DIRECTORIO DE PROFESORES DEL CURSO :

Probabilidad y Estadística: Fundamentos y Aplicaciones 1982.

1. Dr. Octavio A. Rascón Chávez (Coordinador)
Subdirector
Instituto de Ingeniería
U N A M
México , D.F.
548 54 79

2. M. en I. Augusto Villarreal Aranda
Gerente de Operaciones
Grupo VEA
Asia No. 31
México 21, D.F.
554 45 31 y 554 41 31

'edcs.

PROBABILIDAD Y ESTADISTICA : FUNDAMENTOS Y APLICACIONES

8 de Febrero al 19 de Marzo de 1982

Fecha	Tema	Horario	Profesor
8 de Febrero al 5 de marzo	1. INTRODUCCION Probabilidad. Estadística descriptiva e inferencia estadística.	18:15 a 21:15 h c/día	Dr. Octavio A. Rascón Chávez
	2. ESTADISTICA DESCRIPTIVA Obtención de datos: muestreo aleatorio simple Procesamiento de información. Tabla de frecuencias. Histogramas. Polígonos de frecuencias. Medidas de tendencia central y de dispersión. Fractiles. Distribución conjunta de frecuencias. Regresión y correlación lineal. Análisis de series en el tiempo; predicción. Ejemplos y aplicaciones		
	3. PROBABILIDAD Eventos. Teoría de conjuntos. Espacio de eventos Probabilidad condicional. Independencia. Teorema de Bayes. Variables aleatorias continuas y discretas. Densidad de probabilidades. Función de distribución. Momentos y esperanzas. Distribuciones de Bernoulli, hipergeométrica, binominal y de Poisson, Proceso de Poisson simple. Distribuciones uniforme, exponencial y normal. Ejemplos y Aplicaciones.		M. en I. Augusto Villarreal Aranda.
Marzo 5 al 19	4. INFERENCIA ESTADISTICA Estimación puntual de los parámetros de una distribución de probabilidades. Estimación por intervalos Distribuciones muestrales. Pruebas de hipótesis que involucren medias, variancias o proporciones. Prueba de bondad de ajuste en regresión lineal y en distribuciones de probabilidades. Ejemplos y Aplicaciones	18 a 21 h	Dr. Octavio A. Rascón Chávez

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

PROBABILIDAD Y ESTADISTICA

PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA EN REGRESION LINEAL

DR. OCTAVIO A. RASCON CHAVEZ

MARZO 1982

PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA

EN REGRESION LINEAL

SI EL MODELO QUE RELACIONA A Y CON X ES LINEAL, ENTONCES

$$Y = MX + B$$

SI NO SE CONOCEN M Y B, ES NECESARIO ESTIMARLOS CON BASE EN UNA MUESTRA, CON LO CUAL SE OBTIENE

$$\tilde{Y} = mX + b$$

EN DONDE m ES EL ESTIMADOR DE M, Y b, EL DE B. SEA $\sigma^2_{y|x}$ LA VARIANCIA DE LA ESTIMACION DE Y CON BASE EN X.

SE PUEDE DEMOSTRAR QUE, SI SE CONOCE $\sigma^2_{y|x}$, ENTONCES:

$$Var(m) = \sigma^2_m = \sigma^2_{y|x} / \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma^2_{y|x}/nS^2_x$$

$$Var(b) = \sigma^2_b = \sigma^2_{y|x}/n + \frac{\bar{x}^2 \sigma^2_{y|x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sigma^2_{y|x}\left(\frac{1}{n} + \frac{\bar{x}^2}{nS^2_x}\right)$$

$$Var(mx + b) = \sigma^2_{y|x}/n + \frac{\sigma^2_{y|x}(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sigma^2_{\tilde{y}} = \sigma^2_{y|x}\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{nS^2_x}\right)$$

SI $\sigma^2_{y|x}$ NO SE CONOCE, SE PUEDE OBTENER UNA <u>ESTIMACION</u> <u>INSESGADA</u> DE ELLA MEDIANTE LA ECUACION

$$S^2_{y|x} = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

INTERVALOS DE CONFIANZA: $\sigma_{y|x}$ CONOCIDA

a. PARA LA ORDENADA EN EL ORIGEN, B,

$$b \pm z_c \sigma_b$$

DONDE $z_c = P(Z < z_c) = 1 - \alpha/2$; α = NIVEL DE SIGNIFICANCIA

b. PARA LA PENDIENTE, M:

$$m \pm z_c \sigma_m$$

c. PARA LA PREDICCION, Y_i:

$$\tilde{Y}_i \pm z_c \sigma_{\tilde{y}}$$

EN CASO DE QUE $\sigma_{y|x}$ SEA DESCONOCIDA (ES LO USUAL), DEBE ESTIMARSE A PARTIR DE LA MUESTRA MEDIANTE $S_{y|x}$. EN TAL CASO LOS INTERVALOS DE CONFIANZA CAMBIAN A:

a. PARA LA ORDENADA EN EL ORIGEN, B: $b \pm t_c \sigma_b$

$$b \pm t_c S_{y|x}\sqrt{\frac{\bar{x}^2}{nS^2_x}}$$

DONDE t_c ES EL VALOR CRITICO DE UN NIVEL DE SIGNIFICANCIA α, CORRESPONDIENTE A UNA DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD, Y S^2_x ES LA VARIANCIA (SESGADA) DE LA MUESTRA DE x.

b. PARA LA PENDIENTE, M: $m \pm t_c \sigma_m$

$$m \pm t_c S_{y|x}/\sqrt{nS^2_x} \quad \text{O} \quad m \pm t_c \frac{S_{y|x}}{S_x\sqrt{n}}$$

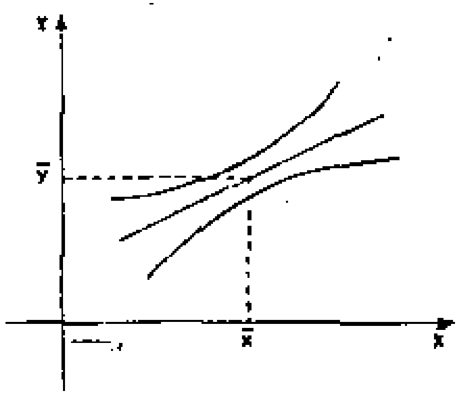
c. PARA LA PREDICCION, Y_i: $\hat{Y}_i \pm t_c \sigma_{\hat{y}}$

$$\hat{Y}_i \pm t_c\, s_{y|x} \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{nS_x^2}}$$

SI x_i ESTA DENTRO DEL RANGO DE LA MUESTRA, O

$$\hat{Y}_i \pm t_c\, s_{y|x} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{nS_x^2}}$$

SI x_i ESTA FUERA DEL RANGO.



EJEMPLO

LA FORMACION DEL ALCOHOL EN UN PROCESO DE FERMENTACION SE RELACIONA CON LA TEMPERATURA. EN UNA SERIE DE SEIS MEDICIONES A DISTINTAS TEMPERATURAS SE OBTUVO LO SIGUIENTE:

TEMPERATURA, x, °C	35	40	45	50	55	60
ALCOHOL, y, lt	20.2	23.1	23.2	23.6	25.8	26.3

SI SE AJUSTA UNA RECTA POR MINIMOS CUADRADOS SE OBTIENE

$$\tilde{Y} = 0.225\, X + 13.01$$

$(\bar{x} = 47.5,\ \bar{y} = 23.7)$

INTERVALOS DE CONFIANZA CON $\sigma^2_{y|x} = 0.8$ (CONOCIDA). $\alpha = 0.05$.

$$\sigma_b = \sqrt{\frac{0.8}{6} + \frac{0.8 \times 47.5^2}{437.5}} = 1.845$$

DONDE $\sum_{i=1}^{6} (x_i - \bar{x})^2 = 437.5$

$$b \pm z_c \sigma_b = 13.01 \pm 1.96 \times 1.845 = (9.39,\ 16.63)$$

$$\sigma_m = \sqrt{\frac{0.8}{437.5}} = 0.0428$$

$$m \pm z_c \sigma_m = 0.225 \pm 1.96 \times 0.0428 = 0.225 \pm 0.084 =$$

$$= (0.141,\ 0.309)$$

EJERCICIO

PARA LOS DATOS DE X y Y PRESENTADOS EN LA SIGUIENTE TABLA, CALCULAR $S_{y|x}$ Y LOS INTERVALOS DE CONFIANZA DE B y M PARA $\alpha = 0.05$, Y PARA Y CORRESPONDIENTE A X=50

Temp. x, °C	Alcohol lts.	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$x_i-\bar{x}_i$	$(x_i-\bar{x})^2$
35	20.2	20.9	-0.7	0.49	-12.5	156.2
40	23.1	22.0	1.1	1.21	- 7.5	56.2
45	23.2	23.1	0.1	0.01	- 2.5	6.2
50	23.6	24.3	-0.7	0.49	2.5	6.2
55	25.8	25.4	0.4	0.16	7.5	56.2
60	26.3	26.5	-0.2	0.04	12.5	156.2
Σ=285				Σ=2.40		Σ=437.2

DADOS QUE $\hat{y} = 0.225\,x + 13.01$

$\hat{y}(35)=0.225(35)+13.01=20.9,\qquad \hat{y}(40)=0.225\,(40)+13.01=22.0$, etc.

$\bar{x} = \frac{285}{6} = 47.5,\qquad S_x^2 = \frac{437.2}{6} = 72.9$

INTERVALOS DE CONFIANZA:

a) PARA B:
$$13.01 \pm t_c\, S_{y|x}\sqrt{\frac{\bar{x}^2}{nS_x^2} + \frac{1}{n}}$$

$t_c = t_{0.975,4} = 2.776,\qquad S_{y|x}^2 = \frac{1}{n-2}\ \Sigma(y_i-\hat{y}_i)^2 = \frac{1}{4}\ 2.4 = 0.6,$

$S_{y|x} = \sqrt{0.6} = 0.77$

$$13.01 \pm 2.776 \times 0.77\sqrt{\frac{47.5^2}{6(72.9)} + \frac{1}{6}} = 13.01 \pm 4.93 = (8.08, 17.94)$$

B) PARA M:
$$0.225 \pm t_c \frac{S_{y|x}}{\sqrt{nS_x^2}} = 0.225 \pm 2.776\,\frac{0.77}{\sqrt{6\ (72.9)}} =$$

$$= 0.225 \pm 0.102 = (0.123, 0.327)$$

C) PARA $y_i(x=50)$: $y_i(50) = 24.3$

$$24.3 \pm t_c S_{y|x}\sqrt{\frac{1}{n} + \frac{(x_i-\bar{x})^2}{n\,S_x^2}} = 24.3 \pm 2.776 \times 0.77\sqrt{\frac{1}{6} + \frac{(50-47.5)^2}{6(72.9)}} =$$

$$= 24.3 \pm 0.9 = (23.4, 25.2)$$

TAREA: HACER ESTIMACIONES DE INTERVALOS DE CONFIANZA PARA $\alpha = 0.05$ Y $\alpha = 0.01$, DE b, m Y Y_1, ESTE ULTIMO PARA UN $x = x_1$ QUE SELECCIONE CADA QUIEN. UTILIZAR UNO DE LOS PROBLEMAS DE REGRESION DEJADOS COMO TAREA ANTERIORMENTE.

PRUEBAS DE HIPOTESIS

a. PARA LA ORDENADA EN EL ORIGEN

SE DEMUESTRA QUE

$$\frac{B - b_0}{S_{y|x}\sqrt{\frac{\overline{x^2}}{nS_x^2}}} = \frac{B - b_0}{\frac{S_{y|x}}{S_x}\sqrt{\frac{\overline{x^2}}{n}}} = T$$

TIENE DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD.

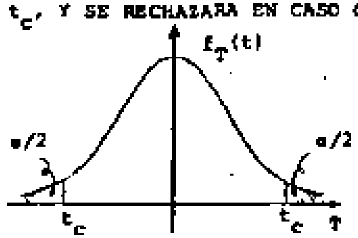
SI SE DESA PROBAR LA HIPOTESIS

$$H_0 : B = b_0$$
$$H_1 : B \neq b_0$$

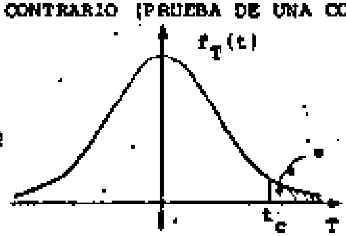
BASTA SUSTITUIR A $B = b_0$ EN LA ECUACION ANTERIOR Y EVALUAR $T = t$, ES DECIR,

$$t = \frac{b - b_0}{\frac{S_{y|x}}{S_x}\sqrt{\frac{\overline{x^2}}{n}}}$$

SE ACEPTARA H_0 SI $|t| < |t_c|$; EN CASO CONTRARIO SE RECHAZARA (PRUEBA DE DOS COLAS). SI H_1 FUERA $B > b_0$, SE ACEPTARA SI $t < t_c$, Y SE RECHAZARA EN CASO CONTRARIO (PRUEBA DE UNA COLA)



PRUEBA DE DOS COLAS



PRUEBA DE UNA COLA

b. PARA LA PENDIENTE, M

ANALOGAMENTE, PARA M, LA ESTADISTICA

$$\frac{M - m_0}{S_{y|x}/\sqrt{nS_x^2}} = \frac{M - m_0}{\frac{S_{y|x}}{S_x\sqrt{n}}} = T,$$

DONDE m_0 = VALOR DE M BAJO LA HIPOTESIS NULA $H_0 : M = m_0$.

TAMBIEN TIENE DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD: $t = \frac{m - m_0}{\frac{S_{y|x}}{S_x\sqrt{n}}}$

EJEMPLO

CONSIDERE LOS DATOS SIGUIENTES:

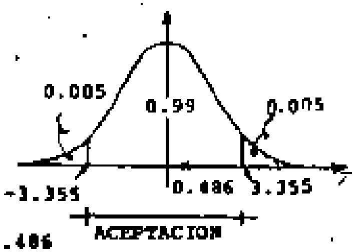
x	0	1	2	3	4	5	6	7	8	9
y	0.16	0.09	0.08	0.23	0.60	0.39	0.55	0.75	0.81	0.85

$m = 0.093$, $b = 0.032$, $S^2_{y|x} = 0.01258$

$S_x^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 82.50/10 = 8.25$; $\Sigma x_i^2 = 285$, $\overline{x^2} = \frac{285}{10} = 28.5$

a. PROBAR LA HIPOTESIS DE QUE $B = 0$

b. PROBAR LA HIPOTESIS DE QUE $M = 0.1$ CON $\alpha = 0.01$ Y $S_{y|x}$ DESCONOCIDA.



a. $H_0 : B = 0$; $H_1 : B \neq 0$

$$t = \frac{b - b_0}{S_{y|x}\sqrt{\frac{\overline{x^2}}{nS_x^2}}} = \frac{0.032 - 0}{\sqrt{0.01258}\sqrt{\frac{28.5}{10 \times 8.25}}} = 0.486$$

$t_c = t_{0.995,\,8} = 3.355 > 0.486 \therefore$ SE ACEPTA H_0.

b. $H_0 : M = 0.1; \quad H_1 : M \neq 0.1$

$$t = \frac{m - m_0}{\frac{S_{y|x}}{s_x\sqrt{n}}} = \frac{0.093 - 0.1}{\frac{\sqrt{0.01258}}{\sqrt{8.25 \times 10}}} = 0.567 < 3.355$$

SE ACEPTA H_0 CON 99% DE NIVEL DE CONFIANZA.

PRUEBA DE HIPOTESIS PARA EL COEFICIENTE DE CORRELACION, ρ_{xy}

PRUEBA

$H_0 : \rho_{xy} = 0 ; \quad H_1 : \rho_{xy} \gtrless\neq 0$

SE DEMUESTRA QUE EN CASO DE QUE X Y Y SON INDEPENDIENTES ($\rho = 0$), LA ESTADISTICA

$$T = r_{xy}\sqrt{\frac{n - 2}{1 - r_{xy}^2}}$$

TIENE DISTRIBUCION t DE STUDENT CON $n - 2$ GRADOS DE LIBERTAD.

EJEMPLO

EN BASE A UNA MUESTRA ALEATORIA DE 30 DATOS SOBRE LA TEMPERATURA MEDIA DURANTE UN MES, X, Y EL PESO MEDIO DE LOS TOMATES PISCADOS, Y, SE OBTUVO UN COEFICIENTE DE CORRELACION $r_{xy} = 0.931$.

PROBAR LA HIPOTESIS DE QUE $\rho_{xy} = 0$. USAR $\alpha = 0.05$.

$H_0 : \rho_{xy} = 0 ; \quad H_1 : \rho_{xy} \neq 0$

$$t = r_{xy}\sqrt{\frac{n - 2}{1 - r_{xy}^2}} = 0.931\sqrt{\frac{30 - 2}{1 - 0.931^2}} = 13.448$$

$t_c = t_{0.975,\ 28} = 2.048 < 13.448$

$\therefore$ SE RECHAZA H_0 A UN NIVEL DE CONFIANZA DEL 95%.

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

PROBABILIDAD Y ESTADISTICA

EJERCICIOS

DR. OCTAVIO A. RASCON CHAVEZ

MARZO 1982

EJEMPLO

EN UNA PRUEBA DE APTITUD QUE SE APLICO A 17 ALUMNOS DE PRIMERO DE SECUNDARIA, SELECCIONADOS AL AZAR DE LAS SECUNDARIAS DE UNA CIUDAD, SE OBTUVO UN PROMEDIO DE LAS CALIFICACIONES IGUAL A 34.35 PUNTOS, Y UNA DESVIACION ESTANDAR DE 3.9 PUNTOS. PROBAR LA HIPOTESIS DE QUE LA MEDIA DE LA VARIABLE ALEATORIA "CALIFICACION EN LA PRUEBA DE APTITUD DE LOS ALUMNOS DE 1° DE SECUNDARIA DE ESA CIUDAD" ES DE 38 PUNTOS, CONTRA LA DE QUE ES MENOR QUE 38. TOMAR $\alpha = 0.05$.

$$H_0: \mu = 38$$

$$H_1: \mu < 38$$

$$T = \frac{\bar{x} - \mu}{S_x}\sqrt{n-1} = \frac{34.35 - 38}{3.9}\sqrt{17-1} = -3.74$$

$$t_{c,0.05,16} = -1.746 > -3.74$$

PUESTO QUE $T < t_{c,0.05,16}$ SE RECHAZA LA HIPOTESIS NULA, CON 95% DE NIVEL DE CONFIANZA.

EJEMPLO

EN UN ESTUDIO DE MERCADOTECNIA SE TOMO UNA MUESTRA DE 20 PRECIOS DE CARNE EN 20 TIENDAS DISTINTAS PARA ESTIMAR SU VARIABILIDAD. LOS DATOS ARROJARON UN PROMEDIO $\bar{X}$ = \$92.00 Y UNA DESVIACION ESTANDAR S = \$8.00. CALCULAR EL INTERVALO DEL 95% DE NIVEL DE CONFIANZA DE LA VARIANCIA.

$$I.C. = \left(\frac{20(8)^2}{32.9}, \frac{20(8)^2}{8.91}\right) = (38.91, 143.66)\ \$^2$$

EL INTERVALO DE CONFIANZA PARA LA DESVIACION ESTANDAR ES

$$(\sqrt{38.91}, \sqrt{143.66}) = (6.24, 11.99)\ \$$$

EJERCICIO

LA DURACION DE LOS TRANSFORMADORES PRODUCIDOS EN UNA FABRICA FUE MEDIDA EN UNA MUESTRA DE 50 ELEMENTOS TOMADOS AL AZAR, OBTENIENDOSE LA SIGUIENTE DISTRIBUCION DE FRECUENCIAS:

INTERVALO N°	1	2	3	4
INTERVALO DE TIEMPO, AÑOS	$0 \leq t < 1$	$1 \leq t < 2$	$2 \leq t < 3$	$t \geq 3$
FRECUENCIA	21	16	9	4

PROBAR LA HIPOTESIS DE QUE LA DISTRIBUCION DE PROBABILIDADES DE LA VARIABLE ALEATORIA "DURACION DE LOS TRANSFORMADORES" ES EXPONENCIAL CON PARAMETRO $\lambda = 0.45$ AÑOS^{-1}. USAR $\alpha = 0.05$.

SOLUCION

LAS FRECUENCIAS ESPERADAS SON: $nP(x_1 \leq X < x_2)$

DONDE n = TAMAÑO DE LA MUESTRA

$$p_1 = P(0 \leq X < 1) = \int_0^1 (0.45)e^{-0.45t}\,dt = 0.362; \quad 50p_1 = 18.10$$

$$p_2 = P(1 \leq X < 2) = \int_1^2 (0.45)e^{-0.455}\,dt = 0.232; \quad 50p_2 = 11.60$$

$$p_3 = P(2 \leq X < 3) = \int_2^3 (0.45)e^{-0.45t}\,dt = 0.145; \quad 50p_3 = 7.25$$

$$p_4 = P(X > 4) = \int_4^\infty (0.45)e^{-0.45t}\,dt = \underline{0.259} \quad ; \quad 50p_4 \doteq \underline{12.95}$$

$$\Sigma = 0.998 \doteq 1 \qquad \Sigma = 49.90 \doteq 50$$

$$\chi^2 = \frac{(21-18.10)^2}{18.10} + \frac{(16-11.6)^2}{11.6} + \frac{(9-7.25)^2}{7.25} + \frac{(4-12.95)^2}{12.95} = \underline{8.71}$$

$$\chi^2_{0.95,3} = 7.81 < 8.74$$

POR LO QUE SE RECHAZA LA HIPOTESIS NULA CON UN 5% DE NIVEL DE SIGNIFICANCIA.

EJERCICIO

SE PIENSA QUE LA EMISION DE PARTICULAS RADIOACTIVAS DE CIERTA FUENTE OCURRE SEGUN UNA DISTRIBUCION DE PROBABILIDADES DE POISSON. EL NUMERO DE PARTICULAS EMITIDAS EN 100 INTERVALOS CONSECUTIVOS DE 10 SEG QUEDO DISTRIBUIDO DE LA SIGUIENTE MANERA

N° DE PARTICULAS	0	1	2	3	4	>4
N° DE INTERVALOS (FRECUENCIA)	11	30	25	20	10	4

PROBAR LA HIPOTESIS DE QUE EFECTIVAMENTE SE TRATA DE UNA DISTRIBUCION DE POISSON. USAR α = 0.01.

SOLUCION

PUESTO QUE NO NOS INDICAN UN VALOR DEL PARAMETRO DE LA DISTRIBUCION NECESITAMOS ESTIMARLO A PARTIR DE LA INFORMACION DADA ARRIBA:

$\lambda = [(0 \times 11) + (1 \times 30) + (2 \times 25) + (3 \times 20) + (4 \times 10) + (5 \times 4)]/100$

$= 2.00$ PARTICULAS/INTERVALO

LA DISTRIBUCION DE POISSON ES ENTONCES:

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^x e^{-2}}{x!} = P(X = x)$$

$p_1 = f_X(0) = 2^0 e^{-2}/0! = 0.135; \; np_1 = 100 \times 0.135 = 13.5$

$p_2 = f_X(1) = 2^1 e^{-2}/1! = 0.270; \; np_2 = 100 \times 0.270 = 27.0$

$p_3 = f_X(2) = 2^2 e^{-2}/2! = 0.270; \; np_3 = 27.0$

$p_4 = f_X(3) = 2^3 e^{-2}/3! = 0.180; \; np_4 = 18.0$

$p_5 = f_X(4) = 2^4 e^{-2}/4! = 0.090; \; np_5 = 9.0$

$p_6 = P(X \geq 5) = 1 - F_X(4) = 0.055; \; np_6 = 5.5$

$$\chi^2 = \frac{(11-13.5)^2}{13.5} + \frac{(30-27.0)^2}{27.0} + \frac{(25-27.0)^2}{27.0} + \frac{(20-18)^2}{18} + \frac{(10-9.0)^2}{9.0} + \frac{(4-5.5)^2}{5.5}$$

$= \underline{1.687}$ $\quad \nu = 6 - 1 - 1 = 4$ $\quad (\nu = N - r - 1)$; r = N° DE ESTIMACIONES HECHA CON LOS DATOS)

$\chi^2_{0.99,\,4} = 13.277 > 1.687$ $\therefore$ SE ACEPTA LA HIPOTESIS NULA

7

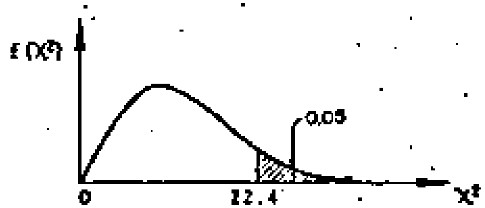
EJERCICIO

EN UN ESTUDIO CON FINES ANTROPOLOGICOS SE OBTUVO UNA MUESTRA ALEATORIA DEL TAMAÑO DE LA CABEZA DE LOS INDIGENAS ORIGINARIOS DE CIERTA REGION TROPICAL. LOS DATOS AGRUPADOS SE PRESENTAN EN LA SIGUIENTE TABLA. PROBAR LA HIPOTESIS DE QUE ESTOS DATOS CORRESPONDEN A UNA VARIABLE CON DISTRIBUCION NORMAL.

INTERVALO DE VALORES, mm	FRECUENCIA OBSERVADA, f_i	FRECUENCIA ESPERADA, e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
<171.5	0	0.4	0.4	0.16	0.40
171.5-175.5	3	2.4	0.6	0.36	0.15
175.5-179.5	9	10.5	-1.5	1.25	0.12
179.5-183.5	29	33.1	-4.1	16.81	0.51
183.5-187.5	76	71.3	4.7	22.09	0.31
187.5-191.5	104	104.2	-0.2	0.04	0.00
191.5-195.5	110	108.8	1.8	3.24	0.03
195.5-199.5	88	77.3	10.7	114.49	1.48
199.5-203.5	30	37.5	-7.5	56.25	1.50
203.5-207.5	6	13.0	-7.0	49.00	3.77
207.5-211.5	4	3.0	1.0	1.00	0.33
211.5-215.5	2	0.5082	1.4918	2.23	4.58
215.5-219.5	1	0.0462	0.9538	0.910	19.69
>219.5	0	≐ 0.		TOTAL:	32.67

$$\chi^2 = \sum_{i=1}^{N} \frac{(f_i - e_i)^2}{e_i}, \qquad \chi^2 = 32.67 > 22.4 = \chi^2_{0.95,13} = \chi^2_c$$

POR LO QUE LA HIPOTESIS NULA NO PUEDE RECHAZARSE CON UN 95% DE NIVEL DE CONFIANZA.



8

EJERCICIO

SACAR UNA MUESTRA DE 50 NUMEROS DE UNA TABLA DE NUMEROS ALEATORIOS Y PROBAR LA HIPOTESIS DE QUE PROVIENEN DE UNA DISTRIBUCION UNIFORME DE 0 A 1, PREVIA REDUCCION A DECIMALES. USAR $\alpha = 0.05$.

SOLUCION

UTILIZANDO LOS RENGLONES 1, 3, 5, 7, 9 DE LA TABLA DE NUMEROS ALEATORIOS PRESENTADA EN EL VOL. 1 DE ESTADISTICA DESCRITIVA*, MULTIPLICANDO x 10^{-5} CADA NUMERO Y ELIMINANDO LOS 3 ULTIMOS DIGITOS SE OBTIENE LA SIGUIENTE MUESTRA:

0.16 - 0.81 - 0.04 - 0.53 - 0.79 - 0.21 - 0.83 - 0.92 - 0.36 - 0.31
0.59 - 0.73 - 0.47 - 0.47 - 0.87 - 0.99 - 0.00 - 0.88 - 0.71 - 0.18
0.20 - 0.23 - 0.30 - 0.03 - 0.23 - 0.14 - 0.15 - 0.45 - 0.22 - 0.19
0.09 - 0.74 - 0.68 - 0.96 - 0.20 - 0.42 - 0.78 - 0.05 - 0.22 - 0.24
0.54 - 0.35 - 0.19 - 0.11 - 0.31 - 0.76 - 0.17 - 0.03 - 0.44 - 0.64

AGRUPANDO DATOS EN 10 INTERVALOS TENEMOS:

INTERVALO	f_i	e_i	$f_i - e_i$	$(f_i - e_i)^2$	$(f_i - e_i)^2/e_i$
0.000-0.105	6	5	1	1	0.20
0.105-0.205	10	5	5	25	5
0.205-0.305	7	5	2	4	0.80
0.305-0.405	4	5	-1	1	0.20
0.405-0.505	5	5	0	0	0
0.505-0.605	3	5	-2	4	0.80
0.605-0.705	2	5	-3	9	1.80
0.705-0.805	6	5	1	1	0.20
0.805-0.905	4	5	-1	1	0.20
0.905-1.005	3	5	2	4	0.80
					$\Sigma = 10.0$

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i} = 10.0$$

$$\chi^2_{0.95,9} = 16.9 > 10$$

POR LO TANTO SE ACEPTA LA HIPOTESIS DE QUE LOS NUMEROS CORRESPONDEN A UNA DISTRIBUCION UNIFORME, CON UN NIVEL DE SIGNIFICANCIA DE 0.05.

RASCON, O. A., "INTRODUCCION A LA ESTADISTICA DESCRIPTIVA", VOL. I, 3a. ED. (1977), PAG. 71.

It is sometimes difficult to decide whether or not a frequency distribution is sufficiently near to the normal type to be fitted by a normal curve. A preliminary decision in a given case is largely the result of experience—of good guessing. Such a decision, however, can be reinforced by a fairly simple test involving the use of arithmetic probability paper.

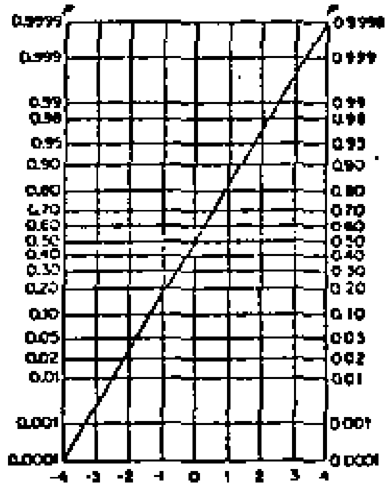


Figure 7-29

Since the area under the normal curve is unity, the partial areas, P_z, represent the percentage cumulative frequencies of a normal curve. For example, if we refer to Figure 7-18, we find that about 2 per cent of the normally distributed z's have values less than -2, about 16 per cent have values less than -1, 50 per cent less than 0, and so on.

We illustrate the use of the paper with the aid of Table 7-2 and Figure 7-30. Table 7-2 contains the familiar data of head lengths. Inasmuch as cumulative frequencies are of prime importance here, we are interested only in boundary values and not mid-values. The last column of values is found from the formula $100 \times \text{cum } f/N$. For example, the fifth number in the last column, 25.3, equals $100 \times 117/462$.

EJERCICIO

CON OBJETO DE VERIFICAR LA CONSISTENCIA INTERNA DE UNA PRUEBA PSICOLOGICA, ESTA SE APLICO DOS VECES A CADA UNA DE DOS MUESTRAS ALEATORIAS. ESTAS MUESTRAS SE EXTRAJERON DE NIÑOS DEL CUARTO GRADO DE DOS ESCUELAS DISTINTAS, "A" y "B". LAS CALIFICACIONES DE LA PRIMERA APLICACION CORRESPONDEN A LA VARIABLE X; LAS DE LA SEGUNDA APLICACION (15 DIAS DESPUES DE LA PRIMERA), CORRESPONDEN A LA VARIABLE Y.

a. CALCULAR EL COEFICIENTE DE CORRELACION DE X y Y PARA CADA ESCUELA, Y PARA LAS DOS ESCUELAS JUNTAS, Y PROBAR LA HIPOTESIS DE QUE $\rho_{XY}>0$ EN CADA CASO.

b. PROBAR LA HIPOTESIS DE QUE $\mu_X = \mu_Y$ PARA AMBAS ESCUELAS JUNTAS, Y PARA CADA ESCUELA POR SEPARADO.

c. PROBAR LA HIPOTESIS DE QUE

1. $\mu_{X_A} = \mu_{X_B}$

2. $\mu_{Y_A} = \mu_{Y_B}$

3. $\sigma^2(X_A) = \sigma^2(X_B)$

4. $\sigma^2(Y_A) = \sigma^2(Y_B)$

FORMULAS

$\bar{X} = \Sigma X_j/n$, $\bar{Y} = \Sigma Y_j/n$, $S^2(X) = \Sigma X^2/n - \bar{X}^2$, $S^2(Y) = \Sigma Y_j^2/n - \bar{Y}^2$,

$S^2(d) = \Sigma d_i^2/n - \bar{d}^2$, $t_d = \dfrac{(\bar{X}-\bar{Y})\sqrt{n-1}}{S_d}$, $t_\rho = r_{XY}\sqrt{\dfrac{n-2}{1-r^2}}$

$$t_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{n_X S^2(X) + n_Y S^2(Y)}{n_X+n_Y-2}\left(\dfrac{1}{n_X}+\dfrac{1}{n_Y}\right)}}, \quad F = \frac{S_M^2}{S_m^2}$$

DONDE S_M^2 y S_m^2 SON ESTIMACIONES INSESGADAS DE LAS VARIANCIAS MAYOR Y MENOR, RESPECTIVAMENTE, DE LAS DOS QUE SE ESTAN COMPARANDO.

RESPUESTAS A LOS INCISOS a y b

ESCUELA A

X	Y	X^2	Y^2	XY	d=x-y	d^2
34	35	1156	1225	1190	-1	1
39	36	1521	1296	1404	3	9
40	40	1600	1600	1600	0	0
35	38	1225	1444	1330	-3	9
30	29	900	841	870	1	1
28	26	784	676	728	2	4
33	34	1089	1156	1122	-1	1
38	40	1444	1600	1520	-2	4
32	39	1024	1521	1248	-7	49
37	35	1369	1225	1295	2	4
26	26	676	676	676	0	0
40	39	1600	1521	1560	1	1
32	30	1024	900	960	2	4
33	34	1089	1156	1122	-1	1
38	33	1444	1089	1254	5	25
34	39	1156	1521	1326	-5	25
35	37	1225	1369	1295	-2	4
584	590	20326	20816	20500	-6	142

$\bar{X} = \frac{584}{17} = 34.352941$; $\bar{X}^2 = 1180.1245$

$\bar{Y} = \frac{590}{17} = 34.705882$; $\bar{Y}^2 = 1204.4982$

$\bar{d} = -6/17 = -0.3529411$; $\bar{d}^2 = 0.1245674$

$S^2(x) = \frac{20326}{17} - 1180.1245 = 15.5225$; $S(x) = 3.9398604$

$S^2(Y) = \frac{20816}{17} - 1204.4982 = 19.9723$; $S(Y) = 4.4690379$

$S_d^2 = \frac{142}{17} - 0.1245674 = 8.2283737$; $S_d = 2.8685141$

$$r_{XY} = \frac{(20500/17)-(34.352941)(34.705882)}{(3.9398604)(4.4690379)} = 0.7742882$$

$H_0: \rho = 0$, $H_1: \rho \neq 0$, $t = t_{0.975,15} = 2.13$

$$t_\rho = 0.774 \sqrt{\frac{17-2}{1-0.744^2}} = 0.774 \times 6.116 = 4.73 > 2.13$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy} = 0$ CON UN NIVEL DE CONFIANZA DEL 95%. $t_c = t_{0.975,16} = 2.12$

$$t_d = \left|\frac{(34.353-34.706)\sqrt{16}}{2.869}\right| = 0.492 < 2.12$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_x = \mu_y$ CON UN 95% DE NIVEL DE CONFIANZA.

ESCUELA B

X	Y	X^2	Y^2	XY	d=X-Y	d^2
39	41	1521	1681	1599	-2	4
27	36	729	1296	972	-9	81
33	31	1089	961	1023	2	4
37	36	1369	1296	1332	1	4
35	36	1225	1296	1260	-1	1
31	33	961	1089	1023	-2	4
33	32	1089	1024	1056	1	1
39	40	1521	1600	1560	-1	1
39	35	1521	1225	1365	4	16
27	29	891	841	783	-2	4
32	36	1024	1296	1152	-4	16
34	35	1156	1225	1190	-1	1
35	34	1225	1156	1190	1	1
36	42	1296	1764	1512	-6	36
34	34	1156	1156	1156	0	0
29	31	841	961	899	-2	4
540	561	18614	19867	19072	-21	175

$\bar{X} = \frac{545}{16} = 33.75$; $\bar{X}^2 = 1139.0625$; $\bar{d} = -1.3125$; $\bar{d}^2 = 1.7226562$

$\bar{Y} = \frac{561}{16} = 35.0625$; $\bar{Y}^2 = 1229.3789$

$S_x^2(x) = \frac{18614}{16} - 1139.0625 = 24,3125$; $S(x) = 4,9307707$

$S^2(Y) = \frac{19867}{16} - 1229.3789 = 12.3086$; $S(y) = 3.5083614$

$S_d^2 = \frac{175}{16} - 1.7226562 = 9.214844$ $S_d = 3.0355961$

$$r_{xy} = \frac{(19072/16)-(33.75)(35.0625)}{(4.9307707)(3.5083614)} = \underline{\underline{0.4994934}}$$

$$t_\rho = 0.499\sqrt{\frac{14}{0.751}} = 2.154 < 2.15$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy} = 0$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_d = \left|\frac{(33.75-35.063)\sqrt{15}}{3.036}\right| = 1.67 < 2.13$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_x = \mu_y$, CON UN 95% DE NIVEL DE CONFIANZA.

AMBAS ESCUELAS JUNTAS

$\Sigma x_i = 1124$, $\Sigma y_i = 1151$, $\Sigma x_i^2 = 38940$, $\Sigma y_i^2 = 39572$, $\Sigma d_i = -27$, $\Sigma d_i^2 = 317$

$\bar{X} = \frac{1124}{33} = 34.060606$; $\bar{X}^2 = 1160.1248$; $\bar{d} = \frac{-27}{33} = -0.8181818$; $\bar{d}^2 = 0.6694214$

$\bar{Y} = \frac{1151}{33} = 34.878787$; $\bar{Y}^2 = 1216.5297$

$S^2(X) = \frac{38940}{33} - 1160.1248 = 19.8752$; $S(X) = 4.458161$

$S^2(Y) = \frac{40683}{33} - 1216.5297 = 16.2884$; $S(Y) = 4.0358889$

$S_d^2 = \frac{317}{33} - 0.6694214 = 8.9366392$; $S_d = 2.9894212$

$$r_{xy} = \frac{(39572/33)-(34.060606)(34.878787)}{(4.458161)(4.0358889)} = \underline{\underline{0.6201924}}$$

$$t_d = \left| \frac{(33.061 - 34.879)\sqrt{32}}{2.989} \right| = 1.548 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_X = \mu_Y$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_0 = 0.620\sqrt{\frac{31}{0.616}} = 4.398 > 2.04$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy} = 0$, CON UN NIVEL DE CONFIANZA DEL 95%.

RESPUESTAS AL INCISO c

$$t_{\bar{X}_A - \bar{X}_B} = \frac{34.35 - 33.75}{\sqrt{\frac{17 \times 15.52 + 16 \times 24.31}{31}\left[\frac{1}{17} + \frac{1}{16}\right]}} = \frac{0.6}{\sqrt{\frac{263.84 + 388.96}{31}(0.121)}} =$$

$$= 0.368 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_{X_A} = \mu_{X_B}$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_{\bar{Y}_A - \bar{Y}_B} = \frac{|34.71 - 35.06|}{\sqrt{\frac{17 \times 19.97 + 16 \times 12.31}{31}(0.121)}} = \frac{|-0.35|}{\sqrt{\frac{339.49 + 196.96}{31}(0.121)}}$$

$$= -0.34 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_{Y_A} = \mu_{Y_B}$, CON UN 95% DE NIVEL DE CONFIANZA.

PARA LA PRUEBA DE IGUALDAD DE VARIANCIAS USAREMOS

$$F = \frac{24.31\sqrt{\frac{16}{15}}}{15.52\sqrt{\frac{17}{16}}} = 1.57 < 3.41 = F_{0.01}(15,16)$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\sigma^2_{X_A} = \sigma^2_{X_B}$ CON UN 98% DE NIVEL DE CONFIANZA.

$$F = \frac{19.97\sqrt{\frac{17}{16}}}{12.31\sqrt{\frac{16}{15}}} = \frac{20.58}{12.71} = 1.62 < 3.41$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\sigma^2_{Y_A} = \sigma^2_{Y_B}$ CON UN 98% DE NIVEL DE CONFIANZA.

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

PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES

INFERENCIA ESTADISTICA

M. en I. Augusto Villarreal Aranda

FEBRERO, 1982

lacio de Minería Calle de Tacuba 5 primer piso México 1, D. F. Tel: 521-40-20 Apdo. Postal M-2285

INFERENCIA ESTADISTICA

Por: M en I Augusto Villarreal Aranda

1. Introducción

La parte de la estadística que proporciona las reglas para inferir ciertas características de una población a partir de muestras extraídas de ella, junto con indicaciones probabilísticas de la veracidad de tales inferencias, se llama *inferencia estadística*.

En la inferencia estadística se estudian las relaciones existentes entre una población, las muestras obtenidas de ella, y las técnicas para estimar parámetros, tales como la media y la varianza, o bien para determinar si las diferencias entre dos muestras son debidas al azar, etc.

2. Distribuciones muestrales

Si se consideran todas las muestras posibles de tamaño n que pueden extraerse de una población, y para cada una se calcula el valor del promedio aritmético, este seguramente variará de una muestra a otra, ya que depende de los valores de los datos que se hayan obtenido en cada muestra. Por lo tanto, el promedio aritmético es en sí una variable aleatoria, como también lo son, por la misma razón, el rango y la variancia de la muestra.

A todo elemento que es función de los valores de los datos que se tienen en una muestra se le denomina *estadística*; toda estadística es, entonces, una variable aleatoria cuya distribución de probabilidades se conoce como *distribución muestral*. Si, por ejemplo, la estadística considerada es la varíancia de la muestra, su densidad de probabilidades se llama *distribución muestral de la variancia*.

En forma similar se pueden obtener las distribuciones muestrales de la desviación estándar, del rango, etc., cada una de las cuales tendrá sus propios parámetros, lo que permite hablar de la media y la desviación estándar de la variancia, etc.

3. Muestreo con y sin remplazo

Cuando se efectúa un muestreo en una población de tal manera que cada elemento de la misma se pueda escoger más de una vez, se dice que el muestreo es *con remplazo*; en caso contrario, el muestreo es *sin remplazo*. Si de una urna se quiere extraer una muestra de bolas de colores, se puede proceder de dos maneras: se saca al azar una bola, se anota su color y se regresa a la urna antes de obtener otra, y así sucesivamente; en este caso el muestreo es *con remplazo*. La segunda forma consiste en extraer al azar todas las bolas que constituyen la muestra sin regresarlas a la urna, siendo entonces un muestreo *sin remplazo*.

4. Distribución muestral del promedio aritmético

Supóngase que se extraen sin remplazo todas las muestras posibles de tamaño n de una población finita de tamaño $N_p > n$. Si la media y la desviación estándar de la distribución muestral del promedio aritmético se denotan con $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$, y la media y la desviación estándar de la población con μ y σ, respectivamente, entonces es posible demostrar que se cumplen las siguientes ecuaciones

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Además, si la población es infinita (o el muestreo es con remplazo), los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

puesto que

$$\lim_{N_p \to \infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{\sigma}{\sqrt{n}}$$

Para valores grandes de n ($n \geq 30$) se demuestra, empleando el teorema del límite central, que la distribución muestral del promedio aritmético es aproximadamente una distribución normal con media $\mu_{\bar{X}}$ y desviación estándar $\sigma_{\bar{X}}$, independientemente de cuál sea la densidad de probabilidades de X, la variable aleatoria asociada a la población. Si esta variable tiene distribución normal, la distribución muestral del promedio aritmético también es normal, aun para valores pequeños de n ($n < 30$).

Ejemplo 4.1

Supóngase que se tiene una población finita formada por los datos 1,2,3,4,5. Se desea conocer la media y la desviación estándar de la distribución muestral del promedio aritmético, considerando las muestras de tamaño 3 obtenidas sin remplazo.

Primer procedimiento.

Siendo la población finita y el muestreo sin remplazo, es posible obtener la distribución muestral correspondiente para calcular después sus parámetros, considerando que el número total de muestras distintas de tamaño 3 que pueden obtenerse a partir de una población de 5 elementos es

$$\frac{5!}{3!(5-3)!} = 10$$

Dichas muestras son las siguientes, junto con sus promedios aritméticos correspondientes:

	$\bar{x}_i$		$\bar{x}_i$
1, 2, 3	6/3	3, 4, 5	12/3
1, 2, 4	7/3	3, 4, 1	8/3
1, 2, 5	8/3	4, 5, 1	10/3
2, 3, 4	9/3	4, 5, 2	11/3
2, 3, 5	10/3	5, 1, 3	9/3

Para calcular la media y la desviación estándar, se emplea la siguiente tabla

$\bar{x}_i$	6/3	7/3	8/3	8/3	9/3	9/3	10/3	10/3	11/3	12/3
$\bar{x}_i^2$	36/9	49/9	64/9	64/9	81/9	81/9	100/9	100/9	121/9	144/9

$$\sum_{i=1}^{10} \bar{x}_i = 90/3 \qquad \sum_{i=1}^{10} \bar{x}_i^2 = 840/9$$

$$\mu_{\bar{x}} = \bar{\bar{x}} = \frac{1}{10} \sum_{i=1}^{10} \bar{x}_i = \frac{1}{10} \cdot \frac{90}{3} = 3$$

$$\sigma_{\bar{x}}^2 = \frac{1}{10} \sum_{i=1}^{10} \bar{x}_i^2 - \bar{\bar{x}}^2 = \frac{1}{10} \cdot \frac{840}{9} - (3)^2 =$$

$$= 9.333 - 9.000 = 0.333 \Rightarrow \sigma_{\bar{x}} = \sqrt{0.333} = 0.577$$

Es decir, $\mu_{\bar{x}} = 3$ y $\sigma_{\bar{x}} = 0.577$

Segundo procedimiento.

Por tratarse de una población finita, se verifica que

$$\mu_{\bar{x}} = \mu \qquad \text{y} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en donde $N_p = 5$, $n = 3$ y $\mu = 3$.

El valor de σ^2 de la población es

$$\sigma^2 = \frac{1+4+9+16+25}{5} - (3)^2 = \frac{55}{5} - 9 = 11-9 = 2$$

Por lo tanto, $\sigma = \sqrt{2} = 1.4145$ y

$$\sigma_{\bar{x}} = \frac{1.4145}{\sqrt{3}} \sqrt{\frac{5-3}{5-1}} = (0.8164)(0.7071) = 0.577$$

Es decir, $\mu_{\bar{x}} = 3$ y $\sigma_{\bar{x}} = 0.577$

Comparando los resultados, se puede observar que ambos procedimientos conducen a la obtención de los mismos valores de $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$ para la distribución muestral del promedio aritmético.

Ejemplo 4.2

En una bodega se tienen cinco mil varillas de acero; el valor medio del peso, X, de cada varilla es de 5.02 kg, y la desviación estándar 0.3 kg. Hallar la probabilidad de que una muestra de cien varillas, escogida al azar, tenga un peso total

a. entre 496 y 500 kg

b. de más de 510 kg.

Para la distribución muestral del promedio, se tiene que $\mu_{\bar{X}} = \mu = 5.02$ kg y, por tratarse de una población finita,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{0.30}{\sqrt{100}}\sqrt{\frac{5000 - 100}{5000 - 1}} = 0.027$$

a. El peso total de la muestra estará entre 496 y 500 kg si el peso promedio de las cien varillas se encuentra entre 4.96 y 5.00 kg. Puesto que la muestra es mayor de 30 elementos se puede considerar como aproximadamente normal a la distribución muestral, y los valores estándar correspondientes a $\bar{X} = 4.96$ y a $\bar{X} = 5.00$ se obtienen mediante la transformación

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

es decir,

$$Z_1 = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$Z_2 = \frac{5.00 - 5.02}{0.027} = -0.74$$

En la fig 4.1 se puede apreciar que

$$P[496 \le X \le 500] = P[-2.22 \le Z \le -0.74] =$$

$$= P[-2.22 \le Z \le 0] - P[-0.74 \le Z \le 0]$$

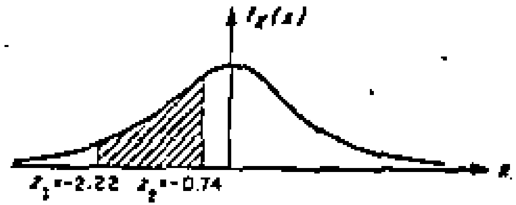


Fig 4.1 Distribución normal correspondiente al ejemplo

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y Z queda finalmente

$$P[496 \le X \le 500] = 0.4868 - 0.2704 = 0.2164$$

b. El peso total de la muestra excederá de 510 kg si el peso promedio de las cien varillas pasa de 5.10 kg.

Estandarizando dicho valor, queda

$$Z_3 = \frac{5.10 - 5.02}{0.027} = 2.96$$

Calculando el área bajo la curva normal a la derecha de este valor (fig 4.2), se tiene que

$$P[X \ge 510] = P[Z \ge 2.96] = P[Z \ge 0] - P[0 \le Z \le 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

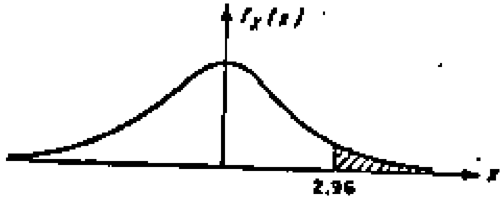


Fig 4.2 Distribución normal correspondiente al ejemplo

5. Distribución muestral de diferencias de promedios aritméticos

Con frecuencia se presenta el caso en el que se tienen datos de dos poblaciones con variables aleatorias asociadas X y Y, respectivamente, surgiendo la duda de si estas se pueden considerar como una sola, es decir, si X = Y. Para probar estadísticamente esta hipótesis (como se verá más adelante), es necesario obtener las distribuciones muestrales de la diferencia de los promedios y de las varianzas de las muestras de ambas variables.

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de muestras aleatorias de tamaño n_X y n_Y de dos poblaciones con características X y Y, respectivamente. Se puede demostrar que la distribución muestral de la diferencia de los promedios correspondientes a poblaciones infinitas con medias μ_X y μ_Y y desviaciones estándar σ_X y σ_Y, tiene los siguientes parámetros:

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X} + \frac{\sigma^2_Y}{n_Y}}$$

si las muestras son independientes.

Esta distribución también es aplicable a poblaciones finitas si el muestreo es con remplazo. Para el caso de poblaciones finitas en las cuales el muestreo se hace sin remplazo, los parámetros de la distribución muestral de la diferencia de los promedios aritméticos son

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X}\,\frac{N_X - n_X}{N_X - 1} + \frac{\sigma^2_Y}{n_Y}\,\frac{N_Y - n_Y}{N_Y - 1}}$$

suponiendo que las muestras sean independientes.

Ejemplo 5.1

Considérese que de una población X se obtienen tres muestras posibles, cuyos correspondientes promedios aritméticos son 3, 7 y 8. De otra población Y se extraen dos muestras posibles, con promedios 2 y 4, respectivamente. Se deben obtener los parámetros de la distribución muestral de las diferencias de los promedios aritméticos.

Primer procedimiento

Todas las posibles diferencias de promedios aritméticos de X con los de Y serían

$$\begin{array}{ccccccc} 3-2 & 7-2 & 8-2 & \longrightarrow & 1 & 5 & 6 \\ 3-4 & 7-4 & 8-4 & & -1 & 3 & 4 \end{array}$$

Es decir,

$$\mu_{\bar{X}-\bar{Y}} = \frac{-1+1+3+4+5+6}{6} = \frac{18}{6} = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{(-1-3)^2 + (1-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (6-3)^2}{6} =$$

$$= \frac{34}{6} = \frac{17}{3}$$

gundo procedimiento

Se sabe que

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} \;;\; \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}$$

Por ello,

$$\mu_{\bar{X}} = \frac{3+7+8}{3} = \frac{18}{3} = 6$$

$$\mu_{\bar{Y}} = \frac{2+4}{2} = \frac{6}{2} = 3$$

$$\sigma^2_{\bar{X}} = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

$$\sigma^2_{\bar{Y}} = \frac{(2-3)^2 + (4-3)^2}{2} = \frac{2}{2} = 1$$

$$\mu_{\bar{X}-\bar{Y}} = 6 - 3 = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{14}{3} + 1 = \frac{17}{3}$$

Se observa que ambos procedimientos conducen a los mismos resultados.

Ejemplo 5.2

Las varillas de acero que fabrica una compañía A tienen un peso medio de 6.5 kg y una desviación estándar de 0.4, en tanto que las producidas por una empresa B tienen un peso medio de 6.3 kg y una desviación estándar de 0.3 kg. Si se toman muestras aleatorias de 100 varillas de cada fábrica, ¿cuál es la probabilidad de que las de la compañía A tengan un peso promedio de por lo menos

a. 0.35 kg

b. 0.10 kg

mayor que el de la compañía B?

Se puede suponer en este caso que las distribuciones muestrales involucradas son normales, en virtud de que el tamaño de ambas muestras es mayor de 30 elementos. También se puede suponer que ambas poblaciones son infinitas, y siendo $\bar{X}_A$ y $\bar{X}_B$ los pesos promedios de las muestras de las fábricas A y B, respectivamente, entonces

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 6.5 - 6.3 = 0.20 \text{ kg}$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} = 0.05 \text{ kg}$$

La variable estandarizada de la diferencia de los promedios es

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_{\bar{X}_A - \bar{X}_B}}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0.20}{0.05}$$

a. Estandarizando la diferencia de 0.35 kg se llega a

$$Z_1 = \frac{0.35 - 0.20}{0.05} = \frac{0.15}{0.05} = 3$$

La probabilidad deseada es el área bajo la curva normal a la derecha de $Z = 3$, es decir,

$$P\left[\bar{X}_A > \bar{X}_B + 0.35\right] = P\left[Z > 3\right] = 0.500 - 0.4987 = 0.0013$$

b. Al estandarizar la diferencia de 0.10 kg, la variable Z resulta

$$Z_2 = \frac{0.10 - 0.20}{0.05} = \frac{-0.1}{0.05} = -2$$

La probabilidad requerida es el área bajo la curva normal a la derecha de $Z = -2$, es decir

$$P\left[\bar{X}_A > \bar{X}_B + 0.10\right] = P\left[Z > -2\right] = 0.5 + 0.4772 = 0.9772$$

6. Teoría estadística de la estimación

En la práctica profesional a menudo resulta necesario inferir información acerca de una población mediante el uso de muestras extraídas de ella; una parte básica de dicha inferencia consiste en *estimar* los valores de los parámetros de la población (media, variancia, etc.) a partir de las estadísticas correspondientes de la muestra, como se explica a continuación.

7. Estimadores puntuales. Clasificación

Si un estimador de un parámetro de la población consiste en un solo valor de una estadística, se le conoce como *estimador puntual* del parámetro.

Cuando la media de la distribución muestral de una estadística es igual al parámetro que se está estimando de la población, entonces la estadística se conoce como *estimador insesgado* del parámetro; si no sucede así, entonces se denomina *estimador sesgado*. Ambos estimadores son puntuales, y sus valores correspondientes se llaman estimaciones insesgadas o sesgadas, respectivamente. Dicho de otra manera, si S es una estadística cuya distribución muestral tiene media μ_S, y el parámetro correspondiente de la población es θ, se dice que S es un estimador insesgado de θ si

$$\mu_S = \theta$$

Por otra parte, si la estadística S_m de la muestra tiende a ser igual al parámetro θ de la población a medida que se hace más grande el tamaño de la muestra, entonces la estadística recibe el nombre de *estimador consistente* del parámetro.

Empleando símbolos, si

$$\lim_{n\to\infty} S_n = \theta$$

resulta que la estadística S_n es un estimador consistente. Por ejemplo, el promedio aritmético es un estimador insesgado y consistente de la media, y la variancia de la muestra es un estimador sesgado y consistente de la variancia de la población.

Si las distribuciones muestrales de varias estadísticas tienen el mismo valor de la media, se dice que la estadística que cuenta con la menor variancia es un *estimador eficiente* de dicha media, en tanto que las estadísticas restantes se conocen como *estimadores ineficientes* del parámetro.

Por ejemplo, las distribuciones muestrales del promedio aritmético y de la mediana cuentan con medias que son, en ambos casos, iguales a la media de la población. Sin embargo, la variancia de la distribución muestral del promedio aritmético es menor que la de la distribución de la mediana, por lo que el promedio aritmético obtenido de una muestra aleatoria proporciona un estimador eficiente de la media de la población, en tanto que la mediana obtenida de la muestra proporciona un estimador ineficiente de dicho parámetro.

8. Estimación de intervalos de confianza para los parámetros de una población

La estimación de un parámetro de una población mediante un par de números entre los cuales se encuentra, con cierta probabilidad, el valor de dicho parámetro, se llama estimación del intervalo del mismo.

Sea S una estadística obtenida de una muestra de tamaño n para estimar el valor del parámetro θ, y sea σ_S la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1-\alpha$, de que el valor de θ se localice en el intervalo de $S - z_c\,\sigma_S$ a $S + z_c\,\sigma_S$, donde z_c es una constante, se escribe en la forma

$$P[S - z_c\,\sigma_S \le \theta \le S + z_c\,\sigma_S] = 1-\alpha$$

Si se fija el valor de $1-\alpha$, se puede obtener el valor de z_c necesario para que se satisfaga la ecuación anterior, con lo cual queda definido el *intervalo de confianza* del parámetro θ, $(S - z_c\,\sigma_S,\ S + z_c\,\sigma_S)$, correspondiente al nivel de *confianza* $1-\alpha$.

La constante z_c que fija el intervalo de confianza se conoce como *valor crítico*. Si la distribución de S es normal, el valor de z_c correspondiente a uno de α se obtiene de la tabla de áreas bajo la curva normal o de la tabla 8.1 siguiente.

/...

TABLA 8.1 VALORES DE z_c PARA DISTINTOS NIVELES DE CONFIANZA

Nivel de confianza, en porcentaje	z_c
99.73	3.00
99.00	2.58
98.00	2.33
96.00	2.05
95.45	2.00
95.00	1.96
90.00	1.64
80.00	1.28
68.27	1.00
50.00	0.674

Ejemplo 8.1

Sea el promedio aritmético $\bar{X}$ una estadística con distribución normal. Las probabilidades o niveles de confianza de que $\mu_{\bar{X}}$ (o μ de la población) se encuentre localizada entre los límites $\bar{X} \pm \sigma_{\bar{X}}$, $\bar{X} \pm 2\,\sigma_{\bar{X}}$ y $\bar{X} \pm 3\,\sigma_{\bar{X}}$ son 68.26, 95.44 y 99.73%, respectivamente, obteniéndose dichos valores de la tabla de área bajo la curva normal. Lo anterior significa que el intervalo $\bar{X} \pm 3\,\sigma_{\bar{X}}$ contendrá a $\mu_{\bar{X}}$ en el 99.73 por ciento de las muestras de tamaño n, por lo que los intervalos de confianza de 68.26, 95.44 y 99.73 por ciento para estimar a μ son $(\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}})$ $(\bar{X} - 2\,\sigma_{\bar{X}}, \bar{X} + 2\,\sigma_{\bar{X}})$ y $(\bar{X} - 3\,\sigma_{\bar{X}}, \bar{X} + 3\,\sigma_{\bar{X}})$, lo cual se aprecia en la fig 8.1 siguiente.

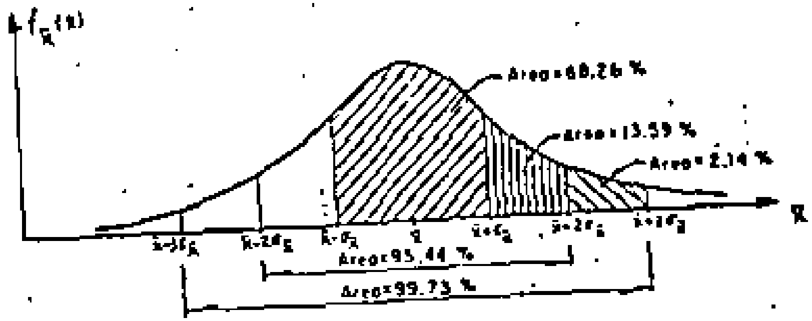


Fig 8.1

3. Estimación de intervalos de confianza para la media

Los límites de confianza para la media de una población con variable aleatoria X asociada están dados por

$$\bar{X} \pm z_c\,\sigma_{\bar{X}}$$

en donde z_c depende del nivel de confianza deseado. Si $\bar{X}$ tiene distribución normal, z_c puede obtenerse en forma directa de la tabla 8.1. Por ejemplo, los límites de confianza de 95 y 99 por ciento para estimar la media, μ, de la población son $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\,\sigma_{\bar{X}}$, respectivamente. Al obtener estos límites hay que usar el valor calculado de $\bar{X}$ para la muestra correspondiente.

Entonces, los límites de confianza para la media de la población quedan dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}}$$

en caso de que el muestreo se haga a partir de una población infinita o de que se efectúe con remplazo a partir de una población finita, o por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

si el muestreo es sin remplazo a partir de una población finita de tamaño N_p.

Ejemplo 9.1

Las mediciones de los diámetros de una muestra aleatoria de 100 tubos de albañal mostraron una media de 32 cm y una desviación estándar de 2 cm. Obténganse los límites de confianza de

a. 95 por ciento
b. 97 por ciento

para el diámetro medio de todos los tubos.

a. De la tabla 8.1, los límites de confianza del 95 por ciento son

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 32 \pm 1.96(2/\sqrt{100}) = 32 \pm 0.392 \text{ cm}$$

o sea 31.608 y 32.392, en donde se ha empleado el valor de S_X para estimar el de σ de la población, puesto que la muestra es suficientemente grande (mayor de 30 elementos). Esto significa que con una probabilidad de 95 por ciento, el valor de μ_X se encuentra entre 31.608 y 32.392 cm.

b. Si $Z = z_c$ es tal que el área bajo la curva normal a la derecha de z_c es el 1.5 por ciento del área total, entonces el área entre 0 y z_c es 0.5 - 0.015 = 0.485, por lo que de la tabla de áreas bajo la curva normal se obtiene $z_c = 2.17$. Por lo tanto, los límites de confianza del 97 por ciento son:

$$\bar{X} \pm 2.17\sigma/\sqrt{n} = 32 \pm 2.17(2/\sqrt{100}) = 32 \pm 0.434 \text{ cm}$$

y el intervalo de confianza respectivo es (31.566 cm, 32.434 cm).

Ejemplo 9.2

Una muestra aleatoria de 50 calificaciones de cierto examen de admisión tiene un promedio aritmético de 72 puntos, con desviación estándar igual a 10. Si el examen se aplicó a 1018 personas, obtener

a. El intervalo de confianza del 95% para la media del total de calificaciones.
b. El tamaño de muestra necesario para que el error en la estimación de la media no exceda de 2 puntos, considerando el mismo nivel de confianza.
c. El nivel de confianza para el cual la media de la población sea 72 ± 1 puntos.

a. Si se estima a σ de la población con S_X de la muestra y se considera que la población es finita, los límites de confianza son, puesto que $\bar{X} = 72$, $Z_c = 1.96$, $S_X = 10$, $N_p = 1018$ y $n = 50$,

$$72 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm 1.96\,(1.4142).(0.9755)$$

$$72 \pm 2.704$$

y el intervalo de confianza respectivo es

$$(69.296,\ \ 74.704)$$

b. Puesto que el error en la estimación de la media es, para población finita,

$$\text{Error en la estimación} = Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en este caso se tendría

$$Z_c \ \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} < 2$$

o sea, para un nivel de confianza de 95%,

$$1.96 \ \frac{10}{\sqrt{n}} . \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

$$\frac{19.6}{\sqrt{n}} \ \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

Elevando al cuadrado la desigualdad, queda

$$\frac{384.16}{n} \quad \frac{1018 - n}{1017} < 4$$

o sea

$$87.85 < n$$

Por lo cual, se requieren al menos 88 elementos en la muestra para que el error en la estimación no exceda de 2 puntos, para $1 - \alpha = 0.95$.

c. Los límites de confianza son, en este caso

$$72 \pm Z_c \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm Z_c \ (1.4142)(0.9755)$$

o sea

$$72 \pm 1.3795 \ Z_c$$

Puesto que se desea que el valor de la media sea 72 ± 1 puntos, se verifica que

$$1 = 1.3795 \ Z_c$$

Es decir

$$Z_c = \frac{1}{1.3795} = 0.725$$

El área bajo la curva normal estándar entre 0 y Z_c = 0.725 es, por interpolación lineal, igual a 0.2657. Por lo tanto, el nivel de confianza es igual al doble del área anterior, es decir, 2(0.2657) = 0.5314 (o 53.14%), tal como se muestra en la fig 9.1.

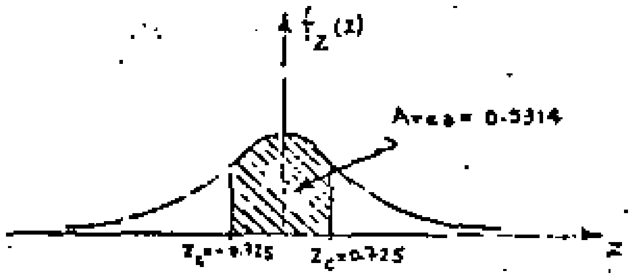


Fig 9.1

10. Intervalos de confianza para diferencias de medias

Los límites de confianza para la diferencia de las medias cuando las poblaciones X y Y son infinitas, o cuando el muestreo se realiza con remplazo de poblaciones finitas, se encuentran dados por

$$\bar{X} - \bar{Y} \pm Z_c \, \sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm Z_c \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

donde $\bar{X}$, n_X y $\bar{Y}$, n_Y son los respectivos promedios aritméticos y tamaños de las dos muestras extraídas de las poblaciones, y σ_X y σ_Y las desviaciones estándar de estas últimas.

En el caso de que las poblaciones X y Y sean finitas y el muestreo sin remplazo, los límites de confianza son

$$\bar{X} - \bar{Y} \pm Z_c \sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm Z_c \sqrt{\frac{\sigma_X^2}{n_X} \, \frac{N_X - n_X}{N_X - 1} + \frac{\sigma_Y^2}{n_Y} \, \frac{N_Y - n_Y}{N_Y - 1}}$$

en donde N_X y N_Y son los tamaños de las poblaciones X y Y, respectivamente.

Las dos ecuaciones anteriores son válidas únicamente si las muestras aleatorias seleccionadas son independientes.

Ejemplo 10.1

Para el ejemplo de las varillas tratado anteriormente (5.2), encontrar el intervalo de confianza del 95.45% para las diferencias de las medias de las poblaciones.

Siendo $\bar{X}_A = \mu_A = 6.5$ kg, $\sigma_A = 0.4$ kg, $\bar{X}_B = \mu_B = 6.3$ kg, $\sigma_B = 0.3$ kg y $n_A = n_B = 100$, los límites de confianza para la diferencia de las medias son, empleando la tabla 8.1

$$\bar{X}_A - \bar{X}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 6.5 - 6.3 \pm 2 \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} =$$

$$= 0.2 \pm 0.1$$

Por lo tanto, el intervalo de confianza respectivo es (0.1, 0.3).

Ejemplo 10.2

Se tienen en una bodega 3000 focos de marca X, y 5000 de marca Y. Se extrae una muestra aleatoria de 150 focos de la marca X, y se obtiene una duración promedio de 1400 horas, con desviación estándar igual a 120 horas. Otra muestra aleatoria de 200 focos de la marca Y tuvo una duración promedio de 1200 horas, con desviación estándar igual a 80 horas. Obtener intervalos de confianza de

a. 95%

b. 99%

para la diferencia de los tiempos medios de duración de los focos de ambas marcas.

a. Puesto que se trata de poblaciones finitas y $\bar{X} = 1400$ h, $S_X = 120$ h, $N_X = 3000$, $n_X = 150$, $\bar{Y} = 1200$ h, $S_Y = 80$ h $N_Y = 5000$ y $n_Y = 200$, se obtiene, estimando a σ_X y σ_Y con S_X y S_Y, respectivamente

$$1400 - 1200 \pm 1.96 \sqrt{\frac{(120)^2}{150} \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \frac{5000 - 200}{5000 - 1}}$$

$$200 \pm 1.96\ (11.04)$$

$$200 \pm 21.638$$

o sea, (178.362, 221.638), puesto que de la tabla 8.1, para un nivel de confianza de 95%, $Z_c = 1.96$.

b. En este caso, al emplear la tabla 8.1 se obtiene $Z_c = 2.58$ para un nivel de confianza de 99%, por lo cual los límites son

$$1400 - 1200 \pm 2.58 \sqrt{\frac{(120)^2}{150} \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \frac{5000 - 2000}{5000 - 1}}$$

$$200 \pm 2.58\ (11.04)$$

$$200 \pm 28.483$$

y el intervalo de confianza es

$$(171.517,\ 228.483)$$

11. Pruebas de hipótesis

Supóngase que una empresa armadora de automóviles está en la disyuntiva de emplear una nueva marca de bujías en sus unidades o la que regularmente utiliza, y que su departamento de control de calidad debe decidir, con base en la información de las muestras de las dos marcas distintas. Las decisiones de este tipo, es decir, que se basan en estudios estadísticos, reciben el nombre de *decisiones estadísticas*, y a los procedimientos que permiten decidir si se acepta o rechaza una hipótesis se les llama *pruebas de hipótesis*, *pruebas de significancia* o *reglas de decisión*.

Al tomar decisiones estadísticas, es necesario postular las diversas alternativas o cursos de acción que pueden adoptarse.

En el caso particular de una *prueba de hipótesis* solamente se tienen dos cursos de acción posibles, los que se denotarán como H_0 y H_1. A la acción H_0 se le llama *hipótesis nula*, y a la H_1, *hipótesis alternativa*. Por ejemplo, si la hipótesis nula establece que $\mu_1 = \mu_2$, la hipótesis alternativa puede ser una de las siguientes:

$$\mu_1 > \mu_2, \mu_1 < \mu_2 \text{ o } \mu_1 \neq \mu_2$$

Al realizar una prueba de hipótesis, se prueba siempre la verdad de la hipótesis nula H_0, aun cuando de antemano se desea rechazarla.

12. Errores de los tipos I y II. Nivel de significancia

En muchas ocasiones se presenta el caso de que se rechaza una hipótesis nula cuando en realidad debería ser aceptada; cuando esto sucede se dice que se ha cometido un *error de tipo I*. En otras ocasiones se acepta una hipótesis nula siendo en realidad falsa; en este caso se dice que se ha cometido un *error de tipo II*.

Al probar una hipótesis nula, a la máxima probabilidad con la que se está dispuesto a cometer un error del tipo I se le llama *nivel de significancia*, α, de la prueba, el cual dentro de la práctica se acostumbra establecer de 5 por ciento (0.05) o 10 por ciento (0.1). El complemento del nivel de significancia, $1 - \alpha$, se conoce como *nivel de confianza*.

Si, por ejemplo, al realizar una prueba de hipótesis se escoge un nivel de significancia de 10 por ciento, significa que existen 10 posibilidades en 100 de que se rechace ésta cuando debería ser aceptada; es decir, que se rechaza a un nivel de significancia del 10 por ciento, y que la probabilidad de que la decisión haya sido errónea es de 0.1.

13. Comportamiento de los errores tipos I y II

Supóngase que se trata de probar la hipótesis nula de que la media, μ_S, de la distribución muestral de la estadística S es μ_1, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

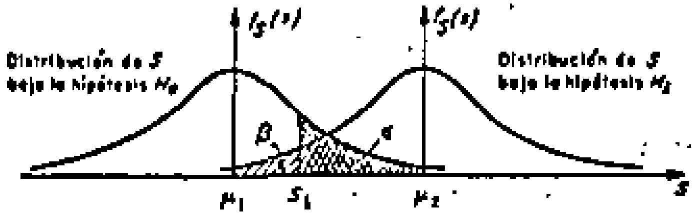
$$H_1 : \mu_S = \mu_2$$

En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar H_0 es la siguiente:

> *Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico S_1, rechácese H_0; en caso contrario, acéptese.*

Es evidente que si H_0 es verdadera, entonces α (área con rayado doble) es la probabilidad de que $S > S_1$, o sea la de rechazar a H_0 siendo verdadera (error tipo I). Por otro lado, si H_1 es verdadera, entonces β (área con rayado sencillo) es la probabilidad de que $S < S_1$, o sea la de aceptar H_0 siendo falsa (error tipo II).

Obsérvese que si se aumenta el valor de S_1 se reduce la probabilidad α, pero se incrementa la β; lo contrario sucede si se disminuye el valor de S_1.



$P[S>S_1] = \alpha$ (error tipo I)

$P[S<S_1] = \beta$ (error tipo II)

Fig 13.1 Probabilidades de los errores tipos I y II en pruebas de hipótesis.

En realidad, la única forma posible en la cual se pueden minimizar simultáneamente los errores de tipos I y II es aumentando el tamaño de la muestra, para hacer más "picudas" las distribuciones muestrales de la estadística bajo las hipótesis H_0 y H_1.

Al observar la fig 13.2 siguiente, es posible concluir que el tamaño de los errores I y II es menor para un tamaño de muestra igual a 100 que para un tamaño igual a 50, considerando la misma regla de decisión anterior.

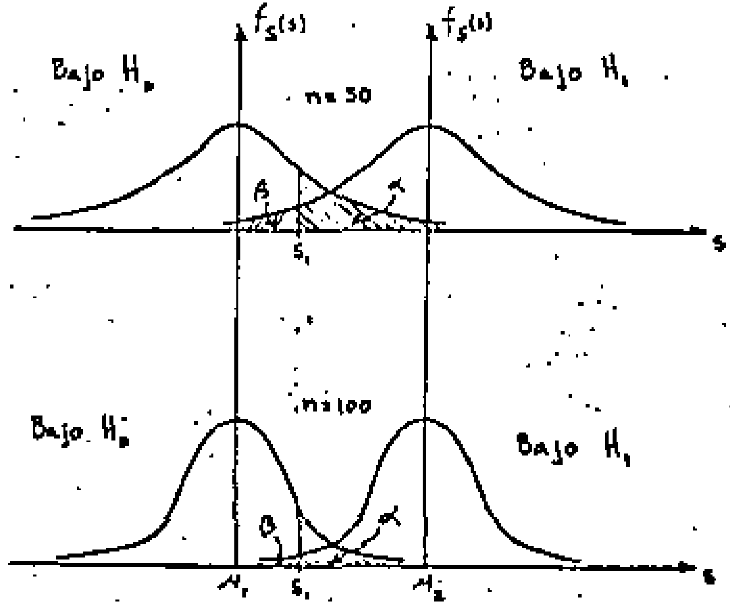


Fig 13.2

Sin embargo, esta técnica de reducción simultánea de los tipos de errores no siempre puede ponerse en práctica, debido razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo ó de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del α por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia α*.

Al conjunto de los valores de la estadística en el que se rechaza la hipótesis nula se le denomina *región crítica, de rechazo, o de significancia*. Por el contrario, al conjunto de los valores de la estadística en que se acepta la hipótesis, se le llama *región de aceptación*.

Considérese que la distribución muestral de la estadística S es normal con desviación estándar σ_S, que la variable Z resulta de estandarizar a S, que la hipótesis nula, H_0, es que la media de S vale μ_S, y que la hipótesis alternativa H_1 es que dicha media es diferente de μ_S, es decir, que

$$Z = \frac{S - \mu_S}{\sigma_S}$$

H_0: media de la distribución muestral de $S = \mu_S$

H_1: media de la distribución muestral de $S \neq \mu_S$

Si se adopta la regla de decisión de aceptar la hipótesis H_0, si el valor de S cae dentro del intervalo central que encierra al 99 por ciento del área de la distribución de probabilidades, entonces H_0 se aceptará en el caso en que

$$-2.58 \leq Z \leq 2.58$$

empleando la tabla de áreas bajo la curva normal estándar. Pero si el valor estandarizado de la estadística se encuentra fuera de dicho intervalo, se concluye que el evento puede ocurrir con probabilidad de 0.01 si la hipótesis H_0 es verdadera (área rayada total de la *fig 14.1*). En tal caso, el valor Z de la variable estándar difiere *significativamente* del que se podría esperar de acuerdo con la hipótesis nula, lo cual inclina a rechazarla a un nivel de confianza del 99 por ciento.

De lo anterior se deduce que el área total rayada de la *fig 14.1* es el nivel de significancia α de la prueba, y representa la probabilidad de cometer un error del tipo I. Por ello, la región de aceptación de H_0 es $-2.58 \leq Z \leq 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

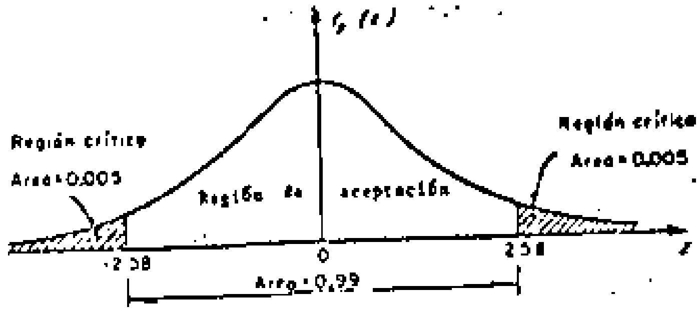


Fig 14.1 Región de significancia

En la tabla 14.1 se presentan los valores de la variable estandarizada, Z, que limitan las regiones de aceptación y de rechazo para el caso en el que la estadística involucrada en la prueba tenga distribución muestral normal. Cuando en alguna prueba de hipótesis se consideren niveles de significancia diferentes a los que aparecen en la tabla mencionada, resulta necesario emplear la de áreas bajo la curva normal estándar.

TABLA 14.1 VALORES CRITICOS DE z

Nivel de significancia, α	Valores de z para pruebas de una cola	Valores de z para pruebas de dos colas
0.1	−1.281 o 1.281	−1.645 y 1.645
0.05	−1.645 o 1.645	−1.960 y 1.960
0.01	−2.326 o 2.326	−2.575 y 2.575
0.005	−2.575 o 2.575	−2.810 y 2.810

15. Pruebas de una y de dos colas

En la prueba de hipótesis del ejemplo anterior, la región de rechazo de la hipótesis nula quedó en ambos extremos (colas) de la distribución muestral de la estadística involucrada en la prueba; a las pruebas de este tipo se les denomina *pruebas de dos colas*. Cuando la región de rechazo se encuentra solamente en un extremo de la distribución muestral en cuestión, se les llama *pruebas de una cola*.

Las pruebas de dos colas se presentan cuando en la hipótesis alternativa aparece el signo ≠ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1 ,$$

$$H_1 : \mu_S \neq \mu_1$$

en donde μ_S es la media de la estadística S, y μ_1 es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

y

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

las pruebas resultan de una cola.

16. Pruebas de hipótesis para la media

Para el caso de una población infinita (o finita en que se muestree con remplazo), cuya desviación estándar σ se conoce o se puede estimar adecuadamente, si se tiene que la estadística S obtenida de la muestra es el promedio aritmético, entonces la media de su distribución muestral es $\mu_S = \mu_{\bar{X}} = \mu$, y su desviación estándar es $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ son, respectivamente, la media y la desviación estándar de la variable aleatoria X asociada a la población, y n es el tamaño de la muestra. En tal caso, si $\bar{X}$ tiene distribución normal, la variable estandarizada correspondiente es:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Para el caso de muestreo sin remplazo de población finita, se tiene que $\sigma_{\bar{X}} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$, en donde N_p es el tamaño de la población, por lo que la variable estandarizada será

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}}$$

En los dos casos anteriores, el valor de Z correspondiente al de $\bar{X}$ de la muestra es el que se debe comparar con el valor crítico correspondiente al nivel de significancia fijado, para así aceptar o no la hipótesis nula (prueba de una cola). Si se trata de una prueba de dos colas, el valor de Z se debe comparar con los dos valores críticos que corresponden al valor de α seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de α.

Ejemplo 16.1

Se sabe que el promedio de calificaciones de una muestra aleatoria de tamaño 100 de los estudiantes de tercer año de ingeniería civil es de 7.6, con una desviación estándar de 0.2. Si μ denota la media de la población de esas calificaciones, X, y si se supone que $\bar{X}$ tiene distribución normal, probar la hipótesis $\mu = 7.65$ en contra de la hipótesis alternativa $\mu \neq 7.65$, usando un nivel de significancia de

a. 0.05

b. 0.01

Para la solución se deben considerar las hipótesis

$$H_0 : \mu = 7.65$$

$$H_1 : \mu \neq 7.65$$

Puesto que $\mu \neq 7.65$ incluye valores menores y mayores de 7.65, se trata de una prueba de dos colas.

La estadística bajo consideración es el promedio aritmético, $\bar{X}$, de la muestra, que se supone extraída de una población infinita. La distribución muestral de $\bar{X}$ tiene media $\mu_{\bar{X}} = \mu$, y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ denotan, respectivamente, la media y la desviación estándar de la población de calificaciones.

Bajo la hipótesis H_0 (considerándola verdadera), se tiene que

$$\mu_{\bar{X}} = 7.65 = \mu$$

y utilizando la desviación estándar de la muestra como una estimación de σ, lo cual se supone razonable por tratarse de una muestra grande,

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.2/\sqrt{100} = 0.2/10 = 0.02$$

a. Para la prueba de dos colas a un nivel de significancia de 0.05 se establece la siguiente regla de decisión

> *Aceptar H_0 si el valor Z correspondiente al valor del promedio de la muestra se encuentra dentro del intervalo de -1.96 a 1.96 (tabla 14.1). En caso contrario, rechazar H_0.*

Puesto que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.6 - 7.65}{0.02} = -2.5$$

se encuentra fuera del rango de -1.96 a 1.96, se rechaza la hipótesis H_0 a un nivel de significancia de 0.05.

b. Si el nivel de significancia es 0.01, el intervalo de -1.96 a 1.96 de la regla de decisión del inciso a se remplaza por el de -2.58 a 2.58 tabla (14.1). Entonces, puesto que el valor muestral $Z = -2.5$ se encuentra dentro de este intervalo, se acepta la hipótesis H_0 a un nivel de significancia de 0.01.

Ejemplo 16.2

La resistencia media a la ruptura de cables de acero fabricados por la empresa X es de 905 kg. Una empresa consultora sugiere a X que cambie su proceso de manufactura, con lo cual incrementará la resistencia de sus cables. Se prueba el nuevo proceso, y se extrae una muestra aleatoria de 50 cables, obteniéndose para ellos una resistencia promedio de 926 kg, con desviación estándar igual a 42 kg. ¿Se puede considerar que el nuevo proceso realmente incrementa la resistencia, con un nivel de confianza de 99%?

En este caso, se debe plantear una prueba de hipótesis de una cola, para la cual

$$H_0 : \mu = 905 \text{ kg}$$

$$H_1 : \mu > 905 \text{ kg}$$

Puesto que el tamaño de la muestra es suficientemente grande, se puede aproximar la distribución muestral de la resistencia promedio mediante una normal, y estimar el valor de σ de la población mediante S_X de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a H_0, se tiene que

$$\mu_{\bar{X}} = \mu = 905 \text{ kg}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

Para la prueba de una cola a un nivel de significancia de $\alpha = 1 - (1 - \alpha) = 1 - 0.99 = 0.01$, la regla de decisión es

> *Aceptar H_0 si el valor estandarizado de $\bar{X}$ de la muestra es menor o igual a $Z_c = 2.326$ (tabla 14.1); en caso contrario, rechazar H_0.*

En virtud de que

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{926 - 905}{5.94} = 3.535$$

es mayor de 2.326, se rechaza H_0 a un nivel de significancia de 1%, concluyéndose que en realidad el nuevo proceso sí incrementa la resistencia de los cables.

17. Pruebas de diferencias de medias.

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de dos muestras de tamaños n_X y n_Y, extraídas respectivamente de dos poblaciones con medias μ_X y μ_Y, y desviaciones estándar σ_X y σ_Y. Se trata de probar la hipótesis nula, H_0, de que no existe diferencia entre las medias, es decir, que $\mu_X = \mu_Y$. Si n_X y n_Y son suficientemente grandes (>30), la distribución muestral de las diferencias de los promedios es aproximadamente normal. Dicha distribución muestral es rigurosamente normal si las variables aleatorias X y Y asociadas a la población tienen distribución normal, aunque n_X y n_Y sean menores de 30. Para esta distribución muestral, la variable estandarizada Z, que se compara con los valores críticos correspondientes, se encuentra dada por

$$Z = \frac{\bar{X} - \bar{Y} - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y} - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}$$

con la cual se puede probar la hipótesis nula H_0 en contra de otras hipótesis alternativas, H_1, a un nivel apropiado de significancia.

Ejemplo 17.1

En el laboratorio de pruebas de una empresa fabricante de aparatos electrónicos se ensayaron dos marcas de transistores, A y B, de características similares, con objeto de comprobar su ganancia de voltaje. Se tomaron muestras aleatorias de 100 transistores de cada marca, arrojando una ganancia promedio de 31 decibeles, con desviación estándar de 0.3 decibeles para la marca A, y 30.9 decibeles de ganancia promedio, con desviación estándar de 0.4 decibeles para la otra. ¿Existe una diferencia significativa entre las ganancias en voltaje de los transistores a un nivel de significancia de

a. 0.05

b. 0.01

Si μ_A y μ_B son las medias respectivas de las dos poblaciones infinitas a las que corresponden las muestras, la prueba de hipótesis adopta la forma siguiente:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

Entonces, el valor de Z es, bajo la hipótesis H_0:

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{31 - 30.9}{\sqrt{\frac{(0.3)^2}{100} + \frac{(0.4)^2}{100}}} = 2$$

a. Puesto que se trata de una prueba de dos colas a un nivel de significancia de 0.05, la diferencia es significativa si el valor de Z se encuentra fuera del intervalo de -1.96 a 1.96. Como este es el caso, puede concluirse que efectivamente existe diferencia significativa en la ganancia en voltaje de los transistores.

b. Si la prueba es a un nivel de significancia de 0.01, la diferencia es significativa si Z se encuentra fuera del rango de -2.58 a 2.58. Partiendo del hecho de que $Z = 2$, la diferencia entre las ganancias es producto del azar, y se acepta la hipótesis de que ambos tipos de transistores tienen igual ganancia media en voltaje a un nivel de confianza de 99 por ciento.

Ejemplo 17.2

La estatura promedio de 50 estudiantes varones tomados al azar que participan en actividades deportivas es de 173 cm, con desviación estándar de 6.3 cm. Otra muestra aleatoria de 50 estudiantes varones que no participan en ese tipo de actividades tiene promedio de estatura igual a 171 cm, con desviación estándar igual a 7.1 cm. Probar la hipótesis de que los estudiantes varones que practican deportes son más altos que los que no lo hacen, a un nivel de significancia de 0.05.

Se debe decidir entre las hipótesis

$$H_0 : \mu_X = \mu_Y$$

$$H_1 \quad \mu_X > \mu_Y$$

siendo X la variable aleatoria asociada a la población infinita de estaturas de alumnos que practican deportes, y Y la asociada a la de estudiantes que no lo hacen, que también es infinita.

Bajo la hipótesis H_0, se tiene que

$$\mu_{\bar{X}-\bar{Y}} = 0$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = \sqrt{\frac{(6.3)^2}{50} + \frac{(7.1)^2}{50}} = 1.3424$$

Entonces, el valor de Z es

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{173 - 171}{1.3424} = \frac{2}{1.3424} = 1.489$$

Puesto que se trata de una prueba de hipótesis de una cola, a un nivel $\alpha = 0.05$, se rechazaría H_0 si el valor de Z muestral fuera mayor del valor crítico para dicho nivel, el cual es $Z_c = 1.645$. Puesto que $Z < Z_c$, en este caso se concluye que la diferencia en las estaturas de ambos grupos de estudiantes se debe únicamente al azar.

tiene ordenadas mayores de cero en el lado de las abscisas negativas. De hecho, la estadística S_x^2 se puede estudiar si se consideran muestras aleatorias de tamaño n extraídas de una población normal con desviación estándar σ_x y si para cada muestra se calcula el valor de la estadística.

$$\chi^2 = \frac{n\, S_x^2}{\sigma^2} \qquad (3.14)$$

donde S_x^2 es la varianza de la muestra.

El número de grados de libertad, ν, de una estadística se define como

$$\nu = n - k$$

siendo n el tamaño de la muestra y k el número de parámetros de la población que deben estimarse a partir de ella.

La distribución muestral de la estadística χ^2 está dada por la ecuación

$$f(\chi^2) = U\, \chi^{\nu-1}\, e^{-\frac{1}{2}\chi^2}$$

en la que U es una constante que hace que el área total bajo la curva resulte igual a uno, y $\nu = n - 1$ es el número de grados de libertad. Esta distribución se llama *Ji cuadrada*, misma que se presenta en la fig 21 para distintos valores de ν.

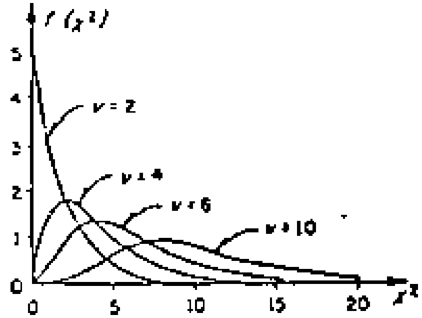


Fig 21. Distribución Ji cuadrada para distintos valores de ν

3.4 Muestras pequeñas

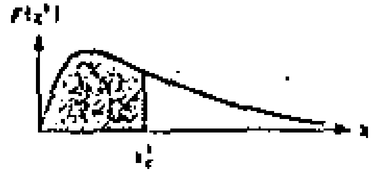
Como ya se indicó, para muestras grandes ($n \geq 30$) las distribuciones muestrales de muchas estadísticas son aproximadamente normales, siendo tanto mejor la aproximación cuanto mayor es el tamaño de n. Sin embargo, cuando se trata de muestras en las que $n < 30$, llamadas *muestras pequeñas*, la aproximación no es suficientemente buena, por lo que resulta necesario introducir una teoría apropiada para su estudio.

Al estudio de las distribuciones muestrales de las estadísticas para muestras pequeñas se le llama *teoría estadística de las muestras pequeñas*. Existen al respecto tres distribuciones importantes: *Ji cuadrada, F y t de Student.*

3.4.1 Distribución Ji cuadrada (χ^2)

Hasta ahora solo se ha tratado la distribución muestral de la media. En esta sección se verá lo concerniente a la distribución muestral de la varianza, S_x^2, para muestras aleatorias extraídas de poblaciones normales. Puesto que S_x no puede ser negativa, es de esperarse que su distribución muestral no sea una curva normal, ya que esta

TABLA 8. VALORES CRITICOS χ_c^2



ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.016	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

No obstante que la distribución ji cuadrada sólo se ha presentado en el estudio de las muestras pequeñas, cabe aclarar que es válida para aquellas mayores de 30 si la variable aleatoria involucrada tiene distribución normal.

3.4.1 Intervalo de confianza para la varianza

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la varianza de la población en términos de la varianza de una muestra extraída de ella, a un nivel de confianza dado $1-\alpha$, si se hace uso de los valores críticos χ_c^2 de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística χ^2, estará dado por

$$\chi^2_{c_1} < \frac{n S_x^2}{\sigma^2} < \chi^2_{c_2}$$

donde $\chi^2_{c_1}$ y $\chi^2_{c_2}$ son los valores críticos para los cuales el $(1-\alpha)/2$ por ciento del área se encuentra en los extremos izquierdo y derecho de la distribución, respectivamente.

Con base en lo anterior, se concluye que

$$\frac{n S_x^2}{\chi^2_{c_2}} < \sigma^2 < \frac{n S_x^2}{\chi^2_{c_1}}$$

es un intervalo de confianza para estimar a σ^2 a un nivel de confianza $1-\alpha$.

3.4.2 Prueba de hipótesis para la varianza

La prueba de hipótesis para la varianza de una población normal se efectúa calculando el valor de la estadística χ^2 y estableciendo las hipótesis H_0 y H_1 apropiadas, es decir, se adoptan reglas de decisión similares a las usadas para la estadística Z.

Ejemplo

La varianza del tiempo de elaboración de cierto producto es igual a 40 min; sin embargo, su proceso de manufactura se modifica y se toma una muestra de veinte tiempos, para la cual la varianza resulta ser igual a 62 min. ¿Es significativo - miento del tiempo de elaboración a un nivel de significancia de

a) 0.05

b) 0.01?

Se debe decidir de entre las hipótesis

$$H_0 : \sigma^2 = 40 \text{ min}$$

$$H_1 : \sigma^2 > 40 \text{ min}$$

Suponiendo que la hipótesis nula es correcta, el valor de la estadística χ^2 para la mue considerada es

$$\chi^2 = \frac{n S_x^2}{\sigma^2} = \frac{(20)(62)}{40} = 31$$

a) Como se trata de una prueba de una cola, la hipótesis H_0 se rechazaría el valor de la estadística χ^2 fuera mayor que el de χ^2 para un nivel de significancia igu a 0.05, el cual, para $\nu = 20 - 1 = 19$ grados de libertad resulta ser 30.1 (tabla 8). Con 31 > 30.1, H_0 se rechaza a un nivel de significancia de 0.05.

b) En este caso, el valor de χ^2 para un nivel de significancia de 0.01 y 19 gra dos de libertad es igual a 36.2. Puesto que 31 < 36.2, se acepta H_0 a un nivel de significan cia de 0.01.

3.4.2 Distribución F

Al efectuar la prueba de hipótesis de igualdad de medias para muestras pequeñas, en la siguiente sección se supondrá que las varianzas de las poblaciones a las que corresponden tales muestras son iguales. Por lo tanto, es necesario probar antes si tal su posición es correcta. Para ello, debe considerarse que si S_X^2, n_X y S_Y^2, n_Y son respectiva- mente la varianza y el tamaño de dos muestras extraídas de poblaciones normales que tienen igual varianza, entonces

$$F = \frac{S_X^2}{S_Y^2} \qquad (3.15)$$

TABLA 9 VALORES F_c PARA $\alpha = 0.01$

[illegible]

resulta ser el valor de una variable aleatoria (estadística) que tiene distribución F, con parámetros $\nu_x = n_x - 1$ y $\nu_y = n_y - 1$. Esta distribución (fig 22) cuenta con dos parámetros, ν_x y ν_y, que son los grados de libertad que corresponden a la variancia del numerador y del denominador de la ec 3.15, respectivamente. Cuando se hace referencia a una distribución F en particular, siempre se dan primero los grados de libertad para la variancia del numerador, es decir, $F(\nu_x, \nu_y)$. En la tabla 9 se presentan los valores críticos F_c para distintos valores de ν_x y ν_y y un nivel de significancia de 0.01. Cuando los grados de libertad ν_x o ν_y no se encuentren en dicha tabla, el valor de F se puede obtener mediante interpolación lineal. Si se desea probar la hipótesis a otros niveles de significancia, es factible emplear las tablas de la distribución F (refs. 9 y 11).

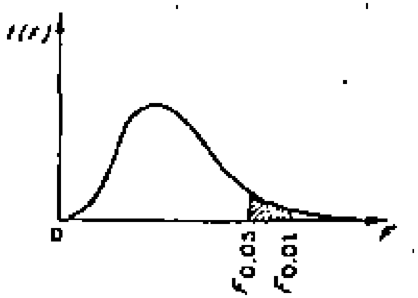


Fig 22. Distribución F.

De acuerdo con lo anterior, se puede probar la hipótesis nula

$$H_0: \sigma_x^2 = \sigma_y^2$$

en contra de alguna hipótesis alternativa adecuada haciendo uso del hecho de que el cociente S_x^2/S_y^2 es una estadística que tiene distribución F.

Ejemplo

Una empresa manufacturera de cartón prensado va a decidir acerca del empleo de una prensadora A o una B a fin de obtener un grosor determinado en su producto. El problema estriba en que ambas prensadoras proporcionan grosores muy similares, es decir, que la variancia de los grosores para las dos máquinas es la misma. Para decidir acertadamente, se toma una muestra aleatoria de 31 cartones prensados por la máquina A y otra de 41 por la B. Como las variancias del grosor para los cartones de las muestras resul-

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 > \sigma_B^2$$

con objeto de probarlas a un nivel de significancia de 0.01.

El valor de la estadística F resulta

$$F = \frac{S_A^2}{S_B^2} = \frac{12}{5} = 2.4$$

puesto que $\nu_A = 31 - 1 = 30$ y $\nu_B = 41 - 1 = 40$, en la tabla 9 se puede ver que para un nivel de significancia de 0.01 el valor, F_c, de $F(30, 40)$ es 2.11. De acuerdo con estos valores, la hipótesis H_0 se rechazaría si el valor de F fuera mayor que $F_c(30, 40)$.

Puesto que lo anterior resulta ser cierto, se rechaza H_0, concluyéndose que la prensadora B sería la mejor elección.

3.4.3 Distribución t de Student

Si se consideran muestras de tamaño n extraídas de una población normal con media μ y variancia desconocida, para cada muestra se puede calcular la estadística T definida mediante la fórmula

$$T = \frac{\bar{X} - \mu}{S_x}\sqrt{n - 1} \qquad (3.16)$$

G

donde $\bar{X}$ es el promedio y S_x la desviación estándar de la muestra.

La distribución muestral de T (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}}$$

γ es ... exponente

en la que U es una constante que hace que el área bajo la curva sea igual a uno, y $\nu = n - 1$ es el número de grados de libertad.

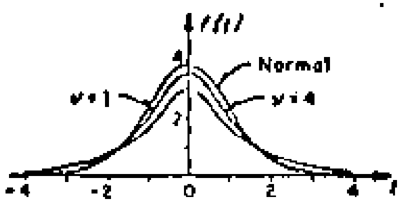


Fig 23. Distribución t de Student para distintos valores de ν

En la fig 23 se aprecia que conforme ν (o n, el tamaño de la muestra) aumenta, la distribución de $f(t)$ se aproxima a la distribución normal.

3.4.3.1 Límites e intervalos de confianza

De manera similar a como se hizo con la distribución normal, es posible estimar los límites de confianza de la media, μ, de una población mediante los valores *críticos*, t_c, de la distribución t, que dependen del tamaño de la muestra y del nivel de confianza deseado, encontrándose dichos valores en la tabla 10.

Así pues,

$$-t_c < \frac{\bar{X} - \mu}{S_x}\sqrt{n-1} < t_c$$

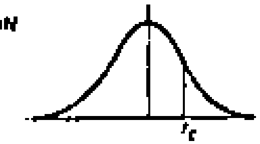
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\bar{X} - t_c\frac{s_x}{\sqrt{n-1}} < \mu < \bar{X} + t_c\frac{s_x}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\bar{X} \pm t_c\frac{s_x}{\sqrt{n-1}}$$

TABLA 10. VALORES t_c PARA LA DISTRIBUCIÓN t DE STUDENT



ν	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.277	.137
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.559	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.35	.870	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.537	.258	.128
15	2.95	2.60	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.127
19	2.86	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.257	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.70	1.31	.855	.684	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.527	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

3.4.3.2 Pruebas de hipótesis

La prueba de hipótesis para la media de una población se puede efectuar con muestras pequeñas en forma análoga a la de muestras de tamaño mayor de 30 si en lugar de utilizar a la estadística Z se emplea la T. Entonces, si se consideran dos muestras aleatorias cuyos tamaños, desviaciones estándar y promedios son n_X, S_X, $\bar{X}$ y n_Y, S_Y, $\bar{Y}$, respectivamente, extraídas de poblaciones normales de igual varianza ($\sigma_X^2 = \sigma_Y^2$), se puede probar la hipótesis, H_0, de que las muestras provienen de una misma población, es decir, de que también sus medias son iguales, utilizando la estadística T definida por

$$T = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \qquad (3.17)$$

donde

$$s = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \qquad (3.18)$$

cuya distribución es la t de Student, con $\nu = n_X + n_Y - 2$ grados de libertad.

Ejemplo

Conforme al plan de desarrollo agrícola de una región, se probó un nuevo fertilizante para maíz. Para ello se escogieron 24 ha de terreno, aplicándose dicho producto a la mitad de ellas. El promedio de producción de maíz en la zona que se usó fertilizante fue de 5.3 ton, con una desviación estándar de 0.40 ton, en tanto que en la otra zona el promedio fue de 5.0 ton, con desviación estándar de 0.36 ton.

De acuerdo con los resultados, ¿se puede concluir que existe un aumento significativo en la producción de maíz al usar fertilizante, si se utiliza un nivel de significancia de

a) 0.01

b) 0.05?

Solución

Para probar la hipótesis de igualdad de medias es indispensable saber primero si las muestras provienen de dos poblaciones normales de igual varianza. En ese caso, si σ_X^2 y σ_Y^2 denotan a las varianzas de la producción de maíz en la zona tratada y en la no tratada, respectivamente, se debe probar la hipótesis nula H_0: $\sigma_X^2 = \sigma_Y^2$ en contra de la hipótesis alternativa H_1: $\sigma_X^2 > \sigma_Y^2$ a los dos niveles de significancia establecidos.

El valor de la estadística F es, de la ec 3.15,

$$F = \frac{S_X^2}{S_Y^2} = \frac{(0.40)^2}{(0.36)^2} = 1.23$$

y el valor crítico de F (11, 11), obtenido de la tabla 9 mediante interpolación lineal, resulta 4.47. Por lo tanto, como 1.23 < 4.47, se acepta la hipótesis nula a un nivel de significancia de 0.01.

El valor crítico de F (11, 11) a un nivel de significancia de 0.05 (ref 9) es 2.82, de ahí que como 1.23 < 2.82, también se acepta la hipótesis H_0.

Con base en lo anterior, se debe decidir entre las hipótesis

H_0: $\mu_X = \mu_Y$ (la diferencia en los promedios se debe al azar)

H_1: $\mu_X > \mu_Y$ (el fertilizante mejora la producción)

Bajo la hipótesis H_0, se tiene que

$$s = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$t = \frac{5.3 - 5.0}{0.397\sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.85$$

a) Puesto que se trata de una prueba de una cola a un nivel de significancia de 0.01, se rechaza la hipótesis H_0 si t es mayor que el valor crítico, t_c, correspondiente a dicho nivel, el cual para $v = n_x + n_y - 2 = 12 + 12 - 2 = 22$ grados de libertad, se obtiene de la tabla 8 como $t_c = 2.51$. Como $t < t_c$, la hipótesis H_0 no se puede rechazar a un nivel de significancia de 0.01.

b) Si el nivel de significancia de la prueba es de 0.05, se rechaza H_0 si t es mayor que el valor t_c respectivo que para 22 grados de libertad es $t_c = 1.72$, por lo que de acuerdo con lo anterior, H_0 se rechaza a un nivel de significancia de 0.05.

Directorio de Alumnos del Curso Probabilidad y Estadística: Fundamentos y Aplicaciones Febrero de 1982.

1. Juan Ricardo Albarrán Luna
UNAM
Instituto de Ingeniería
Coordinador del Lab. de Ing. Ambiental
México 20, D.F.
550 52 15 Ext. 3602

Sta. Cruz 210-22
Coyoacán
000600 México, D.F.
549 87 54

2. Eloy Altamirano Conde
I M S S
Jefe de la Ofic. de Estadística
Pirul 37
Sta. Ma. Insurgentes
México, D.F.
583 06 37

Portal 26
Col. Pastores
Naucalpan, Edo. de Méx.
373 49 68

3. Francisco Alvarez Ledesma
Constructora ICA
Ingeniero Civil
Minería 145
Col. Escandón
México 18, D.F.
516 04 60

Calle Bosavea de Tunez 117
Fracc. Bosque de Aragón
Edo. de México

4. Héctor Basurto Nava
PEMEX
Jefe del Depto. de Análisis
Gerencia de Exploración
Av. Marina Nal. 329
México 17, D.F.

Framboyanes 14
Jardines de Sn. Mateo
Naucalpan, Edo. de Méx.
560 44 84

5. Roberto Bocanegra Guzmán
Policya, S.A.
Ing. de Operaciones
La Presa s/n
Sn. Juan Ixhuatepec
Tlanepantla, Edo. de Méx.
586 08 83

Edif. Antonio Rosales C -12
Tlatelolco
Delegación Cuauhtémoc
México 3, D.F.
583 41 63

6. Oscar Caballero Rojas
SARH
Dir. Gral. de Grande Irrigación

7. Ricardo Cadena Galicia
Jefe de Oficina
SARH
Dir. Gral. de Obr. Hidr. y de
Ingr. Agrícola para el Desarrollo Rural
P. de la Reforma 107-7°
Delegación Cuauhtémoc
06470 México, D.F.

Mar Kara 18
Col. Popotla
Delegación M. Hgo.
México 11400, D.F.

8. Antonio Carrillo Avelar
ENEP Aragón
Ayudante de Profesor

Palacio de Gobeierno 9
Metropolitana
Edo. de Méx.
797 30 89

9. Mauricio Carrillo García
Escuela Nacional de Estudios Profesionales
Aragón
Profesor de Asignatura
Av. Rancho Seco
Impulsora
Netzahualcoyotl
México
796 01 88

La Palmas/n
San Diego Texcoco
Estado de Mexico
4 26 19

10. Marina Castillo Garduño
S A R H
Jefe de Sección de Análisis de Diagnóstico
Reforma 107-7°
San Rafael
D. Cuauhtémoc
México, D.F.
535 02 52

Dr. Durán 39 G
Col. Doctores
D. Cuauhtémoc
México, D.F.

11. Gabriel Castro Medina
I M P
Jefe de Ofi. del Sist. de Inf.
Av. Cien Metros No. 152
México 14, D.F.
567 66 00 Ext. 2457

Sabino 208-14
Sta. Ma. la Ribera
D. Cuauhtémoc
065 00 México, D.F.
547 16 57

12. Romaldo Corral García
S A R H
Programador Analista "B"
P. de la Reforma 107-10° Piso
D. Cuauhtémoc
México, D.F.
535 71 77

Av. Universidad 1900 Ed. 3 Depto. 201
Oxtopulco Universidad
Coyoacán
México, D.F.
658 34 28

13. Leonardo M. Díaz Angeles
Lino Merino No. 847
Juan Escutia
D. Iztapalapa
09100 México, D.F.
797 76 97

POLICYD, S.A.
Ingeniero de Proceso
Av. La pResa s/n
San Juan Ixhuatepec
Tlanepantla, Edo. de Méx.
586 08 88

14. Elizabeth España Gómez
SARH
Dir. Gral. de Obr. Hidr. y de Ing. Agr. Para el Desarrollo Rural
Investigadora Social, Evaluación de Proyectos
Av. P. de la Reforma 107-7°
Col. Cuauhtémoc
06470 México, D.F.
546 16 90

4ta. Cerrada de Retoño 122
Col. El Retoño
D. Iztapalapa
09410 México, D.F.
672 26 77

15. Gloria Flores Calderón
Av. Niños Héroes de Chapultepec No.149-4
Col. Niños Héroes
México, D.F.
696 33 39

16. Raymundo Jorge García Barrientos
Comisión de Vialidad y Transporte Urbano
Aux. de Residente "B"
Jefe de la Sección Control y Supervisión
Av. Universidad No. 800
Narvarte
B. Juárez
03020 México, D.F.
575 97 47

17. Gustavo González Gutiérrez
Secundaria Técnica No. 15
Sistematizador
Calz. Azcapotzalco La Villa
México, D.F.
561 19 30

Edif. 18 A 202
Unidad Cuitláuac
02500 México, D.F.
556 04 97

18. Ricardo López Florido
CABLE VISION S.A.
Analista de Mercados
Dr. Río de la Loza 182
Col. Doctores
D. Cuauhtémoc
06720 México, D.F.
588 15 39

Av. 529 # 180
Aragón
D. G. A. Madero
07000 México, D.F.
760 40 65

19. Fidel López Guzmán
Ing. de Sist. y Transp. Metropolitano
Jefe de Sección
Legaría 252
Col. Pensil
México, D.F.
399 69 22

Pte. 81 # 112-2
COVE
D. M. Hgo.
México, D.F.
399 69 22

20. Miguel Angel López López
Cía. Minera Autlán
Inv. Metalúrgico

Dos Arbolitos 171
Col. B. Juárez
Cda. Nzh.
Estado de México
57000 México, D.F.

21. Francisco López Paredes
Colgate Palmolive, S.A. de C.V.
Ing. Quím. en Lab. de Inv. y Desarrollo
Presa la Angostura 225
Col. Irrigación
D. M. Hgo
11500 México, D.F.
557 00 22

Rinconada Monedas, Solea 201
Pedregal de Carrasco
Coyoacán

22. Ricardo Heriberto Manzano Cañas
Instituto de Ingeniería
Técnico Academico
UNAM
550 52 15 Ext. 3601

Calz. de la Romería 161 Casa K
Colina del Sur

23. Manuel Ramón Martínez Cárdenas
Dir. Gral. de Protección y Ordenación Ecológica
Jefe de Oficina
Reforma 107-2°
San Rafael
Col. Cuauhtémoc
06000 México, D.F.
566 95 59

Castañas 764
Nva. Sta. Ma.
Azcapotzalco
02800 México, D.F.
556 16 55

24. Reynaldo Martínez y Carrera
S H C P
Subsecretaría de Ingresos
Asesor
Dir. Gral. Técnica
Izazaga 89-10° Piso
Centro
D. Cuauhtémoc
06000 México, D.F.
521 65 31

Rincón del Cielo 1
Bosque Residencial del Sur
Xochimilco
16010 México, D.F.
676 80 49

25. Rodolfo Morales Díaz
Secretaría de Agricultura y Rec. Hidráulicos
Analista
Reforma 107-10° Piso
Col. San Rafael
D. Cuautémoc
México, D.F.
546 49 67

E. Zapata 174 Ed. H. 301
San Lázaro 1
542 02 18

26. Ricardo Moreno Chan
Facultad de Medicina Veterinaria
Profesor
UNAM
México 20, D.F.
550 52 15 Ext. 4965
Av. Hda. 92
Villa Coapa
Tlalpán
México 22, D.F.
594 37 36

27. Arturo Orozco Torres
S A H O P
Supervisor
Reforma 77
México, D.F.
Sn. Cosme 23-9
Sta. Ma. la Ribera
D. Cuauhtémoc
México, D.F.
516 91 02

28. Mordejal Zvi Retchkiman Konigsberg
Laboratorista
UNAM
México 20, D.F.
Minería 17-2
Escandón
11800 México, D.F.
516 61 91

29. Georgina Solano Lozano
PEMEX
Analista
Av. Marina Nal. 329
D. M. Hgo.
México 17, D.F.
250 26 11 Ext. 2663
Rosa Violeta 198-2
Molino de Rosas
D. A. Obregón
México, D.F.
651 24 40

30. Jesús Torres Calderón
Colgate Palmolive, S.A. de C.V.
Ing. Químico
Area de Investigación
Presa la Angostura 225
Col. Irrigación
D. M. Hgo.
11500 México, D.F.
557 00 22
Lag. Sn. Cristobal 126-3
Col. Anáhuac
D. M. Hgo.
11320 México, D.F.
250 04 03

31. Alfonso Urban Sánchez
PEMEX
Analista Programador
Marina Nal. 329
Col. Anáhuac
México 17, D.F.
531 66 92
Plaza Hgo. 2
Tultepec, Edo. de México

32. Román Vázquez Berber
Comisión de Vialidad y
Transporte Urbano
Coordinador de Proyecto Geométrico
Av. Universidad 800-3° Piso
Sta. Cruz Atoyac
D. Benito Juárez
03310 México, D.F.

33. Braulio Vázquez
Cía. de Luz y Fza. del Centro, S.A.

34. Manuel Vázquez Portillo
Comisión del Plan Nacional Hidráulico
Jefe de Proyecto
Tepic 40-2° Piso
Col. Roma
D. Cuauhtémoc
067-- México, D.F.
574 14 97
Cda. Bartolache 8-1
Col. del Valle
B. Juárez
03100 México, D.F.
534 68 99

35. Francisco Velasco Castillo
S. H. C. P.
Subsecretaría de Ingresos
Actuario
Dir. Gral. Téc.
Izazaga 89-10
Centro
D. Cuauhtémoc
06000 México, D.F.
585 61 92
Vista Hermosa 140
Col. Portales
D. B. Juárez
03300 México, D.F.
532 55 91

36. José Luis Wala Prasencia
I C A, S.A.
Jefe de Frente C (Técnico)
Minería 145 Edif. B 1° Piso
Col. Escandón
México 18, D.F.
516 04 60 Ext.618
Calle 1505 No. 61
6ta. Secc. Sn. J. de Aragón
D. G. A. Madero
07920 México, D.F.
796 78 30

'edes.