^2}{\epsilon^2} = \frac{1{,}96^2 * 11{,}65^2}{3^2} =$$

$$n = \frac{n_0}{1+\frac{n_0}{N}} = \frac{58}{1+\frac{58}{507}} = 52$$

Asignación de la muestra. Se prueban dos formas de asignación, proporcional y óptima de Neyman

ESTRATOS	*PROPORCIONAL*	*NEYMAN*
Estratos 1	*11*	*10*
Estratos 2	*19*	*19*
Estratos 3	*9*	*10*
Estratos 4	*14*	*13*
TOTAL	*53*	*52*

Selección de la muestra. Se selecciona una muestra aleatoria simple con asignación de Neyman y en forma proporcional al tamaño de cada curso que corresponde al estrato. Ver anexo 1.

Análisis de la información. Para probar la hipótesis de investigación, y seleccionada la muestra al azar de cada estrato, se tienen cuatro subpoblaciones, dos turnos, 9 cursos y dos niveles para sexo.

Análisis de resultados del Pre test. Interesa primero saber si en el Pre test hay alguna diferencia entre los grupos según los factores señalados. Se usa una significación de 5 % Los niveles de autoestima registrados en la primera oportunidad, se los reúne en la siguiente tabla resumen, según factores por grados de libertad para él test de análisis de varianza.

Resultados de Pre Test según factor F calculando del Anova				
Factor	**g.l. numerador**	**g.l. denominador**	**F**	**Significación**
Cursos	8	43	1,96	0,077
Turno	1	50	8,66	0,005
Estrato	3	48	2,83	0,048
Sexo	1	50	3,64	0,096

Los resultados del Pre test muestran diferencia significativa por turno y por estrato. El nivel de autoestima del turno de la mañana es significativamente mayor que del turno de la tarde pues los datos del resumen anterior se comprueba también con una dócima t.

Para el factor Estrato, la diferencia es significativa y por tratarse de 4 grupos, se aplica el Test de Tuckey LSD para determinar entre cuales grupos hay esa diferencia puesto que el mínimo p es de 4,8% ligeramente inferior al de significación. Las probabilidades son las siguientes:

Prueba de Trukey de diferencia de medias según estrato			
Estrato	**Estrato (j)**	**Diferencia**	**Significación**
1	3	7,50	0,042
	4	6,93	0,046
2	3	6,43	0,046
	4	5,86	0,048

Estos resultados explican que el nivel de autoestima del estrato 1 (exactas del turno de la mañana) es mayor que el nivel de autoestima de los estratos 3 y 4 (turno de la tarde). De igual manera este resumen indica que ocurre similar situación entre el estrato 2 (sociales turno de mañana) con referencia de los estratos 3 y 4.

Análisis de resultados del Post test. Estos se presentan en el siguiente resumen:

Resultados de Pre Test según factor F calculando del Anova				
Factor	***g.l. numerador***	***g.l. denominador***	***F***	***Significación***
Cursos	***8***	***43***	***0,414***	***0,906***
Turno	***1***	***50***	***1,95***	***0,546***
Estrato	***3***	***48***	***0,437***	***0,727***
Sexo	***1***	***50***	***3,64***	***0,624***

Dado que ninguna de las probabilidades que miden la significación son menores 5%, se concluye que no hay diferencias significativas para ninguno de los estratos.

Análisis de resultados de la diferencia. En este acápite se analizan las diferencias entre los puntajes obtenidos en el Post test respecto de Pre test.

Resultados de Pre Test según factor F calculando del Anova				
Factor	***g.l. numerador***	***g.l. denominador***	***F***	***Significación***
Cursos	***8***	***43***	***0,432***	***0,836***
Turno	***1***	***50***	***0,549***	***0,427***
Estrato	***3***	***48***	***0,943***	***0,427***
Sexo	***1***	***50***	***0,645***	***0,726***

Estos resultados indican que no hay diferencias significativas en los diferentes factores y que las probabilidades que detectan la significación son todos mayores a 5%.

Conclusiones.

1. Las pruebas de análisis de varianza por turno indican que los niveles de autoestima son significativamente diferentes, el valor de significación esta en 5 por mil.

2. En efecto, es posible cambiar la percepción que tienen los jóvenes respecto de si mismos, de cómo los ve la sociedad y como pueden mejorar su autoestima después de haber recibido charlas, conferencias especializadas y dinámicas de grupo dirigidas con ese propósito.

3. Es de resaltar que los niveles de autoestima por sexo no tienen diferencias significativas, implicando este detalle que los problemas de cambio de niño a adolescente afectan por igual tanto a jóvenes como a niñas.

4. Del análisis de las diferencias se concluye que el aprovechamiento inducido por aspectos motivadores iguala al

comportamiento sobre si mismos en todos los factores considerados, tanto en turno, curso especialidad o sexo.

Anexos.

1. Análisis de varianza por estratos

Resultados de Pre Test según factor F calculando del Anova						
ITEM	***FUENTE***	***SUMA DE CUADRADOS***	***GRADOS DE LIBERTAD***	***CUADRADOS MEDIOS***	***F CALCULADO***	***SIGNIFICACION***
Pre Test	***Entre los grupos***	***547,629***	***3***	***182,543***	***2,83***	***0,048***
	Dentro los grupos	***3096,429***	***48***	***64,509***		
	Total	***3644,058***	***51***			
Post Test	***Entre los grupos***	***425,074***	***3***	***141,691***	***0,437***	***0,727***
	Dentro los grupos	***15558,157***	***48***	***324,128***		
	Total	***15983,231***	***51***			
Diferencia	***Entre los grupos***	***708,629***	***3***	***236,284***	***0,943***	***0,427***
	Dentro los grupos	***12021,206***	***48***	***250,442***		
	Total	***12730,058***	***51***			

2. Estadígrafos.

ITEM	***MINIMO***	***MAXIMO***	***MEDIA***	***DESVIACIÓN***
Pre test	***25***	***62***	***39,13***	***8,457***
Post test	***22***	***94***	***51,77***	***17,70***
Diferencia	***-20***	***50***	***21,7***	***19,67***

BIBLIOGRAFIA

Índice

MODELO DE WINTERS

Nicolás Chávez Quisbert

El modelo de Winters, es un método de suavizamiento muy importante en el campo de la estadística, ya que permite realizar pronósticos, para serie de tiempo que no tienen mucha información histórica, lo cual se da en muchas series económicas de nuestro país. Es por eso que no se puede utilizar los procesos estocásticos, por requerirse de muchas datos, y se ve la necesidad de utilizar técnicas de suavizamientos.

El método de Winters consiste en determinar los tres elementos explicables de la serie de tiempo que son, el elemento estacionario, la tendencia, y la componente estacional, y combinar estos elementos para obtener el pronóstico del siguiente periodo.

Selección del modelo para hacer pronósticos

La selección del modelo usado para el pronóstico se basa en la naturaleza de los datos, y por el comportamiento que estos presenten, por lo tanto la ecuación del modelo multiplicativo usada para hacer el pronóstico es:

$$\boxed{P_{t+m} = (S_t + b_t\ m)\ \ I_{t-L+m}} \qquad (1)$$

Donde:

P_{t+m} = Pronóstico para el período (t+m).

S_t = Valor del elemento constante en el período t.

b_t = Valor de la tendencia en el periodo t.

I_{t-L-m} = Valor del factor estacional en el período (t-L+m), osea L períodos atrás del período para el que se desea el pronóstico.

L = Número de estaciones o períodos en un año.

m = Número de períodos en el futuro en que se requieren los pronósticos.

Actualización de los elementos que integran la serie de tiempo

El procedimiento para obtener el pronóstico y hacer las correcciones a cada uno de los elementos que integran la serie de tiempo, se describe a continuación:

1.- Al final del período t, se registra el valor de la serie de tiempo X_t durante ese período.
2.- El elemento constante se actualiza con la ecuación definida en (2), usando el elemento constante y la tendencia del período anterior, y el factor estacional calculado durante el ciclo.

$$S_t = \alpha (X_t / I_{t-L}) + (1 - \alpha) (S_{t-1} + b_{t-1}) \quad (2)$$

3.- La tendencia se actualiza con la siguiente ecuación.

$$b_t = \beta (S_t - S_{t-1}) + (1 - \beta) b_{t-1} \quad (3)$$

4.- Para actualizar el factor estacional se usa la ecuación.

$$I_t = t (X_t / S_t) + (1 - t) b_{t-L} \quad (4)$$

5.- El pronóstico se realiza mediante la ecuación.

$$P_{t+m} = (S_t / B_t m) I_{t-L+m} \quad (5)$$

Si se observa detalladamente esta ecuación ya fue definida en la ecuación (1), que corresponde a la de un modelo multiplicativo con tendencia aditiva, y factor estacional multiplicativo, este tipo de series son las que se presentan con mas frecuencia en la vida practica.

6.- Realizar el corrido del modelo para seleccionar la mejor combinación de los parámetros convexos α ζ β y T, de tal forma de que el Error cuadrático Medio (ECM) sea mínimo. Por lo tanto tendrá que calcularse, los errores como:

$$e_t = (X_t - P_t) \quad (6)$$

Para luego calcular el Error Cuadrático Medio definido en (7).

$$ECM = 1/N \sum_{t=1}^{N} (X_t - P_t)^2 \quad (7)$$

Carrera de Estadística

Una de las ventajas de este método, es que siempre el modelo podrá ser actualizado con información nueva obtenida.

Aplicación del Modelo

Para una mejor comprensión del Modelo de Winters, se presenta a continuación una aplicación real a la serie:

X_t = Crecimiento Poblacional Estudiantil de la Universidad Mayor de San Andrés para los años 1983 al 2000.

Tomándose en cuenta a los universitarios matriculados en todas las facultades, sin considerar los estudiantes de post grado.

La información fuente fue tomada del CEUB , y software utilizado fue el paquete ***"ECONOMETRIC VIEWS"***

La siguiente tabla muestra los datos de la serie mencionada, y a su lado se puede observar

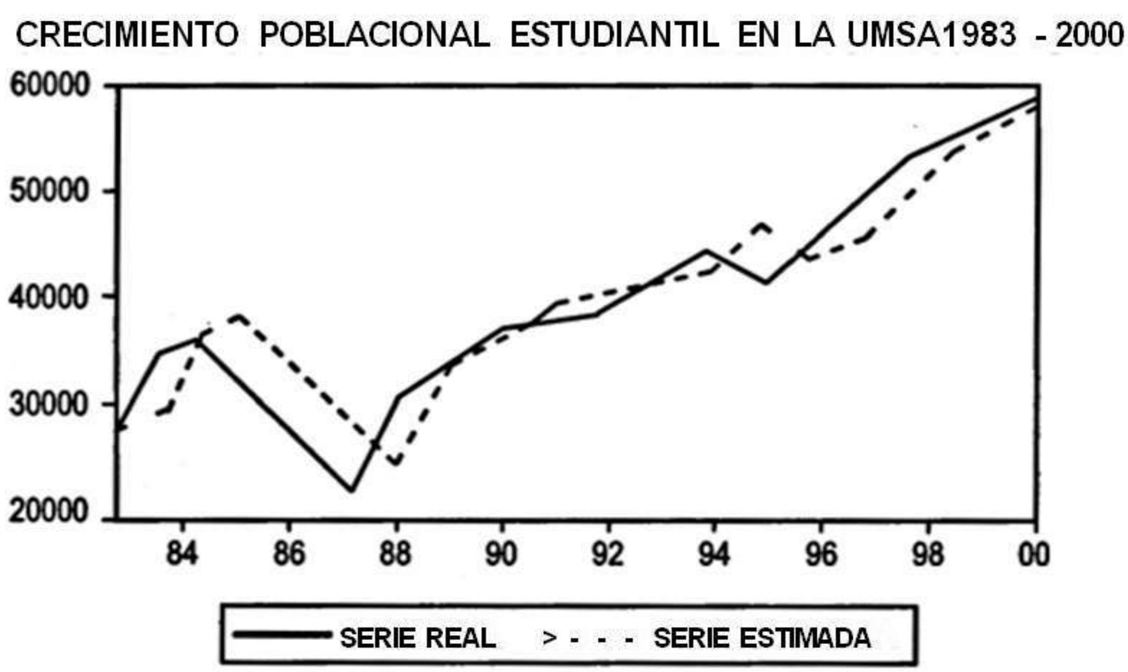
$$\alpha = 0.99,\ \beta = 0.05,\ \tau = 0.00$$

el corrido del modelo con el paquete, obteniéndose los valores de los parámetros:

Estos se determinaron de tal forma de minimizar el Error Cuadrático Medio, definido anteriormente en (7), también se puede decir que no existe influencia estacional al ver el valor del último parámetro.

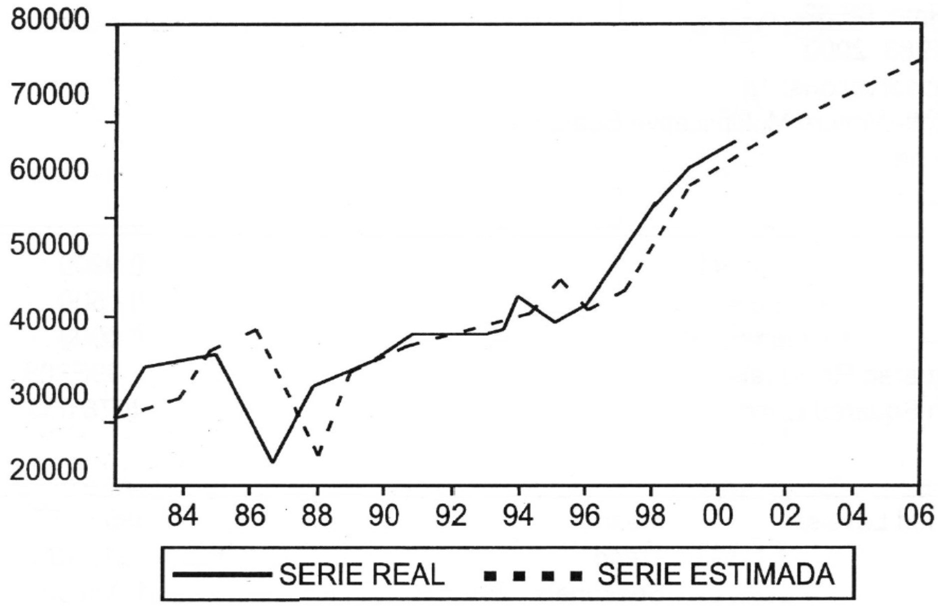
AÑOS	SERIE
1983	26199
1984	34146
1985	35632
1986	29821
1987	22668
1988	31368
1998	33236
1990	35966
1991	36400
1992	37123
1993	38734
1994	42085
1995	39241
1996	41138
1997	47362
1998	52432
1999	55661
2000	58850

Date:// Time:06:32			
Sample: 1983 2000			
Included observations: 18			
Method: Holt-Winters Multiplicativo Seasonal			
Original Seríes: X			
Forecast Series: XSM			
Parameters:	0< Alpha <1		0.9900
	0< Beta <1		0.0500
	0< Gamma <1		0.0000
Sum of Squared Residuals		2.99E+08	
Root Mean Squared Error		4078.033	
End of Períod Levels:	Mean		58837.87
	Trend		2049.193
	Seasonals: 2000		1.000.000



FUENTE: UNIVERSIDAD MAYOR DE SAN ANDRES
ELABORACIÓN DAF-CEUB

ESTIMACION DE LA POBLACION ESTUDIANTIL DE LA UMSA HASTA EL AÑO 2006



Como se puede ver en la grafica entre los años 1983 y 1988 no existe una tendencia fija, sin embargo desde 1989 a 2000 ya se puede determinar una tendencia definida.

El siguiente cuadro muestra las extrapolaciones para 6 años.

AÑOS	PROYECCIONES
2001	60887
2002	62936
2003	64985
2004	67034
2005	69083
2006	71133

BIBLIOGRAFÍA

DAVIS M. H. A.(1977); LINEAR ESTIMATION AND STOCHASTIC CONTROL; JOHN WILEY & SONS; LONDON.
HAMILTONJ. (1991); TIME SERIES ANALISYS; PRINCETON UNIVERSITY PRESS
HARVEYA. C. (1994); TIME SERIES MODELS; PHILIP ALLAN, PUBLISHERS LIMITED; LONDON.
KENDALL M.G., M. A. Se. D„ F.B.A. (1973); TIME-SERIES; CHARLE GRIFFIN & COMPANY LIMITED; LONDON
PULLIDO SAN ROMAN ANTONIO (1989); PREDICCION ECONOMICA Y EMPRESARIAL, EDICIONES PIRÁMIDE; MADRID

Índice

INDICE GINI Y ECONOMETRIA

David Barrera Ojeda

El coeficiente de Gini

El coeficiente de Gini es uno de los índices más usados por sus propiedades geométricas y económicas, a pesar de ser un índice rígido (Martín Alberto M., 1998). Es interesante éste indicador como un indicador de desigualdad muy usado cuando se estudia la variable ingreso.

Sea X una variable aleatoria con una distribución F. Analizaremos brevemente la diferencia de medias y el indice de gini de na función de distribución F. Sea $U_F : \Re \mapsto [0,1]$ función de distribución sobre que tiene una esperanza finita

$$\mu(F) = \int_{-\infty}^{\infty} x dF(x) > 0$$

La Media de diferencia de medias se define como:

$$M(F) = \frac{1}{2} \iint_{\Re} |x - y| dF(x) dF(y)$$

es la distancia euclidena de dos variables aleatorias independientes. El Indice de Gini de F se define como: $IG = M(F) / |\mu(F)|$ esta definición es ventajosa porque, evita que el indice de Gini sea negativo.

Proposición:

Sea $F^{-1}(s) = \inf\{x : F(x) \geq s\}, s \in]0,1]$ denota la inversa de la función de distribución F y Entonces si F(0)=0

$$|\mu(F)| L_F(t) \text{ y } |\mu(F)| (1 - L_F(t-1)),$$

(i)M(F) es igual al área entre los gráficos de las dos funciones.

$$L_F(t) \text{ y } 1 - L_F(1-t), t \in [0,1]$$

(ii) IG(F) es igual área comprendida entre las dos funciones.

Nota.- Esta proposición es interesante porque permite obtener de manera general el Indice de Gini de cualquier función de distribución F, para ello puede usarse ordenares como el maple, matlab o winplot.

En Estadística los estimadores(IG) sino va acompañado de su error estandar ignoramos la confiabilidad, es decir con que precisión se estima, aquí ingresa en juego el tamaño de muestra, obviamente mejor si se obtiene en un intervalo de confianza, así nos permite función de distribución sobre que tiene una realizar test de hipótesis respecto al IG.

Recientemente, Karagiannis y Kovacevic (2000) y Ogwang (2000) ha reconsiderado este problema. En particular las dificultades en el momentode programar para nuestras grandes las dificultades son enormes. Ogwang proporciona una interpretación basado en la regreseión particular para el coeficiente de Gini que no sólo forma la base de su acercamiento, pero sin querer expone el hecho que alli realmente la necesidad de acudir a una técnica más sencilla. Los propósitos de esta nota son exponer éste punto, y para mostrar có,o una regreseión es útil con respecto a varias pruebas de la hipótesis que son de interés práctico. Nosotros ilustramos nuestros resultados con aplicaciones empíricas.

Sea X una variable aleatoria con función de distribución F. Supongamos que realizamos n observaciones independientes de X, y que los datos ordenados están en orden creciente. El Indice de Gini que puede ser expresado como:

$$IG = \frac{n^2 - 1}{6n}\frac{\hat{\beta}}{X}$$

donde β es un estimador mínimo cuadrático de β (MCO) del modelo:

$$X_i = \alpha + \beta\, i + \varepsilon_i \qquad *$$

donde las disturbancias cumple los supuestos de gauss markov, pero también puede demostrase que el IG puede escribirse como:

$$IG = \frac{2\hat{\theta}}{n} - 1 - \frac{1}{n}$$

donde se estima mediante mínimos cuadrados ponderados(WLS) del modelo:

$$i = \theta + \upsilon_i^{**}$$

donde las tisturbancias son hereocedásticos con varianza , de donde se obtiene

$$\hat{\theta} = \frac{\sum_{i=1}^{n} i\, x_i}{\sum_{i=1}^{n} x_i}$$

esto es una principal contribución al Indice de Gini. De ** se obtiene que $Var(IG) = 4\,\hat{\theta}/n^2$ y el error estandar es . el error estándar es suficiente como media de dispersión para, puesto que IG no depende de las unidades que tenga X.

Nota.- En el primer caso se observa donde la

ESTIMADOR DEL INDICE DE GINI PUNTUAL E INTERVALO DE CONFIANZA (95%) DE UNA POBLACIÓN LOGNORMAL CON MEDIA 10 Y VARIANZA 4 "ESTIMADOS POR WLS"

Nro. obs	Gini	ErrorEst de G	limite inf	limit sup
25	0.102	0.110	-0.114	0.318
50	0.087	0.079	-0.067	0.242
100	0.080	0.056	-0.029	0.189
500	0.065	0.025	0.016	0.114
1000	0.067	0.018	0.032	0.102
5000	0.077	0.008	0.062	0.093
10000	0.112	0.006	0.101	0.123

DE UNA POBLACIÓN LOGNORMAL CON MEDIA 10 Y VARIANZA 1
"ESTIMADOS POR WLS"

Nro. obs	Gini	ErrorEst de G	limite inf	limit sup
25	0.018	0.117	-0.211	0.248
50	0.017	0.082	-0.143	0.178
100	0.019	0.058	-0.094	0.132
500	0.024	0.026	0.027	0.074
1000	0.026	0.018	0.010	0.061
5000	0.036	0.008	0.020	0.052
10000	0.056	0.006	0.045	0.067

DE UNA POBLACIÓN LOGNORMAL CON MEDIA 10 Y VARIANZA 1
"ESTIMADOS POR WLS"

Nro. obs	Gini	ErrorEst de G	limite inf	limit sup
25	0.009	0.117	-0.221	0.240
50	0.010	0.082	-0.151	0.172
100	0.011	0.058	-0.102	0.124
500	0.012	0.026	-0.039	0.062
1000	0.012	0.018	-0.023	0.048
5000	0.017	0.008	0.001	0.033
10000	0.028	0.006	0.017	0.040

varianza es pequeña el Índice de gini se incrementa a medida que aumenta el tamaño de muestra, pero el error estandar disminuye es decir se gana precisión. En el segundo caso la varianza es mayor al primero el Índice de gini se aproxima a 1, esto es debido a la dispersión respecto a la media, pero nuevamente se observa que se gana precisión a medida que aumenta el tamaño de muestra. Podemos concluir que para ganar precisión es determinante el tamaño de muestra, empero el IG está determinado por la varianza de la población de donde proviene. Es recomendable para trabajar con el IG, determinar de la distribución que proviene, porque en otros casos puede darse que IG sea negativo, caso concreto de una distribución extrema del tipo I.

Índice

PARADOJAS

Raúl Delgado Álvarez

Son problemas en los cuales se llega a resultados contradictorios habiendo seguido caminos diversos de apariencia lógica

LA PARADOJA DE BERTRAND

Problema.

¿Cuál es la probabilidad de que una cuerda dibujada caprichosamente en un círculo sea mayor que el lado del triángulo equilátero inscrito en dicho círculo?

Primera Solución

El conjunto total de los casos posibles de dibujar una cuerda se agota cuando desde cada uno de los puntos de la circunferencia se traza un recta que vaya a pasar por cada uno de los demás puntos de la curva : cada una de las cuerdas puede ser trazada en dos direcciones.

El conjunto de todos los casos favorables se consigue al unir los puntos de la circunferencia con solo aquellos otros que subtiendan por el lado del arco menor, arcos mayores que $\frac{2\pi}{3}$ también se tomará en cuenta que este último conjunto se agota dos veces , según se traza la cuerda de un lado u otro del diámetro que pase por el punto considerado. Los puntos a los cuáles se pueden trazar cuerdas favorables desde otro fijo, están situados sobre un arco de magnitud

$$2\pi - 2\frac{2\pi}{3} = \frac{2\pi}{3}$$

mientras que los extremos de todas las cuerdas posibles que parten de un punto fijo, pueden ser todos u cada uno de los de la circunferencia 2π. Por lo tanto la probabilidad buscada es igual a

$$\frac{\frac{2\pi}{3}}{2\pi} = \frac{1}{3}$$

Segunda Solución

En cada uno de los puntos de un diámetro levántese una perpendicular: así se obtiene un sistema de cuerdas paralelas y haciendo lo mismo con todos los diámetros posibles , se obtienen, se obtienen todas las cuerdas posibles. Las cuerdas mayores que el lado del triángulo equilátero inscrito se obtendrán levantando perpendiculares a todos los diámetros posibles en los puntos de él situados entre el centro y el punto medio del radio. Puesto que el conjunto de todos los puntos comprendidos entre el centro y dicho punto medio es la mitad del de los puntos que hay en el radio, se deduce que la probabilidad de que una cuerda dibujada a capricho sea mayor que el lado del triángulo equilátero inscrito, es igual a 1/2.

Tercera Solución

La posición de una cuerda que queda determinada unívocamente por la de su punto medio, y por los puntos medios de todas las cuerdas posibles que llenan , la superficie circular. Según esto la hipótesis de la cuerda trazada por un punto cualquiera puede ser sustituida por el supuesto de considerar un punto cualquiera del círculo como punto medio de aquella cuerda. La probabilidad de que ésta sea mayor que el lado del triángulo equilátero inscrito es equivalente a la probabilidad de que el punto medio de la curda supuesta quede dentro de un círculo de radio *½ r* El conjunto de los puntos medios de tales cuerdas mayores es 1/4 π r^2, mientras que el conjunto de los puntos medios de todas las cuerdas posibles es .por lo cuál la probabilidad buscada es

$$\frac{\frac{1}{4}\pi\, r^2}{\pi\, r^2} = \frac{1}{4}$$

La diferencia de resultados depende de la incompleta determinación del problema . Se precisa que cuerdas mayores a las de mayor longitud que el lado del triángulo equilátero inscrito en el círculo, y cuerdas menores a las demás. E problema consiste en hallar la relación entre el número de cuerdas mayores y el de todas las cuerdas posibles . Pero puesto que existe un número infinito de cuerdas mayores y menores , el resultado es en principio $\frac{\infty}{\infty}$

Así que el valor de la expresión indeterminada dependerá de a manera como se disponga el aumento de las cuerdas hasta un número ilimitado.

En la primera solución sucede que : el número de puntos de división equidistantes sobre la circunferencia desde los cuáles se trazan las cuerdas que los unen a todos los demás puntos de división , se supone cada vez mayor.

En la segunda solución se admite que las partes iguales en que se ha dividido todo diámetro, se hacen cada vez menores y por esta razón los pies de las cuerdas perpendiculares al d i á m e t r o se van aproximando unos a otros

En la tercera Solución se considera un círculo densamente lleno de puntos, que son los puntos medios de las cuerdas Describiendo 2m círculos concéntricos equidistantes con el de radio r, de radios $\rho\,,2\rho,.3\rho \ldots .2m\rho = r$ y sobre las circunferencias de éstos círculos márquese puntos que estén distantes unos de otros, y P tomando esta fracción como unidad de longitud . El número de puntos situados en las circunferencias , desde la primera hasta la de radio mp es $\frac{m(m+1)}{2}$ 6 y hasta la de radio $2m\rho \frac{2m(2m+1)}{2} 6$ .Cuanto mayor se suponga a m tanto más se aproxima a 1/4, la relación entre el número de cuerdas mayores y el número de todas ellas expresadas por los valores anteriores halladas.

La disposición de las infinitas cuerdas posibles se supone distinta en las tres soluciones por lo que se han obtenido resultados diferentes

Bibliografía:

Carlos Araujo Notas de Probabilidades CIENES
Enrique Cansado Estadística General
Allyn and Bacon A Survey of Geometry
P.D. Si plantea otra solución háganosla saber

Carrera de Estadística

Índice

ESTIMACIÓN DEL TOTAL EN ENCUESTAS AGROPECUARIAS

Jaime Pinto Ajhuacho

En el presente caso es la aplicación del muestreo en la estimación de las superficies totales de tierras cultivadas (Superficie dedicada a cultivos determinados, etc.), en un país. El muestreo puede hacerse atendiéndose esencialmente al plan de muestreo. El universo de estudio puede ser el país, departamento, provincia, municipio, cantón, etc.

Determinando el Universo, en el Diseño muestral puede considerarse un muestreo bietapico, donde las unidades muéstrales de primera etapa (UPM) pueden ser los segmentos agropecuarios y en la segunda etapa las unidades de muestreo (USM) pueden ser las “Unidades de Observación Agropecuarias”, es decir el diseño plantearía:

1.- Determinación del Universo de Estudio.

2.- División del Universo en unidades primarias de muestreo (UPMs).

3.- Selección de las Unidades de primera etapa de la muestra.

4- División de las Unidades de segunda etapa (USMs).

5.- Selección de las Unidades de segunda etapa de la submuestra.

ESTIMACIÓN.

Si el diseño muestral fuera un estratificado con asignación de Neyman, considerando tres estratos y determinando Unidades Agropecuarias con superficies pequeñas (U.O.N.G.) y las Unidades Agropecuarias con superficies grandes (U.O.G.), se puede estimar la superficie total del Universo de estudio, para ello el procedimiento será el siguiente:

i).- Primero, se calcula la estimación de la superficie de cada una de las UPMs seleccionadas.

ii).- Segundo, se realiza la estimación de la superficie entre las UPMs.

El Estimador del total, su cálculo es, usando la siguiente notación:

h = Subíndice para el estrato.

i = Subíndice para “Segmentos” dentro del estrato h.

(i =1,2,3, ..., Nh).

j = Subíndice para “unidades de observación” dentro del “segmento i” del estrato h,

(j=1.2,3, ..., M hi).

M_{hi} : Total de “Unidades de observación” en el segmento i , estrato h, después de la actualización (LISTADO).

M'_{hi} : Número de “Unidades de Observación No Grandes” (U.O.N.G.), listadas del segmento i, del estrato h.

M''_{hi}: Número de “Unidades de Observación” (U.O.G.), listadas del segmento i, del estrato h.

N_h : Número de “segmentos agropecuarios” del estrato h.

En la muestra sería:

i=1,2,.........., n h
j=1,2,.........., m hi

n_h : Número de segmentos en la muestra, del estrato h.

n^e_h: Número de segmentos en la muestra encuestados, del estrato h.

m'_{hi}: Número de U.O.N.G. seleccionados del segmento i, del estrato h.

m''_{hi}: Número de U.O.G. seleccionados del segmento i, del estrato h.

m^p_{hi}: Número de U.O.N.G. encuestadas del segmento i, del estrato h.

m^G_{hi}: Número de U.O.G. encuestadas del segmento i, del estrato h.

ESTIMACION DEL PARAMETRO DEL TOTAL, EN LA SEGUNDA ETAPA.-

CASO DE UNIDADES DE OBSERVACION NO GRANDES.(U. O. N. G.)

$$F_{2p} = \frac{1}{f_{2p}} = \frac{M'_{hi}}{m'_{hi}} \cdot \frac{m'_{hi}}{m^p_{hi}}$$ Factor de Expansión

$\sum_{j=1}^{m^p{}_{hi}} Y_{hij}$ Suma variable Y en U.O.N.G. encuestadas.

$$\hat{Y}^{NG}{}_{hi} = F_{2p}\left(\sum_{j=1}^{m^p{}_{hi}} Y_{hij}\right)$$

Es el Valor estimado del Total, de la variable Y de las unidades no grandes del segmento i, del estrato h.

CASO DE UNIDADES DE OBSERVACION GRANDES (U.O.G.)

Es un barrido completo, pero puede ocurrir que unidades de observación grandes no reporten datos por distintos motivos; razón por la cual se deberá medir tasas de entrevista.

$$\hat{Y}^G{}_{hi} = \frac{M''_{hi}}{m''_{hi}} \cdot \frac{m''_{hi}}{m^G_{hi}} \cdot \left(\sum_{j=1}^{m^G{}_{hi}} Y_{hij}\right)$$

Es el valor estimado del Total, de la variable Y de las unidades grandes del segmento i, del estrato h.

Donde:

$t_{eG} = \frac{m^G_{hi}}{m''_{hi}}$ =Tasa de entrevistados de la U.O.G.

Por lo tanto, el evalor estimado del total de la variable Y, del segmento i, estrato h, es:

$$\hat{Y}_{hi.} = \hat{Y}^{NG}{}_{hi} + \hat{Y}^G{}_{hi}$$

ESTIMACION DEL PARAMETRO DEL TOTAL EN LA PRIMERA ETAPA.
(ENTRE SEGMENTOS).-

Por lo tanto el valor estimado del total de la varable Y, del estrato h, es:

$$\hat{Y}_{hi.} =. \frac{N_h}{n_{h..}} . \frac{n_{h..}}{n^e{}_{h..}} . \left(\sum_{i=1}^{m^e{}_{hi}} \hat{Y}_{hi}\right)$$

Ejemplo.- Como una aplicación del método estadístico se presenta la estimación del total de superficie cultivada (Has.), realizada en una Encuesta Agropecuaria, para datos capturados en el Universo de Llanos de Chaco (Agrícola) del departamento de Chuquisaca, Bolivia.

1.- Estimación de la superficie cultivada entre las Unidades de Observación Agropecuarias No Grandes (U.O.N.G.) 2DA. ETAPA.

2.- Estimación de la superficie cultivada entre las Unidades de Observación Agropecuarias Grandes (U.O.G.) - 2DA. ETAPA.

3.- Estimación del Total de la superficie cultivada de U.O.N.G. Y U.O.G.

4.- Estimación del total de superficie cultivada a nivel de la Primera Etapa (Segmentos).

DATOS CAPTURADOS EN EL OPERATIVO DE CAMPO EN LAS UNIDADES DE OBSERVACION NO GRANDES (2DA.ETAPA), (VARIABLE SUPERFICIE CULTIVADA - Has.) UNIVERSO: LLANOS DEL CHAC0110 (AGRICOLA) - CHUQUISACA.

N. Sg	Nh	nh	1 Unid.	2 Unid.	3 Unid.	4 Unid.	5 Unid.	6 Unid.	7 Unid.	8 Unid.	9 Unid.	10 Unid.	11 Unid.	12 Unid.	ΣY
1	40	5	0,01	0,05	0,08	0,15	0,50	0,65	1,00						2,44
2			0,30	0,30	0,7	0,8	0,8	1,25	1,5	1,2	1,11	0,06	0,8	0,7	9,52
3			0,05	0,05	0,6	0,6	0,9	0,9	0,75	0,77	0,66	0,7	0,77	0,82	7,57
4			0,1	0,2	0,4	0,6	1,4	1,04	0,08	0,9	0,7	0,77	0,9		7,09
5			0,6	0,6	0,7	0,9	0,1	0,1	0,06	0,88					3,94
1	30	4	0.77	0.08	0.01	0.02	0.01	0.33	0.44	0.56	0.31	0.4			2.93
2			0.33	0.44	0.56	0.77	0.8	0.9	0.9	0.36	0.7	1.6			7.36
3			1.3	1.0	0.9	0.7	0.03	0.03	0.20	0.2	0.2				4.56
4			0.01	0.02	0.3	0.4	0.5	0.3	0.3	0.2	0.02	0.01			2.06
1	11	11	0,5	0,6	0,3	0,7	0,6								2,7
2			1	1,5	0,8	0,7	0,3	0,8	1,2	1,8	2	3			13,1
3			2	3	3,1	0,5	0,6	0,3	0,1	0,1	0,2	0,3	0,3		10,5
4			3	3	2	1	0,9	0,7	0,6	0,3	0,02	0,3			11,82
5			1	1	1,3	3	2,5	1,6	0,6	0,3	0,02	0,1			11,6
6			1,6	2,8	3	0,6	0,3	0,2	0,6	0,4	0,5				10
7			2	1,8	1	1	0,3	0,2	0,2	0,1	0,1	0,11			6,81
8			3	1,5	1,25	0,6	0,5	0,3	0,6	0,7	0,8	0,8			10,05
9			0,05	0,02	0,4	0,6	1	0,8	0,1	0,1	0,3				3,37
10			0,1	0,8	1	1,1	2,01	3	3,01	2	0,5				13,52
11			0,3	0,2	2	2	0,5	0,6	0,7	0,02	0,3	0,8	1	2,2	10,62

DATOS CAPTURADOS EN EL OPERATIVO DE CAMPO EN LAS UNIDADES DE OBSERVACION GRANDES (2DA.ETAPA), (VARIABLE SUPERFICIE CULTIVADA - Has.) UNIVERSO : LLANOS DEL CHACO 110 (AGRICOLA) - CHUQUISACA.

N. Sg	Nh	nh	1 Unid.	2 Unid.	3 Unid.	4 Unid.	5 Unid.	6 Unid.	7 Unid.	8 Unid.	9 Unid.	10 Unid.	11 Unid.	12 Unid.	ΣY
1	40	5	4												4
2			-												-
3			-												-
4			-												-
5			5												9,5
1	30	4	4.6												4,6
2			-												-
3			-												-
4			-												-
1	11	11	5	4											9
2			-												-
3			4	4,2	5										13,2
4			-												-
5			-												-
6			-												-
7			5,1												5,1
8			-												-
9			-												-
10			4,3	4,1											8,4
11			-												-

ESTIMADOR DEL TOTAL (VARIABLE SUPERFICIE CULTIVADA- Has.)
EN LAS UNIDADES DE OBSERVACION (2DAETAPA)
UNIVERSO : LLANOS DEL CHACO - (AGRICOLA) - CHUQUISACA

N. Sg	Nh	nh	M´	m´	m´e	$\sum Y$	$\hat{Y}$ UONG	M´´	m´´	m´´e	$\sum Y$	$\hat{Y}$ UOG	$\hat{Y}_i$ = UONG + UOG
1	40	5	15	7	7	2,44	5,22	1	1	1	4	4	9,22
2			102	12	12	9,52	80,92	-				-	80,92
3			159	12	12	7,57	100,3	-				-	100,3
4			114	12	11	7,09	73,47	-				-	73,47
5			117	12	8	3,94	57,62	2	2	2	9,5	9,5	67,12
													331,03
1	30	4	74	10	10	2,93	21,68	1	1	1	4,6	4,6	26,28
2			98	10	10	7,36	72,12	-					72,12
3			101	12	9	4,56	51,17	-					51,17
4			105	12	10	2,06	21,63	-					21,63
													171,2
1	11	11	48	7	5	2,7	25,92	2	2	2	9	9	34,92
2			101	12	10	13,1	132,3	-					132,3
3			106	12	11	10,5	101,18	3	3	3	13,2	13,2	114,38
4			88	10	10	11,82	104,01	-					104,01
5			100	10	10	11,6	116	-					116
6			69	10	9	10	76,66	-					76,66
7			86	10	10	6,81	58,56	1	1	1	5,1	5,1	63,66
8			108	12	10	10,05	108,54	-					108,54
9			111	12	9	3,37	41,56	-					41,56
10			59	10	9	13,52	88,63	2	2	2	8,4	8,4	97,03
11			104	12	12	10,62	92,04	-					92,04
													981,1

Total Superf. Cultivada (Has.)

CALCULO DE LA VARIANZA DEL ESTIMADOR DE TOTAL.-

BAJO LA AFIJACION DE NEYMAN

$$\hat{V}(Y_{st}) = \sum_{h}^{L} W^2{}_h(1-f_h).\frac{\hat{s}^2{}_h}{n_h}$$

$$\hat{V}(\overline{Y}_{st}) = (0{,}49)^2\left(1-\frac{5}{40}\right)\frac{1169{,}97}{5} + (0{,}37)^2\left(1-\frac{4}{30}\right)\frac{550{,}26}{4} + (0{,}14)^2\left(1-\frac{11}{11}\right)989{,}8$$

$$\hat{V}(\overline{Y}_{st}) = 65{,}4$$

$$\sqrt{\hat{V}(\overline{Y}_{st})} = 8{,}09$$

$$\hat{V}(N\overline{Y}_{st}) = N^2\hat{V}(\overline{Y}_{st}) = (81)^2(65{,}49) = 429679{,}89$$

$$\sqrt{\hat{V}(N.\overline{Y}_{st})} = 655{,}49$$

ERROR DE MUESTREO PARA EL UNIVERSO LLANOS DEL CHACO (USO AGRICOLA), DEL DEPTO DE CHUQUISACA.

METODO MUESTREO ESTRATIFICADO:

Características (Has.)	Valor Estimado	Varianza Estimada	Error Estándar	Error Relativo	Intervalo (t=1,96) Lim. Inf.	Lim. Sup.
Sup. Total Cult.	4913,34	429679,89	655,49	0,13	3628,57	6198,1

En el plan de muestreo antes expuesto, se distinguen tres procedimientos fundamentales.

i).- La construcción de una jerarquía de unidades de muestreo y el empleo de un "muestreo por etapas".
ii).- La selección de una muestra de unidades de muestreo en cada una de las etapas.
iii).- El procedimiento propiamente dicho de la estimación.

Referencias.-

[1] Abad de Servin Adela, Servin Andrade Luis A., Introducción al muestreo, Editorial Limusa, México, 1978.
[2] Miras Julio, Elementos de Muestreo para Poblaciones finitas, Instituto Nacional de Estadística, Madrid, 1985.
[3] Cochran William G., Técnicas de Muestreo, Compañía Editorial Continental, S.A. México, 6ta. Impresión, 1976.
[4] Sánchez - Crespo J.L., Muestreo de Poblaciones Finitas aplicado al Diseño de Encuestas, Instituto Nacional de Estadística , 2da. Edición, Madrid, 1975.

Índice

ALGUNAS CONSIDERACIONES EN EL MUESTREO

Fernando Rivero Suguiura

Tamaño de Muestra Optimo

Un tamaño de muestra optimo para la ejecución de un encuesta, que cuenta con información de la varianza de las variables que se desean estudiar, así como el interés de estimar una proporción de tipo binomial bajo un esquema de selección aleatorio simple (MAS), fue dado por Cochran (1953), bajo el supuesto de una distribución normal en el parámetro, propuso un expresión para calcular el tamaño de muestra dado por:

$$n = \frac{z^2 pq}{\varepsilon^2}$$

que se utiliza con mucha frecuencia y ligereza en investigaciones por muestreo y en los trabajos de tesis de estudiantes de diferentes carreras universitarias.

Cuando no se cuenta con información de la varianza del parámetro de interés -lo cual sucede frecuentemente- la práctica aconseja asumir un valor de el cual maximiza la variabilidad del estimador y genera un tamaño de muestra que en apariencia garantiza la precisión deseada.

Esta forma de proceder llevaría a suponer que la encuesta sólo desea obtener estimaciones de una variable sin considerar que prácticamente todas las encuestas son de propósitos múltiples. Lo cual hace pensar que la variable de interés tenga más bien una distribución multinomial.

El procedimiento sugerido por Cochran puede subestimar el tamaño de muestra, pues no permite fijar en forma simultánea un coeficiente de confianza para todas las categorías en que se distribuye la variable de estudio lo cual inhibe al investigador sobre la posibilidad de controlar la precisión deseada de las estimaciones.

Para resolver este problema, se han propuesto algunos algoritmos de optimización no lineal que permiten calcular el tamaño de muestra óptimo (Kokan 1963 y Khan 1967).

Efecto de Diseño

Para resolver esta situación que se presenta en el cálculo del tamaño de muestra en encuestas complejas, donde no se cuenta con información sobre la varianza de las variables de interés, Kish (1979) p r o p o n e la definición de un factor de ajuste que a partir de una muestra aleatoria simple, permita aproximarse al número de selecciones necesarias para que un diseño de conglomerados proporcione la misma varianza. El factor se conoce como el efecto de diseño (deft) y está dado por:

$$deft = \frac{Var_C(\overline{Y})}{(1-f)S^2/n}$$

donde es la varianza de la variable de interés en el diseño de conglomerados y representa la varianza de un muestreo aleatorio simple, de tal forma que dicha varianza se puede escibir como:

$$(1-f)S^2/n$$

Según Kish, este factor permite evitar la complejidad en el diseño de una muestra compleja, sobre todo los de conglomerados o estatificación.

$$Var_C(\overline{Y}) = (1-f)S^2/n * deft$$

Es cotidiano que en un diseño de muestra estratificado, el efecto de diseño es menor a uno ya que los procedimientos tienden a reducir la varianza del estimador debido a la homogeneidad al interior de los grupos formados. Lo contrario sucede en el caso de un diseño de muestreo por conglomerados, donde generalmente dicho efecto de diseño es mayor a uno.

En caso de que los conglomerados sean en promedio de igual tamaño, el efecto de diseño se puede expresar como:

$deft = [1+\rho(M-1)]$ o $deft = [1+\rho(\overline{M}-1)]$

donde y representan el tamaño del conglomerado y el tamaño medio, respectivamente y es el denominado coeficiente de correlación intraconglomerado y se puede calcular por medio de:

$$\rho = \frac{\sum\sum(Y_{ij}-\overline{Y})(Y_{ik}-\overline{Y})}{m(\overline{M}-1)\overline{M}\sigma^2}$$

donde representa el número de conglomerados en la población, mientras que:

$$\sigma^2 = (M\overline{M}-1)S^2/M\overline{M}$$

Luego, la varianza del estimador se puede escirbir en términos del efecto de diseño como se muestra a continuación:

$$Var_C(\overline{Y}) = (1-f)S^2\left[1+(\overline{M}-1)\rho\right]nM$$

lo cual permite comparar la varianza de un diseño aleatorio simple sin reemplazo con la que se obtiene bajo un diseño de muestreo por conglomerados, lo que se interpreta como una medida de la eficiencia relativa del diseño.

Finalmente, el comentario afirma lo importante que significa la expresión efecto de diseño, pues permite relacionar un diseño complejo de muestreo con un muestreo aleatorio simple y expresar los tamaños de muestra y varianzas de los estimadores de tal forma que no requiera una construcción compleja, sino en términos de un muestreo aleatorio simple.

Índice

TECNICA PARA REDUCIR EL ERROR DE MUETREO

Jorge Troche Luna

Uno de los objetivos de los estudios por muestreo es reducir el intervalo de confianza del estimador de un parámetro . En el muestreo sistemático uno de los más utilizados, puede reducirse dicho intervalo por medio del re-ordenamiento de los elementos de la población adecuadamente con el fin de reducir específicamente la varianza. Para este propósito se considera que tiene que ver mucho la tendencia y comportamiento de la población, que se expresa mediante el orden del marco muestral. Sarndal, Swensson y Wretmam en el libro “Model Assisted Survey Sampling” indican textualmente: La eficiencia del muestreo sistemático depende grandemente del orden particular de los N elementos en el cual la selección sistemática es aplicada. Es decir esta en función de la población ordenada.

El MUESTREO SISTEMÁTICO ofrece un conjunto de procedimientos ventajosos, particularmente en la selección, donde solo un primer elemento es obtenido en forma aleatoria y con igual probabilidad, de los primeros a elementos en el listado de la población. El entero positivo a es fijo llamado intervalo muestral. El resto de la muestra se determina sistemáticamente tomando cada a-ésimo elemento hasta el fin del listado. Así hay solamente a posibles muestras, cada una tiene la misma probabilidad 1/a de ser seleccionada. Esta simplificación de solamente un ensayo aleatorio es una gran ventaja. Es decir la muestra esta conformada por:

$$s = \{k : k = r + (j-1)a \leq N;\ \ j = 1,2,\dots.,,n\}$$

Donde r es llamado arranque aleatorio.y k es la posición del elemento k-esimo

En general si a es un entero positivo, tenemos:

$$\boldsymbol{N = an+c}$$

Por tanto el tamaño de la muestra consideraremos siempre de tamaño n.

La varianza es: $V_{S\overline{Y}}(\hat{t}_\pi) = a(a-1)S_t^2$

donde $S_t^2 = \frac{1}{a-1}\sum_{r=1}^{a}\left(t_{S_r} - \bar{t}\right)^2$

es la varianza del total muestral. La varianza es pequeña si los totales muéstrales son aproximadamente iguales

Para el análisis de varianza, notemos que la sumatoria de la ecuación anterior, es una parte de la siguiente igualdad
Es decir, que la variación total en la población puede ser descompuesto en variación dentro de muestras sistemáticas, y la variación entre muestras sistemáticas, es decir un análisis de varianza. Esto es:

SST =SSW +SSB
SSTotal = SSDentro + SSEntre

donde SST representa la suma de cuadrados del total, SSW representa la suma de cuadrados dentro las muestra y SSB representa la suma de cuadrados entre muestras. Para una población dada, SST =(N - 1) S2yU es fija. Por tanto, un incremento en la variación dentro SSW es acompañado por un correspondiente decremento en la variación entre SSB. Además de una constante multiplicativa fija SSB determina la varianza (1) la cual puede ser escrita como

$$V_{SY}(\hat{t}_\pi) = N \cdot SSB$$

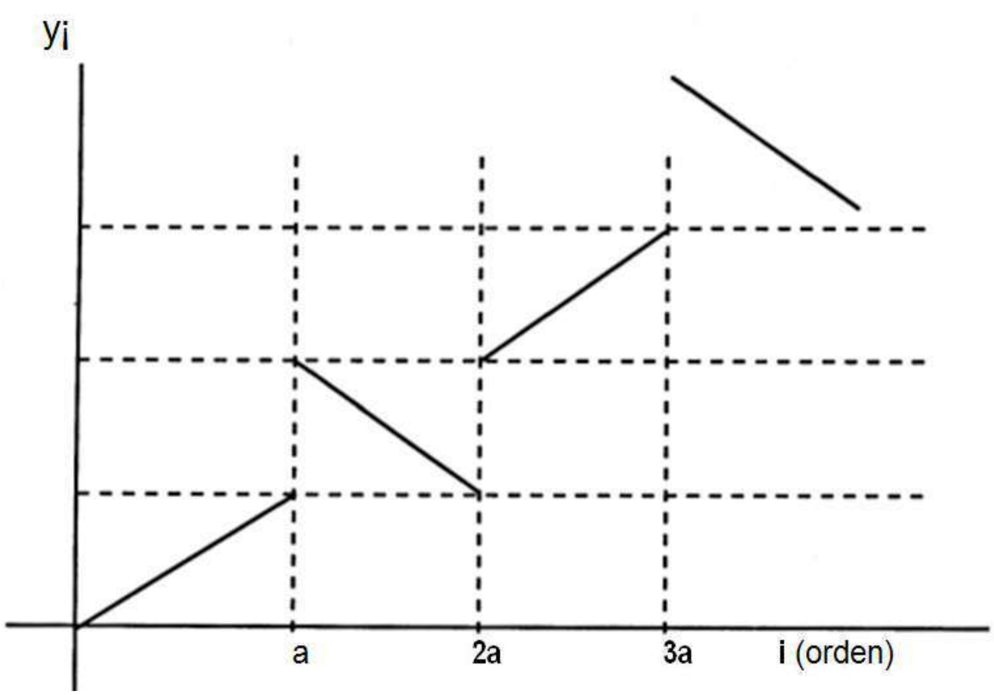
En otras palabras, a más homogeneidad dentro de las muestras es menos eficiente el muestreo sistemático. laa homogeneidad se usa aqui para conocer la tendencia para tener y-valores iguales. así para realizar un favorable ordenamiento de la población para muestras sistemáticas, podemos hacer lo posible, para un ordenamiento que vincula un bajo grado de homogeneidad alrededor de los elementos en la misma muestra sistemática.

La homogenidad mediremos mediante:

$$\rho = 1 - \frac{n}{n-1}\frac{SSW}{SST}$$

la cual es llmada el coeficiente de correlación intraclásica.

ORDEN PROPUESTO



Por ejemplo, consideremos un marco muestral ordenado ascendentemente de tamaño 15 y se desea obtener una muestra de tamaño 5, como sigue:

Número de la observación	Valor de la variable
1	13
2	14
3	17
4	18
5	19
6	20
7	20
8	21
9	22
10	23
11	24
12	25
13	26
14	26
15	27

Calculando tenemos que a (intervalo muestral) es ***a = N/n = 15/5 = 3***, dando como resultado el siguiente orden propuesto.

Número de la observación	Valor de la variable
1	13
2	14
3	17
4	18
5	19
6	20
7	20
8	21
9	22
10	23
11	24
12	25
13	26
14	26
15	27

ALGORITMO PARA EL ORDENAMIENTO DEL MARCO MUESTRAL

El algoritmo que se presenta esta con sintaxis computacional (en Fox Pro concretamente), aunque la adaptación del algoritmo a otros medios de almacenamiento de marcos muéstrales es similar.

```
sort to transí on Edad
use transí
CLEAR
INPUT 'INTRODUZCA  EL TAMAÑO
DE LA POBLACION:'  TO N
INPUT 'INTRODUZCA  EL TAMAÑO
DE LA MUESTRA:'  TO M
goto 1
ini=0
k=N/M
do while not eofQ
for I=ROUND(ini+1tO)  to
ROUND(ini+k, 0)
replace sis te with i
skíp
endfor
b=0
for 1= ROUND(ini+k+1,0)  to
ROUND(iní+2*k,0)
replace síste with ini+2*k-b
skíp
b=b+1
endfor
ini=ini+2*k
enddo
sort to sístem on síste
```

Índice

HUMOR ESTADÍSTICO

Cómo se ponen las notas

Carrera de Estadística: Se colocan los estudiantes por orden alfabético sobre una gráfica distribuida a lo largo de una Gaussiana.

Carrera de Estadística: Las notas son variables aleatorias.

Carrera de Religión: Dios pone las notas (inapelable)

Carrera de Informática: Se usa un generador de números aleatorios

Carrera de Historia: Cada Estudiante recibe la misma nota que el año anterior.

El 97.3 % de las estadísticas han sido claramente inventadas

HISTORIA DE LA ESTADÍSTICA

J. Anibal Angulo A.

I Introducción

Las decisiones que usted, los directores ejecutivos o el Rector toman a diario se basan en estadísticas.

Usted elige un detergente X para su hogar porque la publicidad le dice que blanquea mejor en comparación con otras marcas. El candidato a la elección cambia su campaña en vista de los resultados de una encuesta de opinión. El médico le prescribe un tratamiento por que las estadísticas le informan que el tratamiento cura su enfermedad en más de un 80 por ciento de los casos. Una muestra de sangre es suficiente para decidir miles da casos sobre ia totalidad de la sangre de su cuerpo.

¿Las estadísticas obtenidas a partir de una gota de sangre, permiten siempre tomar decisiones acertadas? La respuesta puede influir decisiones relativas a su vida. No se puede ignorar el uso generalizado de la Estadística en todas las actividades de nuestra sociedad.

Actualmente los gobiernos de cada país recolectan sistemáticamente datos relativos a su población, su economía, sus recursos naturales y su condición política y social para tomar decisiones. Lo mismo en las actividades industriales o comerciales las estadísticas son parte de la organización así como en los sectores agrícolas, donde se requiere predicciones de la producción. En la investigación científica el rol de la estadística es primordial.

II Historia del azar y del desarrollo de la estadística

El adelanto de los medios informáticos trastornó el progreso de la estadística y su enseñanza. Pretendemos ver aquí como y por quienes se desarrollo la estadística, desde la prehistoria hasta nuestros tiempos, es difícil separar la evolución de la Estadística sin considerar las probabilidades. El progreso de ambas disciplinas puede verse como la historia de una única ciencia: la ciencia del azar.

III La prehistoria

La Estadística descriptiva tiene su origen mil o dos años antes de Cristo, en Egipto, China y Mesopotamia, donde se hacían censos para la administración de los imperios. Los egipcios tuvieron el barómetro económico más antiguo: un instrumento llamado “Nilometro”, que medía el caudal del Nilo y servia a para definir un índice de fertilidad, a partir del cual se fijaba el monto de los impuestos. Con la variabilidad del clima ya conocían el concepto de incertidumbre. Paralelamente, el concepto de azar es tan antiguo como los juegos(los dados, los juegos con huesos que en Bolivia llamamos la taba) y motivó desde antaño las reflexiones de los filósofos. En las ideas de Aristóteles se encuentran tres tipos de nociones de probabilidad, que definen más bien actitudes frente al azar y la fortuna, que siguen vigentes hasta nuestros días: (1) el azar no existe y refleja nuestra ignorancia; (2) el azar proviene de causas múltiples y (3) el azar es divino y sobrenatural. Sin embargo, pasó mucho tiempo antes de que alguien intentara cuantificar el azar y sus efectos.

Durante la edad media hubo una gran actividad científica y artística en Oriente y el nombre de Azar parece haber venido desde Siria a Europa. La flor del Azahar, que aparecía con los dados de la época podría ser el origen de la palabra. Las compañías aseguradoras hincaron investigaciones matemáticas desde tiempos antiguos, y en el siglo XVII parecieron los primeros famosos problemas de juegos de azar. En la sociedad francesa, el j u e g o era uno de los entretenimientos más frecuentes. Los juegos cada vez más complicados y las apuestas muy elevadas hicieron sentir la necesidad de calcular las probabilidades de los juegos de manera racional. El caballero de Méré, un jugador apasionado, escribiendo a Blas Pascal sobre ciertos juegos de azar, dio origen a una correspondencia entre algunos matemáticos de la época.

Las preguntas de Méré permitieron, en particular, iniciar una discusión entre Pascal y Pierre Fermat así el desarrollo de la teoría de las probabilidades. En el siglo anterior, los Italianos Tartaglia, Cardano, e incluso el gran Galileo abordaron algunos problemas numéricos de combinaciones de dados.

En cada juego de azar, dados, cartas o ruleta, por ejemplo, cada una de las jugadas debe dar un resultado tomado de un conjunto finito de posibilidades(números de 1 a 6 dado, 52 posibilidades para las cartas o 38 para la ruleta. Si el juego de azar es “correcto”, no se puede predecir de antemano el resultado que se obtendrá en una jugada. Es lo que define el azar del juego. Se observa una cierta simetría en los posibles resultados: son todos igualmente posibles, es decir que el riesgo para un jugador es el mismo cualquiera sea el que juega. De aquí surgió la primera definición de una medida de probabilidad para un determinado suceso:

$$p = \frac{a}{b}$$

donde a es el número de casos favorables (el número de casos que producen el suceso) y b el número de casos posibles. Por ejemplo, la probabilidad de sacar un 6 en el lanzamiento de un dado es 1/6 , de sacar un corazón de un conjunto de 52 cartas es:

$$p = \frac{13}{52} = \frac{1}{4}$$

El Caballero de Méré, que jugaba con frecuencia, había acumulado observaciones en diversos juegos y constató una cierta regularidad en los resultados. Esta regularidad, a pesar de tener su base de un hecho empírico, permitió relacionar la frecuencia relativa de la ocurrencia de sucesos y su probabilidad. S i f es la frecuencia absoluta de un suceso(el número de veces que ocurrió) en n jugadas, como el número de casos favorables debería ser aproximadamente igual a una, $f \approx \frac{an}{b}$

y entonces la probabilidad de que ocurra el suceso será:

$$p = \frac{a}{b} = \frac{f}{n}$$

En un juego, Méré encontraba una contradicción en su interpretación de la probabilidad a partir de la frecuencia relativa que obtuvo empíricamente Pascal y Fermat pudieron mostrarle que sus cálculos eran erróneos y que la interpretación propuesta era correcta. De Méré siguió planteando problemas que no pudieron resolver los matemáticos de su época, sin embargo Jacques Bernoulli, el primero de una famosa familia de matemáticos suizos dio, una demostración de la ley de los Grandes Números y Abraham de Moivre enuncio el teorema de la regla de la multiplicación de la teoría de la probabilidad.

Según Richard Epstein, la ruleta es el juego de casino más antiguo que está todavía en vigencia. No se sabe a quién atribuirlo: puede ser pascal, el matemático Italiano, la primera ruleta se introduce en París 1765.

IV La demografía

Las reglas del cálculo desarrolladas hasta entonces para los juegos de azar vieron sus aplicaciones en otras disciplinas. Los censos demográficos, que se hacían desde la antigüedad, requieren recolectar muchos datos. La demografía y los seguros de vida aprovecharon del desarrollo de la teoría de las probabilidades. Consideremos , por ejemplo, el sexo de una sucesión de niños recién nacidos. Se puede ver como una repetición de lanzamientos de una moneda, con niño y niña en vez de cara y sello. De la misma manera, podemos considerar un conjunto de hombres mayores de 50 años. Al final del año, una cierta proporción sigue viva.

Durante el siglo XVIII Pierre Simón, Marqués de Laplace, estos problemas fueron reconocidos como similares a los de un juego, y se encontraron las correspondientes frecuencias relativas, lo que permitió determinar la probabilidad que nazca una niña, o que un hombre mayor que 50 años 3 muera en el año.

Si bien la extensión de los juegos de azar a la demografía o a la matemática actuarial fue extremadamente importante, su planteamiento tiene grandes limitaciones debido a que considera todos los resultados posibles simétricos. ¿Qué pasa cuando una situación real puede expresarse como un juego de azar? Por ejemplo, Daniel Bernoulli, careciendo de datos sobre la mortalidad producida por la viruela a distintas edades, supuso que el riesgo de morir de la enfermedad era el mismo a toda edad. Lo que evidentemente es muy discutible.

V La teoría de los errores

Durante los siglos XVIII y XIX la estadística se expandió sin interrupción mientras la teoría de las probabilidades no mostró progreso. Una de las aplicaciones importantes fue desarrollada al mismo tiempo por Gauss, Legendre y Laplace: el análisis numérico de los errores de mediciones de física y astronomía. ¿ Cómo determinar el mejor valor leído por un instrumento que entrega diferentes mediciones del mismo fenómeno? Si tenemos $\{x_1, x_2, ..., x_n\}$ n mediciones de un mismo fenómeno deberíamos tener

$$x_1 = x_2 = ... = x_n$$

si no hubiera errores. En su anexo sobre el método de los mínimos cuadrados, de "Nuevo método para la determinación de las órbitas de los cometas", Legendre propone determinar el valor único Z de la medición de manera que una función de los errores

$\varepsilon_j = x_j - z$ sea mínima:

$$Min \sum_{j=1}^{N} (x_i - z)^2$$

la solución es el promedio de las mediciones.

Esta función cuadrática encuentra su justificación en la distribución normal con Gauss y Laplace, aunque la distribución de los errores fue estudiada mucho antes por Thomas Simpson, que hizo los supuestos que esta distribución tenía que ser simétrica y que la probabilidad de errores pequeños debería ser más grande que la de los errores grandes. Sir Francis Galton partió de una distribución discreta y la fue refinando hasta llegar a una distribución continua muy parecida a la distribución normal , Galton trabajo en meteorología y en herencia. Era el primo de Charles Darwin.

VI Nacimiento de la Estadística Moderna

Es con la introducción de nuevas aplicaciones que la teoría de las probabilidades del siglo XVIII funda la Estadística Matemática. El término de Estadística se debe posiblemente a G. Achewall, profesor de la Universidad de Góttingen, tomando del latín la palabra Status.

Aparte de la demografía y la matemática actuarial, otras disciplinas introdujeron la teoría de las probabilidades. Fue el inicio de la mecánica estadística, debido a Maxwell y Bolzmann, quienes dieron también una justificación de la distribución normal en la teoría cinética de los gases.

La estadística se empezó a usar de una manera u otra en todas las disciplinas, a pesar de un estancamiento de la teoría de las probabilidades. En particular, muchos vieron la dificultad de aplicar el concepto de simetría, o de casos igualmente posibles, en todas las aplicaciones. Hubo que esperar a que Andrey Nickolacvich Kolmogorov separara la determinación de los valores de las probabilidades de sus reglas de cálculo.

Los primeros resultados importantes de la Estadística Matemática se deben al Ingles Karl Pearson y a otros investigadores de la escuela biométrica Inglesa.

VII La segunda mitad del siglo XX: la revolución computacional.

Los científicos, especialmente los ingleses, desarrollaron métodos matemáticos para la estadística, pero en la práctica manipularon cifras durante medio siglo sin disponer de verdaderas herramientas de calculo. La llegada de los computadores revolucionó el desarrollo de la Estadística. En Francia (Benzécri) y en los Estados Unidos (Tuckey) fueron los pioneros en repensar la estadística en función de los computadores. Mejoraron, adaptaron y crearon nuevos instrumentos para estudiar grandes volúmenes de datos: nuevas técnicas y herramientas gráficas.

VIII Calculo de probabilidades y Estadística

Algunas palabras para concluir Sí bien la historia de la estadística no se puede separar de la historia del Calculo de las Probabilidades, la Estadística no puede considerarse como una simple aplicación del Calculo de las Probabilidades. Podemos comparar esta situación a la de la geometría y la Mecánica. La mecánica usa conceptos de la geometría, y sin embargo es una ciencia aparte.

Él calculo de las probabilidades es una teoría matemática y la Estadística es una ciencia aplicada donde hay que dar un contenido concreto a la noción de probabilidad. Como ilustración citemos el experimento de Weldon, que lanzo 315672 veces un dado (bajo la supervisión de un Juez) y anotó que 106602 veces salió un 5 o un 6. La frecuencia teórica debería ser 0,3333 si el dado hubiera sido perfectamente equilibrado. La frecuencia observada aquí fue 0.3377. ¿Deberíamos concluir que el dado estaba cargado? Es una pregunta concreta que es razonable considerar. El Calculo de las probabilidades no responde a esta pregunta y es la estadística la que permite hacerlo.

Bibliografía

Principios de Calculo de Probabilidad

Autor Pedro Diez

Editorial Pedro Diez Diez España

Probabilidades Estadística y Muestreo

Autor Trocomz , A, FZ

Editorial Tebar Flores

Historia de la Estadística

José Guillermo Molina

Has oído se chiste de estadística?

Probablemente

Valor es seguro

En realidad, volar en avión es muy seguro. La práctica totalidad de los fallecidos en accidentes aéreos han muerto al llegar al suelo.

LA DISTRIBUCIÓN DE LA POBLACIÓN BOLIVIA EN 311 MUNICIPIOS

Jaime Chumacero

El movimiento de la población tiene una marcada influencia y en grado variado, de las fuerzas sociales, económicas y políticas tanto en las áreas de origen como en las de destino y de los medios de comunicación y transporte. En las decisiones para cambiar de residencia de un municipio a otro , son también significativas las características personales de individuo tales como la edad, el sexo, el estado familiar, la salud y la ocupación. El movimiento de la población afecta obviamente en cualquier tiempo, la concentración o la dispersión de una población dentro de determinadas fronteras. Los aspectos referidos a la migración son los siguientes: 1) la concentración o dispersión de la población dentro de las fronteras nacionales según lo observado en un censo; 2) la migración interna, o el movimiento de personas dentro de una nación debido al cambio de residencia: y 3) la migración internacional, o movimiento de personas a través de las fronteras nacionales.

En Bolivia a partir de la promulgación de la Ley de Participación Popular (1994-1996), se han generado unidades administrativas gubernamentales regionales en todo el ámbito nacional, denominados municipios, en total se han creado 311 secciones municipales.

Sin embargo se ha constatado desigualdades extremas entre estos 311 municipios, algunos de los cuales presentan cuadros de extrema pobreza, los mismos que presentan carencias de capacidad humana básica, con inaccesibilidad de gran parte de la población a los servicios de educación y de salud, con escasa cobertura de servicios básicos, con inseguridad alimentaria, aislamiento en toma de decisiones nacionales, desempleo, injusticia, discriminación, deterioro del medio ambiente y dificultades en sus sistemas productivos.

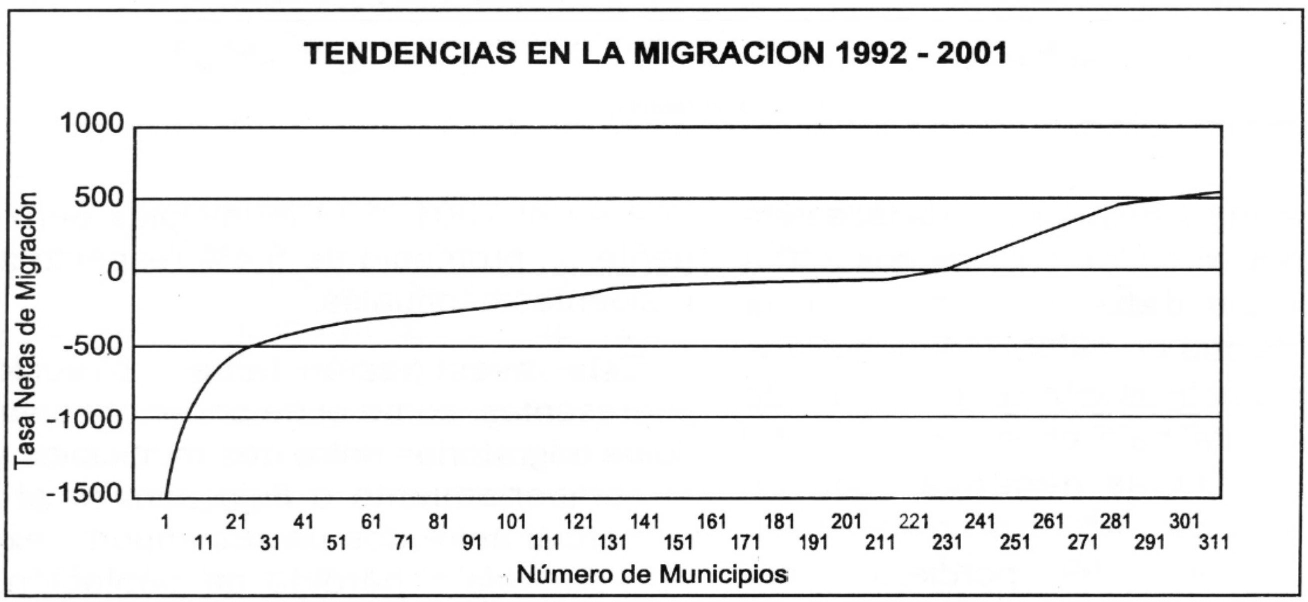
Con el levantamiento de datos realizados con un intervalo de menos de 10 años, por los censos de población y vivienda de 1992 y 2001, podemos contar con información con cobertura total, a nivel de los 311 municipios, si bien actualmente son 314 municipios, en toda Bolivia, para fines comparativos entre ambos censos, hemos tomado en cuenta tan solo a los 311 municipios originales.

Un hecho importante para la realización de esta investigación es el de haber municipalizado por primera vez los códigos referidos a las preguntas de lugar de residencia habitual, lugar de nacimiento y lugar de procedencia de hace 5 años atrás, en el censo de 1992, ya que como se sabe, dicho censo se realizó previa a la creación de las secciones municipales, y esta, tan solo fue municipalizado en el código referente a lugar donde uno fue censado. En el censo del 2001, se cuenta con los códigos ya municipalizados. El error de cobertura máxima en las respuestas de estas tres variables se tiene para el 92 y 2001 de 5.13%, 5.5% respectivamente.

Observemos algunos resultados importantes de la investigación: En el Censo del 92, se detectó a 1’866.923, de personas que migraron del lugar de nacimiento, mientras que en el 2001, se contabilizarón a 2’576,971, existiendo una diferencia de 710,048 o un aumento en un 38% de personas que migraron de sus lugares de nacimiento, la misma equivale, a que anualmente 78,894 nuevas personas salieron de sus lugares de nacimiento con destino a otros municipios.

El siguiente gráfico muestra las tendencias de las Tasas Netas de Migración (inmigrantes menos emigrantes sobre total de residentes habituales) en 1992 y 2001, la misma que muestra un comportamiento similar de migración respecto al lugar de nacimiento, con una cierta ligera disminución de la tasa neta negativa en el 2001 respecto a 1992.

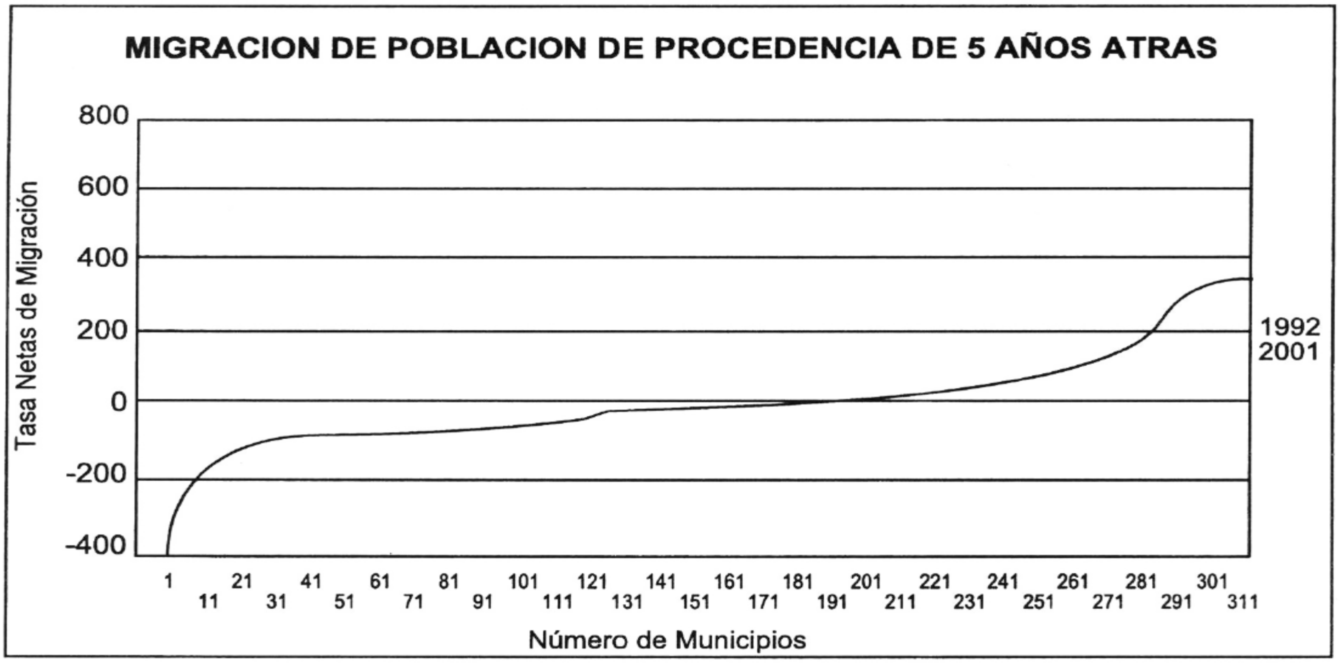
Gráfica 1



Realicemos ahora el análisis correspondiente al flujo migratorio de lugar de procedencia de hace 5 años atrás, respecto a 1992 y 2001; 664425 personas migraron del lugar en la que en 1987 vivían, mientras que 883614 personas migraron del lugar en la que en 1996 vivían, esto quiere decir que la migración reciente se incrementó en 219189, más personas migrantes que en 1992. El error de cobertura para cada uno de los períodos de referencia es de 1,43% para 1992 y 3,42% para el 2001, (este último de manera preliminar).

La siguiente gráfica muestra las tendencias de migración para el lugar de procedencia de 5 años atrás a 1992 y 2001.

Gráfica 2



Se observa una disminución en la tasa neta negativa de aproximadamente menos 300 a menos 200. Un dato i m p o r t a n t e es la observación de que en 1992, 214 municipios tenían tasas negativas con un promedio de menos 5.5%, esta bajó en el 2001 a 199 municipios con tasas negativas con un promedio de menos 5.4%, esto quiere decir que 214 municipios en 1992, perdieron gente en un promedio de 5.5% respecto a su población habitualmente residente, mientras que en el 2001, 199 municipios perdieron gente en promedio de 5.4% respecto a sus residentes habituales.

Esta investigación tiene proyecciones interesantes, como el de poder observar los flujos migratorios entre dos municipios y ver el comportamiento o flujo (similar al flujo comercial entre dos países import - export) de ganancia o pérdida de población, los mismos que en próximos números serán dados a conocer.

Secuestros aéreos

Un hombre tenía miedo de coger un avión por aquellos de los secuestros aéreos. Mirando unas estadísticas, encontró que la probabilidad de que hubiese una bomba en su vuelo era de 1 entre 1000, mientras que la probabilidad de que hubiese dos era 1 entre 100.000. por lo tanto, lo que hizo fue tomar el avión llevando el mismo una bomba.

Índice

LA PERSPECTIVA DE LA ESTADISTICA PARA EL CONOCIMIENTO DE LA REALIDAD NACIONAL

Augusto S. Soliz Sánchez

El propósito de este breve artículo es destacar la importancia de la estadística como instrumento para conocerla realidad nacional. En ese sentido, es conveniente recordar la definición de estadística descriptiva de Hines y Montgomery1 que textualmente dice lo siguiente: "La estadística descriptiva es la rama de la estadística que trata de la organización, el resumen y la presentación de datos"

El hecho de que me refiera a la estadística descriptiva, en principio, no significa que la estadística inferencial o la no paramétrica sean menos importantes. Lo que sucede es que, en mi opinión, el grado de desarrollo de la estadística en nuestro país, con las salvedades del caso, aún se encuentra enfrentando el problema de la organización, el resumen y la presentación de datos.

En este artículo se hace un esbozo del potencial que tiene la estadística como instrumento singular para el conocimiento de la realidad nacional y, por lo tanto, para contribuir a la resolución de los grandes problemas que afectan al país en su conjunto.

Para tal fin, en primer lugar, voy a presentar una breve descripción de la situación actual del Sistema Nacional de Información Estadística a cargo del INE y, en segundo lugar, un conjunto de sugerencias que podrían ayudar a superar la problemática situación actual de la información estadística.

La situación actual de la información estadística

Para abordar el examen de la situación de la información en el país, utilizaré dos puntos de vista: primero, con relación a la cobertura de la información estadística y, segundo, con relación a la calidad de los datos.

• Cobertura de la información

En este punto debo distinguir dos clases de cobertura de la información. A la primera la llamaré cobertura temática y a la otra, quiero decir a la segunda, la llamaré cobertura de la información recolectada. En relación con la cobertura temática se puede señalar que el problema principal es que no se produce la información estadística de acuerdo a las necesidades nacionales del país. Para dar ejemplos de este problema voy a utilizar algunos temas del campo social aunque debo señalar que los ejemplos serían similares en otras áreas. En este sentido, un ejemplo de un tema ignorado en la investigación estadística es la frecuencia del servicio de agua por cañería en los centros urbanos. En determinadas zonas de algunas ciudades el servicio de agua está disponible sólo por algunas horas.

Igualmente, no se produce información estadística del movimiento de la población (estadísticas vitales) sobre la base de registros administrativos

En lo que se refiere a la cobertura de la información puedo señalar dos problemas serios: primero, la cobertura incompleta de la información a niveles subnacionales. Esto quiere decir que cierta información sólo existe en el ámbito nacional y no existen datos a nivel departamental, provincial, municipal y menos local. Segundo, el problema de la cobertura temporal. Esto significa que la información que se produce, en general, no tiene una periodicidad regular definida ni la continuidad deseable en el tiempo. Aquí, mi ejemplo favorito es la irregularidad de los intervalos íntercensales. Aunque el Censo de 1900 fue llamado Primer Censo Decenal, en el siglo XX se han realizado solamente cuatro censos nacionales de población a intervalos irregulares: el primero en 1900, el segundo en 1950, el tercero en 1976 y el último en 1992. El Censo Nacional de Población y Vivienda 2001 corresponde al siglo XXI. Como ejemplo similar se puede mencionar a las tres encuestas nacionales de empleo que se realizaron en 1996 y 1997 y, luego, el tema de empleo, pese a su enorme importancia, quedó olvidado.

Por consiguiente, a manera de resumen, se puede concluir que la Información estadística producida en el país tiene problemas de cobertura temática, porque no se produce toda la información en temas de interés nacional; tiene problemas de cobertura espacial o geográfica, porque no existe información desagregada de acuerdo a la división político-administrativa del país y, finalmente, tiene problemas de cobertura temporal porque la información no se produce con una periodicidad definida ni con la continuidad deseable en el tiempo.

Calidad de los datos

El problema de la calidad de los datos es, en cierta medida, tanto o más grave que el problema de la cobertura de la información. En primer lugar, existen casos en que para un mismo indicador existen dos cifras diferentes para el mismo período de referencia. Este es un problema tan serio que en el Censo de 1992 se han publicado dos cifras diferentes del número de habitantes en los departamentos de Potosí y Santa Cruz. Una cifra fue publicada en el volumen nacional y otra distinta en los volúmenes departamentales correspondientes a los citados departamentos. Un segundo problema que afecta a la calidad de los datos es la falta de oportunidad en la presentación de la información. Por las dificultades que aparentemente existen en el procesamiento de los datos y las restricciones que se imponen a la divulgación de la información estadística, la mayoría de las veces el dato estadístico se convierte en una mera referencia histórica, poco útil para la toma de decisiones o definición de políticas públicas.

Algunas causas de los problemas

Existe una diversidad de causas que se podrían señalar como condicionantes de la situación actual de la información estadística en el país. Sin embargo, para no extenderme mucho en este punto, sólo señalaré las causas que, en mi opinión, son las principales. En primer lugar, se puede señalar que, en general, no existe el hábito o la costumbre de proporcionar, producir y utilizar información estadística, tanto en la propia comunidad como en las entidades privadas y públicas. De aquí se desprende un segundo condicionante que es la escasa prioridad que los gobiernos dan a la producción de información estadística oportuna y confiable. Los gobiernos se interesan en la estadística sólo para mostrar sus logros y, en algunos casos, despiertan sospechas por presentar indicadores que parecen no reflejar adecuadamente la realidad.

En tercer lugar, por la poca costumbre de producir y utilizar datos estadísticos, no existe una adecuada relación entre productores y usuarios de información estadística. Un inventario de productores y usuarios de información estadística, que se hizo en el INE a mediados de los años setenta, hasta donde tengo conocimiento, no ha sido actualizado. A todo lo señalado hay que añadir la falta o la escasez relativa de recursos humanos calificados en el campo de la estadística, la ausencia de coordinación a nivel del sector público del Sistema Nacional de Información Estadística y el desconocimiento de las disposiciones legales con relación a la producción de información estadística.

La perspectiva de la estadística

Ante el desalentador panorama descrito, ¿Cómo podría enfocarse la perspectiva de la estadística como instrumento útil para conocer la realidad nacional?

Desde mi punto de vista, la estadística constituye el instrumento privilegiado en el campo científico para el conocimiento de la realidad nacional. En primer lugar, el extraordinario avance tecnológico que se ha producido en el campo de la informática y las comunicaciones hace que el procesamiento electrónico de datos y la transmisión de información se realicen con una rapidez y eficiencia no conocidas antes. Antes, me refiero a fines de los años setenta, el procesamiento electrónico de datos en el sector público era monopolizado por CENACO, el Centro Nacional de Computación. El Censo Nacional de Población y Vivienda de 1976 fue procesado por CENACO y cuando se requerían tabulaciones especiales de los datos censales era necesario prever un tiempo de espera de semanas y, a veces, de varios meses. En la actualidad, ya no existe el monopolio de CENACO y el uso de computadores ha crecido de manera extraordinaria y el avance tecnológico es tal que la vida útil de los equipos de procesamiento electrónico de datos se ha reducido a pocos años y, en ciertos casos, a meses. Actualmente, los equipos de computación son de mayor capacidad para almacenar y procesar datos, por una parte, y, por otra, cada vez tienen mayor velocidad de procesamiento. Las unidades para medir la capacidad de memoria fija han pasado de kilobytes a megabytes, luego a gigabytes y no es ninguna sorpresa que ya se esté hablando de tirabytes. La velocidad de los computadores personales de hace 15 años era de 33 megahertz y actualmente es común encontrar máquinas con más de 800 megahertz de velocidad.

En segundo lugar, aunque quizá no en la medida adecuada, se ha producido un aumento en la oferta de recursos humanos calificados en el campo de la estadística. Existen varias universidades públicas que, como la Universidad Mayor de San Andrés, forman profesionales en el campo de la estadística. Sin embargo, en mi opinión, estos profesionales aun no han alcanzado un grado de aceptación plena en el mercado de trabajo. No voy a intentar explicar porqué sucede este fenómeno pero creo que sí existe y muchos estarían de acuerdo con mi opinión. Por ahora, sólo quiero señalar que sí existen recursos humanos calificados y hasta podríamos decir sobre calificados para el grado de desarrollo en que se encuentra la producción de información estadística en el país.

Entonces, si existe capital, incluida la tecnología, y trabajo, que son los recursos humanos calificados, la perspectiva de la estadística para el conocimiento de la realidad nacional tiene un enorme potencial que, por diversas razones, aun no lo hemos considerado.

Algunas sugerencias para mejorar la formación profesional en estadística

Estas sugerencias tocan algunos aspectos de orden administrativo específicamente de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés y puede que no se apliquen a otras universidades. Aun así, el propósito de esta sección es presentar sugerencias de algunas acciones para destacar la gran utilidad de la ciencia estadística y que podrían tomarse en el ámbito de la Carrera de Estadística.

En primer lugar, es necesario llevar a la práctica lo que se enseña en la cátedra. En la Carrera de Estadística se debería elaborar estadísticas de la educación de nivel superior con participación de los estudiantes. Me refiero a datos sobre estudiantes (matrícula, asistencia, rendimiento académico), recursos humanos e infraestructura. El paso inicial puede ser la información de la propia Carrera, luego de la Facultad de Ciencia Puras y Naturales de la U.M.S.A. y, soñando un poco, ¿Porqué no de todo el nivel de educación superior del departamento de La Paz?

En segundo lugar, en mi opinión, es necesario incorporar nuevas áreas del conocimiento en la formación profesional de los estadísticos. Esta idea no es nueva pero creo que se debe insistir en ella. Un profesional o una profesional en estadística que tenga un conocimiento sólido del mercado de trabajo estará mejor capacitado o mejor capacitada para el diseño de una muestra para una encuesta orientada a la medición del empleo. Igualmente, un profesional o una profesional en estadística, con ideas claras sobre las variables demográficas, podría trabajar de manera idónea en las encuestas demográficas.

Conclusión

Sin abundaren más consideraciones quiero concluir este artículo señalando que yo veo un potencial verdaderamente atractivo para la ciencia estadística en Bolivia como un instrumento singular para conocer la realidad nacional. Y cuando menciono la realidad nacional me refiero a los complejos problemas que enfrentamos en materia de empleo, salud, educación, vivienda, producción, comercio exterior y las condiciones de pobreza. Se trata de mejorar la producción de información estadística de acuerdo a las necesidades nacionales; se trata de organizar, resumir y presentar adecuadamente los datos disponibles; se trata de proporcionar la información estadística oportuna y confiable. Los y las profesionales nacionales en estadística pueden hacerlo, ¿Verdad?

1 Hiñes, William W. Y Montgomery, Douglas C., Probabilidad y estadística para ingeniería y administración, Compañía Editorial Continental, S.A. de C.V., México, D.F., 1994, p. 3

LA ESTADISTICA EN LA EPOCA DE GLOBALIZACION

Juan Carlos Flores López

La globalización no es ni más ni menos que la extensión del capitalismo a escala global. El planteamiento de la “Globalización” como una fase expansiva del desarrollo del capital, nos lleva a definir su relación con el trabajo, es decir con la fuente del valor.

La crisis de esta etapa globalizante, a su vez contiene un fuerte efecto sobre el trabajo a todo nivel. Tomando en cuenta los últimos acontecimientos económicos, en el ámbito mundial, para nadie son desconocidos en la actualidad los cambios que sufren todos los países del mundo, en el sentido económico, social, académico, etc. en ese sentido cabe hacer la siguiente reflexión respecto al desenvolvimiento que debe tener un profesional en el área de Estadística. Todo profesional Estadístico esta obligado a tener un dominio de utilización de las herramientas que ofrece la tecnología, para trabajos con verdadera credibilidad. Los aspectos vinculados con el trabajo concreto, experimentan una gran variabilidad que se representa en las cadenas de producción, el sistema de racionalización de la producción y el aumento de la productividad a través de un reforzamiento de las tasas de trabajo, pero todo esto se realiza con el apoyo de la manipulación de información estadística la cual lleva a tomar decisiones correctas, en el cual debe intervenir un profesional especializado en Estadística.

La globalización es un proceso con una doble vertiente: extensiva e intensiva; por un lado, abarcar potencialmente todo el espacio físico planetario y por otro, afectar a todas las áreas de la actividad humana. La globalización es, “nos dicen sus defensores”, la creación de un espacio mundial de intercambio económico, productivo, financiero, político, ideológico y cultural (académico).

El término “globalización” hoy tan de moda, se aplica en múltiples sentidos. Por un lado, se utiliza para reflejar la consideración del mundo como un gran hipermercado global en el cual se producen, se adquieren y se comercializan productos en cualquier parte del planeta . Sin embargo , la palabr a globalización no se usa sólo referida a la globalización económica o financiera, sino que abarca muchos más aspectos. Se trata de un proceso que parece integrar o englobar todas las actividades de nuestro planeta, tanto las actividades económicas, como las actividades sociales, culturales, laborales, tecnológicas, ambientales, académicas, etc.

La globalización entraña una interdependencia de las sociedades, parece como si las fronteras geográficas, materiales y espaciales del planeta desaparecieran.

Por lo anteriormente expuesto, sé esta conciente que la manipulación de información a todo nivel esta al orden del día, eso significa por implicación que todo profesional Estadístico debe ofrecer un conocimiento de manejo de técnicas (herramientas), sofisticadas para resolver problemas que piden inmediata solución.

Índice

ALGUNOS PROBLEMAS EN PROBABILIDAD

J. Anibal Angulo A.

1. LOS INCONVENIENTES DE SER DESPISTADO.

La siguiente historia de la casualidad de que es cierta:

Es bien sabido que en cualquier grupo de al menos 23 personas, la probabilidad de que al menos dos de ellas cumplan años el mismo día es mayor del 50 %. Bien, en cierta ocasión un profesor estaba dando clase de probabilidad I en la carrera de estadística, de la UMSA, y estaba explicando la teoría elemental de probabilidad. Explicó a la clase que con 30 personas en lugar de 23, la probabilidad de que al menos dos de ellos cumpliesen años el mismo día sería muchísimo mayor.

Profesor: "Como en esta clase sólo hay diecinueve estudiantes, la probabilidad de que dos de ustedes cumpláis años el mismo día es mucho menor del 50%". En ese momento uno de los alumnos levantó la mano y dijo:

Alumno: "Le apuesto que al menos dos de lo que estamos aquí cumplen años el mismo día".

Profesor: "No estaría bien que aceptes la apuesta, porque las probabilidades estarían claramente a mi favor".

Alumno: "No me importa, ¡se lo apuesto de todas maneras!".

Profesor: De acuerdo".

El profesor aceptó la apuesta, pensando en dar al alumno una buena lección. Procedió a llamar uno a uno a los estudiantes para que dijeran el día de su cumpleaños hasta que, cuando iban por la mitad, tanto la clase como el profesor estallaron en carcajadas motivadas por el despiste del profesor.

El joven que con tanta seguridad había hecho la apuesta no sabía el día de nacimiento de ninguno de los presentes, excepto el suyo propio. ¿ Sabes por qué se mostraba tan seguro?.

EL MISMO N° DE PELOS. La densidad máxima de cabellos del cuero cabelludo humano es de 5 por mm2(generalmente es menor). Teniendo en cuenta que el número de habitantes de una nación es de 40 millones, ¿Cuál es la probabilidad de que dos habitantes de este país, al menos, tengan el mismo número de pelos en la cabeza?.

LAS HERMANAS DE LOS OJOS AZULES.

Sí nos encontramos con dos de las hermanas Flores (lo que presupone que las dos anteriores sean extracciones al azar del conjunto de las hermanas Flores), hay un caso favorable en cada dos de que ambas chicas tengan los ojos azules. ¿Cuál es la predicción más razonable acerca del número de hermanas Flores que tienen los ojos azules?.

Índice

UMSA
FCPN
CARRERA
ESTADÍSTICA

IETA

Instituto de Estadística
Teórica y Aplicada

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

La Paz - Bolivia