

**FACULTAD DE INGENIERIA U.N.A.M.
DIVISION DE EDUCACION CONTINUA**

CURSOS INSTITUCIONALES

MODULO BÁSICO DE TECNOLOGÍA DE LA INFORMACIÓN Y BIOESTADÍSTICA

Análisis Estadístico

Del 19 de Septiembre al 15 de Octubre del 2003

APUNTES GENERALES

CI - 121

Instructor: Ing. Rodolfo González Maldonado

ISSSTE

SEPTIEMBRE DEL 2003

INDICE

Introducción	1
Estimación	1
Estimación por intervalos	2
Prueba de hipótesis, Análisis de regresión	3
Regresión en Excel	4
Técnicas de Muestreo	7
Muestreo aleatorio simple	7
Muestreo estratifico	7
Muestreo conglomerados	7
Muestreo sistemático	7
Muestreo por áreas	7
Definición, Construcción y Análisis de Indicadores	8
Tipos de preguntas	10
Aplicación del concepto "Tablero de Control"	13

I. INTRODUCCION

El estudio de indicadores, a menudo refleja que existe una correlación entre ellos, por lo que es interesante el estudio de las variables y las formas como se relacionan. Adicionalmente y como parte de este curso, se estudiará la forma de realizar los bancos de datos a través de muestreos.

II. ESTIMACION

Para calcular el tamaño de la muestra, es necesario conocer si la población tiene un tamaño conocido o no. en el primer caso, es decir, cuando se conoce el tamaño de la población, el cálculo se hace de la siguiente manera:

Se determina el error máximo que puede aceptarse en los resultados, que por lo general es del 10%, ya que variaciones mayores hacen dudar de la validez de la información.

Se conoce en el mercado la situación que guarda la característica o el fenómeno investigado.

Cuando no se conoce, o no se tiene una clara idea de la situación del mercado, se dan los valores máximos a la probabilidad de que ocurra el evento y a la probabilidad de que no ocurra (0.50 y 0.50) Estos valores se asignan a las literales p y q.

$$n = \frac{1.96^2 pqN}{s^2(N-1) + 1.96^2 pq}$$

$$n = \frac{2.576^2 pqN}{s^2(N-1) + 2.576^2}$$

En donde:

p= probabilidad de que se realice el evento

q= probabilidad de que no se realice el evento

E= error permitido al cuadrado

N= tamaño de la población

III. ESTIMACION POR INTERVALOS

Conforme la población es más grande o tiende a ser infinita, será necesario realizar más entrevistas que si se tiene una población más pequeña o bien finita. Este tipo de cálculo es válido sólo para métodos probabilísticos.

Cuando no se conoce el tamaño de la población, se utiliza la fórmula:

$$n = \frac{1.96^2 pq}{E^2}$$

$$n = \frac{2.576^2 pq}{E^2}$$

IV PRUEBA DE HIPÓTESIS, ANÁLISIS DE REGRESIONES

La regresión es una herramienta que nos permite identificar cuáles y de que forma las variables independientes afectan o explican a la variable dependiente.

Nuestro modelo es:

$$Y = \beta_0 + \beta_1 X_1$$

Donde

y = variable dependiente

X = variable independiente o explicativa

β = coeficiente explicativo .Cambio que se experimenta en Y por una unidad de cambio en X1.. Xn

Desde el punto de vista económico, es posible utilizar el análisis de regresión para determinar la ecuación de demanda de un bien, en función de variables independientes que influyen en la demanda de dicho bien, tales como: ingresos de la población, precios de productos sustitutos precio del producto mismo y gustos y preferencias de los consumidores.

Sin embargo, para que un análisis de regresión sea aún más útil en la estimación de una ecuación de demanda, es preciso complementarlo con otros instrumentos estadísticos como el coeficiente de determinación R cuadrada y las pruebas de hipótesis sobre los parámetros β_1 . β_n . utilizando la distribución estadística "t" de Student y la función de distribución F.

El coeficiente de determinación R^2 indica el porcentaje de cambios en la variable dependiente que son explicados por la o las variables independientes o explicativas utilizadas en el análisis de regresión. El coeficiente de determinación R^2 puede tener valores que van desde 0 hasta 1; mientras más cercano a uno sea el valor de R^2 , mayor será el grado en que las variables independientes explican a la variable dependiente. El valor de R^2 tiende a aumentar a medida que se considera una mayor cantidad de variables independientes en el análisis de regresión. El coeficiente de determinación R^2 ajustado toma en cuenta los grados de libertad del error y el número total de grados de libertad del modelo.

La distribución estadística "t" de Student es una herramienta que permite realizar inferencias de una población a partir de una muestra pequeña (se considera que una muestra es pequeña cuando contiene entre 1 y 30 observaciones). La "t" de Student y sus valores críticos se basan en la suposición de que la población muestreada posee una distribución de probabilidad normal.

La distribución "t" posee una forma que resulta muy cercana a la de la distribución teórica de "t" para poblaciones que si bien no son normales, poseen una distribución de probabilidad de forma acampanada.

La distribución "t" de Student permite realizar una prueba estadística para determinar si los parámetros β_1 , β_2 ,... β_n muestran suficiente evidencia de que las variables X_1 , X_2 , X_n , contribuyen a predecir el valor de la variable dependiente Y.

La distribución F es una familia de distribuciones diferenciadas por dos parámetros(grados de libertad del numerador y grados de libertad del denominador) utilizada principalmente para probar hipótesis con respecto a varianzas, así el cociente F es utilizado para comparar la magnitud de dos estimaciones de la varianza de la población para determinar si las dos estimaciones son aproximadamente iguales. En el planteamiento de la hipótesis nula se supone que las varianzas son iguales

REGRESION EXCELL

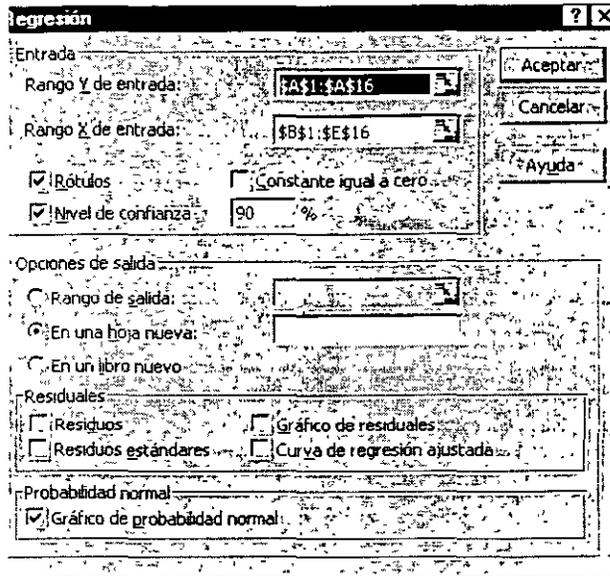
	A	B	C	D	E	F
1	proyectos terminados	nivel del proyecto	personal que participa	responsabilidad del líder	presupuesto asignado (miles)	
2	101	812	15	30	38	
3	60	834	8	20	52	
4	10	377	5	20	27	
5	27	303	10	54	36	
6	89	505	13	52	34	
7	60	401	4	27	45	
8	16	177	6	28	50	
9	164	598	9	52	60	
10	34	412	16	34	44	
11	17	127	2	28	39	
12	78	601	6	42	41	
13	141	297	11	84	58	
14	11	205	4	31	51	
15	104	603	5	38	63	
16	76	484	8	41	30	
17						
18						
19						

Menú de Herramientas: Pulsar

Análisis de datos
Regresión

The screenshot shows the Microsoft Excel interface with the 'Análisis de datos' (Data Analysis) task pane open. The 'Regresión' (Regression) option is selected in the list of functions. The task pane also shows other options like 'Covarianza', 'Estadística descriptiva', 'Suavización exponencial', 'Prueba F para varianzas de dos muestras', 'Análisis de Fourier', 'Histograma', 'Media móvil', 'Generación de números aleatorios', and 'Jerarquía y percentil'. The background shows the same data table as above.

Se llenan las celdas con TODOS los datos de la variable dependiente y la siguiente celda con TODOS los datos de la variable o las variables independientes. Se pueden incluir los títulos de las columnas indicándolo con una paloma.



Resultado de la Regresión como lo brinda EXCEL

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0,43791109
Coefficiente de determinación R ²	0,95
R ² ajustado	95
Error típico	,0054
Observaciones	4

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	2756,206	1378,1036	0.118	0.0001
Residuos	1	11616,543	11616,5435		
Total	3	14372,75			

Varl independ X.	Coefficient es $\hat{\beta}$	Error típico	Estadístic o t	Probabilid ad	Inferior 95%	Superior 95%
Intercepci on	5.2	5.2469E-11	0.9913536	-0.502764E-15	7,1869	8,1466
precio	2.5	4.2497E-14	5.8828E+13	1,0822E-14	2,5	2,5
propagand a	1.2	1.2293E-12	0.9918301	0.50261E-3	1,4401	1,6839

La probabilidad debe ser menor que $\alpha = 0.05$, si es así entonces la variable independiente correspondiente si influye en la explicación de la variable dependiente.

El Coeficiente de determinación R^2 es un valor entre 0 y 1 entre más se acerque a 1 significa que las variables independientes analizadas explican de manera completa a la variable dependiente

	A	B	C	D	E	F	G	H
1	Resumen							
2								
3	<i>Estadística de la regresión</i>							
4	Coefficiente de correlación múltiple	0.849311						
5	Coefficiente de determinación R^2	0.805962						
6	R^2 ajustada	0.729747						
7	Error típico	26.59799						
8	Observaciones	15						
9								
10	ANÁLISIS DE VARIANZA							
11	<i>Grados de libertad de cuadrados de los cu</i>							
12	Regresión	4	24573.87	7393.468	10.45083	0.00135		
13	Residual	10	7074.529	707.4529				
14	Total	14	31648.4					
15								
16	<i>Coefficiente t, estadística t, estadística F, estadística de los cuadrados de los cu</i>							
17	Intercepción	-113.376	35.10104	-3.22999	0.009021	-191.588	-35.1859	
18	nivel del proyecto	0.15196	0.040644	3.734211	0.003883	0.081288	0.242832	
19	porcentaje participa	-1.89938	2.31321	-0.81674	0.433101	-7.04376	3.264943	
20	responsabilidad del líder	2.085828	0.533746	3.907534	0.002925	0.896455	3.2752	
21	propuesta asignada (milz)	1.065307	0.724488	1.465372	0.173542	-0.55452	2.685137	
22								
23								

V. TECNICAS DE MUESTREO

El muestreo es, sin duda, el procedimiento estadístico que la mayor parte de las personas ve con mayor suspicacia, ya que no siempre es fácil entender cómo se puede proyectar a toda una población la información recopilada en un grupo relativamente reducido de la misma.

Cuando no se está familiarizado con el muestreo, existe la creencia de que es mejor entrevistar a toda o a la mayor parte de la población que se desea estudiar, a pesar de los altos costos que ello implica. Sin embargo, la razón de ser del muestreo es la minimización del tiempo y el dinero que se emplean en la labor de campo (o levantamiento de encuestas).

Aún cuando resulta más exacta una medición por medio de entrevistas a toda una población, esta es prácticamente imposible y quizás innecesaria en estudios de mediano y gran alcance, de ahí la necesidad de muestrear.

MUESTREO ALEATORIO SIMPLE

Cada elemento tiene una posibilidad igual de ser seleccionado, y cada combinación de elementos de la muestra n tiene una igual oportunidad de ser seleccionada.

MUESTREO ESTRATIFICADO

Tiene en cuenta la reducción del error estándar sobre el muestreo aleatorio simple. Para seleccionar una muestra estratificada, divide la población en estratos mutuamente excluyentes y colectivamente exhaustivos, y luego seleccione una muestra probabilística de cada estrato. Los estratos deben diseñarse para que sean más homogéneos que la población total.

Una muestra estratificada proporcional es aquella en la que el tamaño de la muestra está distribuido en estratos, en proporción al número de elementos de la población en los estratos.

Los elementos de la muestra también pueden asignarse sobre una base no proporcional con el tamaño de los estratos de la población. En este caso, el objetivo es tomar más de una muestra de los estratos con la mayor variabilidad, con el fin de reducir el error estándar dentro de los estratos.

MUESTREO CONGLOMERADOS

Se selecciona aleatoriamente un conglomerado o grupo de unidades muestrales simultáneamente. Para lograr esto, la población se divide en grupos mutuamente excluyentes y colectivamente exhaustivos.

MUESTREO SISTEMÁTICO

Se selecciona cada k -ésimo elemento del marco, después de un inicio aleatorio en alguna parte dentro de los primeros k elementos.

El intervalo muestral equivale a N / n y es igual al número de posibles muestras en un procedimiento sistemático.

MUESTREO POR AREAS

Comprende la selección de partes geográficas. Una muestra por áreas de múltiples etapas comprende el proceso repetido de enumeración de las unidades muestrales y la selección de una muestra probabilística de esta lista.

VI.- Definición, Construcción y Análisis de Indicadores

Reglas para elaborar cuestionarios.

No reunir 2 o más respuestas

El servicio es bueno y Económico

No asumir la experiencia del entrevistado

Es mejor ISSSTE o IMSS

No obligar a recordar datos

Cuántas veces se enfermó en el año?

Ser claro, no subjetivo o ambiguo

Viene seguido?

No usar preguntas positivas o negativas

Fue malo el servicio?

Inicie preguntando por la identificación del entrevistado, después el perfil y finalmente información básica

Nombre

Pasatiempo

Cuanto ha esperado?

Inicie creando interés, termine con las preguntas difíciles

Cree que se puede mejorar el servicio?

Esta dispuesto a pagar?

TIPOS DE PREGUNTAS

Abiertas

Proporcionan mucha información

Difícil codificación

Poco interés

Hay que motivar al encuestado

Cerradas

Opción Múltiple (No excluyente)

Porque selecciono este lugar?

Cercanía

Economía

Servicio

Dicotómicas

Tiene automóvil

Si ___

No ___

Escala por categoría

La atención fue

Excelente Muy buena Normal Mala

Escala Jerárquica

Que le es mas importante

Marque los 3 principales factores
use 1 para el más importante
escala Likert.

EL TIEMPO DE ESPERA ES ACEPTABLE

Totalmente de acuerdo Medianamente de acuerdo Indeciso Totalmente en desacuerdo

Diferencial Semántico

El hospital es ?

Moderno _____ Anticuado

Sucio _____ Limpio

Capacidad suficiente _____ Insuficiente

Personal desagradable _____ Personal agradable

**VII.- Aplicación del concepto:
"Tablero de Control"**

Casos de fallecimiento en menores

Muertes de niño de 5 a 9 años

0 = niño, 1 = niña

Folio	Casos	Edad	Sexo	Folio	Casos	Edad	Sexo
1	1	5	0	50	1	7	1
2	2	6	0	51	4	8	1
3	4	9	1	52	1	6	0
4	5	5	0	53	1	6	0
5	2	5	1	54	5	8	0
6	1	5	0	55	1	8	0
7	1	6	0	56	1	8	0
8	3	5	0	57	2	6	1
9	1	6	0	58	3	8	1
10	5	7	0	59	4	9	1
11	5	6	0	60	1	9	0
12	1	6	0	61	1	8	0
13	1	5	1	62	2	9	0
14	1	7	0	63	2	6	0
15	3	7	0	64	3	5	0
16	2	8	0	65	5	5	1
17	1	7	0	66	3	6	1
18	3	9	0	67	1	9	0
19	1	8	0	68	5	8	0
20	1	6	0	69	1	5	0
21	3	8	0	70	5	6	0
22	1	5	0	71	3	8	0
23	3	9	0	72	1	8	0
24	1	7	0	73	1	7	0
25	1	5	0	74	1	6	0
26	5	5	0	75	1	7	1
27	2	9	0	76	4	5	0
28	1	8	0	77	1	9	1
29	2	6	1	78	1	6	0
30	1	8	0	79	5	6	0
31	1	5	0	80	1	9	0
32	5	7	0	81	1	6	1
33	1	6	0	82	4	7	0
34	5	5	1	83	1	9	1
35	1	8	0	84	2	8	0
36	1	6	0	85	5	5	1
37	1	5	0	86	1	7	0
38	5	6	0	87	2	7	0
39	4	6	0	88	3	6	0
40	5	5	0	89	1	8	0
41	2	5	0	90	1	5	0
42	3	6	1	91	5	5	0
43	5	6	1	92	1	8	0
44	1	6	0	93	2	5	0
45	2	5	0	94	1	7	0
46	1	6	0	95	1	5	0
47	1	8	0	96	1	5	0
48	2	8	0	97	3	9	1
49	5	7	0	98	4	9	0
				99	5	6	0
				100	5	5	0

Peso de recién nacidos (grms)

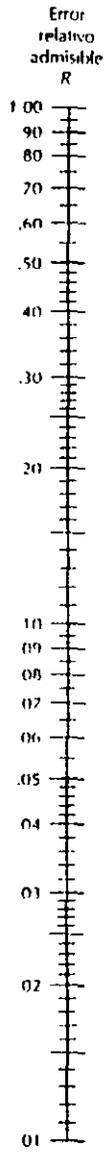
<i>min</i>	<i>max</i>	<i>mediciones</i>
0	499	4,843.00
500	999	17,487.00
1000	1499	23,139.00
1500	1999	49,112.00
2000	2499	160,919.00
2500	2999	597,738.00
3000	3499	1,376,008.00
3500	3999	1,106,634.00
4000	4499	344,390.00
4500	4999	62,769.00
5000	5499	8,236.00

CO₂ Expirado

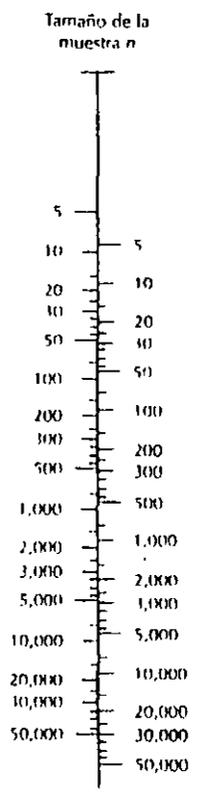
1	2.3
2	2.15
3	3.5
4	2.6
5	2.75
6	2.82
7	4.05
8	2.25
9	2.68
10	3
11	4.02
12	2.85
13	3.38

Ritmo Cardíaco(Latido / min)

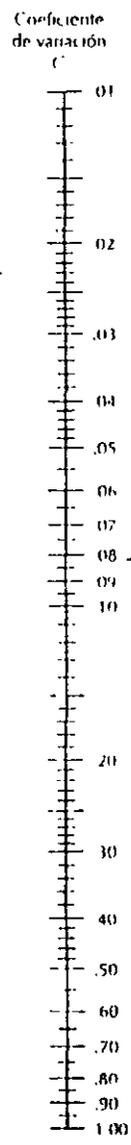
Folio	Medición
1	167
2	150
3	125
4	120
5	150
6	150
7	40
8	136
9	120
10	150



Nivel de confianza del 99%



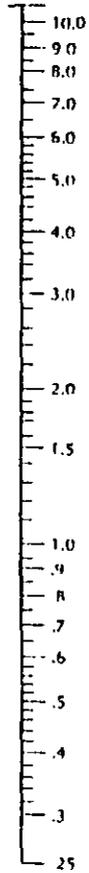
Nivel de confianza del 95%



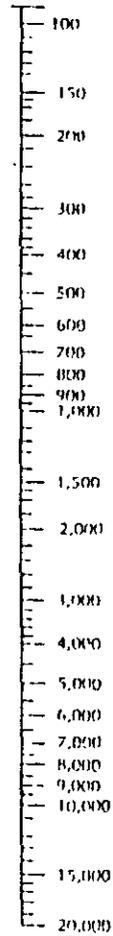
Proporção
 p
(expressada como
porcentagem)



Erro
admissível
(expressado como
porcentagem)



Tamanho da amostra
(para o erro admissível
especificado)



BIBLIOGRAFIA

INTRODUCCION A LA INVESTIGACION DE MERCADOS

MARCELA BENASSINI

PRENTICE HALL

INVESTIGACION DE MERCADOS

KINNEAR / TAYLOR

MC GRAW HILL