VII.7 PRÁCTICA X

ESTIMACIÓN DE PARÁMETROS

Problema No.1.- Suponga que μ = \$ 100 ; $Z_\alpha = \pm 2.58$ y $\sigma_{\bar{x}} = 10$, encontrar:

a) El intervalo de confianza
b) Los límites de confianza
c) El coeficiente de confianza

Problema No. 2 Con los datos del problema 1, suponga que μ es desconocida y $\overline{X}$ = \$105, encontrar:

a) La estimación de punto
b) La estimación de intervalo
c) Interprete los resultados de los dos tipos de estimaciones. Si μ es conocida e igual a \$ 100, ¿está la verdadera media poblacional dentro del intervalo de estimación?

Problema No. 3. Suponga que Π = 45%, $Z_\alpha = \pm 1.645$ y σ_p = 6%, encontrar:

a) El intervalo de confianza
b) Los limites de confianza
c) El coeficiente de confianza

Problema No. 4. Con los datos del problema 3, suponga que Π es desconocida y p = 48%, encontrar:

a) La estimación de punto
b) La estimación del intervalo
c) Interprete los resultados de las dos estimaciones. Si Π es conocida e igual a 45%, ¿Esta la verdadera proporción de la población dentro del intervalo de estimación?

Problema No. 5 Poner un ejemplo numérico de estimadores:

a) Insesgados ; b) Eficientes ; c) Suficientes ; d) Consistentes

VIII. TEORÍA DE LA DECISIÓN ESTADÍSTICA O PRUEBA DE HIPÓTESIS [9]

Conexión entre pruebas y estimaciones de de intervalo

Antes de de iniciar el tema he creído conveniente señalar que existe dicha relación, ya que en una prueba de hipótesis , dado un nivel de significación α , se construye un intervalo de confianza para ξ o nivel de confianza para no rechazar la hipótesis nula.

Así, por ejemplo, si en una prueba de dos extremos utilizamos Z, con un cierto nivel de α de que la Ho: $\mu x = \mu_O$, la condición o intervalo para no rechazar Ho es que la z crítica o valor teórico que se obtiene de tablas sea igual o mayor que la Z real u observada, que es lo mismo, en el caso del intervalo de confianza, ya que: la media muestral, más menos, el producto de z teórica por el error estándar de la media, sea superior a la media hipotética. Ambas desigualdades son lo mismo y cada una tiene una probabilidad 1-α = ξ si es que μ_X es μ_O ; la primera garantiza que la prueba tiene nivel de significación α y la segunda, garantiza que el intervalo de confianza tiene probabilidad 1-α de contener μ_X

Con este enlace ahora procederemos a desarrollar el tema de la prueba de hipótesis

VIII.1 DEFINICIÓN

Podríamos decir que esta es una de las principales aportaciones de la teoría de la probabilidad a la inferencia estadística, ya que al verificar una hipótesis de trabajo con una muestra probabilística, si dicha hipótesis es aceptada, ello es una gran contribución a la investigación que ahora ya dispone de un método estadístico para la toma de decisiones con certidumbre pero, además, contribuye al aumento del acervo de conocimientos en el área que se esté efectuado la verificación de la hipótesis, en virtud de que muchas hipótesis al corroborarse que son ciertas, pasan a formar parte de la ciencia en que se desenvuelve el investigador.

Al respecto, como señala el Profesor L. Kazmier(7) , en la prueba de hipótesis empezamos suponiendo que un parámetro poblacional tienen un determinado valor, tal como la media de la población; enseguida seleccionamos aleatoriamente una muestra, y calculamos su media aritmética para probar con ella que el supuesto valor poblacional es en efecto correcto, sobre la base de los elementos que componen la muestra, es decir, comparando el supuesto valor del parámetro con el valor equivalente de media muestral.

Es importante decir que en cualquier situación de prueba de hipótesis, la exactitud del supuesto valor del parámetro poblacional, es decir, la validez de la hipótesis, no se puede probar directamente. En su lugar, lo que se prueba es la magnitud de la diferencia entre el supuesto valor del parámetro poblaciona y el valor obtenido de una estadística muestral.

La evidencia ideal en apoyo de una hipótesis sería la observación que la diferencia entre los dos valores fuera igual a cero. Esta hipótesis se conoce como hipótesis nula. Así por ejemplo, si en la producción de anillos industriales se requiere que el diámetro medio de cada uno sea 0.575 centimetros y si tomamos una muestra aleatoriamente para verificarlo, si su media es 0.565, se prueba la hipótesis nula en el sentido de verificar la diferencia entre los valores 0.565 y 0.575 centimetros y nos preguntamos ¿ la diferencia de 0.010 centimentros es sifnificativamente diferente de cero ?

Dicha diferencia se juzga considerando su valor absoluto estandarizado, es decir, considerando el error estándar de la media de la muestra.

Al respecto, al generar las distribuciones de muestras, nosotros recordamos que usando el muestreo con o sin reemplazo, se producen distribuciones de muestras, de cuyas medias podemos seleccionar una de ellas aleatoriamente para comparar su valor con el del supuesto valor del parámetro poblacional. Esperamos que si éste último es cierto, los valores de muchas medias muestrales se agruparàn o situarán simétricamente alrededor de su valor. Si hubieramos seleccionado aleatoriamente otra muestra de todas las que están disponibles, otro sería el valor de su media muestral y otra seía la diferencia con el valor del parámetro poblacional.

Por otra parte es importante mencionar que aun cuando no hay un número universal definido para aceptar o rechazar una hipótesis nula, generalmente se usa α = 5%, conocida α **como nivel de significación, cuya interpretación es :** cuando es grande la diferencia entre el valor de la estadística muestral y el supuesto valor del parámetro poblacional, una diferencia de esa magnitud ocurrirá al azar con una probabilidad de 5% o menos cuando el supuesto valor es en realidad correcto; en esa situación rechazamos la hipótesis nula y se considera que la diferencia observada es significativa, en otras palabras, que la gra diferencia no se debe a la selección aleatoria de la muestra, sino que se debe a otras razones.

Con ese razonamiento, si definimos a $\mu o \pm Z\alpha_{\sigma} \quad \bar{x}$, como límites críticos para aceptar o rechazar la hipótesis nula; en este caso decimos que 0.95 es la proporción de medias muestrales comprendidas dentro de los límites crítcos establecidos para rpbar la hipótesis nula, con un nivel de significación del 5%.

También es importante decir que cuando no se específica la dirección en que se desea probar la diferencia entre $\mu o \; y \; \bar{x}$, la prueba se denomina de " dos lados, colas o extremos ". Por el contrario, cuando si se específica la dirección para probar dicha diferencia haci auno u otro lado de la curva, la prueba se llama " de un extremo " y α no se divide entre dos colas o extremos, por lo que $Z\alpha$ toma otro valor (1.64 y no 1.96 , cuando es de dos colas).

Conviene señalar que una hipótesis nula se prueba con unidades originales (medias muestrales) y con unidades Z (número de unidades del error estándar de la media); **este último procedimiento es el más usual.**

Con estas referencias y trabajando con más literales, diremos que :
Hipótesis estadística es una suposición o conjetura concerniente a una característica de la población.
Por ejemplo: $\bar{X} = \mu$
Hipótesis nula H_o: $\bar{X} = \mu$
Hipótesis alternativa H_i: $\bar{X} \neq \mu$.
Reiteramos, cuando se formula una hipótesis, no podemos probarla directamente por que no conocemos μ, lo que se prueba es la diferencia entre μ y $\bar{X}$ tal que:

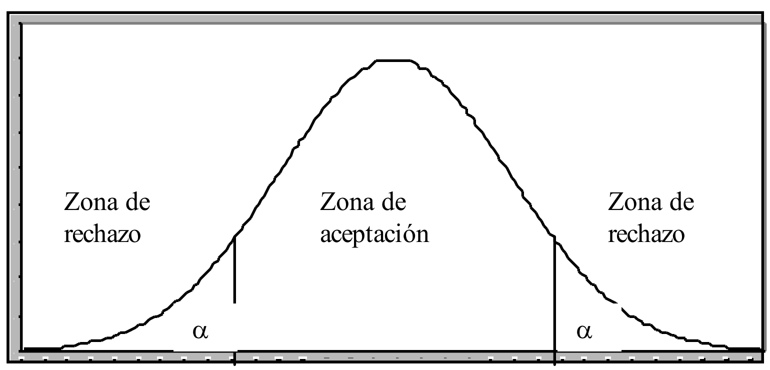
$\overline{X} - \mu = 0$

Ello indica que se prueba que no hay diferencia entre μ y $\overline{X}$, lo cual es llamado **HIPÓTESIS NULA, (Ho).**
Cualquier hipótesis diferente de la nula es llamada HIPÓTESIS ALTERNATIVA, (Hi).

Cuando se hace el planteamiento para tomar una decisión estadística, es factible cometer 2 tipos de errores:
Error 1: Rechazar una hipótesis nula cuando realmente es verdadera; y
Error 2: Es aceptar una hipótesis nula cuando realmente ésta es falsa.

La probabilidad de cometer un error del tipo 1, es usualmente denotada por α (alfa); y la probabilidad de cometer un error del tipo 2, es denotada por β (beta).

Veamos gráficamente La prueba de hipótesiscuando se hace con dos "colas" o extremos:



μo
Dos colas

μ_o = media hipotética

VIII.2 PRUEBA DE HIPÓTESIS CON Z

Asimismo, la prueba de hipótesis se hace con :

$Z = \dfrac{\bar{x} - \mu o}{\sigma \bar{x}}$ cuando se conoce σ y $n > 30$

$\sigma \bar{x} = \dfrac{\sigma}{\sqrt{n}}$ para una población infinita y se hace con: $t = \dfrac{\bar{x} - \mu o}{s\bar{x}} = \dfrac{\bar{x} - \mu o}{\dfrac{s}{\sqrt{n-1}}}$ cuando se desconoce σ y n < 30

Ahora también, cuando Z ó t ≤ Zα ó tα se acepta Ho; y
" Z ó t > Zα ó tα se rechaza Ho.

TABLA Zα PARA PRUEBA DE HIPÓTESIS :

Nivel de significación (α)	0.10	0.05	0.01	0.005	0.00
Valores críticos para ensayos de un extremo	-1.28 1.28	-1.645 1.645	-2.33 2.33	-2.58 2.58	-2.88 2.88
Valores críticos para ensayos de dos extremo	-1.645 y 1.645	-1.96 y 1.96	-2.58 y 2.58	-2.81 y 2.81	-3.00 y 3.00

1.- Diferencia de una media muestral y una poblacional conocido σ, tal que:

$Z = \dfrac{\bar{x} - \mu o}{\sigma\bar{x}}$; $\sigma\bar{x} = \dfrac{\sigma}{\sqrt{n}}$ para una población infinita

Para una diferencia de proporciones:

$Z = \dfrac{p - P}{\sigma_p}$; $\sigma_p = \sqrt{\dfrac{PQ}{n}}$ para una población infinita

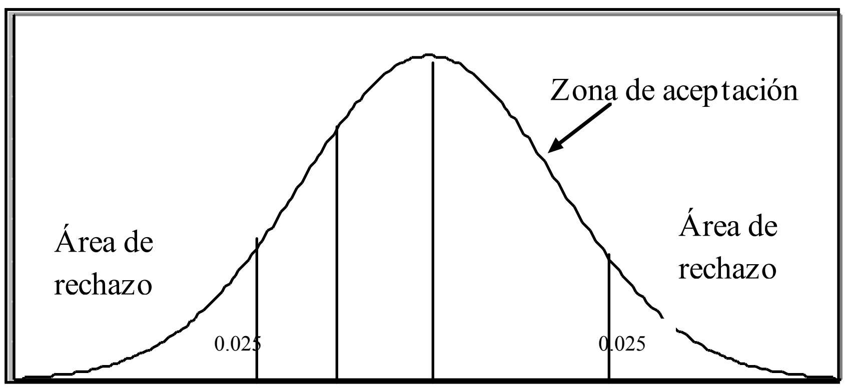
EJEMPLO 1:

Retomando el ejemplo planteado al principio del tema, cuando lo usamos para enfatizar el significado de " las diferencias ", ahora digamos que en la producción de cierto tipo de anillos se requiere que exista una calidad estándar de tal manera que su diámetro medio sea de 0.575 centimetros. Se toma una muestra de 50 anillos, y arrojan un diámetro medio de 0.565 pulgadas. Pruebe la hipótesis de que la media poblacional es igual ala media muestral, si el nivel de significación es de 5%, y se hace un ensayo de dos extremos. La desviación estándar es de 0.50 centimetros.

μ o= 0.575 centimetros

n = 50 anillos industriales

$\overline{X} = 0.565$ centimetros
$\sigma = 0.50$ centimetros
$\alpha = 5$ %
$Z\alpha = \pm 1.96$ = valor critico para aceptar o rechazar Ho.
H_o: $\overline{X} = \mu$; H_i: $\overline{X} \neq \mu$



$\overline{X} = 0.565$ $\mu_o = 0.575$
Z_α=-1.96 Z= - .1414 Z_α=1.96

En donde:

$$Z = \frac{-0.010}{0.0707} = -.1414$$

$$Z = \frac{\overline{x} - \mu o}{\sigma \overline{x}} = \frac{0.565 - 0.575}{0.0707}; \text{y} \quad \sigma \overline{x} = \frac{\sigma}{\sqrt{n}} = \frac{0.050}{\sqrt{50}} = 0.00707$$

Z= -.1414 esta dentro de la zona de aceptación y se acepta H_t.

2.- Diferencia de dos medias muestrales, cuando se conoce σ.
H_o : $\mu_1 - \mu_2 = 0$
$\mu_1 = \mu_2$
H_i : $\mu_1 \neq \mu_2$
En ese caso $Z = \dfrac{\overline{x}_1 - \overline{x}_2}{\sigma(\overline{x}_1 - \overline{x}_2)}$

$$\sigma(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$$

En el caso de una proporción :

$$Z = \frac{p_1 - p_2}{\sigma(p_1 - p_2)}; \ \sigma(p_1 - p_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Cuando N y n son grandes, se pueden usar las desviaciones estándar de las muestras en el error estándar.

$$S(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1}{n_1} + \frac{S_2}{n_2}} \text{ tal que } Z = \frac{\bar{x}_1 - \bar{x}_2}{S(\bar{x}_1 - \bar{x}_2)}$$

VIII.3 DISTRIBUCIÓN t DE STUDENT: [12]

Esta distribución fue elaborada por William S. Gosset, que usaba el nombre de "Student".

Se utiliza para manejar muestras pequeñas, generalmente menores de 30 y cuando no se conoce σ.

t - tiende a la normal igual que Z, de tal manera que:

$$t = \frac{\bar{x} - \mu}{S\bar{x}}$$

Tiene media 0 y desviación estándar 1, es decir, adopta la forma de la distribución normal estándar.

$S_{\bar{x}}$ es el error estándar calculado a partir de la muestra, de tal manera que:

$$S_{\bar{x}} = \frac{S}{\sqrt{n-1}}$$

donde S es la desviación estándar de la muestra;

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Así, en el caso de la prueba de hipótesis cuando no se conoce σ, ésta se estima a partir de S. De esta manera al igual que Z; si: Z ó t son menores o iguales que Z_α ó t_α se acepta la hipótesis nula.

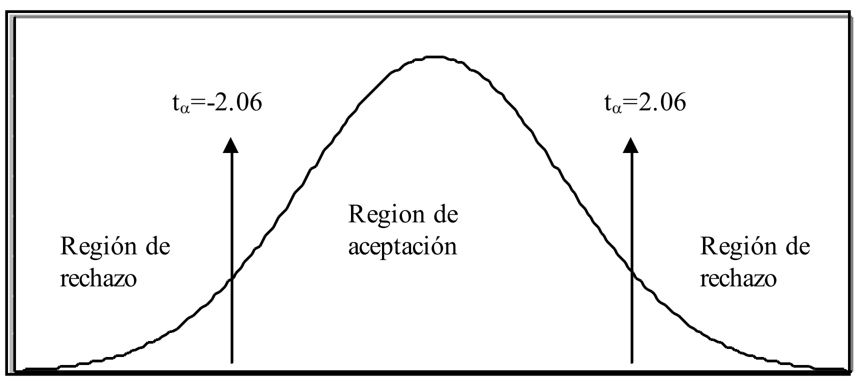
Ejemplo : Se desea probar que el ingreso mensual de la ciudad gama es de \$2,500.00 al mes por trabajador, con alfa α = 5%. Para ello se selecciona una muestra al azar de 26 trabajadores cuyo ingreso medio mensual es de \$3,000.00 mes con una desviación estándar de \$ 100.00.

H_o: μ_o = \$ 2,500.00 μ_o = media hipotética.

tal que $\overline{X}$ - μ =0.
H_1 : $\mu \neq$ \$ 2,500.00
n = 26 trbajadores
$\overline{X}$ = \$3,000.00
S = \$100.00
α = 5%
$t_\alpha = \pm$ 2.06

Como n -1 indica el número de grados de libertad (G.L.), en este caso G.L. = n - 1 = 25.

Con ello, buscamos en el apéndice D, el valor critico de t con α =5% y G.L. = 25, para determinar el área de aceptación y rechazo de la hipótesis, en la cual t = ± 2.06.



$\mu_0 - t_\alpha s_{\bar{x}}$ 0 $\mu_0 + t_\alpha s_{\bar{x}}$

\$2,458.80 μ_o=\$2,500 \$2,541.20

$$t = \frac{\bar{x} - \mu_0}{S\bar{x}}; \text{ donde } S_{\bar{x}} = \frac{S}{\sqrt{n-1}} = \frac{100}{\sqrt{25}} = 20$$

$$S_{\bar{x}} = 20$$

$$t = \frac{3{,}000 - 2{,}500}{20} = \frac{500}{20} = 25$$

Como t = 25 > t_α = 2.06 se rechaza la hipótesis de que el ingreso medio de los trabajadores sea de \$2,500.00 en la ciudad γ. Ello se fundamenta en que la diferencia entre $\overline{X}$ y μ_o es significativa y no puede atribuírsele a la selección aleatoria de la muestra.

De igual manera se puede probar la hipótesis nula de la diferencia de dos medias usando t, es decir cuando no se conoce σ y se trabaja con muestras pequeñas menores de 30.
Si n es grande se puede usar S en lugar de σ en Z, es decir:

$$Z = \frac{\bar{x} - \mu_0}{S\bar{x}} \quad \text{donde } S_{\bar{x}} = \frac{S}{\sqrt{n-1}}$$

H_o: $\mu_1 = \mu_2$ ó $\mu_1 - \mu_2 = 0$; tal que $\bar{x}_1 - \bar{x}_2$ es al azar

Es decir se desea probar que $\bar{x}_1 - \bar{x}_2$ provienen de la misma población por ello:
$\mu_1 = \mu_2$
La hipótesis alternativa será todo lo contrario:
H_1 : $\mu_1 \neq \mu_2$ ó $\mu_1 - \mu_2 \neq 0$ tal que $\bar{x}_1 - \bar{x}_2$ no proviene de la misma población.
Cuando se conoce σ:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma(\bar{x}_1 - \bar{x}_2)} \text{; donde } \sigma(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Sin embargo, cuando no se conoce, se utilizara t dado:

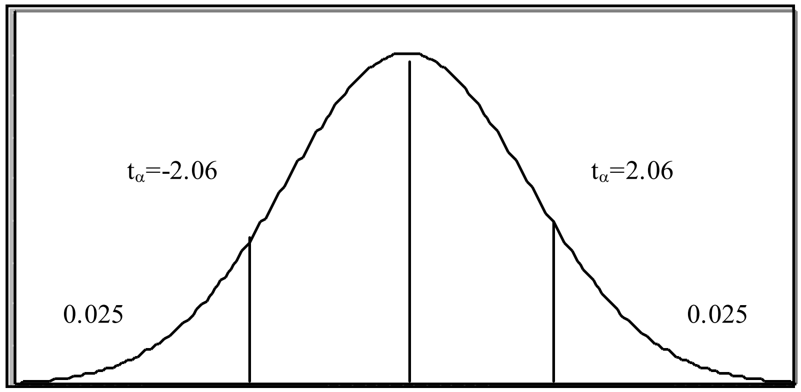
$$t = \frac{\bar{x}_1 - \bar{x}_2}{S(\bar{x}_1 - \bar{x}_2)} \quad \text{donde } S(\bar{x}_1 - \bar{x}_2) = \hat{S}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$\text{y} \quad \hat{S} = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \text{ con } n_1 + n_2 - 2 = \text{G.L.}$$

EJEMPLO 1:
Se desea estimar la diferencia de medias cuando no se conoce σ y son muestras pequeñas.
Se desea probar la hipótesis con α= 5%, de que el ingreso medio por trabajador es el mismo en las colonias Arenal y Tlacotal. Para ello se seleccionan dos muestras al azar, y se obtienen los siguientes datos:

n_1 = 10 familias	n_2 = 17 familias
$\bar{x}_1$ = \$6,200.00/mes	$\bar{x}_2$ = \$5,600.00/mes
S_1 = 690	S_2 = 600

luego, con $n_1 + n_2 - 2 = 10 + 17 - 2 = 25$ G.L.
y α = 5% t α = ± 2.06



$$\left(\overline{X}_1 - \overline{X}_2\right)$$

$$\hat{S} = \sqrt{\frac{10(476{,}100) + 17(360{,}000)}{10 + 7 - 2}} = \sqrt{\frac{4{,}761{,}000 + 6{,}120{,}000}{25}} = \sqrt{\frac{10{,}881{,}600}{25}} = \$600.00$$

$$\hat{S} = 660$$

$$S(\bar{x}_1 - \bar{x}_2) = \hat{S}\sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 660\sqrt{\frac{10 + 17}{10(17)}} = 660\sqrt{\frac{27}{170}} = 660(0.399) = \$263.00$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S(\bar{x}_1 - \bar{x}_2)} = \frac{6{,}200 - 5{,}600}{263} = \frac{600}{263} = 2.28$$

t=2.28

Como t = 2.28 > tα = 2.06, tomamos la decisión de rechazar la hipótesis nula, porque $\bar{x}_1$ y $\bar{x}_2$ difieren significativamente y no se puede atribuir esa diferencia a la selección aleatoria de las dos muestras.

De tal manera que decidimos con una probabilidad de 5% de cometer error tipo I, que el ingreso medio por familia es diferente en las colonias Arenal y Tlacotal.

Por otra parte, cuando n - 1 es grande, casi es n, tal que Z puede usarse en lugar de t. En general cuando n > 30 se usará Z; y cuando n < 30 se usará t.

Dado lo anterior, la estimación del intervalo de confianza con t sería:
Limite de confianza $= \bar{x} \pm t\alpha S\bar{x}$.

EJEMPLO 3:

Si deseamos estimar el gasto mensual promedio por familia en ropa en una región con población homogénea, para ello seleccionamos al azar una muestra de 10 familias cuya media aritmética es igual a \$838.00 con S = \$105.00.

n = 10

$\overline{X}$ = \$ 838.00

S = \$ 105.00

luego:

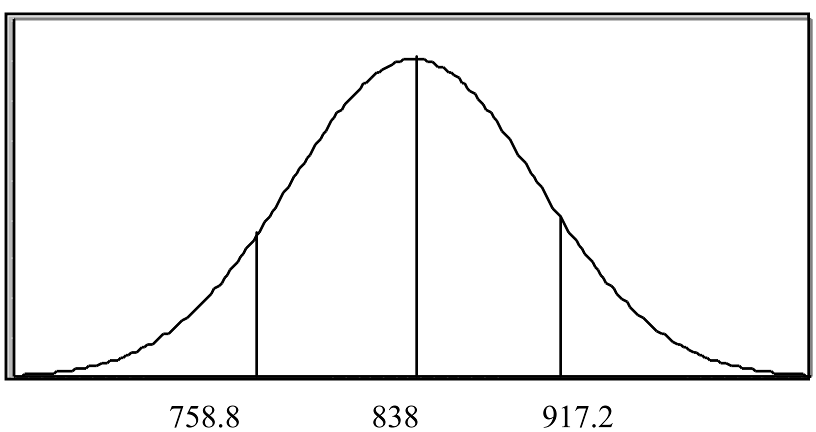
$$S_{\bar{x}} = \frac{S}{\sqrt{n-1}} = \frac{105}{\sqrt{9}} = 105 / 3 = \$35.00$$

Así mismo, con α = 5% y 9 G.L. tα = ± 2.262

Limites de confianza = 838 ± 2.262 (35) = 838 ±79.2

Intervalo de confianza = \$ 758.8 a \$ 917.2

INTERPRETACIÓN: Indica que existe una probabilidad o seguridad del 95% de que el gasto medio mensual de las familias en ropa sea entre \$758.8 y \$917.2.



De manera similar se puede estimar el intervalo de confianza para la diferencia de medias.

Si D = $\bar{x}_1 - \bar{x}_2$ entonces :

Limites de confianza = D ± tα $S_{\bar{x}}$

Tomando los datos del ejemplo de la prueba de hipótesis
D = 6,200 - 5,600 = $600.00
Con α = 5% y G.L. = 25; tα = ± 2.06 y $S_{\bar{x}}$ = $35.00

luego limites de confianza = D ± tα $S_{\bar{x}}$ = 600.00 ± 2.06 (263) = 600 ± 541.7; intervalo de confianza = 58.3 a 1,141.7.

VIII.4 X CUADRADA (X^2)

Con ella podemos comparar frecuencias observadas y esperadas y dos o más conjuntos de frecuencias para ver si difieren significativamente, fórmula.:

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

Nunca es regular.
Grados de libertad : n - n_0 restricciones.

VIII.4.1 PRUEBA DE BONDAD DE AJUSTE

Ejemplo.

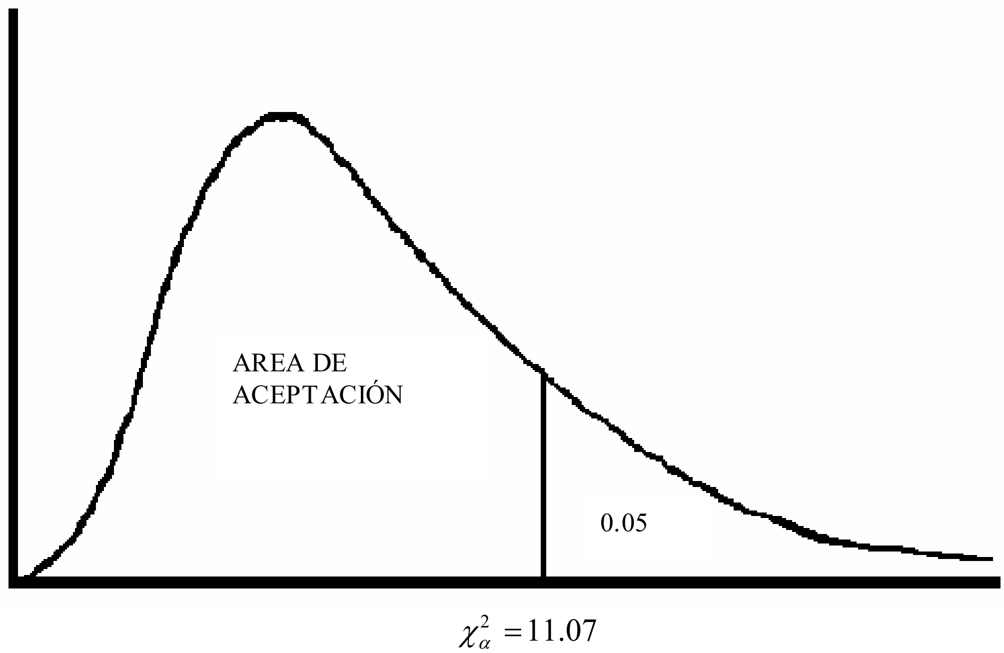
En la venta de un producto, el gerente dividió al país en 6 regiones de venta para obtener pedidos por correo. El gerente espera igual número de pedidos en cada una de las 6 áreas. Después de un breve período, decide probar la eficacia de su campaña; en ese momento ha recibido 60 solicitudes. El establece la hipótesis de que no hay diferencia, que las 6 áreas son iguales, espera 10 solicitudes de cada área. Los resultados son los siguientes :

Área	Nº de pedidos fo	fe	fo-fe	$(fo-fe)^2$	$\frac{(fo-fe)^2}{fe}$
A	6	10	-4	16	1.6
B	15	10	5	25	2.5
C	7	10	-3	9	0.9
D	4	10	-6	36	3.6
E	17	10	7	49	4.9
F	11	10	1	1	0.1
	60				13.6

f_o: frecuencia observada; f_e: frecuencia esperada.

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe} = 13.6$$

Grados de libertad = n - 1 = 5; α = 5%; $\chi^2_\alpha = 11.07$



Se rechaza la hipótesis porque $\chi^2 = 13.6 > \chi^2_\alpha = 11.07$.

Esto es : $\chi^2 = 13.6$ se halla en la zona de rechazo, ya que la zona de aceptación llega hasta $\chi^2_\alpha = 11.07$

VIII.4.2 PRUEBA DE INDEPENDENCIA DE PRINCIPIOS O DE CLASIFICACIÓN EN LAS TABLAS DE CONTINGENCIA.

Cuando tres grupos se puedan clasificar en tres formas, se obtiene una tabla de contingencia.

Clase	A_1	A_2	A_3
B_1	n 11	n 21	n 31
B_2	n 12	n 22	n 32
B_3	n 13	n 23	n 33

Ejemplo de cómo se prueba la hipótesis nula:
Si se envían dos cuestionarios para ser contestados por correo por los subscriptores de una revista: 100 con una moneda en agradecimiento y 200 sin la moneda, la hipótesis es que no influye la moneda, por lo que la clasificación es independiente :

CUESTIONARIO	OBSERVADOS			ESPERADOS		
	Respondieron	No respondieron	Total	Respondieron	No respondieron	Total
Moneda incluida	77	23	100	65.7	34.3	100
Moneda no incluida	120	80	200	131.3	68.7	200
TOTAL	197	103	300	197	103	300

Se esperan 197 normalmente distribuidos entre con y sin moneda.
Cuestionarios esperados : $\frac{197}{300} * 100 = 65.7$ se esperan normalmente distribuidos con moneda.
$\frac{197}{300} * 200 = 131.3$ cuestionarios sin moneda
$\frac{103}{300} * 100 = 34.3$ y $\frac{103}{300} * 200 = 68.7$

Así podremos calcular χ^2.

Celda	Fo	Fe	Fo-Fe	$(Fo-Fe)^2$	$\frac{(Fo-Fe)^2}{Fe}$
1 - 1	77	65.7	11.3	127.7	1.9437
1 - 2	23	34.3	-11.3	127.7	3.7230
2 - 1	120	131.3	-11.3	127.7	0.9726
2 - 2	80	68.7	11.3	127.7	1.8588
					8.4981

$\chi^2 = 8.49 > \chi^2_\alpha = 3.84$
$\chi^2_\alpha = 3.841$ con α=5% y G.L. = 1 por lo tanto se rechaza la hipótesis.
Hay R.C. = (2)2 = 4 celdas y (R - 1)(C - 1) = 1 grados de libertad.

Ejemplo adicional a resolver:

Se tomó una muestra de 200 tornillos producidos por 4 diferentes máquinas para ver la eficacia de los operadores. Para ver si ellos tienden a producir la misma distribución de la calidad del producto de acuerdo con las clasificaciones de calidad previamente definidas.

FRECUENCIAS OBSERVADAS

	OPERADOR				
	1	2	3	4	Total
CALIDAD					
Excelente	40	44	32	24	140
Marginal	7	5	12	16	40
No aceptable	3	11	6	-	20
Totales	50	60	50	40	200

Probar la hipótesis de que no hay diferencia entre los cuatro operadores que producirán la misma calidad, con $\alpha = 5$ %.

Como el numerador es siempre positivo: $\chi^2 > 0$, tal que la prueba de hipótesis es de una sola cola o extremo. Con α y (C-1) (R-1) grados de libertad.
Ejemplo de prueba de hipótesis, sobre la independencia de principios:
$H_o : f_o = f_e$

TABLA DE CONTINGENCIA

Clase	A_1	A_2	A_3
B_1	n 11	n 21	n 31
B_2	n 12	n 22	n 32
B_3	n 13	n 23	n 33

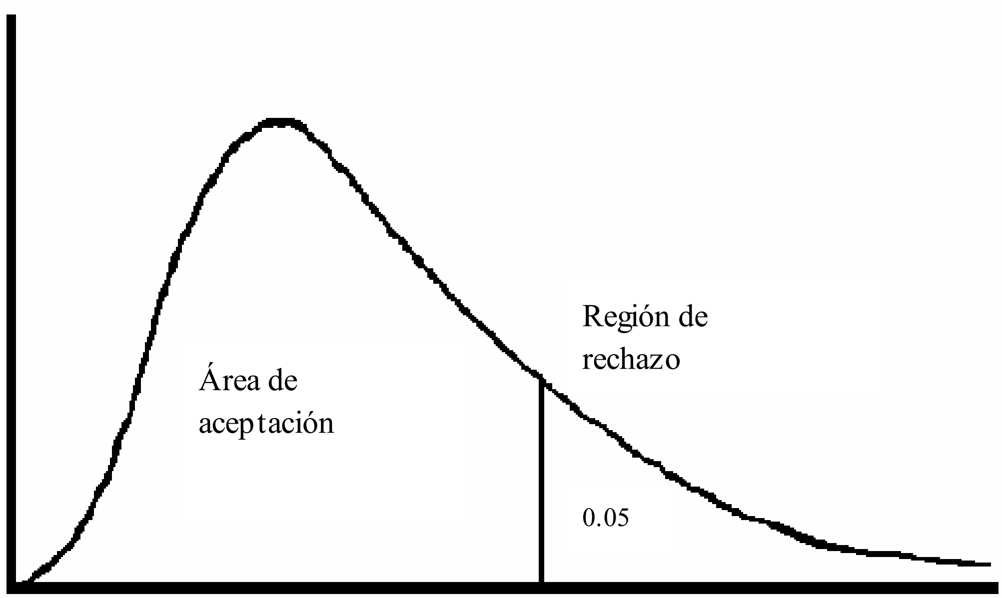
	FRECUENCIA OBSERVADA					FRECUENCIA ESPERADA				
	H1	H2	H3	H4	Total	X1	X2	X3	X4	Total
CALIDAD Excelente Marginal No aceptable	 40 7 3	 44 5 11	 32 12 6	 24 16 -	 140 40 20	 35 10 5	 42 12 6	 35 10 5	 28 8 4	 140 40 20
Totales	50	60	50	40	200	50	60	50	40	200

CALCULO DE LAS FRECUENCIAS ESPERADAS

X_1	X_2	X_3	X_4
5 ÷ 200 * 140 = 35 ; 0	60 ÷ 200 * 140 = 4 ; 2	5 ÷ 200 * 140 = 35 ; 0	40 ÷ 200 * 140 = 28
5 ÷ 200 * 40 = 10 ; 0	60 ÷ 200 * 40 = 1 ; 2	5 ÷ 200 * 40 = 10 ; 0	40 ÷ 200 * 40 = 8
5 ÷ 200 * 20 = 5 ; 0	60 ÷ 200 * 20 = 6 ;	5 ÷ 200 * 20 = 5 ; 0	40 ÷ 200 * 20 = 4

Con α = 0.05 y (C - 1)(R - 1) = (4 -1)(3 - 1) = 3(2) = 6 G.L

$\chi^2 = 12.592$



χ^2=12.595

Celda	Fo	Fe	Fo-Fe	$(Fo-Fe)^2$	$\frac{(Fo-Fe)^2}{Fe}$

1 - 1	40	35	5	25	0.7143
1 - 2	44	42	2	4	0.0952
1 - 3	32	35	-3	9	0.2571
1 - 4	24	28	-4	16	0.5714
2 - 1	7	10	-3	9	0.9000
2 - 2	5	12	-7	49	4.0833
2 - 3	12	10	2	4	0.4000
2 - 4	16	8	8	64	8.0000
3 - 1	3	5	-2	4	0.8000
3 - 2	11	6	5	25	4.1667
3 - 3	6	5	1	1	0.2000
3 - 4	0	4	-4	16	4.0000
					24.1881

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe} = 24.17$$

Como $\chi^2 = 24.17 > \chi^2_\alpha = 12.592$, rechazamos la hipótesis de que los cuatro operadores no difieran en habilidad para producir tornillos.

VIII.5 EVALUACIÓN ESTADÍSTICA DE ENCUESTAS MENSUALES O PERIODICAS

1.- Introducción

El levantamiento mensual de encuestas requiere de una supervisión estadística que permanentemente favorezca la confiabilidad de la información. Para ello es necesaria la aplicación de ciertas técnicas que detecten si existe o no relación entre el tamaño de la muestra y el valor de los indicadores.

Para ilustrar lo anterior se tomó como referencia una encuesta mensual que hace el Instituto de la Pequeña y Mediana Empresa.

La periodicidad de la encuesta requiere la aplicación de técnicas fuertes que permitan eliminar rápidamente los factores irrelevantes y retener los de gran significación en los resultados. A las medidas estadísticas que permitan cumplir con estos objetivos se les denominará

COEFICIENTES DE ASOCIACIÓN.

Puesto que el método de muestreo utilizado es el de proporciones correspondientes a indicadores con distribuciones fuera de cualquier curva definida por funciones matemáticas, se optó por la aplicación de pruebas de asociación no paramétrica de las variables en la pequeña y mediana empresa en conjunto.

2.- Prueba de asociación.

La escasez de recursos humanos y el limitado acceso a la computadora, en esta etapa determinaron el manejo de solo tres medidas de asociación; en la medida que se resuelvan estos problemas y que el personal se familiarice con el análisis estadístico, se aplicarán diseños muestrales y coeficientes de asociación más sofisticados.

Por otra parte, mientras el análisis estadístico no se instrumente en la computadora, mensualmente se evaluará una de las siguientes variables. Para iniciar los trabajos de julio, de la encuesta de junio se analizó:

-Personal ocupado promedio respecto al mes anterior

-Inventario de productos finales

-Fuentes de financiamiento para resolver problemas de liquidez, total industria pequeña y mediana.

PERSONAL OCUPADO PROMEDIO

Se recurre a la χ^2 : Ji Cuadrada basada en las tablas de contingencia para probar la hipótesis de independencia entre el tamaño de la muestra y la opinión de los empresarios. Para ello se comparan las respuestas "reales" de la muestra con las respuestas "esperadas".

Personal Ocupado

Muestra	Respuesta Real			Total
	Aumento	No Aumentó	Disminuyó	
Alimentos . . . . Otros				
Total				419
%				100

Cuando se acepta la hipótesis no es necesario modificar el tamaño de la muestra. En cambio si se rechaza la hipótesis, se identifica que si hay una relación entre el tamaño de la muestra y la opinión de los empresarios; por lo que es necesario hacer un análisis como el que se describe a continuación:

Así por ejemplo: partiendo del rechazo de la hipótesis nula basado en la χ^2, se utilizará la estadística ϕ para cuantificar la relación entre la muestra y las opiniones ; ya que si es baja quizá no valga la pena hacer las revisiones correspondientes; en cambio si es alta de inmediato se hace un análisis de sesgo y cobertura.

Phi (ϕ)

Es una medida de la fuerza de la relación que existe entre las variables descriptivas, la cuantitativa (muestra) y la cualitativa (opinión de los empresarios). Phi toma el valor de 0 cuando no hay relación y + 1 cuando las variables se relacionan a la perfección. Phi hace la corrección en el valor de χ^2 porque éste es directamente proporcional al tamaño de la muestra (n) y por ello su fórmula es :

$$\phi = \left[\frac{\chi^2}{n}\right]^{1/2}$$

V de Cramer

Cuando ϕ se obtiene de tablas de contingencia más grande a la de 2 x 2, como es el caso concreto de la encuesta, su valor no tiene límite superior, por lo que se usa V de Cramer para ajustar ϕ en términos de las columnas o de las hileras, dependiendo cual de ellas es más pequeña.

El valor de la estadística V también oscila entre 0 y +1. Así, un valor alto de V significa que hay un alto grado de asociación.

Su fórmula es:

$$V = \left(\frac{\phi^2}{min(r-1, c-1)}\right)^{1/2}$$

En resumen, si una vez aplicadas las estadísticas χ^2, ϕ y V, se encuentra que el valor de V es alto, entonces se toma la decisión de hacer el análisis de sesgo y cobertura, para lo cual se analiza la información a fin de validarla y determinar si los resultados pueden atribuirse a relaciones o asociaciones legítimas o a la selección aleatoria de la muestra.

Si es ésta última habrá que hacer lo siguiente:

1) recalcular el tamaño de la muestra (cobertura) en los grupos industriales afectados y,
2) mantener el porciento dentro de ciertos limites de control (sesgo).

PROCEDIMIENTO

A continuación se expone un ejemplo completo con datos del mes de junio, empezando por la χ^2, ϕ y V , hasta el análisis de cobertura para el caso extremo en que tuviéramos que recalcular toda la muestra, aplicando el **muestreo simple aleatorio**; así como para el cálculo especifico para algunos grupos industriales, **usando el muestreo estratificado proporcional**.

TABLA DE CONTINGENCIA				
GRUPO INDUSTRIAL	A_1	A_2	A_3	TOTAL
B_1	R_{11}	S_{21}	T_{31}	$V_1 = R_{11} + S_{21} + T_{13}$
B_2	R_{12}	S_{22}	T_{32}	$V_2 = R_{12} + S_{22} + T_{32}$
B_3	R_{13}	S_{23}	T_{33}	$V_3 =$
B_4	R_{14}			$V_4 =$
B_5	R_{15}			$V_5 =$
B_6	R_{16}			$V_6 =$
B_7	R_{17}			$V_7 =$
B_8	R_{18}	S_{28}		$V_8 =$
B_9	R_{19}			$V_9 = R_{19} + S_{29} + T_{39}$

B_{10}	R_{110}		T_{310}	V_{10} =
B_{11}	R_{111}	S_{211}		V_{11} =
B_{12}	R_{112}			V_{12} =
B_{13}	R_{113}			V_{13} =
B_{14}	R_{114}		T_{314}	V_{14} =
B_{15}	R_{115}	S_{215}	T_{315}	V_{15} =
B_{16}	R_{116}	S_{216}	T_{316}	$V_{16} = R_{116} + S_{216} + T_{316}$
TOTAL	R	S	T	V = R + S +T

Construyendo la tabla de contingencia con los resultados observados para el Personal Ocupado en junio, se obtiene la tabla 3X16 que aparece a continuación para las dos variables descriptivas Bi (cuantitativa y A : Cualitativa: opinión de los empresarios).

Donde:

B1: grupo industrial

A_1 : Aumentó

A_2 : No aumentó

A_3 : Disminuyó

$R = \Sigma\ R_i$

$S = \Sigma\ S_i$

$T = \Sigma\ T_i$

$V = \Sigma\ V_i = R + S + T$

$V_i = \Sigma\ (R_i + S_i + T_i)$

i = 1, 2, 3, ..., 16

Personal Ocupado Promedio				
	Respuesta Real			
Muestra	Aumento	No vario	Disminuyo	Total
Fab. de alimentos	10	61	13	84
Industria Textil	3	22	3	28
Fab. de Prendas de Vestir	4	27	9	40
Fab. de Calzado e Ind. del Cuero	5	25	7	37
Ind. y Prod. de Madera y Corcho				0
Excepto Muebles	1	9	5	15

Fab. y Rep. de Muebles de Madera	1	11	9	21
Ind. Editorial de Impresión y Conexas	6	13	1	20
Industria Química	3	11	2	16
Fab. de Prod. de Hule y Plástico	4	19	2	25
Fab. de Productos Minerales no Metalicos	3	24	9	36
Industrias Metalicas Básicas	-	4	1	5
Fab. de Prod. Metalicos	2	27	12	41
Fab. de Maq. y Equipo Excepto los Electricos	9	13	2	24
Fab. de Maq. y Equipo y Aparatos Electricos	-	4	3	7
Construcción de Equipo de Transporte	3	6	5	14
Otras Industrias Menufactureras	2	3	1	6
TOTAL	56 R	279 S	84 T	419 V

Cálculo de las frecuencias esperadas

GRUPO				

INDUSTRIAL	X_1	X_2	X_3	TOTAL
B_1	V_1 (R/V) = 11	V_1 (S/V) = 56	V_1 (T/V) = 17	V_1 = 84
B_2	V_2 (R/V) = 4	V_2 (S/V) = 17	V_2 (T/V) = 6	V_2 = 28
B_3	V_3 (R/V) = 5	V_3 (S/V) = 27	V_3 (T/V) = 9	V_3 = 40
B_4	V_4 (R/V) = 5	V_4 (S/V) = 25	V_4 (T/V) = 7	V_4 = 37
B_5	V_5 (R/V) = 2	V_5 (S/V) = 10	V_5 (T/V) = 3	V_5 = 15
B_6	V_6 (R/V) = 3	V_6 (S/V) = 14	V_6 (T/V) = 4	V_6 = 21
B_7	V_7 (R/V) = 3	V_7 (S/V) = 13	V_7 (T/V) = 4	V_7 = 20
B_8	V_8 (R/V) = 2	V_8 (S/V) = 11	V_8 (T/V) = 3	V_8 = 16
B_9	V_9 (R/V) = 3	V_9 (S/V) = 17	V_9 (T/V) = 5	V_9 = 25
B_{10}	V_{10} (R/V) = 5	V_{10} (S/V) = 24	V_{10} (T/V) = 7	V_{10} = 36
B_{11}	V_{11} (R/V) = 1	V_{11} (S/V) = 3	V_{11} (T/V) = 1	V_{11} = 5
B_{12}	V_{12} (R/V) = 5	V_{12} (S/V) = 28	V_{12} (T/V) = 8	V_{12} = 41
B_{13}	V_{13} (R/V) = 3	V_{13} (S/V) = 16	V_{13} (T/V) = 5	V_{13} = 24
B_{14}	V_{14} (R/V) = 1	V_{14} (S/V) = 5	V_{14} (T/V) = 1	V_{14} = 7
B_{15}	V_{15} (R/V) = 2	V_{15} (S/V) = 9	V_{15} (T/V) = 3	V_{15} = 14
B_{16}	V_{16} (R/V) = 1	V_{16} (S/V) = 4	V_{16} (T/V) = 1	V_{16} = 6
TOTAL	R=56	S=279	T=84	V=419

Agrupándolos por celda, tendremos:

Celda			fr	Fe	fr-fe
1	-	1	10	11	-1
1	-	2	61	56	5
1	-	3	13	17	-4
2	-	1	3	5	-2
2	-	2	22	17	5
2	-	3	3	6	-3
3	-	1	4	4	0
3	-	2	27	27	0
3	-	3	9	9	0
4	-	1	5	5	0
4	-	2	25	25	0
4	-	3	7	7	0
5	-	1	1	2	-1
5	-	2	9	10	-1
5	-	3	5	3	2
6	-	1	1	3	-2
6	-	2	11	14	-3
6	-	3	9	4	5
7	-	1	6	3	3
7	-	2	13	13	0
7	-	3	1	4	-3
8	-	1	3	2	1
8	-	2	11	11	0
8	-	3	2	3	-1
9	-	1	4	3	1
9	-	2	19	17	2
9	-	3	2	5	-3
10	-	1	3	5	-2
10	-	2	24	24	0
10	-	3	9	7	2
11	-	1	0	1	-1
11	-	2	4	3	1
11	-	3	1	1	0
12	-	1	2	5	-3
12	-	2	27	28	-1
12	-	3	12	8	4

Celda			fr	fe	fr-fe
13	-	1	9	3	6
13	-	2	13	16	-3
13	-	3	2	5	-3
14	-	1	0	1	-1
14	-	2	4	5	-1
14	-	3	3	1	2
15	-	1	3	2	1
15	-	2	6	9	-3
15	-	3	5	3	2
16	-	1	2	1	1
16	-	2	3	4	-1
16	-	3	1	1	0

Donde: fr = frecuencia real

fe = frecuencia esperada

Haciendo las comparaciones, entre fr, fe para sustituirlas en la fórmula, se obtiene:

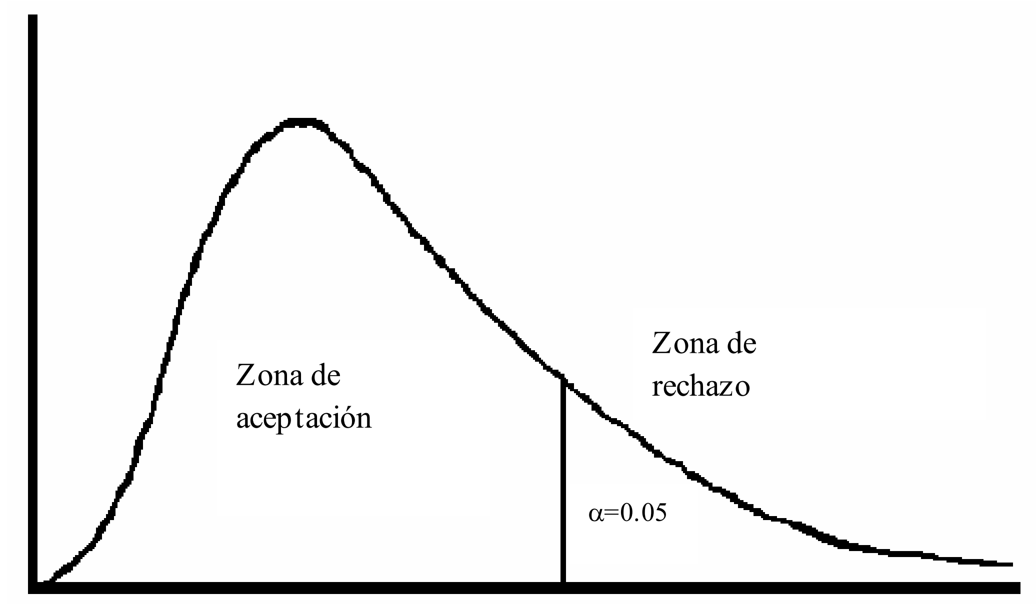
$(fr-fe)^2$	$\frac{(fr-fe)^2}{fe}$
1	0.0909
25	0.4464
16	0.9412
4	0.8000
25	1.4706
9	1.5000
0	0.0000
0	0.0000
0	0.0000
0	0.0000
0	0.0000
0	0.0000
1	0.5000
1	0.1000
4	1.3333
4	1.3333
9	0.6429
29	7.2500
9	3.0000
0	0.0000
9	2.2500
1	0.5000
0	0.0000
1	0.3333
1	0.3333

$(fr-fe)^2$	$\frac{(fr-fe)^2}{fe}$
2	0.1176
9	1.8000
4	0.8000
0	0.0000
4	0.5714
1	1.0000
1	0.3333
0	0.0000
9	1.8000
1	0.0357
16	2.0000
36	12.0000
9	0.5625
9	1.8000
1	1.0000
1	0.2000
4	4.0000
1	0.5000
9	1.0000
4	1.3333
1	1.0000
1	0.2500
0	0.0000
Total	53.90

$$\chi^2 = \frac{(fr-fe)^2}{fe} \qquad \chi^2 = 53.90$$

con $\alpha = 0.05$ y (c-1)(R-1) = (3-1)(16-1) = 30 grados de libertad.

el valor critico de $\chi^2_\alpha = 43.773$ tenemos que



Como $\chi^2 = 53.90 > \chi^2_\alpha = 43.773$ se rechaza la hipótesis nula de que no hay diferencia entre el tamaño de la muestra y la opinión de los empresarios.

Luego se inicia la prueba Phi (ϕ) para cuantificar el grado de asociación entre las dos variables descriptivas.

$$\phi = \left[\frac{\chi^2}{n}\right]^{1/2} = \left(\frac{53.90}{419}\right)^{1/2} = (0.12864)^{1/2} = 0.358$$

La interpretación es que hay una relación sensiblemente significativa.

Como la tabla de contingencia es más grande que una de dos por dos, se aplica la V Cramer para corregir el valor de ϕ.

$$V = \left[\frac{\phi^2}{C-1}\right]^{1/2} = \left[\frac{(0.358)^2}{2}\right]^{1/2} = \left[\frac{0.128164}{2}\right]^{1/2}$$

$V = (0.064082)^{1/2}$

$V = 0.25$

Puesto que el valor de V oscila entre cero y más uno, no se modifica el tamaño de la muestra para el mes de junio porque la asociación no es fuerte.

Si se hubiera tomado la decisión de hacer el análisis de cobertura y sesgo, el procedimiento seria:

Cobertura: Nuevo Tamaño de la Muestra

A) Muestreo simple aleatorio

Antecedentes

Obtener el tamaño de la muestra adecuado para asegurar con una probabilidad igual a 95%, que el error en la estimación del número medio de empresas necesarias no sea mayor del 6%.

Para ella se tomó la muestra aleatoria del mes de junio, la cual fue de 419 empresas distribuidas en 16 grupos industriales de la siguiente manera:

	Concepto	Nº de Empresas (X_i)
	Total	419
1 .-	Fab. de alimentos	84
2 .-	Industria Textil	28
3 .-	Fab. de Prendas de Vestir	40
4 .-	Fab. de Calzado e Ind. del Cuero	37
5 .-	Ind. y Prod. de Madera y Corcho Excepto Muebles	15
6 .-	Fab. y Rep. de Muebles de Madera	21
7 .-	Ind. Editorial de Impresión y Conexas	20
8 .-	Industria Química	16
9 .-	Fab. de Prod. de Hule y Plástico	25
10 .-	Fab. de Productos Minerales no Metalicos	36
11 .-	Industrias Metalicas Básicas	5
12 .-	Fab. de Prod. Metalicos	41
13 .-	Fab. de Maq. y Equipo Excepto los Electricos	24
14 .-	Fab. de Maq. y Equipo y Aparatos Electricos	7
15 .-	Construcción de Equipo de Transporte	14
16 .-	Otras Industrias Menufactureras	6

2 Cálculo

Como no se conocen los valores de los parámetros poblacionales μ y σ^2, es necesario estimarlos a partir de las estadísticas $\bar{x}$ y S^2 de la muestra. Así;

Grupo Industrial	X_i	%	X_i^2
1	84	20	7,056
2	28	7	784
3	40	10	1,600
4	37	9	1,369
5	15	4	225
6	21	5	441
7	20	5	400
8	16	4	256
9	25	6	625
10	36	9	1,296
11	5	1	25
12	41	10	1,681
13	24	6	576
14	7	2	49
15	14	3	196
16	6	1	36
Suma	419	102	16,615

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{16}(419) = 26 \text{ empresas}$$

$$S^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = \frac{1}{16}(16{,}615) - (26)^2 = 1038 - 676 = 362 \text{ empresas}$$

Considerando que el error en la estimación (e) del promedio de empresas no debe ser superior al 6%, y recordando que el estimador de μ = $\bar{x}$ = 26 empresas, se observa que e = 26 (0.06) = 1.56 empresas.

Igualmente, como se desconoce el valor de σ^2 y tomando en cuenta que su estimador proviene de una muestra mayor de 30 empresas, la distribución teórica a la cual se aproxima la distribución de muestreo es a la normal.

En este caso se estima μ de la población con variable aleatoria asociada X mediante el empleo de $\bar{x}$, proveniente de n = 419 con e = 6 % y un nivel de confianza ξ = 95 %, donde Z= desviación correspondiente al nivel de confianza de ξ en la distribución normal; en este caso la probabilidad ξ le corresponde Z_α = ± 1.96.

Considerando a $K\sigma_{\bar{x}_x}$ como $= Z\alpha(\sigma \bar{x})$ este razonamiento para obtener el tamaño de la muestra se basa en el hecho de que:

$P(\bar{x} - k\sigma \leq \mu \leq \bar{x} + k\sigma) = Pk = 1 - \alpha = 95\%$

α = nivel de significación = 5%

En otras palabras $P\left[|\hat{p} - p| \geq 0.06p\right] = 1 - 0.95 = 5\%$

Ello significa que el error en la estimación del valor de μ en valores absolutos es:

|error en la estimación de μ| = $k\sigma$, por lo que

|error máximo admisible| = |error en la estimación de μ| = e

Derivado de lo anterior se puede escribir.

$e = k\sigma_{\bar{x}} = Z_\alpha \sigma_{\bar{x}}$ donde Z_α = variable estandarizada.

donde $\sigma\bar{x} = \dfrac{\sigma}{\sqrt{n}}$,para una población infinita.

Sabiendo que K= Z

Cuando la población es finita $e = k\sqrt{\dfrac{\sigma^2}{n}\dfrac{N-n}{N-1}}$

Como no se conoce σ^2, la estima S^2 y sabiendo que K=Z

$e = Z\sigma_{\bar{x}} = Z\sqrt{\dfrac{S^2}{n}\dfrac{N-n}{N-1}}$

Para obtener el tamaño de la muestra (n), se despeja de la ecuación anterior elevando al cuadrado ambos miembros.

$$e^2 = Z^2 \frac{S^2}{n}\frac{N-n}{N-1}$$

Así: $n = \dfrac{Z^2 * S^2 * N}{e^2 N - e^2 + Z^2 S^2}$

Con e = 6% ; en absolutos e = 26(0.06) = 1.56 empresas;
$\alpha = 5$ %
ξ = 95 %

Z = ± 1.96
S^2 = 362
N = 8,966

$$n = \frac{Z^2 * S^2 * N}{e^2 N - e^2 + Z^2 S^2} = \frac{(1.96)^2 (362)(8,966)}{(1.56)^2 (8,966) + (362)(1.96) - (1.56)^2} = \frac{12,468,650}{21,820 + 1,391 - 2} = \frac{12,468,650}{23,209}$$

n = 537 empresas.
Comprobación del valor de (e)

$$e^2 = Z^2 \frac{S^2}{n} \frac{N-n}{N-1} = (3.84)(0.6741)(0.94)$$

e^2 = 2.43 luego e = 1.56 empresas= error permitido= error de muestreo.

Si deseamos distribuir la muestra de 537 empresas por grupo industrial, se hace con el procedimiento llamado de **afijación proporcional de la muestra, de conformidad con la importancia que tenga cada estrato (Ni) dentro del universo (N).**

Grupo Industrial	Ni/N %	n= 537	n_i
1			
2			
3			
4			
5			

Donde i = 1, 2, 3, 4, 5, ..., 16
por lo que $n_1 + n_2 + n_3 + \ldots + n_{16} = n = 537$

B. muestreo estratificado

Tomando como referencia los datos de este diseño muestral que aplicamos en el inciso en que hablamos de la precisión, donde indicamos que el error de muestreo se mide con el error estándar, entonces digamos ahora que si el error estándar de la proporción proveniente de una distribución de muestreo estratificada finita es:

$$\sigma_p = \sqrt{\sum_{1}^{k} W_i^2 S_i^2 \; \frac{N_i - n_i}{N_i * n_i}}$$

$$\sigma_p^2 = \sum_{1}^{k} W_i^2 S_i^2 \; \frac{N_i - n_i}{N_i * n_i} = \frac{\sum^{k} W_i^2 S_i^2 N_i - n_i \sum^{k} W_i^2 S_i^2}{N_i * n_i}$$

$$\sigma_p^2 (N_i * n_i) = N_i \sum_{1}^{k} W_i^2 S_i^2 - n_i \sum_{1}^{k} W_i^2 S_i^2$$

$$\sigma_p^2 (N_i * n_i) + n_i \sum_{1}^{k} W_i^2 S_i^2 = N_i \sum_{1}^{k} W_i^2 S_i^2$$

Entonces :

$$n_i (\sigma_p^2 N_i + \sum_{1}^{k} W_i^2 S_i^2) = N_i \sum_{1}^{k} W_i^2 S_i^2$$

$$n_i = \frac{N_i \sum^{k} W_i^2 S_i^2}{\sigma_p^2 N_i + \sum_{1}^{k} W_i^2 S_i^2} ; como S^2 = pq$$

$$n_i = \frac{N_i \sum^{k} W_i^2 pq}{\sigma_p^2 N_i + \sum_{1}^{k} W_i^2 pq}$$

Ejemplo:

Estratos	N_i	W_i	n_i muestra	Empresas de la muestra que contestaron	P_i	
1	7,000	0.7	200	160	160 ÷ 200 =	0.8
2	1,000	0.1	100	40	40 ÷ 100 =	0.4
3	2,000	0.2	100	60	60 ÷ 100 =	0.6
	10,000	1	400	260		

Con $\sigma p = 0.025$

Como $\sum_{1}^{k} W_i^2 S_1^2$ =(0.49)(0.16)+(0.01)(0.24)+(0.04)(0.24)=0.0784 + 0.0024 + 0.0096 = 0.904

La muestra para cada estrato se va obteniendo así:

$$n_1 = \frac{7{,}000(0.0904)}{(0.025)^2 7{,}000 + 0.0904}$$

$$n_1 = \frac{633}{4{,}465} = 142;$$

$$n_2 = \frac{1{,}000(0.0904)}{(0.000625)1{,}000 + 0.0904} = \frac{90.4}{0.715} = 126$$

$$n_3 = \frac{2{,}000(0.0904)}{(0.000625)2{,}000 + 0.09041} = 135$$

$n_1 + n_2 + n_3 = n = 402$

Sesgo : Limites de Central

Para el análisis de sesgo se definen limites de control (o de confianza) donde con cierta probabilidad se mantendrá el valor del porciento con un tamaño dado de muestra.

Así limites de control = $p \pm Z\,\sigma p$

Cuando se salga de esos limites de control nuevamente se hará la prueba de X^2 ; si se rechaza la hipótesis nula, nuevamente se revisará la muestra en el grupo y se determinará si el porciento es legitimo o se debe a errores de muestreo, de tal manera que el proceso se vuelve interactivo, en el sentido de que se harán ajustes cuantas veces sea necesario hasta llegar a muestras satisfactorias.

VIII.6 PRUEBA DE HIPÓTESIS CON F: ANÁLISIS DE VARIANCIA[12]
VIII.6.1 MÉTODO LARGO

Con el uso de este instrumental se pueden manejar o comparar las diferencias de más de dos medias muestrales, es decir, cuya homogeneidad puede probarse analizando la variabilidad o variancia entre ellos.

$$F = \frac{Variancia \;\; entre \;\; grupos}{Variancia \;\; dentro \;\; de \;\; grupos}$$

Ejemplo.- La compañía NESTLE desea probar si sus 3 agentes de ventas : Rodríguez, Salinas y Pacheco tienden a vender el mismo volumen de mercancía o si difieren en su habilidad para vender.

Lo anterior lo verifican tomando el promedio de ventas hechas por cada uno.

La semana pasada hicieron 14 llamadas:

◊ Rodríguez hizo 5 llamadas
◊ Salinas hizo 4 "
◊ Pacheco hizo 5 "

y sus ventas fueron como sigue en miles de pesos:

Rodriguez	Salinas	Pacheco
$ 300.00	$ 600.00	$ 700.00
$ 400.00	$ 300.00	$ 300.00
$ 300.00	$ 300.00	$ 400.00
$ 500.00	$ 400.00	$ 600.00
		$ 500.00
$ 1500.00	$ 1600.00	$ 2500.00

$\overline{X}_1 = \$300.00 \quad \overline{X}_2 = \$400.00 \;\; \overline{X}_3 = \$500.00$

La gran media muestral $\overline{\overline{X}} = 400.00$

La pregunta a contestar es: si las tres medias difieren más de lo esperado por la selección aleatoria de la muestra :Ho

Para ello se analiza la relación entre sus variancias.

La idea básica es que la variancia poblacional se puede estimar de la muestra en diversas formas.

A continuación se ilustran tres maneras de calcularla:

i) Se podría estimar calculando las desviaciones de las medias muéstrales con respecto a la gran media, a lo cual llamamos: variación entre grupos (medias $\overline{X}$) que son:
$\overline{X}_1 = \$300.00 \quad \overline{X}_2 = \$400.00 \; \overline{X}_3 = \$500.00, \; \overline{\overline{X}} = 400.00$

ii) También se puede estimar de la muestra comparando cada una de las ventas individuales con la media de su grupo, lo cual llamaremos: variación dentro de los grupos.

El total de observaciones (T) es 14; con n_i en cada grupo,esto es:
$n_1 = 5$; $n_2 = 4$; $n_3 = 5$

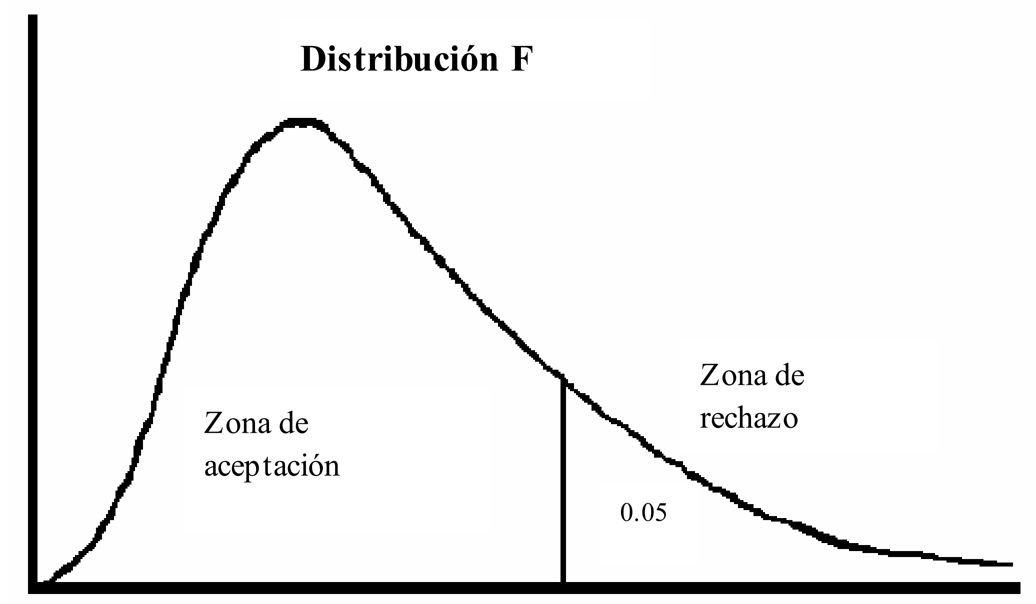
iii) La otra manera de estimarla será comparando cada una de las 14 observaciones con la gran media de las observaciones en los tres grupos a lo cual llamaremos : variación total.

Para realizar la prueba se establece la hipótesis nula de que no hay diferencias entre los vendedores, y las diferencias observadas se deben a la selección de la muestra.

Esta prueba, se realiza con la distribución F, por medio de la cual se determina si dos varianzas difieren más de lo esperado, examinando la razón o cociente entre ellas.

Los grados de libertad $G.L._1$, y $G.L._2$, determinan la forma de la curva; $GL_1 = n_1 - 1$; $GL_2 = n_2 - 2$; $F(n_1 -1, n_2 - 2)$, donde el primer número en el paréntesis son los G.L. del numerador y el segundo los G.L. del denominador.

La mayoría de las pruebas que se realizan con F son de una cola, donde la región de rechazo se halla en la cola derecha.
En general tiene la forma :



Por ello, generalmente la prueba de hipótesis se hace utilizando el extremo o cola derecha.

Debe quedar claro, que las pruebas de análisis de varianza no pretenden probar la significación de las diferencias entre dos varianzas muestrales, su propósito es probar la significación de las diferencias entre medias muestrales con el mecanismo de distribución F.

Ahora procederemos a desarrollar las ecuaciones necesarias para realizar la prueba de análisis de varianza usando los siguientes símbolos :

i : designción de grupo

K : número de grupos (K - 3)

n_i: número de observaciones en el i-ésimo grupo

T : total de observaciones

$$T = \sum_{i=1}^{k} n_i = 5 + 4 + 5 = 14$$

$\overline{X}_i$ Media de i-ésimo grupo

j : numeración seriada de las observaciones dentro de los grupos x_{ij} : una observación. La j-ésima observación en el grupo i-ésimo

$\overline{\overline{X}}$: gran media de las T observaciones

Así :

1o.-El número de grados de libertad es aditivo, esto es: K - 1

2o.-Para la varianza proveniente de dentro de los grupos, los G.L.= (T - K)

3o.-Los G.L. de la varianza de las variaciones totales serán =(T-1)

Por ello es que el número de grados de libertad es aditivo, esto es

(K - 1) + (T - K) = T - 1

Así, primero calculamos las variaciones entre grupos con la fórmula:

$$\sum_{L=1}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2$$

K = 3 ; $\bar{X}_1$ = \$300; n_1=5

$\bar{\bar{X}}$ =\$400 ; $\bar{X}_2$ = \$400; n_2=4

T = 14 ; $\bar{X}_3$ = \$500; n_3=5

$n_1 = (\bar{x}_i - \bar{\bar{x}})^2 = 5(300\text{-}400)^2 = 50{,}000$

$n_2 = (\bar{x}_2 - \bar{\bar{x}})^2 = 4(400\text{-}400)^2 = 0$

$n_3 = (\bar{x}_3 - \bar{\bar{x}})^2 = 5(500\text{-}400)^2 = 50{,}000$

$\sum_{L=1}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2 = 50{,}000 + 0 + 50{,}000 = 100{,}000$

2o.-La variación dentro de los grupos, con la fórmula:

$$\sum_{i=1}^{k} \sum_{j=1}^{ni} (x_{ij} - \bar{x}_j)^2 .$$

Para Rodriguez (i=1) : $\Sigma \ (x_{1j} - 300)^2 = 140{,}000$

Para Salinas (i=2) : $\Sigma \ (x_{2j} - 400)^2 = 60{,}000$

Para Pacheco (i=3) : $\Sigma \ (x_{3j} - 500)^2 = 100{,}000$

$\sum_{i=1}^{k} \sum_{j=1}^{ni} (x_{ij} - \bar{x}_j)^2 . = 140{,}000 + 60{,}000 + 100{,}000 = 300{,}000$

Rodriguez (i=1) Salinas (i=2) Pacheco (i=3)

$\bar{X}_1 = \$300.00$ $\bar{X}_2 = \$400.00$ $\bar{X}_3 = \$500.00$,

j	$x_{ij}-\bar{x}$	$(x_{1j}-\bar{x})^2$	$(x_{2j}-\bar{x}_2)$	$(x_{2j}-\bar{x})^2$	$(x_{3j}-\bar{x}_3)$	$(x_{3j}-\bar{x}_3)^2$
1	0	0	200	40,000	200	40,000
2	100	10,000	-100	10,000	-200	40,000
3	0	0	-100	10,000	-100	10,000
4	200	40,000	0	0	100	10,000
5	-300	90,000	0	0	0	0

		140000		60,000		100,000

3o.-La variación se calcula con :

$$\sum_{i=1}^{k}\sum_{j=1}^{ni}(x_{ij}-\overline{\overline{x}})^2 = \sum_{T}(X_{ij}-\overline{\overline{X}})^2.$$

Cálculo de la variación total :

	RODRIGUEZ		SALINAS		PACHECO	
j	$X_{ij}-\overline{\overline{X}}$	$(X_{ij}-\overline{\overline{X}})^2$	$X_{2j}-\overline{\overline{X}}$	$(X_{2j}-\overline{\overline{X}})^2$	$X_{3j}-\overline{\overline{X}}$	$(X_{3j}-\overline{\overline{X}})^2$
1	-100	10,000	200	40,000	300	90,000
2	0	0	-100	10,000	-100	10,000
3	-100	10,000	-100	10,000	0	0
4	100	10,000	0	0	200	40,000
5	400	160,000	0	0	100	10,000
		190000		60,000		150,000

Luego $\sum_{T}(X_{IJ}-\overline{\overline{X}})^2$ = 190,000 + 60,000 + 150,000 = 400,000

Resumen del análisis de la varianza :

Variación	Grados de libertad	Suma de cuadrados	Varianza
Entre grupos Dentro de grupos	K - 1 = 3 - 1 = 2 T - K = 14 - 3 = 11	100,000 300,000	50,000 27,273
Total	T - 1 = 14 - 1 = 13	400,000	

$$F\ (2,\ 11) = \frac{50,000}{27,273} = 1.83$$

Con $\alpha = 0.05$ y (2 y 11) G.L. $F\alpha = 3.98$

Como $F = 1.83 < F\alpha = 3.98$: concluimos con esta evidencia que hay homogeneidad entre las medias; no es una evidencia de la cual podemos inferir que los vendedores difieren en habilidad para hacerlo.

VIII.6.2 MÉTODO ABREVIADO PARA CALCULAR F

1.-La variación entre los grupos se calcula ahora con :

$$\sum_{I=1}^{K} n_I \bar{x}_i^2 - T\bar{\bar{x}}^2$$

2.-La variación dentro de los grupos :

$$\sum_{T} x_{ij}^2 - \sum_{i=1}^{k} n_i \bar{x}_i^2$$

La variación total:

$$\sum_{T} x_{ij}^2 - T\bar{\bar{x}}_i^2$$

Luego, sólo se calculan 3 términos:

$$\sum_{T} x_{ij}^2 \;; \qquad \sum_{I=1}^{K} n_I \bar{x}_i^2 \qquad T\bar{\bar{x}}^2$$

X_{ij}	x_{ij}^2	x_{ij}	x_{ij}^2	x_{ij}	x_{ij}^2
300	90,000	600	360,000	700	490,000
400	160,000	300	90,000	300	90,000
300	90,000	300	90,000	400	160,000
500	250,000	400	160,000	600	360,000
0	0			500	250,000
	590,000		700,000		1,350,000

Así $\sum_{T} x_{ij}^2$ = 590,000 +700,000 + 1,350,000 = 2,640,000

$$\sum_{I=1}^{K} n_I \bar{x}_i^2 = n_1\bar{x}_1^2 + n_2\bar{x}_2^2 + n_3\bar{x}_3^2 = 5(300)^2 + 4(400)^2 + 5(500)^2 =$$

=450,000 + 640,000 + 1,340,000 = 2,340,000

Finalmente = $T\bar{\bar{x}}^2 = 14(400)^2 = 14(160,000) = 2,240,000$

Luego : 1.- Variación entre grupos: $\sum_{I=1}^{K} n_I \bar{x}_i^2 - T\bar{\bar{x}}^2$ = 2,340,000 - 2,240,000 = 100,000

2.- Variación dentro de los grupos:

$$\sum_{T} x_{ij}^2 - \sum_{i=1}^{k} n_i \bar{x}_i^2 = 2,640,000 - 2,340,000 = 300,000.$$

3.- Variación total:

$$\sum_T x_{ij}^2 - T\bar{x}^2 = 2{,}640{,}000 - 2{,}240{,}000 = 400{,}000$$

Con estos datos se prueba la hipótesis nula y se llega a los mismos resultados que con el método directo.

COMENTARIOS FINALES:

Después de haber expuesto la forma en que se verifica estadísticamente una hipótesis de trabajo, podemos concluir diciendo que este instrumental es muy importante cuando se hacen investigaciones aplicando el método científico, ya que son fundamentales para el desarrollo de los estudios que hagamos; la aceptación o rechazo de la hipótesis nula influye en el cumplimiento de los objetivos establecidos para la solución de un determinado problema.

Si el método científico guía la investigación, dice el Dr. Raúl Rojas Soriano(19), la hipótesis, como estudio especñifico para verificar conjeturas sobre la naturaleza y solución del problema, coadyuva a la obtención de resultados que enriquecen y aceleran el cumplimiento de los objetivos planteados.

Por ello recomienda que debe plantearse con claridad y precisión; sus conceptos deben contar con referencias empíricas y siempre formularse en términos afirmativos para garantizar que sus hallazgos coadyuven a la solución de los problemas.

Se espera que los métodos correspondientes a la estadística descriptiva y a la estadística inductiva, por la manera en que se expuso eilustró su aplicación, corroboren su importancia en el análisis de fenómenos económicos, tanto micro (a nivel de empresa) como macro (a nivel de país) y despierten en el lector el interés por profundizar en su manejo ampliado hacia otros campos del conocimiento.

VIII. PRÁCTICA XI

PRUEBA DE HIPÓTESIS

1.- Establezca la diferencia entre: a) Una hipótesis nula y una hipótesis alternativa; b) Un error tipo I y un error tipo II.

2.- Explique lo siguiente: a) Nivel de significación; b) Valor critico; c) Región de rechazo; Región de aceptación.

3.- Indique la diferencia entre: a) Una prueba de dos extremos y una prueba de un extremo; b) Una prueba de extremo izquierdo y una de extremo derecho.

4.- Describa el procedimiento básico para hacer una prueba de hipótesis, haciendo énfasis cuando el tamaño de la muestra es grande y cuando es pequeña, así como cuando la desviación estándar de la población, σ, es conocida, y cuando σ es desconocida y el error estándar del estadístico, $S\bar{x}$, se estima a partir de una muestra.

Problema 1.- La media y la desviación estándar de la resistencia de cuerdas producidas por una compañía A, fueron 600 libras y 40 libras respectivamente. Se acaba de aplicar una nueva técnica en el proceso de fabricación. Se piensa que la resistencia de las cuerdas puede aumentar con este proceso. Para ello el gerente de producción tomo una muestra de 64 cuerdas cuya media es de 609 libras. ¿Se puede concluir que hay un incremento de la resistencia media con $\alpha = 5\%$?

Problema 2 Los laboratorios de medicina "Anahuac" sostienen que su producto, "Vuelve a la vida" fue 95% efectivo en mitigar los sufrimientos de la fiebre en un período de 5 horas. Una muestra de 150 personas que usaron el producto indica que produjo alivio para 138 personas. ¿Cree que la afirmación hecha por "Anahuac" es valida al nivel de significación de 0.10?

Problema 3 Una muestra de calificaciones de 80 estudiantes en una clase de estadística está dada en las columnas (1 y 2) de la siguiente tabla. El número teórico de estudiantes para cada clase que figura en la columna 3 se obtuvo mediante la curva normal. Determine si hay una diferencia significativa, usando la X^2 entre las calificaciones esperadas o teóricas y las observadas en la clase de estadística con $\alpha = 0.05$.

1	2	3
CALIFICACIONES	NÚMERO DE	

(Intervalo de clase)	ESTUDIANTES Real	Teorico
20 - 29	3	1
30 - 39	6	3
40 - 49	5	8
50 - 59	7	13
60 - 69	10	17
70 - 79	29	16
80 - 89	12	12
90 - 99	8	6
99.5 y más	0	4
Total	80	80

Problema 4.- Se tomo una muestra de los salarios de 13 trabajadores clasificados en electricistas, carpinteros y pintores. Probar si los salarios medios de estas tres categorías de trabajadores difieren significativamente con α = 5% y α = 1%.

Número de trabajadores en cada muestra	Salarios de los trabajadores		
	Electricistas $	Carpinteros $	Pintores $
1	74	75	56
2	65	78	55
3	72	74	53
4	69	76	52
5		72	
Total	280	375	216

SOLUCIÓN PRÁCTICA XI

Solución del problema 1

Datos:

μ = 600 libras de resistencia

σ = 40 libras

n = 64 cuerdas

$\bar{x}$ = 609 libras

α = 5%

H_o: μ > 600 libras de resistencia

H_i: $\mu < 600$ libras
luego $Z_\alpha = \pm\ 1.645$

La prueba se hace con: $Z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ donde $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ porque n >30

Sustituyendo

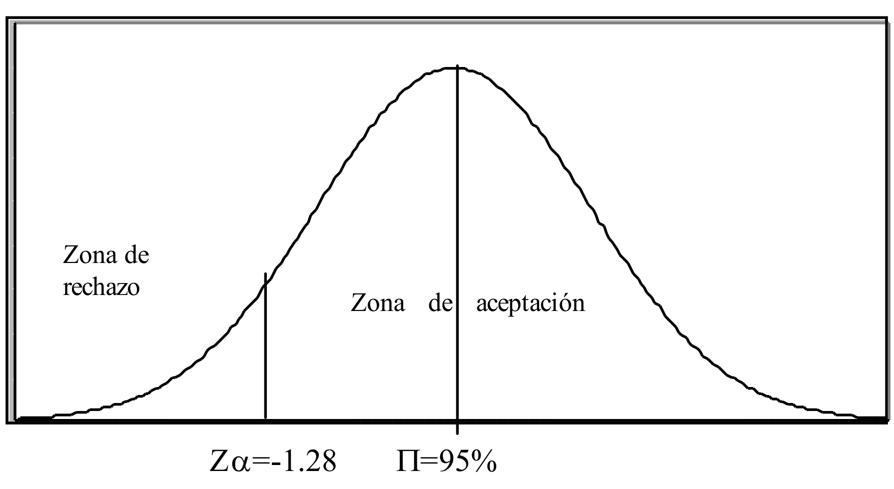
luego $Z = \dfrac{609 - 600}{5} = \dfrac{9}{5} = 1.8$ así $\sigma_{\bar{x}} = \dfrac{40}{\sqrt{64}} = \dfrac{40}{8} = 5$

$\sigma\bar{x} = 5$

Como $Z=1.8 > Z_\alpha=1.645$ se toma la decisión de rechazar la hipótesis nula de que hay un incremento de la resistencia de las cuerdas en la nueva técnica de fabricación es decir, no mejoro su resistencia.

Solución del problema No. 2
Datos:
$\Pi = 95\%$ en un período de 5 horas
$P = 92\% = 138/150$
$n = 150$
$\alpha = 0.10$
luego $Z_\alpha = \pm\ 1.280$
con H_o: $\Pi = 95\%$ en menos de cinco horas

La prueba es de una cola o extremo, puesto que el alivio fue en un período de cinco horas; la hipótesis alternativa es:
H_i: $\Pi = 95\%$ en más de cinco horas: El alivio fue para un período mayor de cinco horas.



La prueba se hace con $Z = \frac{p - \Pi}{\sigma_p}$ donde $\sigma_p = \sqrt{\frac{p * q}{n}}$

porque n>30 luego $\sigma_p = \sqrt{\frac{(0.92)(0.08)}{150}} = \sqrt{0.0004906} = 0.022$

Así tenemos que $Z = \frac{0.92 - 0.95}{0.022}$ por lo tanto Z = -1.36

Decisión: Se toma la decisión de rechazar la hipótesis de que la medicina es efectiva en un 95% en un periodo de 5 horas.

Solución al problema No. 3, de X^2, usando dos métodos.

Antecedentes: Se acostumbra como una regla de seguridad para aplicar la distribución X^2, que la frecuencia esperada o en este caso el número teórico de estudiantes, en cada clase deberá ser cuando menos cinco. Así, cuando hay frecuencias pequeñas en cada clase, éstas deberán ser combinadas para llenar los requisitos.

MÉTODO DEL PROFESOR S. SHAO.

CALIFICACIONES	NÚMERO DE ESTUDIANTES				$\frac{(\text{Real-Teórico})^2}{\text{Teórico}}$
(Intervalo de clase)	Real	Teórico	Real-Teórico	$(\text{Real-Teórico})^2$	
20 - 49	14	12	2	4	0.333333333
= 50 - 59	7	13	-6	36	2.769230769
60 - 69	10	17	-7	49	2.882352941
70 - 79	29	16	13	169	10.5625

80 - 89	12	12	0	0	0
90 y más	8	10	-2	4	0.4
Total	80	80	0	X^2=	16.9474170 4

Grados de libertad= n-1= 6-1= 5; con α = 5%, la $X^2_{\text{TEÓRICA}}$ = 11.070

Luego se rechaza la Ho porque es mrenor que la real u observada.

MÉTODO DEL PROFESOR GENARO SÁNCHEZ BARAJAS

Clases	R - T =	$(R-T)^2$	$\frac{(R-T)^2}{T}$
20 - 29	3 - 1 = 2	4	4
30 - 39	6 - 3 = 3	9	3
40 - 49	5 - 8 = -3	9	1.125
50 - 59	7 - 13 = -6	36	2.7692
60 - 69	10 - 17 = -7	49	2.8824
70 - 79	29 - 16 = 13	169	10.5625
80 - 89	12 - 12 = 0	0	0.0000
90 - 99	8 - 6 = 2	4	0.6667
99.5 y más	0 - 4 = -4	16	4.0000
Total	Σ (R-T) = 0	X^2=	29.0058

Grados de libertad =n-1= 9-1= 8

Luego con α = 5%

X^2_α =15.507

Decisión : se toma la decisión de rechazar la hipótesis nula porque $X^2_\alpha < X^2$.

Como G.L. = n - 1 = 9 - 1 = 8
Y α= 5% tenemos X^2_α = 15.507
Decisión: se toma la decisión de rechazar la hipótesis nula de que las calificaciones reales y las esperadas o teóricos no difieren significativamente, puesto que $X^2_\alpha = 15.507 < X^2 = 29.005$

CONCLUSIÓN: CON LOS DOS MÉTODOS SE TOMA LA MISMA DECISIÓN

Solución del problema No.4, calculando F con el método abreviado.

1. Calculamos $\sum_{t} X^2{}_{IJ},$

X_{ij}	X_{1j}^2	X_{2j}	X_{2j}^2	X_{3j}	X_{3j}^2
74 65 72 69	5,476 4,225 5,184 4,761	75 78 74 76 72	5,625 6,084 5,476 5,776 5,184	56 55 53 52	3,136 3,025 2,809 2,704
	19,646		28,145		11,674

$\overline{X}_1 = 280/5 = 70; \overline{X}_2 = 375/5 = 75; \overline{X}_3 = 216/4 = 54; \overline{\overline{X}} = 871/13 = 67$

Así $\sum_{t} X^2{}_{IJ} = 19{,}646 + 28{,}145 + 11{,}674 = 59{,}465.$

2.- Calculamos:

$\sum_{t}^{K} n_i \overline{x}_i^2 = n_1\overline{x}_1^2 + n_2\overline{x}_2^2 + n_3\overline{x}_3^2 = 4(70)^2 + 5(75)^2 + 4(54)^2 = 19{,}600 + 28{,}125 +$ 11,664 = 59,389.

3.- $T\overline{\overline{x}}^2 = 13(67)^2 = 13(4{,}489) = 58{,}357.$

luego: variación entre grupos:

$$\sum_{t}^{K} n_i \overline{x}_i^2 - T\overline{\overline{x}} = 59{,}389 - 58{,}357 = 1.032$$

Variación dentro de los grupos:

$$\sum_{t} X^2{}_{IJ} - \sum_{i=1}^{k} n_i \overline{x}_i^2 = 59{,}465 - 59{,}389 = 76$$

Variación total:

$$\sum_{t} X^2{}_{IJ} - T\overline{\overline{x}}^2 = 59{,}465 - 58{,}357 = 1{,}108$$

Así			
Variación	Grados de libertad	Suma de cuadrados	Varianza
entre grupos	K - 1 = 3 - 2 = 2	1,032	516

Dentro de Grupos	T - K = 13 - 3 = 10	76	7.6
Total	T - 1 = 13 - 1 = 12	1,108	92.3

$F(2,10) = 516/7.6 = 67.9$

a) En el apéndice F vemos que con α=5% y G.L. (2 y 10) tenemos F_α=4.10, luego se rechaza la hipótesis nula.

b) En el apéndice F vemos que con α = 1% y G.L. (2 y 10) F_α=7.56 también se rechaza la hipótesis nula, porque en ambos casos con $F_{\alpha(2,10)} = 67.9 > 4.10$ y > 7.56.