62 20

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE QUIMICA

UNIDAD DIDACTICA PARA LA ESTADISTICA II
UNDIES II

T E S I S

QUE PARA OBTENER EL TITULO DE

INGENIERO QUIMICO

P R E S E N T A

JUAN MANUEL LOMELIN BAUMEISTER

MEXICO, D. F. 1990

FALLA DE ORIGEN

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

Tabla de Contenido

1.- ESTIMACION

1.1.- Definición

Esta parte del sistema UNDIES II cubre la primera área de la "inferencia estadística" y permite obtener metódicamente la estimación de diversos parámetros a partir de datos básicos de una o varias muestras.

En la mayor parte de este capítulo obtendremos las estimaciones clásicas de parámetros desconocidos de la población, tales como la media, la proporción y la desviación estándar, calculando para ello las estadísticas tomadas de muestras aleatorias y aplicando la teoría de las distribuciones muestrales.

La estimación de un parámetro de la población puede darse en forma de estimación puntual o de estimación por intervalo.

Una estimación puntual de algún parámetro de la población es un valor simple de una estadística.

Una estimación por intervalo de un parámetro de la población es un intervalo de anchura finita, centrado en la estimación puntual del parámetro que se espera, contenga el verdadero valor del parámetro.

Al elegir un estimador se prefiere aquél cuya media sea igual a la del parámetro estimado, o en otras palabras, que sea insesgado

1.2.- Menú

1.2.1.- Estimación de la Media (μ)

1.2.1.1 -> Con σ conocida y n>30

INTERVALO DE CONFIANZA para μ al A(1-α)100% :

$$\boxed{\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

donde $\bar{x}$ es la media de una muestra de tamaño n, tomada de una población con variancia σ^2 conocida y $z_{\alpha/2}$como valor de la distribución normal estándar que deja un área $\alpha/2$ hacia la derecha.

1.2.1.2 -> Con σ desconocida y n<30

INTERVALO DE CONFIANZA para μ al A(1-α)100% :

$$\bar{x} - \frac{t_{\alpha/2} \cdot s}{\sqrt{n}} < \mu < \bar{x} + \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}$$

donde $\bar{x}$ y s son respectivamente la media y la desviación estándar de una muestra de tamaño n<30 , tomada de una población aproximadamente normal, y $t_{\alpha/2}$ es el valor de la distribución t, con $\nu = n-1$ grados de libertad, que deja un área de $\alpha/2$ hacia la derecha.

1.2.2.- Estimación de la Diferencia entre Dos Medias ($\mu_1 - \mu_2$)

1.2.2.1 -> Con σ_1^2 y σ_2^2 conocidas

INTERVALO DE CONFIANZA para $\mu_1 - \mu_2$ al A(1 - α)100% :

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

donde $\bar{x}_1$ y $\bar{x}_2$ son las medias de las muestras aleatorias independientes de tamaños n_1 y n_2, tomadas de poblaciones con variancias conocidas σ_1^2 y σ_2^2 respectivamente, y $z_{\alpha/2}$ es el valor de la distribución normal estándar que deja un área de $\alpha/2$ hacia la derecha.

1.2.2.2 -> Con $\sigma_1^2 = \sigma_2^2$ pero desconocidas

INTERVALO DE CONFIANZA EN MUESTRAS PEQUEÑAS para $\mu_1 - \mu_2$ al A(1 - α)100% :

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras independientes pequeñas de tamaños n_1 y n_2 respectivamente, tomadas de distribuciones aproximadamente normales, s_p es la desviación estándar mancomunada y $t_{\alpha/2}$ es el valor de la distribución t con n_1+n_2-2 grados de libertad que deja un área de $\alpha/2$ hacia la derecha.

Una estimación puntual de la variancia común σ^2 desconocida, puede obtenerse fusionando las variancias de la muestra, para obtener así el estimador mancomunado:

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

1.2.2.3 -> Para n>30 , con $\sigma_1^2 \neq \sigma_2^2$ y desconocidas

INTERVALO DE CONFIANZA para $\mu_1-\mu_2$ al A(1 - α)100% :

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

donde $\bar{x}_1$ y σ_1^2, y $\bar{x}_2$ y σ_2^2, son respectivamente las medias y las variancias de las muestras independientes pequeñas de tamaños n_1 y n_2, tomadas de distribuciones aproximadamente normales y $t_{\alpha/2}$ es el valor de la distribución t con

$$v=\frac{(s_1^2/n_1+s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)]+[(s_2^2/n_2)^2/(n_2-1)]}$$

grados de libertad que deja un área de $\alpha/2$ hacia la derecha.

1.2.2.4 -> Para Observaciones Apareadas ($\mu_1-\mu_2=\mu_d$)

INTERVALO DE CONFIANZA para $\mu_1-\mu_2=\mu_d$ al A(1 - α)100% :

$$\bar{d}-t_{\alpha/2}\cdot\frac{s_d}{\sqrt{n}}<\mu_d<\bar{d}+t_{\alpha/2}\cdot\frac{s_d}{\sqrt{n}}$$

donde $\bar{d}$ y s_d son respectivamente la media y la desviación estándar de las diferencias de n pares de mediciones y $t_{\alpha/2}$ es el valor de la distribución t, con $\nu = n-1$ grados de libertad, que deja un área de $\alpha/2$ hacia la derecha.

> Como ayuda adicional se recuerda que :
> $$s_d^2 = \frac{n \cdot \Sigma d_i^2 - (\Sigma d_i)^2}{n(n-1)}$$

1.2.3.- Estimación de una Proporción o Parámetro Binomial (p)

1.2.3.1 -> Con $n \geq 30$

INTERVALO DE CONFIANZA para p al A(1 - α)100% :

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

donde $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n . $\hat{q} = 1 - \hat{p}$ y $z_{\alpha/2}$ es el valor de la distribución normal estándar que deja un área de $\alpha/2$ hacia la derecha.

1.2.4.- Estimación de la Diferencia entre Dos Proporciones o Parámetros Binomiales ($p_1 - p_2$)

1.2.4.1 -> Con $n_1 \geq 30$ y $n_2 \geq 30$

INTERVALO DE CONFIANZA para $p_1 - p_2$ al A(1 - α)100% :

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

donde $\hat{p}_1$ y $\hat{p}_2$ son la proporción de éxitos en muestras aleatorias de tamaños n_1 y n_2 respectivamente, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$, y $z_{\alpha/2}$ es el valor de la distribución normal estándar que deja un área de $\alpha/2$ hacia la derecha.

1.2.5.- Estimación de la Variancia (σ^2)

1.2.5.- Estimación de la Varianza (σ^2)

1.2.5.1 -> Con población normal

INTERVALO DE CONFIANZA para σ^2 al A(1-α)100% :

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

donde s^2 es la variancia de la muestra aleatoria de tamaño n, y $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son los valores de una distribución chi cuadrada con $v=n-1$ grados de libertad que deja áreas de $\alpha/2$ y $1-\alpha/2$ respectivamente, hacia la derecha.

Para el cálculo de la variancia de la muestra se utiliza:

$$s^2 = \frac{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

1.2.6.- Estimación del Cociente de Dos Variancias (σ_1^2/σ_2^2)

1.2.6.1 -> Con poblaciones normales y muestras independientes.

INTERVALO DE CONFIANZA para σ_1^2/σ_2^2 al A(1-α)100% :

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2}(v_2,v_1)$$

donde s_1^2 y s_2^2 son las variancias de muestras independientes de tamaños n_1 y n_2 respectivamente, tomadas de poblaciones normales, $f_{\alpha/2}(v_1,v_2)$ es el valor de la distribución F con $v_1=n_1-1$ y $v_2=n_2-1$ grados de libertad, dejando un área de $\alpha/2$ hacia la derecha y $f_{\alpha/2}(v_2,v_1)$ es un valor análogo de la distribución F para $v_2=n_2-1$ y $v_1=n_1-1$ grados de libertad .

1.2.7.- Limites de Tolerancia, Errores Máximos, Tamaño de muestras

1.2.7.1 -> Limites de Tolerancia para una Población Normal

Para una distribución normal de mediciones con media μ y desviación estándar σ desconocidas, los límites de tolerancia están dados por :

$$\bar{x} \pm k \cdot s$$

donde k es el Factor de Tolerancia determinado de manera que se puede afirmar que con una tolerancia al $\gamma \cdot$ 100% los límites dados contienen la proporción $1 - \alpha$ de las mediciones.

NOTA: un intervalo de tolerancia necesariamente debe ser más largo que un intervalo de confianza con el mismo grado de confianza, ya que el intervalo de tolerancia se basa en parámetros más generales, mientras que un intervalo de confianza se construye tomando en cuenta más elementos de juicio.

1.2.7.2 -> Error Máximo si se estima la Media (μ) con la Media Muestral ($\bar{x}$)

Si se utiliza $\bar{x}$ como estimador de μ, se puede tener una confianza de $(1 - \alpha)100\%$ de que el error será menor que :

$$e \leq \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$$

1.2.7.3 -> Error Máximo si se estima el Parámetro Binomial (p) con la Proporción Muestral ($\hat{p}$)

Si se utiliza $\hat{p}$ como estimador de p, se puede tener una confianza de $(1 - \alpha)100\%$ de que el error será menor que :

$$e \leq z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

1.2.7.4 -> Tamaño de Muestra mínimo para estimar la Media (μ) a partir de $\bar{x}$ con cierto error máximo fijado

Si se utiliza $\bar{x}$ como estimador de μ, se puede tener una confianza de $(1 - \alpha)100\%$ de que el error será menor que una cantidad especificada e, cuando el tamaño de muestra es:

$$\boxed{n = \left(\frac{z_{\alpha/2}\sigma}{e} \right)^2}$$

1.2.7.5 -> Tamaño de Muestra mínimo para estimar al Parámetro Binomial (p)

Si se utiliza $\hat{p}$ como estimador de p, se puede tener una confianza de $(1 - \alpha)100\%$ de que el error será menor que una cantidad especificada e, cuando el tamaño de muestra es:

$$\boxed{n = \frac{z_{\alpha/2}^2 \hat{p} \hat{q}}{e^2}}$$

1.2.7.6 -> Tamaño de Muestra mínimo para estimar al Parámetro Binomial (p) sin previa estimación de la Proporción Muestral ($\hat{p}$) .

Si se utiliza $\hat{p}$ como estimador de p, se puede tener al menos una confianza de $(1 - \alpha)100\%$ de que el error será menor que una cantidad especificada e, cuando el tamaño de muestra es:

$$\boxed{n = \frac{z_{\alpha/2}^2}{4e^2}}$$

1.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pida.

En los casos en los que existan restricciones, como tamaño de muestra u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menú previo.

1.4.- Problemas Resueltos

1.4.1.- Problema #1

De un grupo de estudiantes se calculó su media y su desviación estándar respecto a cierta puntuación.Establece el intervalo de confianza al 95% para la media de la población estudiantil.

Solución:

Nivel de Confianza	95%
n	36
z para 0.025	1.96
media muestral	2.6
desviación estándar	0.3

El intervalo se encuentra entre 2.502 y 2.698 .

1.4.2.- Problema #2

Los contenidos de metionina en siete frascos similares son 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 y 9.6 gramos. Establece un intervalo de confianza al 95% para la media de todos los recipientes en la fábrica, suponiendo una población aproximadamente normal.

Solución:

Nivel de Confianza	95%
n	7
t al 0.025 y 6º	2.447
media muestral	10.0
desv.est. muestral	0.283

El intervalo se encuentra entre 9.738 y 10.262 .

1.4.3.- Problema #3

Se analizan dos tipos diferentes de preparaciones, 50 de A y 75 de B.

El contenido promedio del ingrediente activo es de 76 en las preparaciones A, con una desviación estándar de 6. En el caso de las B se tienen 82 y 8 respectivamente. Encuentra el intervalo de confianza al 96% para la diferencia de medias entre las preparaciones.

Solución:

Nivel de Confianza	96%
n para A	50
n para B	75
s de A	6
s de B	8
z al 0.02	2.054
promedio A	76
promedio B	82

Ya que se tiene un tamaño de muestra mayor de 30, pueden sustituirse los valores de la desviación estándar de la población por los valores conocidos de la desviación estándar de las muestras.

El intervalo se encuentra entre 3.424 y 8.576 .

1.4.4.- Problema #4

En diversos procesos químicos se comparan dos catalizadores similares para medir su efecto en la reacción resultante.
Se prepara una muestra de 12 experimentos utilizando el catalizador A y una muestra de 10 experimentos empleando el catalizador B. Los experimentos con A arrojaron un resultado promedio de 85 con desviación estándar de 4, mientras que el promedio de la segunda muestra dio un promedio de 81 y una desviación estándar de 5. Establece el intervalo de confianza al 90% para la diferencia entre las medias de las poblaciones, suponiendo que éstas tienen distribuciones aproximadamente normales con variancias iguales.

Solución:

Nivel de Confianza	90%
n para A	12
n para B	10
var. de A	16
var. de B	25
t al 0.05 y 20º	1.725
promedio A	85
promedio B	81

El intervalo se encuentra entre 0.693 y 7.307 .

1.4.5.- Problema #5

Se desea verificar la diferencia en la precisión de dos aparatos de medición de una cierta propiedad física. Ambos aparatos son similares y las corridas se realizan con la misma técnica y con los mismos parámetros fijados.En el aparato A se realizaron 15 mediciones, obteniéndose un promedio de 4.93 con una desviación estándar de 1.14.
En el aparato B se realizaron 10 mediciones de promedio 2.64 y con 0.66 de desviación estándar.
Encuentre un intervalo de confianza del 95% para la diferencia verdadera de las mediciones promedio entre los dos aparatos, suponiendo que las observaciones se han tomado de poblaciones normales con variancias diferentes.

Solución:

Nivel de Confianza	95%
n para A	15
n para B	10
var. de A	1.30
var. de B	0.44
t al 0.025 y 23º	2.069
promedio A	4.93
promedio B	2.64

El intervalo se encuentra entre 1.542 y 3.038 .

1.4.6.- Problema #6

Diez parejas de estudiantes de cociente intelectual aproximadamente igual se eligen para determinar la eficacia de dos sistemas de aprendizaje distintos. Uno de los estudiantes de cada pareja se elige al azar y se asigna al sistema A de material programado, mientras que el otro se asigna al sistema con profesor. Al final del lapso de prueba de los sistemas de enseñanza, los estudiantes presentan el mismo examen.
Los resultados de las calificaciones se presentan a continuación:

Pareja	Material Programado	Profesor
1	76	81
2	60	52
3	85	87
4	58	70
5	91	86
6	75	77
7	82	90
8	64	63
9	79	85
10	88	83

Encuentra un intervalo de confianza al 98% para la diferencia real entre los promedios de calificaciones de los dos sistemas presentados.

Solución:

Nivel de Confianza	98%
n	10
t al 0.01 y 9º	2.821
promedio de dif.	-1.6
s de las dif.	6.38

El intervalo se encuentra entre -7.291 y 4.091 .

1.4.7.- Problema #7

En una muestra aleatoria de 500 familias de Ingenieros Químicos de cierta ciudad se observó que todos poseían televisor, pero sólo 340, de color.
Encuentre un intervalo de confianza al 95% para la proporción real de familias en dicha ciudad con receptores de color.

Solución:

Nivel de Confianza	95%
n	500
proporción muestral	.68
z al 0.025	1.96

El intervalo se encuentra entre 0.639 y 0.721 .

1.4.8.- Problema #8

En el proceso de fabricación de cierta válvula se está considerando cierto cambio. Con objeto de deteminar si el nuevo procedimiento es mejor, se toman muestras tanto del procedimiento existente como del nuevo. Si se detectan 75 válvulas defectuosas de una muestra de 1500 del procedimiento existente, así como 80 de 2000 del nuevo procedimiento, encuentre un intervalo de confianza al 90% para la diferencia real en la fracción de válvulas defectuosas entre ambos procesos.

Solución:

Nivel de Confianza	90%
n para Proc.Exist.	1500
n para Proc.Nuevo	2000
prop.muestr. Exist.	.05
prop.muestr. Exist.	.04
z al 0.05	1.645

El intervalo se encuentra entre -0.00173 y 0.02173 .

1.4.9.- Problema #9

A continuación se listan los valores de contenido de 10 frascos de reactivo QP envasados por cierta compañía : 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2 y 46.0 . Encuentre un intervalo de confianza al 95% para la variancia de dichos frascos distribuidos por esa compañía.

Solución:

Nivel de Confianza	95%
n de frascos	10
var. muestral	0.2862
Ji cuad.al 0.025 y 9º	19.023
Ji cuad.al 0.975 y 9º	2.700

El intervalo se encuentra entre 0.135 y 0.954 .

1.4.10.- Problema #10

Un grupo de 25 varones y 16 mujeres presentan el mismo examen de matemáticas. Los varones obtienen una calificación promedio de 82 con una desviación estándar de 8, mientras que las mujeres obtienen una calificación promedio de 78 con una desviación estándar de 7. Encuentre un intervalo de confianza al 98% para el cociente de variancias de las poblaciones de calificaciones de muchachos y muchachas respectivamente, que hayan presentado dicho examen.

Solución:

Nivel de Confianza	98%
n de varones	25
n de mujeres	16
var.muestr. varones	64
var.muestr. mujeres	49
F al 0.01 y (24,15)º	3.29
F al 0.01 y (15,24)º	2.89

El intervalo se encuentra entre 0.397 y 3.775 .

1.4.11.- Problema #11

Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de estas piezas y se verifica que los diámetros son respectivamente 1.10, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros. Encuentre los límites de tolerancia al 99% que contenga al 95% de las piezas producidas por esta máquina, suponiendo una distribución aproximadamente normal.

Solución:

Nivel de Confianza	95%
n de piezas	9
Límite de Tolerancia	99%
K de 0.99 y al 95%	4.550
promedio muestral	1.0156
desv. estándar muestr.	0.0400

El intervalo de Tolerancia se encuentra entre 0.8336 y 1.1976 .

1.4.12.- Problema #12

Encuentre el error máximo al estimar, al 99% de confianza, la media de una muestra de tamaño 36, con promedio de 2.6 y desviación estándar de 0.3 .

Solución:

Nivel de Confianza	99%
z al 0.005	2.575
n muestral	36
desviación estándar	0.3

El error máximo sería de 0.12875 .

1.4.13.- Problema #13

Encuentre el error máximo al estimar, al 99% de confianza, la proporción de una muestra de tamaño 500, con proporción muestral de 0.72.

Solución:

Nivel de Confianza	99%
z al 0.005	2.575
n muestral	500
prop. muestral	0.72

El error máximo sería de 0.05171 .

1.4.14.- Problema #14

Encuentre el tamaño de muestra mínimo para obtener, con un nivel de confianza del 95%, un error máximo de 0.05 al estimar la media de una población con desviación estándar de 0.4 .

Solución:

Nivel de Confianza	95%
z al 0.025	1.96
desviación estándar	0.4
error máx. fijado	0.05

El tamaño de muestra debe ser de por lo menos 246 .

1.4.15.- Problema #15

Encuentre el tamaño de muestra mínimo para obtener, con un nivel de confianza del 99%, un error máximo de 0.02 al estimar la proporción de una población a partir de una muestra con proporción de 0.72 .

Solución:

Nivel de Confianza	99%
z al 0.005	2.575
prop. muestral	0.72
error máx. fijado	0.02

El tamaño de muestra debe ser de por lo menos 3342 .

1.4.16.- Problema #16

Encuentre el tamaño de muestra mínimo para obtener, con un nivel de confianza del 95%, un error máximo de 0.3 al estimar la proporción de una población, si se desconoce inclusive la proporción muestral.

Solución:

Nivel de Confianza	95%
z al 0.025	1.96
error máx. fijado	0.03

El tamaño de muestra debe ser de por lo menos 1067 .

2.- PRUEBA DE HIPOTESIS

2.1.- Definición

Esta parte del sistema UNDIES II cubre la segunda área de la "inferencia estadística" y nos permite decidir sobre si una hipótesis es verdadera o falsa.

Hipótesis estadística es una suposición o enunciado relativos a una o más poblaciones, que puede o no ser verdadera(o).

La verdad o falsedad de una hipótesis nunca se conoce con certeza si no se examina toda la población; sin embargo, como eso resulta poco práctico la mayor parte de las veces, se toma una muestra aleatoria de la población y se usa la información contenida en ella para tomar la decisión de si la hipótesis es verdadera o falsa.

En este capítulo se utilizan frecuentemente los términos aceptar o rechazar, pero cabe advertir que el rechazo de una hipótesis implica meramente que no se tiene evidencia para creer lo contrario. Debido a esta terminología, se han de establecer siempre como hipótesis a aquellas que se esperan rechazar.

Para la prueba de una hipótesis se fija primeramente una hipótesis nula, (o inicial), contra una hipótesis alterna, (o alternativa). Adicionalmente se establece un valor crítico para una estadística relativa a la hipótesis.

Si el valor calculado de la estadística cae arriba del valor crítico fijado, se dice que se cae en la región crítica, mientras que si cae abajo del valor crítico, se dice que se cae en la región de aceptación.

Si la estadística calculada cae en la región crítica, se rechaza la hipótesis nula en favor de la hipótesis alterna. Si cae en la región de aceptación, la hipótesis nula se acepta.

Este procedimiento de decisión puede conducir a dos tipos de errores llamados error tipo I y error tipo II.

"Se ha cometido un error tipo I si se rechaza la hipótesis nula cuando ésta es verdadera"

"Se ha cometido un error tipo II si se acepta la hipótesis nula cuando ésta es falsa"

A la probabilidad de cometer un error tipo I se le llama nivel de significancia de la prueba y se denota por α .

A la probabilidad de cometer un error tipo II se le denota por β, y $1-\beta$ se conoce como la potencia de la prueba.

-El error tipo I y el error tipo II están relacionados. Una disminución en la probabilidad de uno resulta generalmente en un aumento de la probabilidad del otro.

-El tamaño de la región crítica, y por lo tanto la probabilidad de cometer un error tipo I, puede reducirse siempre ajustando el valor o los valores críticos.

-Un aumento del tamaño de la muestra n reducirá las probabilidades α y β simultáneamente.

-Si la hipótesis nula es falsa, β es un máximo cuando el verdadero valor de un parámetro se acerca al valor supuesto. Cuanto mayor sea la distancia entre el valor verdadero y el valor supuesto, menor será β .

Una prueba de cualquier hipótesis estadística en que la alternativa sea unilateral, tal como :

$H_0:\theta=\theta_0$ $H_1:\theta>\theta_0$	$H_0:\theta=\theta_0$ $H_1:\theta<\theta_0$

se llama prueba de una cola o unilateral.
Las regiones críticas se encuentran totalmente a la derecha o izquierda de la distribución.

Una prueba de cualquier hipótesis estadística en que la alternativa es bilateral como :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

se llama prueba de dos colas o bilateral.
La región crítica corresponde a ambos extremos de la distribución.

Al resolver un problema de prueba de hipótesis conviene seguir un procedimiento ordenado para llegar sistemáticamente a un adecuado resultado.
El procedimiento recomendado se le conoce como el de los "6 pasos" :

1º Establecer la Hipótesis Nula $H_0 : \theta = \theta_0$.
2º Establecer la Hipótesis Alterna $H_1 : \theta < \theta_0$, $\theta > \theta_0$ ó $\theta \neq \theta_0$

3º Escoger un nivel de significancia igual a α .
4º Seleccionar la estadística apropiada y establecer región crítica.
5º Calcular la estadística a partir de una muestra de tamaño n.
6º Llevar a cabo la decisión de aceptación o rechazo.

2.2.- Menú

2.2.1.- Prueba de Hipótesis de la Media ($\mu = \mu_0$)

2.2.1.1 -> Con σ conocida y n>30

Hipótesis Nula H_0	Estadística de Prueba
$\mu = \mu_0$	$Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$\mu < \mu_0$	$Z < -z_\alpha$
$\mu > \mu_0$	$Z > z_\alpha$
$\mu \neq \mu_0$	$Z < -z_{\alpha/2}$
	y $Z > z_{\alpha/2}$

2.2.1.2 -> Con σ desconocida y n<30

Hipótesis Nula H_0	Estadística de Prueba
$\mu = \mu_0$	$T = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$\mu < \mu_0$	$T < -t_\alpha$
$\mu > \mu_0$	$T > t_\alpha$
$\mu \neq \mu_0$	$T < -t_{\alpha/2}$
	y $T > t_{\alpha/2}$

con $v = n - 1$ grados de libertad.

2.2.2.- Prueba de Hipótesis de la Igualdad de Dos Medias $(\mu_1 - \mu_2 = d_0)$

2.2.2.1 -> Con σ_1^2 y σ_2^2 conocidas

Hipótesis Nula H_0	Estadística de Prueba
$\mu_1 - \mu_2 = d_0$	$Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$
Hipótesis Alternas H_1	**Regiones Criticas**
$\mu_1 - \mu_2 < d_0$	$Z < -z_\alpha$
$\mu_1 - \mu_2 > d_0$	$Z > z_\alpha$
$\mu_1 - \mu_2 \neq d_0$	$Z < -z_{\alpha/2}$
	y $Z > z_{\alpha/2}$

2.2.2.2 -> Con $\sigma_1^2 = \sigma_2^2$ pero desconocidas

Hipótesis Nula H_0	Estadística de Prueba
$\mu_1 - \mu_2 = d_0$	$T = \dfrac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p\sqrt{(1/n_1) + (1/n_2)}}$
Hipótesis Alternas H_1	**Regiones Criticas**
$\mu_1 - \mu_2 < d_0$	$T < -t_\alpha$
$\mu_1 - \mu_2 > d_0$	$T > t_\alpha$
$\mu_1 - \mu_2 \neq d_0$	$T < -t_{\alpha/2}$
	y $T > t_{\alpha/2}$

con $v = n_1 + n_2 - 2$ grados de libertad.

Una estimación puntual de la variancia común σ^2 desconocida, puede obtenerse fusionando las variancias de la muestra, para obtener así el estimador mancomunado:

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

2.2.2.3 -> Para n>30 , con $\sigma_1^2 \neq \sigma_2^2$ y desconocidas

Hipótesis Nula H_0	Estadística de Prueba
$\mu_1-\mu_2=d_0$	$T' = \frac{(\bar{X}_1-\bar{X}_2)-d_0}{\sqrt{(S_1^2/n_1)+(S_2^2/n_2)}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$\mu_1-\mu_2<d_0$	$T'<-t_\alpha$
$\mu_1-\mu_2>d_0$	$T'>t_\alpha$
$\mu_1-\mu_2\neq d_0$	$T'<-t_{\alpha/2}$ y $T'>t_{\alpha/2}$

$$\nu = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{\left[(s_1^2/n_1)^2/(n_1-1)\right]+\left[(s_2^2/n_2)^2/(n_2-1)\right]}$$

2.2.2.4 -> Para Observaciones Apareadas ($\mu_0 = d_0$)

Hipótesis Nula H_0	Estadística de Prueba
$\mu_D = d_0$	$T = \dfrac{\bar{D} - d_0}{S_d / \sqrt{n}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$\mu_D < d_0$	$T < -t_\alpha$
$\mu_D > d_0$	$T > t_\alpha$
$\mu_D \neq d_0$	$T < -t_{\alpha/2}$
	y $T > t_{\alpha/2}$

con $\nu = n - 1$ grados de libertad.

2.2.3.- Prueba de Hipótesis de una Proporción o Parámetro Binomial ($p = p_0$)

2.2.3.1 -> Con $n \geq 30$

Hipótesis Nula H_0	Estadística de Prueba
$p = p_0$	$Z = \dfrac{x - n p_0}{\sqrt{n p_0 q_0}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$p < p_0$	$Z < -z_\alpha$
$p > p_0$	$Z > z_\alpha$
$p \neq p_0$	$Z < -z_{\alpha/2}$
	y $Z > z_{\alpha/2}$

donde x es el número de éxitos en una muestra.

2.2.4.- Prueba de Hipótesis de la Igualdad de Dos Proporciones o Parámetros Binomiales ($p_1 = p_2$)

2.2.4.1 -> Con $n_1 \geq 30$ y $n_2 \geq 30$

Hipótesis Nula H_0	Estadística de Prueba
$p_1 = p_2$	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}[(1/n_1)+(1/n_2)]}}$
Hipótesis Alternas H_1	**Regiones Críticas**
$p_1 < p_2$	$Z < -z_\alpha$
$p_1 > p_2$	$Z > z_\alpha$
$p_1 \neq p_2$	$Z < -z_{\alpha/2}$
	y $Z > z_{\alpha/2}$

donde x_1 y x_2 son el número de éxitos de las dos muestras y, por lo tanto:

$$\hat{p}_1 = x_1/n_1$$

$$\hat{p}_2 = x_2/n_2$$

y

$$\hat{p} = \frac{(x_1 + x_2)}{(n_1 + n_2)}$$

$$\hat{q} = 1 - \hat{p}$$

2.2.5.- Prueba de Hipótesis de la Variancia ($\sigma^2 = \sigma_0^2$)

2.2.5.1 -> Con población normal

Hipótesis Nula H_0	Estadística de Prueba
$\sigma^2 = \sigma_0^2$	$X^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$
Hipótesis Alternas H_1	**Regiones Críticas**
$\sigma^2 < \sigma_0^2$	$X^2 < \chi^2_{1-\alpha}$
$\sigma^2 > \sigma_0^2$	$X^2 > \chi^2_{\alpha}$
$\sigma^2 \neq \sigma_0^2$	$X^2 < \chi^2_{1-\alpha/2}$
	y $X^2 > \chi^2_{\alpha/2}$

con $v = n-1$ grados de libertad.

2.2.6.- Prueba de Hipótesis de la Igualdad de Variancias ($\sigma_1 = \sigma_2$)

2.2.6.1 -> Con poblaciones normales y muestras independientes.

Hipótesis Nula H_0	Estadística de Prueba
$\sigma_1^2 = \sigma_2^2$	$F = S_1^2 / S_2^2$
Hipótesis Alternas H_1	**Regiones Críticas**
$\sigma_1^2 < \sigma_2^2$	$F < f_{1-\alpha}(v_1, v_2)$
$\sigma_1^2 > \sigma_2^2$	$F > f_{\alpha}(v_1, v_2)$
$\sigma_1^2 \neq \sigma_2^2$	$F < f_{1-\alpha/2}(v_1, v_2)$
	y $F > f_{\alpha/2}(v_1, v_2)$

con $\nu_1 = n_1 - 1$

y $\nu_2 = n_2 - 1$

grados de libertad.

Recordatorio: la función de probabilidad F posee la siguiente propiedad:

$$f_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{f_\alpha(\nu_2, \nu_1)}$$

2.2.7.- Tamaños de Muestra para Prueba de Medias

Esta sección está diseñada para el cálculo del tamaño de muestra necesario para reducir β la probabilidad de cometer un error tipo II o, en otras palabras, para regular la *potencia* de la prueba.

2.2.7.1 -> Con desviación estándar (σ) conocida.

Prueba de Una Cola :	Prueba de Dos Colas :
$n = \frac{(Z_\alpha + Z_\beta)^2 \cdot \sigma^2}{\delta^2}$	$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \cdot \sigma^2}{\delta^2}$

2.2.7.2 -> Con desviaciones estándar (σ_1 y σ_2) conocidas.

Prueba de Una Cola :	Prueba de Dos Colas :
$n = \frac{(Z_\alpha + Z_\beta)^2 \cdot (\sigma_1^2 + \sigma_2^2)}{\delta^2}$	$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \cdot (\sigma_1^2 + \sigma_2^2)}{\delta^2}$

2.2.7.3 -> Con desviación estándar (σ) desconocida.

Para obtener el tamaño de muestra en estas condiciones se requiere de la distribución *t de Student no central*. Existen tablas, como la IX al final del manual, en las que se leen los tamaños de muestra para Prueba de Hipótesis de Media para una y dos colas.
Pero para obtener el dato, se requiere de algunos datos :

α = nivel de significancia
β = 1 - *potencia de la prueba*

$$\lambda = \frac{\mu - \mu_0}{\sigma} = \frac{\delta}{\sigma}$$

donde δ y σ son estimadas o la primera es múltiplo de la segunda, para poder hacer el cálculo del cociente y hacer la lectura de las tablas.

2.2.7.4 -> Con desviaciones estándar (σ_1 y σ_2) desconocidas.

Este caso se resuelve por medio de tablas, igual que el anterior, pero considerando los siguientes datos :

α = nivel de significancia
β = 1 - *potencia de la prueba*

$$\lambda = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\delta}{\sigma}$$

2.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capitulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pidan.

En los casos en los que hay restricciones, como tamaño de muestra u otras, deben ser cumplidas éstas. De otra manera, el programa no continuará y regresará al menu previo.

2.4.- Problemas Resueltos

2.4.1.- Problema #1

Un fabricante de fibras ópticas ha diseñado un nuevo material que tiene una resistencia promedio de 80 N a la ruptura y una desviación estándar de 5 N . Pruebe la hipótesis de que la media es igual a 80 N contra la alternativa de que la media no es igual a 80 N, si se toma una muestra aleatoria de 50 cuerdas y se encuentra que tienen una resistencia promedio a la ruptura de 78 N. Emplee un nivel de significancia de 0.01 .

Solución :

Hipótesis Nula	media = 80 N
Hipótesis Alterna	media <> 80 N
Nivel de Significancia	0.01
n de la muestra	50
media hipotética	80 N
z al 0.005	-2.58 y 2.58
media muestral	78
desviación estándar	5

La Z calculada es de -2.83, por lo que se rechaza la hipótesis nula y se concluye que la resistencia es menor.

2.4.2.- Problema #2

El tiempo promedio empleado para registrar a un nuevo cliente, en una compañía de la industria química, ha sido de 50 minutos, con una desviación estándar de 10 minutos. Se desea adquirir un sistema en PC's para este registro, y para ello se toma a prueba. Si una muestra aleatoria de 12 clientes obtiene un tiempo promedio en ser registrados de 42 minutos con una desviación estándar de 11.9 minutos con el nuevo sistema, pruebe la hipótesis de que la media de la población es ahora menor de 50 minutos, usando un nivel de significancia de a)0.05 y de b)0.01 . Suponga que la población de los tiempos es normal.

Solución :

Hipótesis Nula	media = 50 min
Hipótesis Alterna	media < 50 min
Nivel de Significancia	0.01 y 0.05
media muestral	42
desv. est. muestral	11.9
t al 0.01 con 11 g.l.	2.718
t al 0.05 con 11 g.l.	1.796

El valor t calculado es de -2.33, que resulta mayor que -2.718 y menor que -1.796, por lo que se rechaza la hipótesis nula al nivel de significancia de 0.05, pero no al 0.01, con lo que se deduce que efectivamente el tiempo de registro es menor, pero no lo suficientemente significativo como para un cambio de sistema.

2.4.3.- Problema #3

Se efectuó un experimento para comparar el desgaste por abrasión en dos diferentes materiales laminados. Con una máquina para medir el desgaste se probaron 12 piezas del material A, así como 10 piezas del material B, en la misma forma. En ambos casos se observó el espesor del desgaste. Las muestras del material A arrojaron un resultado de desgaste promedio de 85 unidades con una desviación estándar de 4, mientras que las del material B, 81 y 5 respectivamente. Pruebe la hipótesis de que los dos tipos de materiales presentan el mismo desgaste al nivel de significancia 0.01 . Suponga que las poblaciones son aproximadamente normales con variancias iguales

Solución :

Hipótesis Nula	$mu_1 - mu_2 = 0$
Hipótesis Alterna	$mu_1 - mu_2 <> 0$
Nivel de Significancia	0.01
n de A	12
n de B	10
var. muestr. A	16
var. muestr. B	25
t a 0.005 con 20 g.l.	1.725
media muestral A	85
media muestral B	81

El valor calculado es 2.09, por lo que se rechaza la hipótesis nula de igual desgaste.

2.4.4.- Problema #4

Para averiguar si existe una diferencia entre un análisis químico y un análisis con rayos X que permita conocer la cantidad de hierro en determinado material, se emplean cinco muestras de una substancia ferrosa. Cada muestra se divide en dos y se les aplican ambos tipos de análisis. Los datos obtenidos que indican el análisis de contenido de hierro son:

	Muestra				
Análisis	1 .	2	3	4	5
Rayos X	2.0	2.0	2.3	2.1	2.4
Químico	2.2	1.9	2.5	2.3	2.4

Suponiendo que la población es normal, pruebe al nivel de significancia 0.05 si los dos mètodos de análisis dan, en promedio, el mismo resultado

Solución :

Hipótesis Nula	$mu_d = 0$
Hipótesis Alterna	$mu_d <> 0$
Nivel de Significancia	0.05
n de la muestra	5
t a 0.05, con 4 g.l.	2.776
De la tabla se obtienen aparte los valores	
desv. est. de las dif's	0.14142
dif. muestral promedio	-0.1

El valor calculado es de -1.58, por lo que se acepta la hipótesis nula y se concluye que los métodos no son significativamente diferentes.

2.4.5.- Problema #5

Un fabricante de catalizadores asegura que la vida de su producto tiene una distribución aproximadamente normal, con una desviación estándar igual a 0.9 años. Si una muestra aleatoria de 10 catalizadores tiene una desviación estándar de 1.2 años, ¿ pensaría usted que la desviación estándar de la población es mayor de 0.9 años ? Utilice un nivel de significancia de 0.05.

Solución :

Hipótesis Nula:	variancia = 0.81
Hipótesis Alterna:	variancia > 0.81
Nivel de Significancia :	0.05
n de la muestra:	10
ji cuadr. 0.05, con 9 g.l.	16.919
variancia muestral	1.44

El valor calculado de la distribución es 16, por lo que no hay razón para pensar que la desviación estándar es mayor de 0.9 .

2.4.6.- Problema #6

Si se desea probar la hipótesis de que la media de una población es igual a 68, contra la alternativa de que sea mayor, con un nivel de significancia de 0.05 y una desviación estándar de 5, encuentre el tamaño de muestra requerido si la potencia de la prueba es de 0.95 cuando la media verdadera es 69.

Solución :

Nivel de Significancia :	0.05
número de colas :	1
z al 0.05	1.645

potencia de la prueba :	0.95
z al 0.05	1.645
desviación estándar	5
delta de la media alterna	1

El valor resultante es de 270.6, por lo que se necesita un tamaño de muestra de 271 para hacer la prueba con una potencia de 0.95 .

2.4.7.- Problema #7

Una compañia manufacturera de aparatos de medición ha trabajado mediante cierto proceso que le garantiza un mínimo de 90% de piezas no defectuosas. Se está considerando la conveniencia de introducir una modificación en él para disminuir la proporción de artículos defectuosos a menos del 10% actual. En un experimento de 100 artículos producidos con el nuevo proceso, se encuentran 5 defec-
tuosos. ¿ Es ésta, evidencia suficiente para afirmar que el procedimiento modificado es mejor ? Utilice un nivel de significancia de 0.05 .

Solución :

Hipótesis Nula :	$p = 0.9$
Hipótesis Alterna :	$p > 0.9$
Nivel de Significancia :	0.05
z al 0.05	1.645
n de la muestra :	100
proporción muestral :	0.95
z al 0.05	1.645

El valor calculado de la distribución es 1.667, por lo que se rechaza la hipótesis nula en favor de la alterna y se puede asegurar que el nuevo procedimiento ha reducido el número de defectuosos.

2.4.8.- Problema #8

Para decidir si se construye una planta química propuesta, se hace una encuesta entre los residentes de una ciudad y del territorio circundante. El sitio de la construcción está dentro de los límites de la ciudad y, por eso, muchos habitantes del territorio piensan que una gran proporción de los residentes la apoyarán. Para averiguar si hay una diferencia significativa en la proporción de unos y otros, se realiza una encuesta. Si 120 de 200 residentes de la ciudad y 240 de 500 habitantes del territorio están a favor de la construcción, ¿ se puede considerar que la proporción de residentes de la ciudad que apoyan la construcción de la planta química es mayor que la de los habitantes del territorio ? Utilice un nivel de significancia de 0.025 .

Solución :

Hipótesis Nula :	$p_1 = p_2$
Hipótesis Alterna :	$p_1 > p_2$
Nivel de Significancia :	0.05
n_1 muestral :	200
n_2 muestral :	500
proporción muestral 1 :	0.60
proporción muestral 2 :	0.48
z al 0.025	1.96

El valor calculado de la distribución es 2.87, por lo que se rechaza la hipótesis nula en favor de la alterna y se concluye que la proporción de la ciudad es mayor a la del territorio.

2.4.9.- Problema #9

Una muestra aleatoria de ocho cigarrillos de cierta marca, contiene un promedio de alquitrán de 18.6 mg y una desviación estándar de 2.4 mg. El anuncio del fabricante declara que el contenido promedio de alquitrán no excede de 17.5 mg. Si suponemos que la distribución del contenido de alquitrán es normal, pruebe la hipótesis de que la variancia es igual a 2.3 contra la alternativa de que la variancia es diferente a 2.3. Utilice un nivel de significancia de 0.05 .

Solución :

Hipótesis Nula :	variancia = 2.3
Hipótesis Alterna :	variancia <> 2.3
Nivel de Significancia :	0.05
número de colas :	2
ji cuadr. al 0.975, 7 g.l.	1.690
ji cuadr. al 0.025, 7 g.l.	16.013
variancia muestral :	5.76

El valor calculado de la distribución es 17.53, por lo que se rechaza la hipótesis alterna y se puede asegurar que la variancia es mayor a 2.3 .

2.4.10.- Problema #10

Dadas dos muestras aleatorias de tamaños 11 y 16, tomadas de poblaciones normales independientes, con promedios 75 y 60, y desviaciones típicas 6.1 y 5.3 respectivamente, pruebe la hipótesis de igualdad de variancias contra la alternativa de que la primera sea mayor que la segunda.Utilice un nivel de significancia de 0.01 .

Solución :

Hipótesis Nula :	$var_1 = var_2$
Hipótesis Alterna :	$var_1 > var_2$
Nivel de Significancia :	0.01
n_1 muestral	11
n_2 muestral	16
var_1 muestral	37.21
var_2 muestral	28.09
f al 0.01, (10,15) g.l.	3.8

El valor calculado de la distribución resulta de 1.325, por lo que se acepta la hipótesis nula y se asegura que no hay razón para pensar en la desigualdad de variancias.

2.4.11.- Problema #11

Un fabricante está tratando de decidir la compra de los neumáticos, para sus nuevos modelos de automóviles, entre la marca A y la B. Para tomar una decisión efectúa una prueba empleando 13 neumáticos de cada marca y los usa hasta su desgaste total. Los neumáticos de la marca A soportaron un promedio de 37 900 Km con desviación típica de 5100, mientras que los de la marca B alcanzaron un promedio de 39 800 Km con desviación típica de 5900. Pruebe a un nivel de significancia de 0.05 la hipótesis de que la desviación estándar de la población de los neumáticos A es igual a la de los B, contra la alternativa de que la primera sea menor a la segunda.

Solución :

Hipótesis Nula :	$var_1 = var_2$
Hipótesis Alterna :	$var_1 < var_2$
Nivel de Significancia :	0.05
$n_1 = n_2$	13
var_1 muestral	26 010 000
var_2 muestral	34 810 000
f al 0.95, (12,12) g.l.	0.3717

El valor calculado de la distribución es 0.747, por lo que se acepta la hipótesis nula y se asegura que no hay razón para pensar en la desigualdad de las variancias.

2.4.12.- Problema #12

¿ De qué tamaño deberá ser la muestra de una prueba de hipótesis de que la media sea igual a 2.22, contra la alterna de que sea menor, para que la potencia de la misma sea de 0.90 cuando la media verdadera es 2.13 ?
Utilice un nivel de significancia de 0.05 .

Solución :

Nivel de Significancia :	0.05
número de colas :	1
z al 0.05	1.645
potencia de la prueba	0.9
z al 0.1	1.28
desv. estándar muestral	0.142
delta de la media alterna :	0.09

El resultado es de 21.3, por lo que el tamaño de muestra debe ser 21.

2.4.13.- Problema #13

Un fabricante anuncia que la resistencia a la tensión promedio de las cuerdas A excede a la clase de las cuerdas B en 12 Kg como mínimo. Para demostrarlo, se prueban bajo condiciones similares 50 piezas de cada tipo de cuerda. Las cuerdas A tienen una resistencia a la tensión de 86.7Kg con una desviación típica de 6.28 Kg, mientras que las cuerdas B tienen una resistencia a la tensión promedio de 77.8 Kg, con una desviación típica de 5.61 Kg. ¿ De qué tamaño deben ser las muestras para que la potencia de la prueba sea de 0.95 cuando la diferencia verdadera entre las cuerdas de los tipos A y B es de 8 Kg ? Trabaje con un nivel de significancia de 0.05 .

Solución :

Nivel de Significancia :	0.05
número de colas :	1
z al 0.05	1.645
potencia de la prueba	0.95
desv. estándar de A	6.28
desv. estándar de B	5.61
delta de la media alterna	8

El valor calculado es 11.99, por lo que el tamaño de muestra debe ser 12.

3.- REGRESION LINEAL

3.1.- Definición

En la Industria Química y en la vida diaria de todo tipo de profesionistas se presenta la necesidad de resolver problemas en que intervienen conjuntos de variables, entre las que se cree que existe alguna relación intrínseca. Casos como éstos se presentan cotidianamente, como cuando se desea idear un método de predicción o proce-
dimiento de estimación de la concentración de un subproducto al final de un reactor químico continuo, en relación a la temperatura de entrada del mismo, a partir de información experimental.

El aspecto estadístico del problema consiste en llegar a la mejor estimación de la relación entre las variables.

En el ejemplo anterior y en la mayoría de las aplicaciones, existe una diferencia clara entre las variables en lo que respecta a su papel en el proceso experimental.

A menudo se tiene una variable dependiente sencilla o respuesta Y, que no está regulada en el experimento. Esta depende de una o más variables independientes, como $x_1, x_2, \ldots, x_k$, que se miden con error insignificante y que de hecho, se fija o regula a menudo en el experimento. Así, las variables independientes $x_1, x_2, \ldots, x_k$ no son variables aleatorias sino cantidades fijas seleccionadas de antemano por el investigador y que no tienen propiedades distribucionales. En el ejemplo citado, la temperatura de entrada es la variable independiente y la concentración de subproducto es la respuesta Y. La relación ajustada al conjunto de datos experimentales, se caracteriza por una ecuación de predicción llamada ecuación de regresión .

En el caso de una Y y una x sencillas, la situación se convierte en una regresión de Y sobre x. Para k variables independientes, se habla en términos de una Y sobre $x_1, x_2, \ldots, x_k$. Así, a un ingeniero químico le puede interesar la cantidad de hidrógeno perdido en las muestras de un metal, cuando éste se almacena. En este caso se pueden tener dos datos iniciales: el tiempo de almacenaje x_1 en horas y la temperatura de almacenaje x_2 en grados centígrados. La respuesta será entonces la pérdida de hidrógeno Y en partes por millón.

En este capítulo se tratará la parte correspondiente a la regresión lineal simple, considerando únicamente el caso de una sola variable independiente. Para el caso de dos o más variables independientes, el usuario debe echar mano del siguiente capitulo.

3.2.- Menú

3.2.1.- Línea Recta

$$y = a + b * x$$

3.2.1.1 -> Estimación de los parámetros a y b.

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$a = y - b\bar{x}$$

donde x_i y y_i son las coordenadas de los datos, $\bar{x}$ y $\bar{y}$ son los valores promedio de las coordenadas de la muestra

3.2.1.2 -> Obtención de valores a partir de la regresión hecha

Una vez hecha la regresión lineal de los datos alimentados, se pueden tabular valores con la ecuación establecida:

$$y = a + b * x$$

Sólo se necesita seleccionar esta opción e ir alimentando los valores de x, para obtener los valores de y.

3.2.1.3 -> Intervalo de Confianza para b

INTERVALO DE CONFIANZA para el parámetro b en la línea de regresión $y = a + b * x$, al A(1-α)100% :

$$b - \frac{t_{\alpha/2} \cdot s}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha/2} \cdot s}{\sqrt{S_{xx}}}$$

donde β es el valor real de la pendiente, b es la pendiente calculada, s la desviación estándar de los datos tomados de una población aproximadamente normal, y $t_{\alpha/2}$ es el valor de la distribución t, con $v=n-1$ grados de libertad, que deja un área de $\alpha/2$ hacia la derecha.

Nota: A continuación se presentan las ecuaciones auxiliares de suma de cuadrados y de la variancia, que se utilizan para este inciso y los restantes.

$$S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}$$

$$S_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}$$

$$S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_i^2y_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{n}$$

El Coeficiente de correlación se calcula:

$$r=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

siendo el cuadrado del mismo, el Coeficiente de Determinación.

$$s^2=\frac{S_{yy}-bS_{xy}}{n-2}$$

3.2.1.4 -> Intervalo de Confianza para a

INTERVALO DE CONFIANZA para el parámetro α en la línea de regresión $y=a+bx$, al A(1-α)100% :

$$a-\frac{t_{\alpha/2}\cdot s\sqrt{\sum_{i=1}^{n}x_i^2}}{\sqrt{nS_{xx}}}<\alpha<a+\frac{t_{\alpha/2}\cdot s\sqrt{\sum_{i=1}^{n}x_i^2}}{\sqrt{nS_{xx}}}$$

donde α es el valor real de la ordenada al origen, a es la ordenada al origen calculada, s la desviación estándar de los n datos tomados de una población aproximadamente normal, y $t_{\alpha/2}$ es el valor de la distribución t, con $v=n-1$ grados de libertad, que deja un área de $\alpha/2$ hacia la derecha.

3.2.1.5 -> Intervalo de Confianza para la Respuesta Media

En muchos casos se desea graficar la línea recta regresionada junto con las cotas formadas a partir de intervalos de confianza de datos clave. El cordón formado recibe el nombre de Respuesta Media.

INTERVALO DE CONFIANZA para $\mu_{Y|x_0}$ al A(1-α)100% :

$$\hat{y}_0-t_{\alpha/2}s\sqrt{\frac{1}{n}+\frac{(\bar{x}_0-\bar{x})^2}{S_{xx}}}<\mu_{Y|x_0}<\hat{y}_0+t_{\alpha/2}s\sqrt{\frac{1}{n}+\frac{(\bar{x}_0-\bar{x})^2}{S_{xx}}}$$

donde $\bar{x}_0$ y $\hat{y}_0$ son las coordenadas de alguno de los puntos sobre la recta regresionada.

3.2.1.6 -> Intervalo de Confianza para una Respuesta Sencilla

En otros casos, el investigador se interesa más concretamente por el intervalo de confianza para una Respuesta Sencilla.

INTERVALO DE CONFIANZA para y_0 al A(1-α)100% :

$$\hat{y}_0-t_{\alpha/2}s\sqrt{1+\frac{1}{n}+\frac{(\bar{x}_0-\bar{x})^2}{S_{xx}}}<y_0<\hat{y}_0+t_{\alpha/2}s\sqrt{1+\frac{1}{n}+\frac{(\bar{x}_0-\bar{x})^2}{S_{xx}}}$$

donde x_0 y y_0 son las coordenadas de uno de los puntos sobre la recta regresionada.

3.2.1.7 -> Cálculo de la t,
para la Prueba de Hipótesis b = b0

Una vez establecidos los intervalos de confianza antes expuestos, puede probarse la hipótesis de que b tenga un valor específico o que simplemente sea 0 . El estadístico necesario para fijar si se sobrepasa el punto crítico es :

$$t = \frac{b - b_0}{s / \sqrt{S_{xx}}}$$

donde b es el valor obtenido de la regresión y b0 es el valor propuesto para la prueba.

El punto crítico se fija con n-2 grados de libertad.

3.2.1.8 -> Cálculo de la t,
para la Prueba de Hipótesis a = a0

Otra prueba que puede hacerse resulta para a respecto a algún valor a0 fijado. El estadístico necesario para fijar si se sobrepasa el punto crítico es :

$$t = \frac{a - a_0}{s \sqrt{\sum_{i=1}^{n} x_i^2 / n S_{xx}}}$$

El punto crítico se fija con n-2 grados de libertad.

3.2.1.9 -> Cálculo de la z,
para la Prueba de Hipótesis r = r0

Una vez calculado el coeficiente de correlación, puede probarse si éste tiene un valor específico r0, o simplemente es igual a cero, lo que significaría que no existe relación lineal entre los datos procesados. El estadístico apropiado para fijar si se sobrepasa el punto crítico es :

$$z = \sqrt{\frac{n-3}{2}} \ln\left[\frac{(1+r)}{(1-r)} \frac{1-r_0}{1+r_0}\right]$$

donde r es el coeficiente de regresión calculado, r0 es el valor propuesto y n el número de datos utilizados para llevar a cabo la regresión.

3.2.1.10 -> Análisis de Variancia en la Regresión Lineal

Cálculo de la F, para la Prueba de Hipótesis b = 0

Este procedimiento sirve para analizar la calidad de una línea de regresión estimada, en el que se subdivide la variación total de la variable dependiente en componentes significativos llamados Suma de Cuadrados de la Regresión y Suma de Cuadrados del Error .

El resultado final del Análisis de Variancia es una f calculada que fija el punto crítico para la prueba de hipótesis deseada:

Análisis de Variancia para probar b = 0

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Regresión	SCR	1	SCR	SCR/s^2
Error	SCE	n-2	$s^2 = \frac{SCE}{n-2}$	
Total	SCT	n-1		

3.2.1.11 -> Opciones Posteriores

Una vez realizado el ajuste deseado y calculados los estadísticos necesarios, el programa ofrece las opciones de:

- Terminar (regresar al Menú Principal de la UNDIES II
- Volver a calcular Estadísticos
- Llevar a cabo un nuevo ajuste con los mismos datos
- LLevar a cabo un nuevo ajuste con nuevos datos

3.2.2.- Curva Exponencial

$$y = a * e^{b*x}, \qquad a > 0$$

Para lograr la regresión de una curva exponencial, se requiere la transformación de la siguiente manera :

$$\boxed{\ln y = \ln a + b * x}$$

Por lo tanto, antes de poder procesar los datos se convierten las coordenadas de y en su logaritmo natural.

Una vez realizada la regresión se reconvierten los valores de los coeficientes para expresarlos según el modelo al que corresponden.

Este sistema recibe datos originales y da como resultado los coeficientes correspondientes al modelo propuesto. No se requiere de transformación previa por parte del usuario.

3.2.3.- Curva Logarítmica

$$y = a + b * \ln x$$

Para lograr la regresión de una curva logarítmica, se toma la ecuación tal cual,pero antes de poder procesar los datos se convierten las coordenadas de **x** en su logaritmo natural.

Este sistema recibe datos originales y da como resultado los coeficientes correspondientes al modelo propuesto. No se requiere de transformación previa por parte del usuario.

3.2.4.- Curva Geométrica

$$y = a * x^{b}, \qquad a > 0$$

Para lograr la regresión de una curva geométrica, se requiere la transformación de la siguiente manera :

$$\ln y = \ln a + b \cdot \ln x$$

Por lo tanto, antes de poder procesar los datos se convierten las coordenadas de x y y en su logaritmo natural.

Una vez realizada la regresión se reconvierten los valores de los coeficientes para expresarlos según el modelo al que corresponden.

Este sistema recibe datos originales y da como resultado los coeficientes correspondientes al modelo propuesto. No se requiere de transformación previa por parte del usuario.

3.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pidan.

En los casos en los que hay restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menú previo.

3.4.- Problemas Resueltos

3.4.1.- Problema #1

En cierto sistema de intercambio de energía se corrieron los siguientes datos experimentales, en donde "x" es la diferencia de temperatura y "y" la rapidez de transmisión de calor :

Variable/Corrida	1	2	3	4	5	6	7	8	9
"x"	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
"y"	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13.6	15.3

a) Estime la línea de regresión lineal según modelo de recta para los datos del sistema de intercambio de energía.
b)Calcule la rapidez de transmisión para una diferencia de temperatura de 1.5 y 5.0 .

Solución :

a) Los valores de regresión son

$a = 0.257$

$b = 2.93$

$r^2 = 0.9823$

b) Para una temperatura de 1.5 $y=4.65$

5.0 $y=14.91$

3.4.2.- Problema #2

En cierto tipo de prueba, se sabe que el esfuerzo normal en una probeta está relacionado funcionalmente con la resistencia al corte. El conjunto de los datos experimentales relativos a ambas variables es el siguiente:

x = esfuerzo normal	y = resistencia al corte
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5
26.9	21.7
27.4	21.4
22.6	25.8
25.6	24.9

a) Estime la ecuación de la recta de regresión $y = a + b*x$.

b) Estime la resistencia al corte para un esfuerzo normal de 24.5 .

c) Obtenga el intervalo de confianza al 99% para el coeficiente a.

d) Obtenga el intervalo de confianza al 99% para el coeficiente b.

Si $x = 24.5$,

e) Construya el intervalo de confianza al 95% para la respuesta media.

f) Construya el intervalo de confianza al 95% para el valor simple de pronosticación y_0 .

Solución:

a) Los resultados de la regresión

$a = 42.58$

$b = -0.656$
$r^2 = 0.4298$

b) La resistencia para x=24.5 $y = 25.77$
c) t al 0.005 y 10 g.l. 3.169
$21.96 < a < 63.20$
d) $-1.478 < b < 0.106$
e) t al 0.025 y 10 g.l. 2.228
$24.44 < mu\ y/x < 27.11$
f) $21.88 < y_0 < 29.66$

3.4.3.- **Problema #3**

Se tienen los siguientes datos de una recta, que se supone pasa por el origen:

x	0.5	1.5	3.2	4.2	5.1	6.5
y	1.3	3.4	6.7	8.0	10.0	13.2

Estime el modelo lineal para una recta y = a + b*x y pruebe la hipótesis de que la ordenada al origen es efectivamente igual a 0, contra la alternativa de que sea desigual, utilizando un nivel de significancia de 0.01 .

Solución:

Los valores de regresión son
$a = 0.349$
$b = 1.929$
$r^2 = 0.9958$

Para realizar la Prueba de Hipótesis se calcula la t de Student correspondiente para una $a_0 = 0$ ---> $t = 1.378$
t al 0.01 y 4 g.l. 3.747
por lo que, se puede aceptar la hipótesis nula de ordenada=0 .

3.4.4.- **Problema #4**

Ajustar el siguiente conjunto de datos con una línea recta

x_i	40.5	38.6	37.9	36.2	35.1	34.6
y_i	104.5	102	100	97.5	95.5	94

Solución:

Los valores de regresión son

$a = 33.53$
$b = 1.760$
$r^2 = 0.9909$

3.4.5.- Problema #5

Ajustar el siguiente conjunto de datos con una curva exponencial

x_i	0.72	1.31	1.95	2.58	3.14
y_i	2.16	1.61	1.16	0.85	0.50

Solución:

Los valores de regresión son

$a = 3.45$
$b = -0.582$
$r^2 = 0.9803$

3.4.6.- Problema #6

Ajustar el siguiente conjunto de datos con una curva logarítmica

x_i	3	4	6	10	12
y_i	1.5	9.3	23.4	45.8	60.1

Solución:

Los valores de regresión son

$a = -47.39$
$b = 41.39$
$r^2 = 0.9798$

3.4.7.- Problema #7

Se tiene el siguiente conjunto de datos:Ajustar el siguiente conjunto de datos con uña curva geométrica

x_i	10	12	15	17	20	22	25	27	30	32	35
y_i	0.95	1.05	1.25	1.41	1.73	2.00	2.53	2.98	3.85	4.59	6.02

a) Ajustar los datos con una curva geométrica
b) Verifique si algún otro modelo de curva mejora el ajuste.

Solución:

a) Los valores de regresión son

$$a = 0.0262$$
$$b = 1.456$$
$$r^2 = 0.9355$$

b) Al realizar la regresión para los mismos datos, pero para el modelo exponencial, se obtiene un coeficiente de determinación mejor = 0.9928 .

4.- Regresión Lineal Múltiple

4.1.- Definición

En el capitulo anterior se trataron los problemas de regresión con una variable independiente. Sin embargo, en la mayoría de los problemas prácticos y/o de investigación se necesita más de una variable independiente en el modelo de regresión.

La complejidad de los mecanismos científicos es tal, que para poder predecir una respuesta importante, hay que valerse de un modelo de regresión múltiple.

Cuando el modelo es lineal en sus coeficientes se llama modelo de regresión lineal múltiple, o sea un modelo en el cual la regresión lineal simple se extiende para poder tener en cuenta dos o más variables independientes.

A continuación se presentan dos modelos de este tipo de regresión. En el primero de los casos se trata de un modelo lineal multivariable; en el segundo, de un modelo polinómico, en el cual los factores de la variable, elevada a diferentes exponentes, se encuentran en relación lineal, a pesar de que la ecuación represente una curva no lineal.

En la ciencia y en la ingeniería hay muchos fenómenos cuya naturaleza es intrínsecamente no lineal, y cuando se conoce su estructura verdadera debe intentarse ajustarlos al modelo real, que en muchos casos es uno polinómico.

Hay casos en los que se plantean modelos combinados de dos o más variables que se elevan indistintamente a diferentes exponentes. Si el usuario se enfrenta a uno de esos casos, deberá establecer una tabla en la que calcule los valores de cada uno de los términos de los modelos y alimentarlos al programa como si se tratara de variables diferentes e independientes.

4.2.- Menú

4.2.1.- Modelo Lineal Multivariable

$$y = b_0 + b_1 x_1 + \ldots + b_i x_i$$

4.2.1.1 -> Regresión

La solución de un modelo multivariable parte de una serie de n datos, mayor o por lo menos igual al número k de variables, que se representan como

$$\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i); \qquad i = 1, 2, \ldots, n \quad \text{y } n > k\}$$

donde y_i es la respuesta observada para los valores $x_{1i}, x_{2i}, \ldots, x_{ki}$ de las k variables independientes $x_1, x_2, \ldots, x_k$.

Cada una de las observaciones $(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i)$ satisface la ecuación

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki}$$

Pero para obtener la ecuación de predicción o <u>de regresión</u> se puede establecer un sistema de $k+1$ ecuaciones con igual número de variables, valiéndose para ello del método de mínimos cuadrados:

$$n b_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} + \ldots + b_k \sum_{i=1}^{n} x_{ki} = \sum_{i=1}^{n} y_i$$

$$b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i} x_{2i} + \ldots + b_k \sum_{i=1}^{n} x_{1i} x_{ki} = \sum_{i=1}^{n} x_{1i} y_i$$

$$\cdot \quad \cdot \quad \cdot \qquad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \qquad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \qquad \cdot \quad \cdot$$

$$b_0 \sum_{i=1}^{n} x_{ki} + b_1 \sum_{i=1}^{n} x_{ki} x_{1i} + b_2 \sum_{i=1}^{n} x_{ki} x_{2i} + \ldots + b_k \sum_{i=1}^{n} x_{ki}^2 = \sum_{i=1}^{n} x_{ki} y_i$$

Este sistema puede resolverse se desglosa en matrices para quedar de a forma

donde A , b y g son las ma ces con los elementos:

$$A = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{ki}x_{1i} & \sum_{i=1}^{n} x_{ki}x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \\ \\ b_k \end{bmatrix}$$

, y

$$g = \begin{bmatrix} g_0 = \sum_{i=1}^{n} y_i \\ g_1 = \sum_{i=1}^{n} x_{1i} y_i \\ \vdots \\ g_k = \sum_{i=1}^{n} x_{ki} y_i \end{bmatrix}$$

La solución consiste entonces de los valores para los coeficientes de regresión incluidos en la matriz:

$$b = A^{-1} g$$

que se resuelve como un conjunto de k+1 ecuaciones con igual número de incógnitas. En el trayecto de resolución se invierte necesariamente la matriz A. Este sistema realiza la inversión por si mismo.

Adicionalmente a los coeficientes de regresión, se puede obtener una estimación de la variancia para los datos de regresión y el coeficiente de determinación múltiple, con las siguientes ecuaciones:

$$s^2 = \frac{SCE}{n-k-1}$$

$$R^2 = \frac{SCR}{SCT}$$

Para las cuales se requieren las ecuaciones complement- arias :

$$SCT = S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SCR = \sum_{j=0}^{k} b_j g_j - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SCE = SCT - SCR$$

4.2.1.2 -> Tabulación

Una vez realizada la regresión lineal múltiple, se pueden tabular las respuestas sencillas a partir de los valores que se alimenten a voluntad. El programa pide los valores númericos de cada una de las variables que el modelo contenga y da el resultado inmediatamente.

4.2.1.3 -> Intervalo de Confianza para la respuesta media

INTERVALO DE CONFIANZA para $\mu_{Y|x_{10},x_{20},\ldots,x_{k0}}$ al A(1-α)100% :

$$\hat{y}_0 - t_{\alpha/2} s \sqrt{x'_0 A^{-1} x_0} < \mu_{Y|x_{10},x_{20},\ldots,x_{k0}} < \hat{y}_0 + t_{\alpha/2} s \sqrt{x'_0 A^{-1} x_0}$$

donde $\hat{y}_0$ es la respuesta estimada, s la desviación típica de los datos, x_0 la matriz de valores originales para las variables, x'_0 la transversa de la anterior, A^{-1} la matriz inversa de los datos originales de regresión y $t_{\alpha/2}$ es el valor de la distribución t de Student con n-k-1 grados de libertad y que deja un área de α/2 hacia la derecha.

4.2.1.4 -> Intervalo de Confianza para una respuesta sencilla

INTERVALO DE CONFIANZA para y_0 al A(1-α)100% ;

$$\hat{y}_0 - t_{\alpha/2} s \sqrt{1 + x'_0 A^{-1} x_0} < y_0 < \hat{y}_0 + t_{\alpha/2} s \sqrt{1 + x'_0 A^{-1} x_0}$$

donde $\hat{y}_0$ es la respuesta estimada, s la desviación típica de los datos, x_0 la matriz de valores originales para las variables, x'_0 la transversa de la anterior, A^{-1} la matriz inversa de los datos originales de regresión y $t_{\alpha/2}$ es el valor de la distribución t de Student con n-k-1 grados de libertad y que deja un área de α/2 hacia la derecha.

4.2.1.5 -> t calculada para un coeficiente de regresión

Cálculo de la t,
para la Prueba de Hipótesis bj = bj0

Una vez establecidos los intervalos de confianza antes expuestos, puede probarse la hipótesis de que b tenga un valor específico o que simplemente sea 0 . El estadístico necesario para fijar si se sobrepasa el punto crítico es :

$$\boxed{t = \frac{b - b_0}{s/\sqrt{c_{jj}}}}$$

donde bj es el valor obtenido de la regresión, bj0 es el valor propuesto para la prueba, s la variancia de los datos y cjj es el elemento correspondiente de la matriz inversa.

El punto crítico se fija con n-k-1 grados de libertad.

4.2.2.- Modelo Polinómico

$$y = b_0 + b_1 x + b_2 x^2 + \ldots + b_k x^k$$

Para realizar la regresión de un modelo polinómico, se parte de una serie de parejas x/y ,de las cuales se toma el valor de la variable independiente x y se eleva a las diferentes potencias que el modelo exige, para formar una matriz análoga a la <u>multivariable</u> de la forma :

$$A = \begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^k \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \cdots & \sum_{i=1}^{n} x^{k+1} x_i \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_i^k & \sum_{i=1}^{n} x_i^{k+1} & \sum_{i=1}^{n} x_i^{k+2} & \cdots & \sum_{i=1}^{n} x_i^{2k} \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

, y

$$g = \begin{bmatrix} g_0 & = & \sum_{i=1}^{n} y_i \\ g_1 & = & \sum_{i=1}^{n} x_i y_i \\ & & . \\ & & . \\ & & . \\ g_k & = & \sum_{i=1}^{n} x_i^k y_i \end{bmatrix}$$

Es decir, los diferentes términos elevados a su respectiva potencia se tratan como si fueran términos ajenos a la variable sencilla, para obtener un resultado con el mismo procedimiento que se utilizó en el caso de varias variables. Este programa facilita la regresión, en cuanto que se encarga por si solo de formar los términos exponenciales.

Comó es también de esperarse, se pueden obtener las inferencias análogas al caso <u>multivariable</u>.

4.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capitulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pida.

En los casos en los que existan restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menú previo.

4.4.- Problemas Resueltos .

4.4.1.- Problema #1

En cierto experimento se midió la concentración del producto A, después de la reacción en diferentes combinaciones de concentración de tres reactivos B, C y D. Se listan a continuación sus porcentajes:

Concentraciones

A	B	C	D
25.5	1.74	5.30	10.80
31.2	6.32	5.42	9.40
25.9	6.22	8.41	7.20
38.4	10.52	4.63	8.50
18.4	1.19	11.60	9.40
26.7	1.22	5.85	9.90
26.4	4.10	6.62	8.00
25.9	6.32	8.72	9.10
32.0	4.08	4.42	8.70
25.2	4.15	7.60	9.20
39.7	10.15	4.83	9.40
35.7	1.72	3.12	7.60
26.5	1.70	5.30	8.20

a) Estime el modelo de regresión múltiple para los datos dados.
b) Establezca un intervalo de confianza del 95% para la respuesta media cuando B=3, C=8 y D=9 .
c) Establezca un intervalo de confianza del 95% para la respuesta sencilla si B=3, C=8 y D=9 .
d) Pruebe la Hipótesis de que el tercer coeficiente de la ecuación de regresión (β_3) es igual a cero.
e) Pruebe la Hipótesis de que el segundo coeficiente de la ecuación de regresión (β_2) sea igual a -2.5, contra la alternativa de que no lo sea.

Solución:

a) núm. de var. independ. 3
núm. de juegos 13
ecuación obtenida:

$$y = 39.16 + 1.02x_1 - 1.86x_2 - 0.34x_3$$

variancia 4.297
coef.determ.múltiple 0.9117

b) nivel de confianza 95%
t al 0.025 con 9 g.l. 2.262

El intervalo se encuentra entre 22.49 y 25.95

c) nivel de confianza 95%
t al 0.025 con 9 g.l. 2.262

El intervalo se encuentra entre 19.23 y 29.22 .

d) Hipótesis Nula $b_3 = 0$

Hipótesis Alterna $b_3 <> 0$
b propuesta 0
t al 0.025 con 9 g.l. 2.262

El valor calculado resulta -0.556, por lo que se acepta la hipótesis nula.

e) Hipótesis Nula $b_2 = -2.5$
Hipótesis Alterna $b_2 > -2.5$
b propuesta -2.5
t al 0.05 con 9 g.l. 1.833

El valor calculado resulta 2.388, por lo que se rechaza la hipótesis nula en favor de la alterna.

4.4.2.- Problema #2

Se desea ajustar el siguiente conjunto de datos a una ecuación de tercer grado :

i	1	2	3	4	5
x	0.8	1.0	1.2	1.4	1.6
y	24.0	20.0	10.0	13.0	12.0

Encuentre la ecuación de regresión apropiada, el coeficiente de determinación múltiple y la variancia del conjunto.

Solución:

grado del polinomio 3
núm. de juegos de datos 5
ecuación obtenida:

$$y = 47.94 - 9.76x - 41.1x^2 + 20.8x^3$$

variancia 18.51
coef.determ.múltiple 0.8685

4.4.3.- Problema #3

Los datos siguientes son resultado de 15 repeticiones de un experimento efectuado sobre cuatro variables independientes y una respuesta sencilla y .

Valores

y	x_1	x_2	x_3	x_4
14.8	7.8	4.3	11.5	6.3
12.1	6.9	3.9	14.3	7.4
19.0	9.3	8.4	9.4	5.9
14.5	6.8	10.3	15.2	8.7
16.6	11.7	6.4	8.8	9.1
17.2	8.5	5.7	9.8	5.6
17.5	12.6	6.8	11.2	6.8
14.1	7.5	4.2	10.9	7.4
13.8	8.4	7.3	14.7	8.2
14.7	11.3	8.8	15.1	9.2
17.7	10.7	3.6	8.7	4.7
17.0	7.3	4.9	8.6	5.5
17.6	8.4	7.3	9.3	6.6
16.3	6.7	9.7	10.8	8.7
18.2	9.6	8.4	11.9	5.4

a) Estime el modelo de regresión lineal múltiple que relaciona a x_1, x_2, x_3, x_4 con y, variancia y coeficiente de determinación múltiple.
b) Construya el intervalo de confianza para la respuesta sencilla cuando $x_1=8.2$, $x_2=6.0$, $x_3=10.3$ y $x_4=5.8$

Solución:

a) núm. de var.independientes 4
núm. de juegos de datos 15

$$y = 19.99 + 0.30x_1 + 0.60x_2 - 0.50x_3 - 0.70x_4$$

variancia 0.000995
coef.determ.múltiple 0.9998

b) nivel de confianza 95%
t al 0.025 con 10 g.l. 2.228

El intervalo se encuentra entre 16.826 y 16.877 .

5.- Análisis de Datos Categorizados

5.1.- Definición

Muy frecuentemente, la información recolectada por un investigador, en diversos campos del conocimiento humano, se clasifica en una estructura inicial en forma de datos categorizados.

Este término se refiere a la estructura inicial en la que se clasifican los datos, generalmente en categorías, y dentro de la cual se cuenta con la frecuencia con la que se presentan los mencionados datos.

Casos como el anterior se presentan abundantemente en casi cualquier campo donde se realicen estudios cuantitativos. Así pues puede interesar establecer alguna conclusión referente a los gustos de los compradores de cierto producto, a la opinión de las personas en alguna encuesta o a los resultados obtenidos en un estudio biológico, etc.

En todos los ejemplos anteriores, cada categoría se especifica en base a una característica cualitativa, por lo que puede ser fijada arbitrariamente partiendo de intervalos originalmente obtenidos de escalas numéricas de medición.

Así pues pueden ser fijadas categorías reconocidas con apelativos como 'alta', 'media' o 'baja', o calificativos como 'pesado', 'moderado' y 'ligero', aunque lo que en realidad se maneje sean datos de frecuencia completamente cuantitativos.

La forma más simple de datos categorizados contiene únicamente dos categorías y es manejada con la ayuda de la distribución binomial. En esta forma se declara si un evento forma parte de una clasificación o no.

La intención se este capítulo es la de manejar procedimientos utilizables para estudiar datos agrupados en varias categorías.

La primera parte del capítulo se ocupa de la prueba de bondad de ajuste, que se basa en el grado de ajuste que hay entre la frecuencia de ocurrencia de los eventos en una muestra observada y las frecuencias esperadas que se obtienen de una distribución hipotética.

La segunda parte del capitulo trata los problemas de estudio en los que los datos se clasifican al mismo tiempo dentro de dos tipos de tratamiento, cada uno de los cuales conlleva su número de categorías, con lo que se establecen tablas de contingencia, en las que generalmente desea probarse la independencia entre los dos tratamientos en relación a los datos obtenidos.

5.2.- Menú

5.2.1.- Prueba de Bondad de Ajuste

5.2.1.1 -> General

La prueba de bondad de ajuste, (o Prueba de Pearson), se aplica a datos obtenidos por medio de ensayos Bernoulli, los cuales deben cumplir las siguientes características:

a) Los resultado de cada ensayo pertenece a una de las categorías disponibles o celdas, que se identifican de 1 a k.

b) La probabilidad p_i de cada celda, sumada a las demás deberá dar 1, o sea $\sum_{i=1}^{k} p_i = 1$

, y deberá ser la misma para todos los ensayos.

c) Los ensayos son independientes.

d) ¡ La frecuencia mínima de cada evento debe ser 5 !

La prueba se lleva a cabo en la forma siguiente:

Hipótesis Nula

$$H_0: p_1 = p_{10}, p_2 = p_{20}, \ldots, p_k = p_{k0}$$

Estadística de Prueba

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n p_{i0})^2}{n p_{i0}} = \sum_{celdas} \frac{(O_i - E_i)^2}{E_i}$$

donde n_i es la frecuencia de cada categoria, p_i la probabilidad de evento para la celda i, O la frecuencia observada, E la frecuencia esperada y χ^2 es el valor de una distribución ji cuadrada con $g.l. = k - 1 = \text{No. de celdas} - 1$ grados de libertad.

Región Crítica

$$\chi^2 \geq \chi^2_\alpha$$

donde χ^2_α es el extremo derecho de la distribución ji cuadrada antes especificada.

5.2.1.2 -> Frecuencia Esperada = Frecuencia Observada

En algunos casos, se establece que la frecuencia esperada coincide con la observada, con el fin de probar la validez del sistema u objeto con el que la prueba se está realizando.

Las condiciones para este caso son las mismas que para el anterior, pero con la siguiente ecuación para la distribución

$$\chi^2 = \frac{n \Sigma O_i^2}{\Sigma O_i} - \Sigma O_i$$

5.2.1.3 -> $H_0: p_1 = p_2 = \ldots = p_k$

En este caso se manejan problemas donde se conoce la totalidad de las observaciones de todas las categorías y se desea probar que la probabilidad de las frecuencias esperadas es la misma para todas las categorías.

La ecuación utilizable es la general, con la consideración de que la probabilidad para todas las categorías es la misma.

5.2.1.4 -> Adecuación a un Modelo tipo Poisson

Cuando un conjunto de datos no se rige de antemano por algún modelo, se puede proponer uno de acuerdo a alguna distribución.

En este caso se manejan problemas donde el conjunto de datos consiste en valores numéricos, (número de éxitos), atribuibles a una distribución Poisson con sus respectivas frecuencias.

La media *m* de la distribución se obtiene a partir de los datos mismos por la ecuacion

$$m = \frac{\sum_{i=1}^{k} (x_i \times n_i)}{\sum_{i=1}^{k} n_i}$$

donde x_i es el número de éxitos de la celda *i* , n_i es la frecuencia de la celda *i*, y *k* es el número de celdas o categorías del ensayo. Cabe señalar que el valor *m* se redondea para poder realizar los cálculos de la distribución de probabilidad Poisson :

$$p(x;m) = \frac{e^{(-m)} m^x}{x!}$$

donde *x* es el número de éxitos del ensayo, *m* es la media antes calculada y *e* es el número de Euler (2.71828...).

En los casos donde se presentan valores de frecuencia menores a 5, se opta por agruparlos con los valores adyacentes o junto con otras frecuencias consecutivas también menores para formar valores de por lo menos 5.

Una vez calculadas las probabilidades de Poisson, se multiplican por la suma de frecuencias y se obtienen así los valores esperados.

El resto del cálculo se realiza entonces según la ecuación general, pero el valor teórico de la distribución χ^2 se calcula con lo siguientes grados de libertad:

$$\boxed{\nu = \text{num. de categorias} - 1 - \text{num. de parametros calculados}}$$

en el caso de este sistema, el único parámetro calculado es m.

5.2.2.- Tablas de Contingencia

Cuando se tienen dos tratamientos para cada elemento de la muestra del ensayo, las frecuencias observadas se clasifican en renglones y columnas según la categoría que les corresponda en cada tratamiento.

De esta manera se obtienen tablas de clasificación simultánea, llamadas tablas de contingencia, también conocidas como de clasificación cruzada.

El número de renglones se representa por r y el número de columnas por c, por lo que, al definir una tabla de contingencia, se acompaña de estos valores en la forma $r \times c$.

Para un tratamiento A y uno B, se tiene entonces la siguiente tabla

Categoría	B_1	B_2	...	B_c	Total por Renglón
A_1	n_{11}	n_{12}	...	n_{1c}	n_{10}
A_2	n_{21}	n_{22}	...	n_{2c}	n_{20}
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
A_r	n_{r1}	n_{r2}	...	n_{rc}	n_{r0}
Total por Columna	n_{01}	n_{02}	...	n_{0c}	n

donde n_{ij} es la frecuencia de cada celda, n_{i0} el total de renglón o frecuencia de A, n_{0j} el total de columna o frecuencia de B y n el gran total.

A partir de los datos de frecuencia, se calcula la probabilidad de cada evento dividiendo cada celda entre el gran total, obteniéndose la tabla de probabilidades de celda :

Categoría	B_1	B_2	...	B_c	Total por Renglón
A_1	p_{11}	p_{12}	...	p_{1c}	p_{10}
A_2	p_{21}	p_{22}	...	p_{2c}	p_{20}
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
A_r	p_{r1}	p_{r2}	...	p_{rc}	p_{r0}
Total por Columna	p_{01}	p_{02}	...	p_{0c}	1

donde p_{ij} es la probabilidad del evento A_i y B_j, p_{i0} la probabilidad total en el i-ésimo renglón y p_{0j} la probabilidad total en la j-ésima columna.

La hipótesis nula de independencia a probar es entonces, para cada celda:

$$H_0: p_{ij} = p_{i0} p_{0j}$$

Para poder probar una independencia entre los tratamientos, se necesita tener algun indicio de cómo deberían ser los datos si hubiera una asociación entre ellos, por lo que se calculan datos esperados para cada celda y se prosigue con una Prueba de Bondad de Ajuste, con ayuda de la distribución ji cuadrada, tal como se manejó en los incisos anteriores.

La forma de probar la hipótesis resulta ser entonces

Estadística de Prueba

$$\chi^2 = \sum_{\text{todas las rc celdas}} \frac{(n_{ij}-E_{ij})^2}{E_{ij}} \quad , E_{ij} = \frac{n_{i0}n_{0j}}{n}$$

donde n_{ij} es la frecuencia de cada celda, E la frecuencia esperada y χ^2 es el valor de una distribución ji cuadrada con $g.l. = (r-1) \times (c-1)$ grados de libertad.

Región Crítica

$$\chi^2 \geq \chi^2_\alpha$$

donde χ^2_α es el extremo derecho de la distribución ji cuadrada antes especificada.

Una vez comprobada o rechazada la independencia entre tratamientos, puede medirse el nivel de asociación que existe entre ellos por medio del coeficiente de Pearson (Q), el cual refleja, a mayores valores, una mayor asociación :

$$Q = \sqrt{\frac{\chi^2}{n+\chi^2}} \qquad ,0 \leq Q \leq \sqrt{\frac{q-1}{q}}$$

donde q representa el número de renglones o columnas, (lo que resulte menor), y n el tamaño de muestra, (gran total).

5.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pidan.

En los casos en los que hay restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menú previo.

5.4.- Problemas Resueltos

5.4.1.- Problema #1

El resultado obtenido al cruzar dos determinados tipos de plantas puede ser de cualquiera de los genotipos A, B o C. Según ciertas consideraciones, la relación que debe existir entre los resultados de genotipos es de 1:2:1 . Para lograr una verificación experimental, se producen 90 plantas a partir de los dos tipos de plantas mencionados y se obtiene la siguiente tabla con los genotipos obtenidos

A	B	C	Total
18	44	28	90

Pruebe si estos datos confirman o contradicen el modelo de relación.

Solución:

número de categorías	3
frecuencias esperadas	22.5 : 45 : 22.5
ji cuadr. 0.05, 2 g.l.	5.991

El valor calculado es de 2.27, por lo que se confirma el modelo.

5.4.2.- Problema #2

Se lanza un dado 120 veces con la intención de probar si está bien hecho. Los resultados del experimento se muestran en la tabla a continuación.

número obtenido	1	2	3	4	5	6
frecuencia Obs.	25	17	15	23	24	16

Pruebe a un nivel de significancia de 0.05 si el dado está bien hecho, considerando que las frecuencias esperadas son iguales a las observadas.

Solución:

número de categorías	6
ji cuadr. 0.05, 5 g.l.	11.07

El valor calculado es de 5, por lo que se acepta la distribución y se puede afirmar que el dado está bien hecho.

5.4.3.- Problema #3

Una máquina lavadora se vende en cinco diferentes colores. En base a los datos recabados por el departamento de mercadotecnia de una muestra de 300 clientes, pruebe la hipótesis, con un nivel de significancia de 0.05, de que los cinco colores son igualmente preferidos por los consumidores.

Aguacate	Marrón	Rojo	Azul	Blanco	Total
88	65	52	40	55	300

Solución:
La proporción de todas las celdas se espera que sea la misma.

número de categorías	5
ji cuadr. 0.05, 4 g.l.	9.488

El valor calculado es de 21.63, por lo que rebasa al teórico y se rechaza la hipótesis de igualdad de popularidad.

5.4.4.- Problema #4

Un investigador actuarial desea estudiar la frecuencia de reclamaciones al derecho de hospitalización hechas por personas aseguradas, miembros de familias con dos hijos y padres de edades menores a 50 años.
Para tal efecto se escogen al azar 200 familias con las características antes mencionadas y se registra el número de reclamaciones a lo largo de cuatro años. A partir de la tabla correspondiente y a un nivel de significancia de 0.05, pruebe si los datos se ajustan o no a una distribución tipo Poisson.

Número de Reclama-ciones	0	1	2	3	4	5	6	7	Total
Frecuencia	22	53	58	39	20	5	2	1	200

Solución:

número de categorías	8
ji cuadr. 0.05, 6 g.l.	12.592

El valor calculado resulta de 2.59, por lo que no hay razón para rechazar la hipótesis de que los datos siguen una distribución tipo Poisson.
Nota: En este caso no se agruparon las frecuencias menores a 5, para poder determinar en forma más precisa la media *m*, sin embargo el resultado sigue siendo válido. En el caso que se desee agrupar los datos, debe hacerse una vez calculadas las probabilidades Poisson. En este problema se agrupan las frecuencias observadas de los valores 5, 6 y 7, obteniéndose un total de 8. Las frecuencias espe-
radas correspondientes dan un total de 10.6 y el valor de la distribución ji cuadrada resulta entonces de 2.33, con lo que se sigue aceptando el modelo propuesto.

5.4.5.- Problema #5

En un estudio efectuado en un taller, se reunió información para averiguar si la proporción de piezas defectuosas producidas por los obreros fue la misma para los turnos de trabajo matutino, vespertino y nocturno. De las piezas producidas se obtuvieron los datos siguientes:

	Turno		
	Diurno	Vespertino	Nocturno
Defectuosas	45	55	70
No defectuosas	905	890	870

Utilizando un nivel de significancia de 0.025, pruebe la hipótesis de que el número de piezas defectuosas es independiente del turno de trabajo.

Solución:

número de renglones	2
número de columnas	3
ji cuadr. 0.025, 2 g.l.	7.378

El valor calculado resulta de 6.23, por lo que se acepta la hipótesis de independencia.

5.4.6.- Problema #6

Una muestra aleatoria de 200 hombres casados se clasifica conforme a sus estudios y al número de hijos:

	Número de Hijos		
Estudios	0 - 1	2 - 3	más de 3
primaria	14	37	32
secundaria	19	42	17
bachillerato	12	17	10

Pruebe la hipótesis, al nivel de significancia de 0.05, de que el tamaño de la familia es independiente del nivel de educación alcanzado por el padre.

Solución:

número de renglones	3
número de columnas	3
ji cuadr. 0.05, 4 g.l.	9.488

El valor calculado es 7.46, por lo que se acepta la hipótesis de independencia.

6.- Análisis de Varianza

6.1.- Definición

En los capítulos de estimación y de prueba de hipótesis se ha restringido hasta ahora, a dos parámetros únicamente. En el caso de dos muestras provenientes de dos poblaciones distintas se tenía inclusive la necesidad de una estimación mancomunada de la variancia. Por lo tanto, se requiere de una técnica en la que se pueda manejar, por ejemplo, k medias poblacionales.

Esta técnica requerida no se aleja de la de regresión. En esta última se utiliza para hacer una partición de la suma de cuadrados total en una parte debida a la regresión y otra debida al error. En algunos casos se hace la partición en SCR (suma de cuadrados) de componentes significativos con el propósito de probar hipótesis relevantes sobre los parámetros del modelo.

El término análisis de variancia describe una técnica con la que se puede analizar o dividir la variación total en componentes de variación significativos. El grado de dificultad del análisis depende de la complejidad del problema.

Una aplicación del análisis de variancia se presenta cuando se tienen k tratamientos (diferentes mezclas, analistas, sustancias, regiones de un país, etc.), y cada tratamiento consta de n muestras. Por lo general, se desea probar la hipótesis de que la diversidad de tratamientos no influye en los resultados, o sea que

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_1: \quad \text{al menos dos de las medias no son iguales}$$

Además, tal vez se quiera hacer comparaciones individuales entre las diferentes medias poblacionales. En el procedimiento del análisis de variancia, se supone que cualquier variación existente entre el promedio de los tratamientos se atribuya a 1) la variación entre los valores de las observaciones del mismo tratamiento y a 2) la variación de cada tipo de tratamiento.

La variación propia de la muestra se llama variación aleatoria y parte del objetivo del análisis de variancia es determinar si las diferencias entre los diversos tratamientos son solamente debidos a la variación aleatoria o si efectivamente también hay una contribución de la variación sistemática atribuida a los tipos de tratamiento.

Por lo tanto, el procedimiento requerido consiste esencialmente en separar la variabilidad total en los dos componentes importantes siguientes:

1. La variabilidad entre los tratamientos, midiendo para ello la variación sistemática y la variación aleatoria.
2. La variabilidad propia del tratamiento, midiendo para ello la variación aleatoria

Adicionalmente se determina si alguno de los componentes es significativamente mayor que algún otro.

6.2.- Menú

6.2.1.- Análisis de Varianza con un Criterio de Clasificación

Las muestras aleatorias de tamaño n se seleccionan de cada una de las k poblaciones. Las k poblaciones diferentes se clasifican frecuentemente de acuerdo con los diferentes tratamientos o grupos.

Al realizar un análisis de varianza, es conveniente ordenar los datos, para así facilitar su estudio. Un ejemplo de clasificación es:

	Tratamiento						
	1	2	...	i	...	k	
	y_{11}	y_{21}	...	y_{i1}	...	y_{k1}	
	y_{12}	y_{22}	...	y_{i2}	...	y_{k2}	
	.	.		.		.	
	.	.		.		.	
	.	.		.		.	
	y_{1n}	y_{2n}	...	y_{in}	...	y_{kn}	
Total	$T_{1.}$	$T_{2.}$	...	$T_{i.}$	...	$T_{k.}$	$T_{..}$
Media	$m_{1.}$	$m_{2.}$	...	$m_{i.}$	...	$m_{k.}$	$m_{..}$

Para realizar el análisis de varianza es necesario hacer el cálculo de las sumas de cuadrados debidas a los tratamientos, al error y su total.

Suma de Cuadrados Total
$$SCT=\sum_{i=1}^{k}\sum_{j=1}^{n}y_{ij}^2-\frac{T_{..}^2}{nk}$$

Suma de Cuadrados debidos a los Tratamientos
$$SCA=\frac{\sum_{i=1}^{k}T_{i.}^2}{n}-\frac{T_{..}^2}{nk}$$

Suma de Cuadrados debidos al Error
$$SCE=SCT-SCA$$

Una vez calculados los datos necesarios, se sumarizan los resultados en una tabla como la que se muestra:

Análisis de Variancia para un criterio de clasificación

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Tratamientos	SCA	k-1	$s^2=\frac{SCA}{(k-1)}$	s_1^2/s^2
Error	SCE	k(n-1)	$s^2=\frac{SCE}{k(n-1)}$	
Total	SCT	. nk-1		

Teniendo este <u>análisis de variancia</u>, se puede probar la hipótesis nula H_0: $\mu_1=\mu_2=\ldots=\mu_k$ contra la alternativa H_1: al menos dos de las medias no son iguales con una región crítica $F>f_\alpha[k-1,k(n-1)]$

6.2.2.- Pruebas para la Igualdad de varias Variancias

En la sección anterior se trabajó con diferentes muestras de tamaños iguales. En esos casos se confía en que las desviaciones de la suposición de igualdad de variancias para las k poblaciones son ligeras, y con esto, que la razón f del análisis de variancia no se ve afectada.

Sin embargo, cuando el tamaño de cada muestra varía o si una varian-
cia es mucho mayor que las otras, la razón f se ve alterada.

Para verificar en un estudio que la estimación de f será la adecuada, conviene probar la hipótesis de igualdad de varias variancias H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, contra la alternativa H_1: no todas las variancias son iguales.

La prueba utilizada más comúnmente se conoce como la Prueba de Bartlett y se basa en una estadística cuya distribución muestral se aproxima bastante mediante la distribución ji cuadrada cuando las k muestras aleatorias se toman de poblaciones normales independientes

La prueba consta de un cálculo referido a la variancia mancomunada del experimento, obtenida ésta a su vez de las variancias de cada tratamiento

El estadístico que se calcula es

$$b = 2.3026 \cdot \frac{q}{h}$$

que es un valor de la variable aleatoria B (Bartlett) que tiene aproximadamente la distribución ji cuadrada con k-1 grados de libertad.

Los valores adicionales que se requieren son

$$q = (N-k)\log s_p^2 - \sum_{i=1}^{k}(n_i - 1)\log s_i^2$$

$$h = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N-k}\right)$$

$$s_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)\cdot s_i^2}{N-k}$$

donde s_p^2 es la variancia mancomunada de las <u>k</u> variancias muestrales $s_1^2, s_2^2, \ldots, s_k^2$ de las muestras de tamaños $n_1, n_2, \ldots, n_k$ con $\sum_{i=1}^{k} n_i = N$

.

La cantidad <u>q</u> es grande cuando las variancias muestrales difieren mucho y es igual a cero cuando todas son iguales. Por lo tanto, al nivel de significancia α, H_0 se rechaza sólo cuando $b > \chi^2_\alpha$.

El riesgo de aplicar la <u>prueba de Bartlett</u> es que es muy sensible a la suposición de normalidad.

6.2.3.- Análisis de Variancia por Bloques al Azar Completos

Conocido también como <u>"modelo con dos criterios de clasificación, sin interacciones"</u>

En los casos hasta ahora estudiados se manejaba un criterio de clasificación, sin embargo, se presentan muchas situaciones en las que se desea probar la influencia de distintos tratamientos y <u>al mismo tiempo</u> la influencia en los resultados de alguna causa que afecte integralmente al conjunto de tratamientos. Los experimentos que se diseñan de esta manera se les conoce por <u>bloques al azar completos</u>.

En otras palabras, cada tratamiento se asigna una vez al azar a todos los <u>bloques</u> o condiciones del desarrollo experimental

Un acomodo de los datos de un diseño experimental como el descrito puede ser

Bloques								
Tratamientos	1	2	...	j	...	b	Total	Media
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1b}	$T_{1.}$	$m_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2b}	$T_{2.}$	$m_{2.}$
.	.	.		.		.	.	.
i	.	.		y_{ij}		y_{ib}	$T_{i.}$	$m_{i.}$
.	.	.		.		.	.	.
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kb}	$T_{k.}$	$m_{k.}$
Total	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.b}$	$T_{..}$	
Media	$m_{.1}$	$m_{.2}$	...	$m_{.j}$	...	$m_{.b}$	$m_{..}$	

Para realizar este análisis de varianza es necesario hacer el cálculo de las sumas de cuadrados debidas a los tratamientos, a los bloques, al error y su total.

Suma de Cuadrados Total
$$SCT = \sum_{i=1}^{k}\sum_{j=1}^{b} y_{ij}^2 - \frac{T^2}{bk}$$

Suma de Cuadrados debidos a los Tratamientos
$$SCA = \frac{\sum_{i=1}^{k} T_{i.}^2}{b} - \frac{T^2}{bk}$$

Suma de Cuadrados debidos a los Bloques
$$SCB = \frac{\sum_{j=1}^{b} T_{.j}^2}{k} - \frac{T^2}{bk}$$

Suma de Cuadrados debidos al Error
$SCE = SCT - SCA - SCB$

Una vez calculados los datos necesarios, se sumarizan los resultados en una tabla como la que se muestra:

Análisis de Variancia para dos criterios de clasificación

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Tratamientos	SCA	k-1	$s_1^2 = \frac{SCA}{(k-1)}$	$f_1 = \frac{s_1^2}{s^2}$
Bloques	SCB	b - 1	$s_2^2 = \frac{SCB}{(b-1)}$	$f_2 = \frac{s_2^2}{s^2}$
Error	SCE	(k-1)(b-1)	$s^2 = \frac{SCE}{(k-1)(n-1)}$	
Total	SCT	bk-1		

Teniendo este <u>análisis de variancia</u>, se puede probar la hipótesis nula H_0: los efectos de los tratamientos son cero contra la alternativa H_1: al menos uno no es cero con una región crítica $F_1 > f_\alpha[k-1,(k-1)(b-1)]$

.

Igualmente se puede probar la hipótesis nula H_0: los efectos de los bloques son cero contra la alternativa H_1: al menos uno no es cero con una región crítica $F_2 > f_\alpha[b-1,(k-1)(b-1)]$

.

6.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pida.

En los casos en los que existan restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menu previo.

6.4.- Problemas Resueltos

6.4.1.- Problema #1

En un experimento industrial, en el que un ingeniero desea conocer la variación de absorción de humedad de cierta resina, se toman cinco mezclas diferentes de resina. Las muestras se exponen a la humedad durante 48 horas. Se toma la decisión de probar 6 muestras de cada mezcla, con lo que se tienen 30 muestras para realizar el estudio.

Los resultados fueron los siguientes

	Mezcla					
	1	2	3	4	5	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	2791	3320	3416	3663	3664	16854
Media	553.33	569.33	610.50	465.17	610.67	561.80

Pruebe la Hipótesis de igualdad de variancias, H_0: $\mu_1 = \mu_2 = \ldots = \mu_5$, al nivel de significancia de 0.05, para los datos de la tabla anterior sobre la absorción de humedad de diferentes mezclas de resinas.

Solución:

número de tratamientos	5
núm. de elementos por trat.	6
Hipótesis Nula	igualdad de medias
Hipótesis Alterna	al menos 2 no lo son
Nivel de Significancia	0.05
f al 0.05 con (4,25) g.l.	2.60

Análisis de Variancia para un criterio de clasificacion

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	f calculada
Mezclas	85356	4	21339	4.30
Error	124020	25	4961	
Total	209377	29		

Se rechaza la hipótesis nula y se concluye que las diferentes mezclas no tienen la misma absorción media de humedad.

6.4.2.- Problema #2

Se tienen los siguientes datos de cierto experimento

	Muestra A	Muestra B	Muestra C	
	4	5	8	
	7	1	6	
	6	3	8	
	6	5	9	
		3	5	
		4		
Total	23	21	36	80

Con la prueba de Bartlett pruebe la hipótesis de igualdad de variancias para las tres poblaciones de las que se seleccionaron los datos expuestos

Solución:

número de tratamientos	3
núm. elem. primer trat.	4
núm. elem. segundo trat.	6
núm. elem. tercer trat.	5
Hipótesis Nula	igualdad de variancias
Hipótesis Alterna	no todas son iguales
Nivel de Significancia	0.05
ji cuadr 0.05 con 2 g.l.	5.991

la *b* calculada resulta de 0.218, por lo que se acepta la hipótesis nula y se concluye que las variancias de las tres poblaciones son iguales.

6.4.3.- Problema #3

Para el ensamble de cierto medidor se consideran cuatro máquinas diferentes. En un experimento para compararlas se decide emplear a seis operadores diferentes. El manejo de las máquinas requiere destreza física y se sabe que entre los operadores hay una diferencia en la velocidad a que pueden operar las máquinas. Las mediciones básicas con que se comparan las máquinas son los tiempos de terminación del ensamble.

Tiempo de ensamble

	Operador						
Máquina	1	2	3	4	5	6	**Total**
1	42.5	39.3	39.6	39.9	42.9	43.6	247.8
2	39.8	40.1	40.5	42.3	42.5	43.1	248.3
3	40.2	40.5	41.3	43.4	44.9	45.1	255.4
4	41.3	42.2	43.5	44.2	45.9	42.3	259.4
Total	163.8	162.1	164.9	169.8	176.2	174.1	1010.9

a) Pruebe la Hipótesis H'_0, al nivel de significancia 0.05, de que las máquinas trabajan a la misma velocidad media.
b) Pruebe la Hipótesis H''_0, al nivel de significancia 0.05, de que los operadores trabajan a la misma velocidad media.

Solución:

número de bloques	6
número de tratamientos	4

a)

Hipótesis Nula	efecto de máquinas nulo
Hipótesis Alterna	un efecto <> 0
Nivel de Significancia	0.05
f al 0.05 con (3,15) g.l.	3.29

b)

Hipótesis Nula	efecto operadores nulo
Hipótesis Alterna	un efecto <> 0
Nivel de Significancia	0.05
f al 0.05 con (5,15) g.l.	2.90

Análisis de Variancia para dos criterios de clasificación

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	*f* calculada
Máquinas	15.92	3	5.31	3.34
Operadores	42.09	5	8.42	5.29
Error	23.85	15	1.59	
Total	81.86	23		

Se rechazan ambas hipótesis nulas y se concluye que ni las máquinas ni los trabajadores trabajan a la misma velocidad media.

7.- Experimentos Factoriales

7.1.- Definición

Muchos experimentos incluyen el estudio del efecto de dos o más factores. Por lo general los diseños factoriales son los más eficientes para estos experimentos.

Si consideramos un caso de experimento químico, en el que se varían simultáneamente la presión y la temperatura para estudiar el avance de una reacción, se dice que es del tipo factorial, si en cada corrida, ensayo o intento se investigan todas las combinaciones posibles entre los niveles de cada uno de los factores.

El término factor se emplea en general para indicar cualquier aspecto del experimento como temperatura, tiempo o presión y que puede modificarse entre un ensayo y otro.

Los niveles de un factor son los valores reales usados en el experimento.

Cuando se tiene una respuesta o serie de ellas a un factor y a un nivel determinado, se tiene una celda

El efecto de un factor se define como el cambio en la respuesta o resultado, debido al cambio en el nivel del factor. Se le conoce también como un efecto principal debido a que se refiere a los factores primarios de interés en el experimento.

Además de estos efectos principales, pueden encontrarse efectos secundarios conocidos como interacción entre los factores. Esto sucede cuando la respuesta entre los niveles de un factor no es la misma a diferentes niveles de los demás factores.

Cuando se presenta la mencionada interacción, se enmascara muchas veces el efecto principal de alguno de los factores. El investigador debe en estos casos fijar los valores de todos los factores adicionales si desea conocer ese efecto principal.

Las ventajas que ofrecen los experimentos factoriales pueden apreciarse al comparar este método de investigación con el que varía un solo factor a la vez y requiere por lógica un mayor número de ensayos. En el experimento factorial se realizan los ensayos diseñados para cada celda y las mismas respuestas en ellas sirven para sacar conclusiones sobre los efectos principales de cada factor y sobre la interacción que pueda existir entre ellos.

Comúnmente se plantea un modelo general para el fenómeno que se desea estudiar

$$y_{ijk\ldots} = \mu + \alpha_i + \beta_j + \ldots + (\alpha\beta)_{ij} + \ldots + \epsilon_{ijk\ldots}$$

en el cual se consideran los efectos principales y las interacciones que pudieran existir entre todos los efectos.

Posteriormente se establecen Hipótesis Nulas para cada término de la anterior ecuación donde se plantea que los efectos de cada celda correspondiente son iguales a cero, con la Hipótesis Alterna de que por lo menos uno de esos efectos tiene un valor diferente de cero.

Este sistema se encarga de establecer la tabla de análisis de variancia para poder concluir si los efectos principales de los factores y la interacción entre ellos, son significativos o no.

Para poder estimar la suma de cuadrados debidos al error, deben tenerse por lo menos 2 réplicas en cada *celda*.

Para obtener un modelo matemático del fenómeno a estudiar puede utilizarse el capítulo de **Regresión Múltiple** .

7.2.- Menú

7.2.1.- Dos Factores

$$y_{ijk} = \mu + \alpha_i + \beta_j + + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

donde μ es el efecto global, α_i es el efecto iésimo del factor A, β_j es el j-ésimo efecto del factor B, $(\alpha\beta)_{ij}$ es el efecto de interacción entre A y B, y ϵ_{ijk} es el componente de error aleatorio.

Las Hipótesis Nulas que se plantean son entonces:

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$

$$H_1: por\ lo\ menos\ una\ \alpha_i \neq 0$$

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_a = 0$$

$$H_1: por\ lo\ menos\ una\ \beta_i \neq 0$$

$$H_0: (\alpha\beta)_{ij} = 0 \qquad para\ toda\ ij$$

$$H_1: por\ lo\ menos\ una\ (\alpha\beta)_{ij} \neq 0$$

Arreglo de Datos para un Experimento 2 Factorial

Factor B

		1	2	...	b
Factor A	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$
	⋮				
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

Tabla de Análisis de Variancia 2 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

Para el cálculo de la Tabla de Análisis de Variancia se requiere además de las siguientes ecuaciones relacionadas:

$$y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}, \qquad y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}, \qquad y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$$

$$SS_{subtotal(AB)} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn}, \qquad y_{ij.} = \sum_{k=1}^{n} y_{ijk}$$

$$SS_{AB} = SS_{subtotal(AB)} - SS_A - SS_B$$

$$SS_T \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B = SS_T - SS_{Subtotal(AB)}$$

7.2.2.- Tres Factores

En forma análoga al inciso anterior se plantea un modelo adecuado al fenómeno afectado por tres factores y las Hipótesis Nulas correspondientes.

Tabla de Análisis de Variancia 3 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	c-1	$MS_C = \frac{SS_C}{c-1}$	$F_0 = \frac{MS_C}{MS_E}$
AB	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	(a-1)(c-1)	$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_E}$
CB	SS_{BC}	(b-1)(c-1)	$MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	(a-1)(b-1)(c-1)	$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_E}$
Error	SS_E	abc (n-1)	$MS_E = \frac{SS_E}{abc(n-1)}$	
Total	SS_T	abcn-1		

Para el cálculo de la Tabla de Análisis de Variancia se requiere además de las siguientes ecuaciones relacionadas:

$$y_{....} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i\ldots}^2}{bcn} - \frac{y_{\ldots\ldots}^2}{abcn}, \qquad y_{i\ldots} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j..}^2}{acn} - \frac{y_{\ldots\ldots}^2}{abcn}, \qquad y_{.j..} = \sum_{i=1}^{a} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_C = \sum_{k=1}^{c} \frac{y_{..k.}^2}{abn} - \frac{y_{\ldots\ldots}^2}{abcn}, \qquad y_{..k.} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij..}^2}{cn} - \frac{y_{\ldots\ldots}^2}{abcn} - SS_A - SS_B, \qquad y_{ij..} = \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_{AC} = \sum_{i=1}^{a} \sum_{k=1}^{c} \frac{y_{i.k.}^2}{bn} - \frac{y_{\ldots\ldots}^2}{abcn} - SS_A - SS_C, \qquad y_{i.k.} = \sum_{j=1}^{b} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_{BC} = \sum_{j=1}^{b} \sum_{k=1}^{c} \frac{y_{.jk.}^2}{an} - \frac{y_{\ldots\ldots}^2}{abcn} - SS_B - SS_C, \qquad y_{.jk.} = \sum_{i=1}^{a} \sum_{l=1}^{n} y_{ijkl}$$

$$SS_{subtotal(ABC)} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \frac{y_{ijk.}^2}{n} - \frac{y_{\ldots\ldots}^2}{abcn}, \qquad y_{ijk.} = \sum_{l=1}^{n} y_{ijkl}$$

$$SS_{ABC} = SS_{subtotal(ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{\ldots\ldots}^2}{abcn}$$

$$SS_E = SS_T - SS_{subtotal(ABC)}$$

7.2.3.- Cuatro Factores

En forma análoga al inciso anterior se plantea un modelo adecuado al fenómeno afectado por cuatro factores y las Hipótesis Nulas correspondientes.

Tabla de Análisis de Variancia 4 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	c-1	$MS_C = \frac{SS_C}{c-1}$	$F_0 = \frac{MS_C}{MS_E}$
D	SS_D	d-1	$MS_D = \frac{SS_D}{d-1}$	$F_0 = \frac{MS_D}{MS_E}$
AB	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	(a-1)(c-1)	$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_E}$
AD	SS_{AD}	(a-1)(d-1)	$MS_{AD} = \frac{SS_{AD}}{(a-1)(d-1)}$	$F_0 = \frac{MS_{AD}}{MS_E}$
BC	SS_{BC}	(b-1)(c-1)	$MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_E}$
BD	SS_{BD}	(b-1)(d-1)	$MS_{BD} = \frac{SS_{BD}}{(b-1)(d-1)}$	$F_0 = \frac{MS_{BD}}{MS_E}$
CD	SS_{CD}	(c-1)(d-1)	$MS_{CD} = \frac{SS_{CD}}{(c-1)(d-1)}$	$F_0 = \frac{MS_{CD}}{MS_E}$
ABC	SS_{ABC}	(a-1)(b-1)(c-1)	$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_E}$

ABD	SS_{ABD}	(a-1)(b-1)(d-1)	$MS_{ABD}=\frac{SS_{ABD}}{(a-1)(b-1)(d-1)}$	$F_0=\frac{MS_{ABD}}{MS_E}$
ACD	SS_{ACD}	(a-1)(c-1)(d-1)	$MS_{ACD}=\frac{SS_{ACD}}{(a-1)(c-1)(d-1)}$	$F_0=\frac{MS_{ACD}}{MS_E}$
BCD	SS_{BCD}	(b-1)(c-1)(d-1)	$MS_{BCD}=\frac{SS_{BCD}}{(b-1)(c-1)(d-1)}$	$F_0=\frac{MS_{BCD}}{MS_E}$
ABCD	SS_{ABCD}	(a-1)(b-1)(c-1)(d-1)	$MS_{ABCD}=\frac{SS_{ABCD}}{(a-1)(b-1)(c-1)(d-1)}$	$F_0=\frac{MS_{ABCD}}{MS_E}$
Error	SS_E	abcd (n-1)	$MS_E=\frac{SS_E}{abcd(n-1)}$	
Total	SS_T	abcdn-1		

Para el cálculo de la Tabla de Análisis de Variancia se requiere además de las siguientes ecuaciones relacionadas:

$$y_{.....}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_A=\sum_{i=1}^{a}\frac{y_{i....}^2}{bcdn}-\frac{y_{.....}^2}{abcdn}, \qquad y_{i....}=\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_B=\sum_{j=1}^{b}\frac{y_{.j...}^2}{acdn}-\frac{y_{.....}^2}{abcdn}, \qquad y_{.j...}=\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_C=\sum_{k=1}^{c}\frac{y_{..k..}^2}{abdn}-\frac{y_{.....}^2}{abcdn}, \qquad y_{..k..}=\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_D=\sum_{l=1}^{d}\frac{y_{...l.}^2}{abcn}-\frac{y_{.....}^2}{abcdn}, \qquad y_{...l.}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_{AB}=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij...}^2}{cdn}-\frac{y_{.....}^2}{abcdn}-SS_A-SS_B, \qquad y_{ij...}=\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n}y_{ijklm}$$

$$SS_{AC} = \sum_{i=1}^{a}\sum_{k=1}^{c}\frac{y_{i.k..}^2}{bdn} - \frac{y_{.....}^2}{abcdn} - SS_A - SS_C . \qquad y_{i.k..} = \sum_{j=1}^{b}\sum_{l=1}^{d}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{AD} = \sum_{i=1}^{a}\sum_{l=1}^{d}\frac{y_{i..l.}^2}{bcn} - \frac{y_{.....}^2}{abcdn} - SS_A - SS_D . \qquad y_{i..l.} = \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{BC} = \sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{.jk..}^2}{adn} - \frac{y_{.....}^2}{abcdn} - SS_B - SS_C . \qquad y_{.jk..} = \sum_{i=1}^{a}\sum_{l=1}^{d}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{BD} = \sum_{j=1}^{b}\sum_{l=1}^{d}\frac{y_{.j.l.}^2}{acn} - \frac{y_{.....}^2}{abcdn} - SS_B - SS_D . \qquad y_{.j.l.} = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{CD} = \sum_{k=1}^{c}\sum_{l=1}^{d}\frac{y_{..kl.}^2}{abn} - \frac{y_{.....}^2}{abcdn} - SS_C - SS_D . \qquad y_{..kl.} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{ABC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y_{ijk..}^2}{dn} - \frac{y_{.....}^2}{abcdn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} .$$

$$y_{ijk..} = \sum_{l=1}^{d}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{ABD} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{d}\frac{y_{ij.l.}^2}{cn} - \frac{y_{.....}^2}{abcdn} - SS_A - SS_B - SS_D - SS_{AB} - SS_{AD} - SS_{BD} .$$

$$y_{ij.l.} = \sum_{k=1}^{c}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{ACD} = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\frac{y_{i.kl.}^2}{bn} - \frac{y_{.....}^2}{abcdn} - SS_A - SS_C - SS_D - SS_{AC} - SS_{AD} - SS_{CD} .$$

$$y_{i.kl.} = \sum_{j=1}^{b}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{BCD} = \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\frac{y_{.jkl.}^2}{an} - \frac{y_{.....}^2}{abcdn} - SS_B - SS_C - SS_D - SS_{BC} - SS_{BD} - SS_{CD},$$

$$y_{.jkl.} = \sum_{i=1}^{a}\sum_{m=1}^{n} y_{ijklm}$$

$$SS_{subtotal(ABCD)} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\frac{y_{ijkl.}^2}{n} - \frac{y_{.....}^2}{abcdn}, \qquad y_{ijkl.} = \sum_{m=1}^{n} y_{ijklm}$$

$$SS_{ABCD} = SS_{subtotal(ABCD)} - SS_A - SS_B - SS_C - SS_D - SS_{AB} - SS_{AC} - SS_{AD} - SS_{BC} - SS_{BD} - SS_{CD} - SS_{ABC} - SS_{ABD} - SS_{ACD} - SS_{BCD}$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{n} y_{ijklm}^2 - \frac{y_{.....}^2}{abcdn}$$

$$SS_E = SS_T - SS_{subtotal(ABCD)}$$

7.2.4.- Cinco Factores

En forma análoga al inciso anterior se plantea un modelo adecuado al fenómeno afectado por cinco factores y las Hipótesis Nulas correspondientes.

Tabla de Análisis de Variancia 5 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	SS_A	a-1	$MS_A = \frac{SS_A}{(a-1)}$	$F_0 = \frac{MS_A}{MS_{ER}}$
B	SS_B	b-1	$MS_B = \frac{SS_B}{(b-1)}$	$F_0 = \frac{MS_B}{MS_{ER}}$
C	SS_C	c-1	$MS_C = \frac{SS_C}{(c-1)}$	$F_0 = \frac{MS_C}{MS_{ER}}$
D	SS_D	d-1	$MS_D = \frac{SS_D}{(d-1)}$	$F_0 = \frac{MS_D}{MS_{ER}}$
E	SS_E	e-1	$MS_E = \frac{SS_E}{(e-1)}$	$F_0 = \frac{MS_E}{MS_{ER}}$
AB	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_{ER}}$
AC	SS_{AC}	(a-1)(c-1)	$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_{ER}}$
AD	SS_{AD}	(a-1)(d-1)	$MS_{AD} = \frac{SS_{AD}}{(a-1)(d-1)}$	$F_0 = \frac{MS_{AD}}{MS_{ER}}$
AE	SS_{AE}	(a-1)(e-1)	$MS_{AE} = \frac{SS_{AE}}{(a-1)(e-1)}$	$F_0 = \frac{MS_{AE}}{MS_{ER}}$
BC	SS_{BC}	(b-1)(c-1)	$MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_{ER}}$
BD	SS_{BD}	(b-1)(d-1)	$MS_{BD} = \frac{SS_{BD}}{(b-1)(d-1)}$	$F_0 = \frac{MS_{BD}}{MS_{ER}}$

BE	SS_{BE}	(b-1)(e-1)	$MS_{BE} = \frac{SS_{BE}}{(b-1)(e-1)}$	$F_0 = \frac{MS_{BE}}{MS_{ER}}$
CD	SS_{CD}	(c-1)(d-1)	$MS_{CD} = \frac{SS_{CD}}{(c-1)(d-1)}$	$F_0 = \frac{MS_{CD}}{MS_{ER}}$
CE	SS_{CE}	(c-1)(e-1)	$MS_{CE} = \frac{SS_{CE}}{(c-1)(e-1)}$	$F_0 = \frac{MS_{CE}}{MS_{ER}}$
DE	SS_{DE}	(d-1)(e-1)	$MS_{DE} = \frac{SS_{DE}}{(d-1)(e-1)}$	$F_0 = \frac{MS_{DE}}{MS_{ER}}$
ABC	SS_{ABC}	(a-1)(b-1)(c-1)	$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_{ER}}$
ABD	SS_{ABD}	(a-1)(b-1)(d-1)	$MS_{ABD} = \frac{SS_{ABD}}{(a-1)(b-1)(d-1)}$	$F_0 = \frac{MS_{ABD}}{MS_{ER}}$
ABE	SS_{ABE}	(a-1)(b-1)(e-1)	$MS_{ABE} = \frac{SS_{ABE}}{(a-1)(b-1)(e-1)}$	$F_0 = \frac{MS_{ABE}}{MS_{ER}}$
ACD	SS_{ACD}	(a-1)(c-1)(d-1)	$MS_{ACD} = \frac{SS_{ACD}}{(a-1)(c-1)(d-1)}$	$F_0 = \frac{MS_{ACD}}{MS_{ER}}$
ACE	SS_{ACE}	(a-1)(c-1)(e-1)	$MS_{ACE} = \frac{SS_{ACE}}{(a-1)(c-1)(e-1)}$	$F_0 = \frac{MS_{ACE}}{MS_{ER}}$
ADE	SS_{ADE}	(a-1)(d-1)(e-1)	$MS_{ADE} = \frac{SS_{ADE}}{(a-1)(d-1)(e-1)}$	$F_0 = \frac{MS_{ADE}}{MS_{ER}}$
BCD	SS_{BCD}	(b-1)(c-1)(d-1)	$MS_{BCD} = \frac{SS_{BCD}}{(b-1)(c-1)(d-1)}$	$F_0 = \frac{MS_{BCD}}{MS_{ER}}$
BCE	SS_{BCE}	(b-1)(c-1)(e-1)	$MS_{BCE} = \frac{SS_{BCE}}{(b-1)(c-1)(e-1)}$	$F_0 = \frac{MS_{BCE}}{MS_{ER}}$
BDE	SS_{BDE}	(b-1)(d-1)(e-1)	$MS_{BDE} = \frac{SS_{BDE}}{(b-1)(d-1)(e-1)}$	$F_0 = \frac{MS_{BDE}}{MS_{ER}}$
CDE	SS_{CDE}	(c-1)(d-1)(e-1)	$MS_{CDE} = \frac{SS_{CDE}}{(c-1)(d-1)(e-1)}$	$F_0 = \frac{MS_{CDE}}{MS_{ER}}$
ABCD	SS_{ABCD}	(a-1)(b-1) (c-1)(d-1)	$MS_{ABCD} = \frac{SS_{ABCD}}{(a-1)(b-1)(c-1)(d-1)}$	$F_0 = \frac{MS_{ABCD}}{MS_{ER}}$

ABCE	SS_{ABCE}	(a-1)(b-1)(c-1)(e-1)	$MS_{ABCE}=\frac{SS_{ABCE}}{(a-1)(b-1)(c-1)(e-1)}$	$F_0=\frac{MS_{ABCE}}{MS_{ER}}$
ABDE	SS_{ABDE}	(a-1)(b-1)(d-1)(e-1)	$MS_{ABDE}=\frac{SS_{ABDE}}{(a-1)(b-1)(d-1)(e-1)}$	$F_0=\frac{MS_{ABDE}}{MS_{ER}}$
ACDE	SS_{ACDE}	(a-1)(c-1)(d-1)(e-1)	$MS_{ACDE}=\frac{SS_{ACDE}}{(a-1)(c-1)(d-1)(e-1)}$	$F_0=\frac{MS_{ACDE}}{MS_{ER}}$
BCDE	SS_{BCDE}	(b-1)(c-1)(d-1)(e-1)	$MS_{BCDE}=\frac{SS_{BCDE}}{(b-1)(c-1)(d-1)(e-1)}$	$F_0=\frac{MS_{BCDE}}{MS_{ER}}$
ABCDE	SS_{ABCDE}	(a-1)(b-1)(c-1)(d-1)(e-1)	$MS_{ABCDE}=\frac{SS_{ABCDE}}{(a-1)(b-1)(c-1)(d-1)(e-1)}$	$F_0=\frac{MS_{ABCDE}}{MS_{ER}}$
Error	SS_{ER}	abcde (n-1)	$MS_{ER}=\frac{SS_{ER}}{abcde(n-1)}$	
Total	SS_T	abcden-1		

Para el cálculo de la Tabla de Análisis de Variancia se requiere además de las siguientes ecuaciones relacionadas:

$$y_{......}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_A=\sum_{i=1}^{a}\frac{y_{i.....}^2}{bcden}-\frac{y_{......}^2}{abcden}\qquad y_{i.....}=\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_B=\sum_{j=1}^{b}\frac{y_{.j....}^2}{acden}-\frac{y_{......}^2}{abcden}\qquad y_{.j....}=\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_C=\sum_{k=1}^{c}\frac{y_{..k...}^2}{abden}-\frac{y_{......}^2}{abcden}\qquad y_{..k...}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_D=\sum_{l=1}^{d}\frac{y_{...l..}^2}{abcen}-\frac{y_{......}^2}{abcden}\qquad y_{...l..}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_E = \sum_{m=1}^{e} \frac{y^2_{....m.}}{abcdn} - \frac{y^2_{......}}{abcden}, \qquad y_{....m.} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y^2_{ij....}}{cden} - \frac{y^2_{......}}{abcden} - SS_A - SS_B,$$

$$y_{ij....} = \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{AC} = \sum_{i=1}^{a} \sum_{k=1}^{c} \frac{y^2_{i.k...}}{bden} - \frac{y^2_{......}}{abcden} - SS_A - SS_C,$$

$$y_{i.k...} = \sum_{j=1}^{b} \sum_{l=1}^{d} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{AD} = \sum_{i=1}^{a} \sum_{l=1}^{d} \frac{y^2_{i..l..}}{bcen} - \frac{y^2_{......}}{abcden} - SS_A - SS_D,$$

$$y_{i..l..} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{AE} = \sum_{i=1}^{a} \sum_{m=1}^{e} \frac{y^2_{i...m.}}{bcdn} - \frac{y^2_{......}}{abcden} - SS_A - SS_E,$$

$$y_{i...m.} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BC} = \sum_{j=1}^{b} \sum_{k=1}^{c} \frac{y^2_{.jk...}}{aden} - \frac{y^2_{......}}{abcden} - SS_B - SS_C,$$

$$y_{.jk...} = \sum_{i=1}^{a} \sum_{l=1}^{d} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BD} = \sum_{j=1}^{b} \sum_{l=1}^{d} \frac{y^2_{.j.l..}}{acen} - \frac{y^2_{......}}{abcden} - SS_B - SS_D.$$

$$y_{.j..m.}=\sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{CD}=\sum_{k=1}^{c}\sum_{l=1}^{d}\frac{y^2_{..kl..}}{aben}-\frac{y^2_{......}}{abcden}-SS_C-SS_D,$$

$$y_{..kl..}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{CE}=\sum_{k=1}^{c}\sum_{m=1}^{e}\frac{y^2_{..k.m.}}{abdn}-\frac{y^2_{......}}{abcden}-SS_C-SS_E,$$

$$y_{..k.m.}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{d}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{DE}=\sum_{l=1}^{d}\sum_{m=1}^{e}\frac{y^2_{...lm.}}{abcn}-\frac{y^2_{......}}{abcden}-SS_D-SS_E,$$

$$y_{...lm.}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{ABC}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{y^2_{ijk...}}{den}-\frac{y^2_{......}}{abcden}-SS_A-SS_B-SS_C-SS_{AB}-SS_{AC}-SS_{BC},$$

$$y_{ijk...}=\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{ABD}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{d}\frac{y^2_{ij.l..}}{cen}-\frac{y^2_{......}}{abcden}-SS_A-SS_B-SS_C-SS_{AB}-SS_{AD}-SS_{BD},$$

$$y_{ij.l..}=\sum_{k=1}^{c}\sum_{m=1}^{e}\sum_{p=1}^{n}y_{ijklmp}$$

$$SS_{ABE}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{e}\frac{y^2_{ij..m.}}{cdn}-\frac{y^2_{......}}{abcden}-SS_A-SS_B-SS_E-SS_{AB}-SS_{AE}-SS_{BE}.$$

$$y_{ij..m.} = \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ACD} = \sum_{i=1}^{a} \sum_{k=1}^{c} \sum_{l=1}^{d} \frac{y^2_{i.kl..}}{ben} - \frac{y^2_{......}}{abcden} - SS_A - SS_C - SS_D - SS_{AC} - SS_{AD} - SS_{CD},$$

$$y_{i.kl..} = \sum_{j=1}^{b} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ACE} = \sum_{i=1}^{a} \sum_{k=1}^{c} \sum_{m=1}^{e} \frac{y^2_{i.k.m.}}{bdn} - \frac{y^2_{......}}{abcden} - SS_A - SS_C - SS_E - SS_{AC} - SS_{AE} - SS_{CE},$$

$$y_{i.k.m.} = \sum_{j=1}^{b} \sum_{l=1}^{d} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ADE} = \sum_{i=1}^{a} \sum_{l=1}^{d} \sum_{m=1}^{e} \frac{y^2_{i..lm.}}{bcn} - \frac{y^2_{......}}{abcden} - SS_A - SS_D - SS_E - SS_{AD} - SS_{AE} - SS_{DE},$$

$$y_{i..lm.} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BCD} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{d} \frac{y^2_{.jkl..}}{aen} - \frac{y^2_{......}}{abcden} - SS_B - SS_C - SS_D - SS_{BC} - SS_{BD} - SS_{CD}.$$

$$y_{.jkl..} = \sum_{i=1}^{a} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BCE} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{m=1}^{e} \frac{y^2_{.jk.m.}}{adn} - \frac{y^2_{......}}{abcden} - SS_B - SS_C - SS_E - SS_{BC} - SS_{BE} - SS_{CE},$$

$$y_{.jk.m.} = \sum_{i=1}^{a} \sum_{l=1}^{d} \sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BDE} = \sum_{j=1}^{b} \sum_{l=1}^{d} \sum_{m=1}^{e} \frac{y^2_{.j.lm.}}{acn} - \frac{y^2_{......}}{abcden} - SS_B - SS_D - SS_E - SS_{BD} - SS_{BE} - SS_{DE},$$

$$y_{.j.lm.} = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{CDE} = \sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e} \frac{y^2_{..klm.}}{abn} - \frac{y^2_{......}}{abcden} - SS_C - SS_D - SS_E - SS_{CD} - SS_{CE} - SS_{DE}.$$

$$y_{..klm.} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ABCD} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d} \frac{y^2_{ijkl..}}{en} - \frac{y^2_{......}}{abcden} - SS_A - SS_B - SS_C - SS_D - SS_{AB} - SS_{AC} - SS_{AD}$$
$$- SS_{BC} - SS_{BD} - SS_{CD} - SS_{ABC} - SS_{ABD} - SS_{ACD} - SS_{BCD}$$

$$y_{ijkl..} = \sum_{m=1}^{e}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ABCE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{e}\frac{y^2_{ijk.m.}}{dn} - \frac{y^2_{.....}}{abcden} - SS_A - SS_B - SS_C - SS_E - SS_{AB} - SS_{AC} - SS_{AE}$$

$$- SS_{BC} - SS_{BE} - SS_{CE} - SS_{ABC} - SS_{ABE} - SS_{ACE} - SS_{BCE}$$

$$y_{ijk.m.} = \sum_{l=1}^{d}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ABDE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{d}\sum_{m=1}^{e}\frac{y^2_{ij.lm.}}{cn} - \frac{y^2_{.....}}{abcden} - SS_A - SS_B - SS_D - SS_E - SS_{AB} - SS_{AD} - SS_{AE}$$

$$- SS_{BD} - SS_{BE} - SS_{DE} - SS_{ABD} - SS_{ABE} - SS_{ADE} - SS_{BDE}$$

$$y_{ij.ml.} = \sum_{k=1}^{c}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{ACDE} = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\frac{y^2_{i.klm.}}{bn} - \frac{y^2_{.....}}{abcden} - SS_A - SS_C - SS_D - SS_E - SS_{AC} - SS_{AD} - SS_{AE}$$

$$- SS_{CD} - SS_{CE} - SS_{DE} - SS_{ACD} - SS_{ACE} - SS_{ADE} - SS_{CDE}$$

$$y_{i.klm.} = \sum_{j=1}^{b}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{BCDE} = \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\frac{y^2_{.jklm}}{an} - \frac{y^2_{.....}}{abcden} - SS_B - SS_C - SS_D - SS_E - SS_{BC} - SS_{BD} - SS_{BE}$$

$$- SS_{CD} - SS_{CE} - SS_{DE} - SS_{BCD} - SS_{BCE} - SS_{BDE} - SS_{CDE}$$

$$y_{.jklm} = \sum_{i=1}^{a}\sum_{p=1}^{n} y_{ijklmp}$$

$$SS_{subtotal(ABCDE)} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\frac{y^2_{ijklm.}}{n} - \frac{y^2_{.....}}{abcden}q. \qquad y_{ijklm.} = \sum_{p=1}^{n} y_{ijklmp}$$

$$\begin{aligned} SS_{ABCDE} = {} & SS_{subtotal(ABCDE)} - SS_A - SS_B - SS_C - SS_D - SS_E - SS_{AB} \\ & - SS_{AC} - SS_{AD} - SS_{AE} - SS_{BC} - SS_{BD} - SS_{BE} - SS_{CD} - SS_{CE} \\ & - SS_{DE} - SS_{ABC} - SS_{ABD} - SS_{ABE} - SS_{ACD} - SS_{ACE} - SS_{ADE} \\ & - SS_{BCD} - SS_{BCE} - SS_{BDE} - SS_{CDE} - SS_{ABCD} - SS_{ABCE} \\ & - SS_{ABDE} - SS_{ACDE} - SS_{BCDE} \end{aligned}$$

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{m=1}^{e} \sum_{p=1}^{n} y_{ijklmp}^2$$

$$SS_E = SS_T - SS_{subtotal(ABCDE)}$$

7.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pidan.

En los casos en los que existan restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menú previo.

La capacidad de resolución de problemas de experimentos factoriales depende de la capacidad de memoria disponible en el ordenador. Por lo general, bastan 64 K de memoria para la resolución de problemas con dos y tres factores con dos réplicas en cada caso. A partir de cuatro factores se requiere de una capacidad mayor.

7.4.- Problemas Resueltos

7.4.1.- Problema #1

Se piensa que el voltaje máximo de salida de un tipo especial de batería se ve influenciado por el material utilizado en las placas internas y por la temperatura del lugar donde es instalada. Se corre un experimento para corroborar este hecho, obteniéndose cuatro réplicas para cada celda del mismo, los cuales se presentan en la tabla:

Tipo de Material (A)	Temperatura (B) 50ºF		65ºF		80ºF	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Verifique si existe una interacción entre el tipo de material y la temperatura de trabajo de las baterías.

Solución:

Tabla de Análisis de Variancia 2 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	10 683.72	2	5 341.86	7.91
B	39 118.72	2	19 559.36	28.97
AB	9 613.78	4	2 403.44	3.56
Error	18 230.75	27	675.21	
Total	77 646.97	35		

7.4.2.- Problema #2

Un embotellador de refrescos desea estudiar el efecto de (A) el porcentaje de carbonatación, (B) la presión de operación en la máquina llenadora y (C) la velocidad en la línea de envasado en el volumen de llenado del liquido . Se realiza entonces un experimento factorial con dos réplicas a tres niveles de carbonatación y a dos niveles de presión y velocidad. El orden de las 24 observaciones necesarias se toma al azar. La codificación del experimento se presenta como sigue:

	Presión de Operación (B)			
	25 psi		30 psi	
Porcentaje (A)	Velocidad (C)		Velocidad (C)	
Carbonatación	200	250	200	250
10	-3	-1	-1	1
	-1	0	0	1
12	0	2	2	6
	1	1	3	5
14	5	7	7	10
	4	6	9	11

Analice la información anterior e indique si la relación entre los tres factores afecta realmente el volumen de llenado en las botellas.

Solución:

Tabla de Análisis de Variancia 3 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	252.75	2	126.38	178.41
B	45.38	1	45.38	64.06
C	22.04	1	22.04	31.12
AB	5.25	2	2.63	3.71
AC	.58	2	.29	.41
BC	1.04	1	1.04	1.47
ABC	1.08	2	.54	.76
Error	8.5	12	.71	
Total	336.63	23		

7.4.3.- Problema #3

Se desea conocer el efecto de cuatro factores en el avance de una reacción química. Para tal efecto de diseña un experimento factorial en el que se obtienen dos réplicas por cada celda resultante. Los datos obtenidos a los diferentes niveles de cada factor se presentan como s sigue :

	Factor B											
	25				30				35			
	Factor C				Factor C				Factor C			
	200		250		200		250		200		250	
	Factor D		Factor D		Factor D		Factor D		Factor D		Factor D	
Factor A	40	90	40	90	40	90	40	90	40	90	40	90
10	1 3	4 5	2 5	4 3	1 4	5 4	3 2	1 1	4 5	3 1	2 4	5 1
12	4 7	8 5	5 9	6 6	10 8	9 7	6 5	4 6	7 8	10 7	9 10	8 9
14	9 11	12 10	11 9	10 10	10 11	11 13	9 12	13 15	10 12	13 16	14 15	15 16

Establezca la Tabla de Análisis de Varianza para los datos anteriores y verifique si existe una interacción entre los cuatro factores.

Solución:

Tabla de Análisis de Variancia 4 Factores

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F_0
A	955.44	2	477.72	221.91
B	45.86	2	22.93	10.65
C	0.12	1	0.12	0.06
D	5.01	1	5.01	2.33
AB	31.14	4	7.78	3.62
AC	9.00	2	4.50	2.09
AD	13.78	2	6.89	3.20
BC	16.58	2	8.29	3.85
BD	0.36	2	0.18	0.08
CD	4.01	1	4.01	1.86
ABC	16.92	4	4.23	1.96
ABD	11.47	4	2.87	1.33
ACD	0.44	2	0.22	0.10
BCD	1.86	2	0.93	0.43
ABCD	15.81	4	3.95	1.84
Error	77.50	36	2.15	
Total	1205.32	71		

8.- Control de Calidad

8.1.- Definición

Este capítulo se ocupa del control estadístico de la calidad y trata de un método estadístico utilizado primordialmente en la industria para regular los medios de producción en cuanto a su calidad y confiabilidad.

El control estadístico de la calidad abarca tanto las cartas de control (contenido de esta parte del sistema) como los planes de muestreo de aceptación.

El punto de partida para la elaboración de cartas de control es la consideración de que '...la calidad medida de un producto manufacturado está siempre sujeta a una cierta variación fortuita. Algún sistema estable de causas fortuitas es inherente a cualquier esquema particular de producción e inspección. La variación propia de este modelo estable es inevitable. Las razones para la variación fuera de este modelo estable pueden ser descubiertas y corregidas...'.

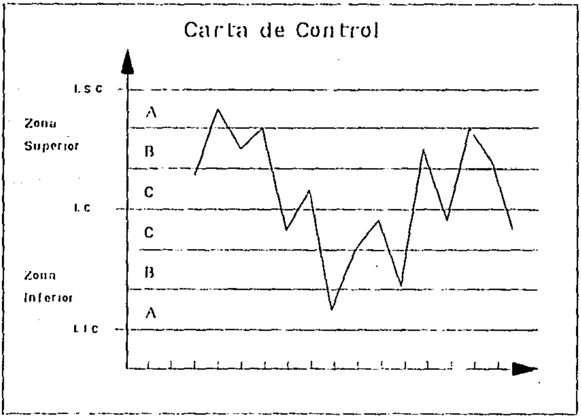
Se buscan, por lo tanto, pruebas para detectar modelos no naturales en la representación gráfica de datos.

La carta de control fue concebida por 'Shewhart' y se trata de un simple dispositivo gráfico para detectar modelos no naturales de variación en los datos resultantes de procesos repetitivos, o sea, resulta ser un criterio para detectar deficiencias con el control estadístico.

Cuando un proceso está actuando bajo un sistema constante de causas fortuitas, se dice que está bajo control estadístico.

Una carta de control se construye en un sistema de coordenadas, graficando de manera secuencial puntos con abscisa según el número de muestra y ordenada según la característica bajo investigación. La gráfica posee una zona delimitada por una horizontal superior llamada límite superior de control (LSC) y por otra denominada límite inferior de control (LIC). Esta zona delimitada es partida por una línea central (generalmente ubicada a la mitad), la que divide a su vez en dos zonas superior e inferior a la anterior.

Ahora bien, cada zona se vuelve a subdividir en tres zonas A, B y C, las cuales constituyen la tercera parte del área entre la línea central y el límite correspondiente de control, quedando las zonas A junto al límite de control, las B en medio y las C junto a la línea central.



Los puntos graficados se unen por una línea, como ayuda para la interpretación visual.

Con referencia a las zonas **A**, **B** y **C** , el modelo de variación observado se dice que es no natural o que el proceso está fuera de control si uno o más de los siguientes eventos ocurre:

1) Un sólo punto cae fuera del límite de control, o sea más allá de la zona **A**.
2) Dos de tres puntos sucesivos, caen en la zona A o más allá.
3) Cuatro de cinco puntos sucesivos, caen en la zona B o más allá.
4) Ocho de nueve puntos sucesivos caen en la zona C, o más allá.

Debe hacerse notar que las pruebas mencionadas se aplican a ambas mitades de la carta de control, pero se aplican separadamente para cada mitad, y nunca a las dos mitades en combinación.

Cabe mencionar que estas pruebas pueden usarse cuando los dos límites de control están colocados con razonable simetría con respecto a la línea central. Si los límites son decididamente asimétricos, la prueba debe ser modificada según métodos más sofisticados.

Al trabajar con cartas de control debe tenerse en mente que:
1) La razón primordial del uso de las cartas de control es la de proporcionar una señal de que alguna acción es necesaria.
2) Antes de que puedan calcularse los límites de control con alguna seguridad de su confiabilidad, se deberá disponer de cuando menos 20 subgrupos o muestras.
3) Antes de que los límites de control, calculados a partir de registros anteriores de producción, sean usados para supervisar la producción futura, el proceso deberá estar bajo control.
4) Si un proceso está bajo control y un punto cae fuera de los límites de control, el supervisor apreciará que se comete un error Tipo I.
5) La carta de control no es la solución universal para los problemas de producción, solamente es una herramienta muy útil.

8.2.- Menú

8.2.1.- Análisis de la Muestra Completa (Población Normal)

Esta carta de control se utiliza, por lo general, cuando se manejan mediciones. Se puede optar por el estudio de las medias o por el de los intervalos.

Para trazar las tres líneas principales de la carta se requiere de las ecuaciones siguientes:

Carta de Medias

$$LC = \bar{\bar{X}}$$

$$LSC = \bar{\bar{X}} + Q\bar{R}$$

$$LIC = \bar{\bar{X}} - Q\bar{R}$$

Carta de Intervalos

Es importante trabajar sólo con números positivos (valores absolutos) para el buen funcionamiento del sistema.

$$LC = \bar{R}$$

$$LSC = S_2 \bar{R}$$

$$LIC = S_1 \bar{R}$$

donde $\bar{X}$ es la media global de la muestra, $\bar{R}$ el intervalo promedio de la muestra y Q, S_1 y S_2 las constantes según la tabla:

Número de observaciones	Valor de	Valor de	Valor de
n	Q	S_1	S_2
2	1.880	0	3.267
3	1.023	0	2.575
4	0.729	0	2.282
5	0.577	0	2.115
6	0.483	0	2.004
7	0.419	0.076	1.924
8	0.373	0.136	1.864
9	0.337	0.184	1.816
10	0.308	0.223	1.777
11	0.285	0.256	1.744
12	0.266	0.284	1.716
13	0.249	0.308	1.692
14	0.235	0.329	1.671
15	0.223	0.348	1.652
16	0.212	0.364	1.636
17	0.203	0.379	1.621
18	0.194	0.392	1.608
19	0.187	0.404	1.596
20	0.180	0.414	1.586
21	0.173	0.425	1.575
22	0.167	0.434	1.566
23	0.162	0.443	1.557
24	0.157	0.452	1.548
25	0.153	0.459	1.541

Nota: cuando n>25 entonces Q, S_1 y S_2 adquieren un valor igual a $\frac{3}{\sqrt{n}}$

8.2.2.- Proporción de Defectuosos en una Muestra (Binomial)

Esta carta de control se utiliza, por lo general, cuando se desea manejar enumeraciones, como lo es el número de defectuosos en una muestra.

Para trazar las tres líneas principales de la carta se requiere de las ecuaciones siguientes:

$$LC = \overline{p}$$
$$LSC = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$
$$LIC = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

donde $\overline{p}$ es la proporción promedio de defectuosos en la muestra y n el tamaño de muestra.

8.2.3.- Número de Defectuosos por Unidad de Muestra (Poisson)

Esta carta de control se utiliza, por lo general, cuando se manejan enumeraciones, como lo es el número de defectos en un objeto. Es muy útil cuando se trata de un número grande de defectuosos en una muestra.

Para trazar las tres líneas principales de la carta se requiere de las ecuaciones siguientes:

$$LC = \bar{c}$$
$$LSC = \bar{c} + 3\sqrt{\bar{c}}$$
$$LIC = \bar{c} - 3\sqrt{\bar{c}}$$

donde $\bar{c}$ es el número promedio de defectos por unidad.

8.3.- Forma de Trabajar

Al elegir una opción dentro de este u otro capítulo, debe consultarse la ecuación o fórmula correspondiente, reunir los valores de las variables afectadas y alimentarlas al programa en la secuencia que se pidan.

En los casos en los que hay restricciones, como tamaño de muestra, valores diferentes de cero u otras, deben ser cumplidas. De otra manera, el programa no continuará y regresará al menu previo.

8.4.- Problemas Resueltos

8.4.1.- Problema #1

En un proceso de producción de piezas de seguridad se midieron los siguientes valores de resistencia a la compresión:

Número de	Valores Individuales				
la muestra	X_1	X_2	X_3	X_4	X_5
1	11.1	9.4	11.2	10.4	10.1
2	9.6	10.8	10.1	10.8	11.0
3	9.7	10.0	10.0	9.8	10.4
4	10.1	8.4	10.2	9.4	11.0
5	12.4	10.0	10.7	10.1	11.3
6	10.1	10.2	10.2	11.2	10.1
7	11.0	11.5	11.8	11.0	11.3
8	11.2	10.0	10.9	11.2	11.0
9	10.6	10.4	10.5	10.5	10.9
10	8.3	10.2	9.8	9.5	9.8
11	10.6	9.9	10.7	10.2	11.4
12	10.8	10.2	10.5	8.4	9.9
13	10.7	10.7	10.8	8.6	11.4
14	11.3	11.4	10.4	10.6	11.1
15	11.4	11.2	11.4	10.1	11.6
16	10.1	10.1	9.7	9.8	10.5
17	10.7	12.8	11.2	11.2	11.3
18	11.9	11.9	11.6	12.4	11.4
19	10.8	12.1	11.8	9.4	11.6
20	12.4	11.1	10.8	11.0	11.9

En base a los datos anteriores, establezca la carta de control y obtenga conclusiones respecto al proceso.

Solución:

Cálculo de promedios

Número de Muestra	Media $\bar{X}$	Intervalo $\bar{R}$
1	10.44	1.8
2	10.46	1.4
3	9.98	0.7
4	9.82	2.6
5	10.90	2.4
6	10.36	1.1
7	11.32	0.8
8	10.86	1.2
9	10.58	0.5
10	9.52	1.9
11	10.56	1.5
12	9.96	2.4
13	10.44	2.8
14	10.96	1.0
15	11.14	1.5
16	10.04	0.8
17	11.44	2.1
18	11.84	1.0
19	11.14	2.7
20	11.44	1.6
Promedio	10.44	1.59

Una vez hecha la carta de control para medias puede apreciarse

Para la Zona Superior de la Tabla de Control, por lo menos:

Un punto más allá de la zona A
Dos puntos consecutivos más allá de la zona B
Cuatro puntos más allá de la zona C

Para la Zona Inferior de la Tabla de Control,

Un punto más allá de la zona A
Dos puntos consecutivos más allá de la zona B
Cuatro puntos más allá de la zona C

Conclusión: El proceso se encuentra fuera de control.

Una vez hecha la carta de control para intervalos puede apreciarse

Para la Zona Superior de la Tabla de Control,
por lo menos:

Un punto más allá de la zona A
Dos puntos consecutivos más allá de la zona B
Cuatro puntos más allá de la zona C

Para la Zona Inferior de la Tabla de Control,

Un punto más allá de la zona A
Dos puntos consecutivos más allá de la zona B
Cuatro puntos más allá de la zona C

Conclusión: El proceso se encuentra fuera de control.

8.4.2.- Problema #2

Durante el proceso de producción de tubos de ensayo, se tomaron 40 muestras de tamaño 50 en tiempos al azar y se contaron (enumeraron) los defectuosos en cada una de ellas:

Número de Muestra	Número de defectuosos

1	2
2	1
3	2
4	0
5	2
6	3
7	4
8	2
9	0
10	3
11	0
12	1
13	2
14	2
15	3
16	5
17	1
18	2
19	3
20	1
21	1
22	1
23	4
24	2
25	2
26	4
27	1
28	3
29	3
30	2
31	3
32	6
33	2
34	3
35	2
36	3
37	1
38	0
39	2
40	0

En base a los datos anteriores, construya la carta de control correspondiente y obtenga conclusiones pertinentes.

Solución:

Cálculo de proporciones y su promedio

Número de Muestra	Fracción defectuosa (p)
1	0.04
2	0.02
3	0.04
4	0.00
5	0.04
6	0.06
7	0.08
8	0.04
9	0.00
10	0.06
11	0.00
12	0.02
13	0.04
14	0.04
15	0.06
16	0.10
17	0.02
18	0.04
19	0.06
20	0.02
21	0.02
22	0.02
23	0.08
24	0.04
25	0.04
26	0.08
27	0.02
28	0.06
29	0.06
30	0.04
31	0.06
32	0.12
33	0.04
34	0.06
35	0.04
36	0.06
37	0.02
38	0.00
39	0.04
40	0.00
Promedio	0.042

Una vez hecha la carta de control para proporciones puede apreciarse

Para la Zona Superior de la Tabla de Control :

¡Esta zona se encuentra bajo control!

Para la Zona Inferior de la Tabla de Control,

¡Esta zona se encuentra bajo control!

Conclusión: El proceso se encuentra bajo control.

8.4.3.- Problema #3

En un proceso de fabricación de reactores, se hizo el conteo de defectos en una junta soldada. Cada muestra se tomó de un total de 8:

Número de Muestra	Número de defectos (c)
1	2
2	4
3	7
4	3
5	1
6	4
7	8
8	9
9	5
10	3
11	7
12	11
13	6
14	4
15	9
16	9
17	6
18	4
19	3
20	9
21	7
22	4
23	7
24	12

En base a las observaciones anteriores, construya la carta de control apropiada y obtenga conclusiones pertinentes.

Una vez hecha la carta de control para conteo puede apreciarse

Para la Zona Superior de la Tabla de Control :

¡Esta zona se encuentra bajo control!

Para la Zona Inferior de la Tabla de Control,

¡Esta zona se encuentra bajo control!

Conclusión: El proceso se encuentra bajo control.

C.- DISCUSION FINAL

Una vez terminada la UNDIES II, el panorama alrededor de sus características, aplicaciones y alcances es positivo.

a.- Aspecto Didáctico

Como unidad didáctica, la UNDIES II cumple con la finalidad de proveer una herramienta sencilla para practicar los temas desarrollados en ella.

El usuario, sin conocer de lleno el tema de cada capitulo, puede ilustrarse con el tipo de problemas que pueden resolverse y con la secuencia en la que se resuelven. Con la utilización del sistema y el apoyo de la parte escrita se tiene en pocos minutos una impresión global del tema, sin perderse en detalles laterales, aunque quizás importantes, que distraen la atención de la idea central. O sea, el estudiante puede conocer alguna aplicación de la Estadística, aún cuando todavía requiera lidiar con bases supuestamente aprendidas, las cuales no domina, y que son requisito para los problemas que se le están presentando.

Lo anterior tiene dos ventajas. Una, el alumno se percata del lado práctico del capitulo. Otra, el alumno se apercibe de sus deficiencias con las bases de la Estadística y tiene la oportunidad de superarlas.

b.- Aspecto Operativo

La UNDIES II quedó configurada, de tal manera, que efectivamente preve posibles 'teclazos' fuera de lugar (por diversas causas) e indica errores cometidos, evita fallas en los cálculos y en el sistema, ilustra y fomenta en el usuario un mayor detenimiento con los problemas.

Debido a la extensión real del sistema, de la cual me hube percatado a lo largo de la elaboración, la atención mayor se centró, como era de esperarse, en que los cálculos resultaran correctos y no se pudo cubrir en su totalidad el aspecto de infalibilidad del sistema.

Expresado lo anterior, puede suceder que en alguno de los capitulos, no se hayan previsto todos los posibles errores en su utilización y pueda provocarse un paro en el sistema. Sin embargo, se confía en que la unidad sea 'infalible' en muchos casos.

c.- Actualidad del Sistema

Cuando se comenzó la elaboración de la UNDIES II se contaba en el Departamento Académico de Informática únicamente con ordenadores Apple y dos PC's Hewlett Packard. Durante el tiempo de estructuración del sistema el desarrollo de las computadoras tuvo, como era de esperarse, un impulso muy fuerte en nuestro país, por lo que, al término del trabajo se cuenta con máquinas más poderosas que las Apple inicialmente adquiridas.

Por lo tanto, aunque las **UNDIES I** y **UNDIES II** conservan su utilidad como herramientas didácticas, el equipo que las soporta queda en gran desventaja con lo actual.

Por otro lado, no sólo hubo cambios en el equipo a disposición, sino que también los hubo en el aspecto del Plan de Estudios de nuestra Facultad.

El cambio que nos incumbe, reside en la unificación de las dos asignaturas anteriores Estadística I y Estadística II, para cursar ahora únicamente Estadística en un sólo semestre.

En este aspecto, las **UNDIES** adquieren una importancia especial, pues pueden apoyar al alumno del nuevo plan de estudios agilizando su aprendizaje, aún cuando el tiempo de docencia programado de los temas estadísticos se haya reducido a la mitad.

d.- Futuro del Sistema

Como todo buen sistema, la **UNDIES II** es factible de mejoras y adaptaciones.

Una vez que la unidad se haya puesto a prueba y se hayan reconocido ampliamente las limitaciones de su forma actual, se podrá corregir, ampliar y traducir a lenguajes más poderosos.Entre otras cosas, puede complementarse con graficaciones, que en Computadores Personales se simplifica enormemente.

Lo anterior puede servir de base para una futura tesis, en la cual se extiendan los temas y sus aplicaciones, y la cual se adapte a computadoras más modernas.

Agosto 1989

Cuadrados y raíces cuadradas

n	n^2	$\sqrt{n}$	$\sqrt{10n}$	n	n^2	$\sqrt{n}$	$\sqrt{10n}$
1.0	1.00	1.000	3.162	5.5	30.25	2.345	7.416
1.1	1.21	1.049	3.317	5.6	31.36	2.366	7.483
1.2	1.44	1.095	3.464	5.7	32.49	2.387	7.550
1.3	1.69	1.140	3.606	5.8	33.64	2.408	7.616
1.4	1.96	1.183	3.742	5.9	34.81	2.429	7.681
1.5	2.25	1.225	3.873	6.0	36.00	2.449	7.746
1.6	2.56	1.265	4.000	6.1	37.21	2.470	7.810
1.7	2.89	1.304	4.123	6.2	38.44	2.490	7.874
1.8	3.24	1.342	4.243	6.3	39.69	2.510	7.937
1.9	3.61	1.378	4.359	6.4	40.96	2.530	8.000
2.0	4.00	1.414	4.472	6.5	42.25	2.550	8.062
2.1	4.41	1.449	4.583	6.6	43.56	2.569	8.124
2.2	4.84	1.483	4.690	6.7	44.89	2.588	8.185
2.3	5.29	1.517	4.796	6.8	46.24	2.608	8.246
2.4	5.76	1.549	4.899	6.9	47.61	2.627	8.307
2.5	6.25	1.581	5.000	7.0	49.00	2.646	8.367
2.6	6.76	1.612	5.099	7.1	50.41	2.665	8.426
2.7	7.29	1.643	5.196	7.2	51.84	2.683	8.485
2.8	7.84	1.673	5.292	7.3	53.29	2.702	8.544
2.9	8.41	1.703	5.385	7.4	54.76	2.720	8.602
3.0	9.00	1.732	5.477	7.5	56.25	2.739	8.660
3.1	9.61	1.761	5.568	7.6	57.76	2.757	8.718
3.2	10.24	1.789	5.657	7.7	59.29	2.775	8.775
3.3	10.89	1.817	5.745	7.8	60.84	2.793	8.832
3.4	11.56	1.844	5.831	7.9	62.41	2.811	8.888
3.5	12.25	1.871	5.916	8.0	64.00	2.828	8.944
3.6	12.96	1.897	6.000	8.1	65.61	2.846	9.000
3.7	13.69	1.924	6.083	8.2	67.24	2.864	9.055
3.8	14.44	1.949	6.164	8.3	68.89	2.881	9.110
3.9	15.21	1.975	6.245	8.4	70.56	2.898	9.165
4.0	16.00	2.000	6.325	8.5	72.25	2.915	9.220
4.1	16.81	2.025	6.403	8.6	73.96	2.933	9.274
4.2	17.64	2.049	6.481	8.7	75.69	2.950	9.327
4.3	18.49	2.074	6.557	8.8	77.44	2.966	9.381
4.4	19.36	2.098	6.633	8.9	79.21	2.983	9.434
4.5	20.25	2.121	6.708	9.0	81.00	3.000	9.487
4.6	21.16	2.145	6.782	9.1	82.81	3.017	9.539
4.7	22.09	2.168	6.856	9.2	84.64	3.033	9.592
4.8	23.04	2.191	6.928	9.3	86.49	3.050	9.644
4.9	24.01	2.214	7.000	9.4	88.36	3.066	9.695

Sumas de probabilidad binomial $\sum_{x=0}^{r} b(x; n, p)$

n	r	p 0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0064	0.0001
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7721	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

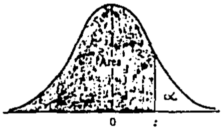
Sumas de probabilidades de Poisson $\sum_{x=0}^{r} p(x;\mu)$ *(continúa)*

r	μ 5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
0	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885
6	0.6860	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649
7	0.8095	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687
8	0.8944	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918
9	0.9462	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218
10	0.9747	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453
11	0.9890	0.9799	0.9661	0.9466	0.9208	0.8881	0.8487	0.8030	0.7520
12	0.9955	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364
13	0.9983	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981
14	0.9994	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400
15	0.9998	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665
16	0.9999	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823
17	1.0000	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911
18		1.0000	0.9999	0.9999	0.9997	0.9994	0.9987	0.9976	0.9957
19			1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980
20					1.0000	0.9999	0.9998	0.9996	0.9991
21						1.0000	0.9999	0.9998	0.9996
22							1.0000	0.9999	0.9999
23								1.0000	0.9999
24									1.0000

Sumas de probabilidades de Poisson $\sum_{x=0}^{r} p(x;\mu)$ *(continúa)*

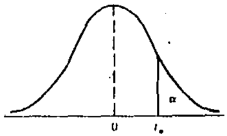
r	μ 10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
3	0.0103	0.0049	0.0023	0.0010	0.0005	0.0002	0.0001	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3750
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317
25		0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554
26		1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718
27			0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827
28			1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897
29				1.0000	0.9999	0.9996	0.9989	0.9973	0.9941
30					0.9999	0.9998	0.9994	0.9986	0.9967
31					1.0000	0.9999	0.9997	0.9993	0.9982
32						1.0000	0.9999	0.9996	0.9990
33							0.9999	0.9998	0.9995
34							1.0000	0.9999	0.9998
35								1.0000	0.9999
36									0.9999
37									1.0000

Areas bajo la curva normal



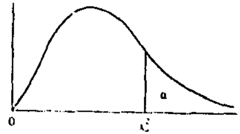
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8439	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Valores críticos de la distribución t



ν	α = 0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576

Valores críticos
de la distribución ji cuadrada



ν	α = 0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	3.841	5.024	6.635	7.879
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672

Valores críticos
de la distribución F

$f_{0.05}(\nu_1, \nu_2)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Valores críticos de la distribución F *(continúa)*

$f_{0.05}(\nu_1, \nu_2)$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Valores críticos de la distribución *F* (*continúa*)

$f_{0.01}(\nu_1, \nu_2)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Valores críticos de la distribución *F* (*continúa*)

$f_{0.01}(\nu_1, \nu_2)$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Tamaño de la muestra para la prueba de la media

	Nivel de la prueba t																				
Prueba unilateral	$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
Prueba bilateral	$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
0.05																					0.05
0.10																					0.10
0.15																				122	0.15
0.20										139					99					70	0.20
0.25					110					90				128	64			139	101	45	0.25
0.30				134	78				115	63			119	90	45		122	97	71	32	0.30
0.35			125	99	58			109	85	47		109	88	67	34		90	72	52	24	0.35
0.40		115	97	77	45		101	85	66	37	117	84	68	51	26	101	70	55	40	19	0.40
0.45		92	77	62	37	110	81	68	53	30	93	67	54	41	21	80	55	44	33	15	0.45
0.50	100	75	63	51	30	90	66	55	43	25	76	54	44	34	18	65	45	36	27	13	0.50
0.55	83	63	53	42	26	75	55	46	36	21	63	45	37	28	15	54	38	30	22	11	0.55
0.60	71	53	45	36	22	63	47	39	31	18	53	38	32	24	13	45	32	26	19	9	0.60
0.65	61	46	39	31	20	55	41	34	27	16	46	33	27	21	12	39	28	22	17	8	0.65
0.70	53	40	34	28	17	47	35	30	24	14	40	29	24	19	10	34	24	19	15	8	0.70
0.75	47	36	30	25	16	42	31	27	21	13	35	26	21	16	9	30	21	17	13	7	0.75
0.80	41	32	27	22	14	37	28	24	19	12	31	22	19	15	9	27	19	15	12	6	0.80
0.85	37	29	24	20	13	33	25	21	17	11	28	21	17	13	8	24	17	14	11	6	0.85
0.90	34	26	22	18	12	29	23	19	16	10	25	19	16	12	7	21	15	13	10	5	0.90
Valor de 0.95	31	24	20	17	11	27	21	18	14	9	23	17	14	11	7	19	14	11	9	5	0.95
$\Delta = \frac{\mu - \mu_0}{\sigma}$ 1.00	28	22	19	16	10	25	19	16	13	9	21	16	13	10	6	18	13	11	8	5	1.00
1.1	24	19	16	14	9	21	16	14	12	8	18	13	11	9	6	15	11	9	7		1.1
1.2	21	16	14	12	8	18	14	12	10	7	15	12	10	8	5	13	10	8	6		1.2
1.3	18	15	13	11	8	16	13	11	9	6	14	10	9	7		11	8	7	6		1.3
1.4	16	13	12	10	7	14	11	10	9	6	12	9	8	7		10	8	7	5		1.4
1.5	15	12	11	9	7	13	10	9	8	6	11	8	7	6		9	7	6			1.5
1.6	13	11	10	8	6	12	10	9	7	5	10	8	7	6		8	6	6			1.6
1.7	12	10	9	8	6	11	9	8	7		9	7	6	5		8	6	5			1.7
1.8	12	10	9	8	6	10	8	7	7		8	7	6			7	6				1.8
1.9	11	9	8	7	6	10	8	7	6		8	6	6			7	5				1.9
2.0	10	8	8	7	5	9	7	7	6		7	6	5			6					2.0
2.1	10	8	7	7		8	7	6	6		7	6				6					2.1
2.2	9	8	7	6		8	7	6	5		7	6				6					2.2
2.3	9	7	7	6		8	6	6			6	5				5					2.3
2.4	8	7	7	6		7	6	6			6										2.4
2.5	8	7	6	6		7	6	6			6										2.5
3.0	7	6	6	5		6	5	5			5										3.0
3.5	6	5	5			5															3.5
4.0	6																				4.0

Tamaño de la muestra para la prueba t de la diferencia entre dos medias

	Nivel de la prueba t																				
Prueba unilateral	$\alpha = 0.005$					$\alpha = 0.01$					$\alpha = 0.025$					$\alpha = 0.05$					
Prueba vilateral	$\alpha = 0.01$					$\alpha = 0.02$					$\alpha = 0.05$					$\alpha = 0.1$					
$\beta =$	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	
Valor de $\Delta = \frac{\mu_1 - \mu_2}{\sigma}$																					
0.05																					0.05
0.10																					0.10
0.15																					0.15
0.20																				137	0.20
0.25															124					88	0.25
0.30										123					87					61	0.30
0.35					110					90					64				102	45	0.35
0.40					85					70				100	50			108	78	35	0.40
0.45				118	68				101	55			105	79	39		108	88	62	29	0.45
0.50				96	55			106	82	45		106	88	64	32		88	70	51	23	0.50
0.55			101	79	46		106	88	68	39		87	71	53	27	112	73	58	42	19	0.55
0.60		101	85	67	39		90	74	58	32	104	74	60	45	23	89	61	49	36	16	0.60
0.65		87	73	57	34	104	77	64	49	27	88	63	51	39	20	76	52	42	30	14	0.65
0.70	100	75	63	50	29	90	66	55	43	24	76	55	44	34	17	66	45	36	26	12	0.70
0.75	88	66	55	44	26	79	58	46	38	21	67	48	39	29	15	57	40	32	23	11	0.75
0.80	77	58	49	39	23	70	51	43	33	19	59	42	34	26	14	50	35	29	21	10	0.80
0.85	69	51	43	35	21	62	46	38	30	17	52	37	31	23	12	45	31	25	18	9	0.85
0.90	62	46	39	31	19	55	41	34	27	15	47	34	27	21	11	40	28	22	16	8	0.90
0.95	55	42	35	28	17	50	37	31	24	14	42	30	25	19	10	36	25	20	15	7	0.95
1.00	50	38	32	26	15	45	33	28	22	13	38	27	23	17	9	33	23	18	14	7	1.00
1.1	42	32	27	22	13	38	28	23	19	11	32	23	19	14	8	27	19	15	12	6	1.1
1.2	36	27	23	18	11	32	24	20	16	0	27	20	16	12	7	23	16	13	10	5	1.2
1.3	31	23	20	16	10	28	21	17	14	8	23	17	14	11	8	20	14	11	9	5	1.3
1.4	27	20	17	14	9	24	18	15	12	8	20	15	12	10	6	17	12	10	8	4	1.4
1.5	24	18	15	13	8	21	16	14	11	7	18	13	11	9	5	15	11	9	7	4	1.5
1.6	21	16	14	11	7	19	14	12	10	6	16	12	10	8	5	14	10	8	6	4	1.6
1.7	19	15	13	10	7	17	13	11	9	6	14	11	9	7	4	12	9	7	6	3	1.7
1.8	17	13	11	10	6	15	12	10	8	5	13	10	8	6	4	11	8	7	6		1.8
1.9	16	12	11	9	6	14	11	9	8	5	12	9	7	6	4	10	7	6	5		1.9
2.0	14	11	10	8	6	13	10	9	7	5	11	8	7	6	4	9	7	6	4		2.0
2.1	13	10	9	8	5	12	9	8	7	5	10	8	6	5	3	8	6	5	4		2.1
2.2	12	10	8	7	5	11	9	7	6	4	9	7	6	5		8	6	5	4		2.2
2.3	11	9	8	7	5	10	8	7	6	4	9	7	6	5		7	5	5	4		2.3
2.4	11	9	8	6	5	10	8	7	6	4	8	6	5	4		7	5	4	4		2.4
2.5	10	8	7	6	4	9	7	6	5	4	8	6	5	4		6	5	4	3		2.5
3.0	8	6	6	5	4	7	6	5	4	3	6	5	4	4		5	4	2			3.0
3.5	6	5	5	4	3	6	5	4	4		5	4	4	3		4	3				3.5
4.0	6	5	4	4		5	4	4	3		4	4	3			4					4.0

Factores de tolerancia para distribuciones normales

$\gamma = 0.95$

n	$1-\alpha$ 0.90	0.95	0.99
2	32.019	37.674	48.430
3	8.380	9.916	12.861
4	5.369	6.370	8.299
5	4.275	5.079	6.634
6	3.712	4.414	5.775
7	3.369	4.007	5.248
8	3.136	3.732	4.891
9	2.967	3.532	4.631
10	2.839	3.379	4.433
11	2.737	3.259	4.277
12	2.655	3.162	4.150
13	2.587	3.081	4.044
14	2.529	3.012	3.955
15	2.480	2.954	3.878
16	2.437	2.903	3.812
17	2.400	2.858	3.754
18	2.366	2.819	3.702
19	2.337	2.784	3.656
20	2.310	2.752	3.615
25	2.208	2.631	3.457
30	2.140	2.549	3.350
35	2.090	2.490	3.272
40	2.052	2.445	3.213

$\gamma = 0.99$

n	$1-\alpha$ 0.90	0.95	0.99
2	160.193	188.491	242.300
3	18.930	22.401	29.055
4	9.398	11.150	14.527
5	6.612	7.855	10.260
6	5.337	6.345	8.301
7	4.613	5.488	7.187
8	4.147	4.936	6.468
9	3.822	4.550	5.966
10	3.582	4.265	5.594
11	3.397	4.045	5.308
12	3.250	3.870	5.079
13	3.130	3.727	4.893
14	3.029	3.608	4.737
15	2.945	3.507	4.605
16	2.872	3.421	4.492
17	2.808	3.345	4.393
18	2.753	3.279	4.307
19	2.703	3.221	4.230
20	2.659	3.168	4.161
25	2.494	2.972	3.904
30	2.385	2.841	3.733
35	2.306	2.748	3.611
40	2.247	2.677	3.518

$\gamma = 0.95$

n	$1-\alpha$ 0.90	0.95	0.99
45	2.021	2.408	3.165
50	1.996	2.379	3.126
55	1.976	2.354	3.094
60	1.958	2.333	3.066
65	1.943	2.315	3.042
70	1.929	2.299	3.021
75	1.917	2.285	3.002
80	1.907	2.272	2.986
85	1.897	2.261	2.971
90	1.889	2.251	2.958
95	1.881	2.241	2.945
100	1.874	2.233	2.934
150	1.825	2.175	2.859
200	1.798	2.143	2.816
250	1.760	2.121	2.788
300	1.767	2.106	2.767
400	1.749	2.084	2.739
500	1.737	2.070	2.721
600	1.729	2.060	2.707
700	1.722	2.052	2.697
800	1.717	2.046	2.688
900	1.712	2.040	2.682
1000	1.709	2.036	2.676
∞	1.645	1.960	2.576

$\gamma = 0.99$

n	$1-\alpha$ 0.90	0.95	0.99
45	2.200	2.621	3.444
50	2.162	2.576	3.385
55	2.130	2.538	3.335
60	2.103	2.506	3.293
65	2.080	2.478	3.257
70	2.060	2.454	3.225
75	2.042	2.433	3.197
80	2.026	2.414	3.173
85	2.012	2.397	3.150
90	1.999	2.382	3.130
95	1.987	2.368	3.112
100	1.977	2.355	3.096
150	1.905	2.270	2.983
200	1.865	2.222	2.921
250	1.839	2.191	2.880
300	1.820	2.169	2.850
400	1.794	2.138	2.809
500	1.777	2.117	2.783
600	1.764	2.102	2.763
700	1.755	2.091	2.748
800	1.747	2.082	2.736
900	1.741	2.075	2.726
1000	1.736	2.068	2.718
∞	1.645	1.960	2.576

BIBLIOGRAFIA

BHATTACHARYYA, Gouri K.
JOHNSON, Richard A.
Statistical Concepts and Methods
Primera Edición, John Wiley & Sons, USA, 1977

DE LA PARRA, Ma. Enriqueta
Nociones de Estadística
Primera Edición, Universidad Nacional Autónoma de México, México, 1984

FRY, Thornton C.
Probability and its Engineering Uses
Segunda Edición, D. van Nostrand Company, USA, 1965

LIPSCHUTZ, Seymour
Probabilidad
Primera Edición, Mc Graw Hill, México, 1980

MONTGOMERY, Douglas C.
Design and Analysis of Experiments
Segunda Edición, John Wiley & Sons, USA, 1984

OSTLE, Bernard
Statistics in Research
Segunda Edición, The Iowa State University Press, USA, 1964

SPIEGEL, Murray R.
Estadística
Primera Edición, Mc Graw Hill, México, 1970

WALPOLE, Ronald E.
Probabilidad y Estadística para Ingenieros
Segunda Edición, Interamericana, México, 1982

YAMANE, Taro
Estadística
Tercera Edición, Harla, México, 1979

Conjunto Estadístico
Manual del Módulo de Aplicación, Hewlett Packard

UNDIES II

INDICE de Términos de la Estadística II

UNDIES II