DIRECTORIO DE PROFESORES DEL CURSO: PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES - Febrero-Marzo, 1984

1. Dr. Octavio A. Rascón Chávez (COORDINADOR)
Director
Facultad de Ingeniería,
U N A M
Ciudad Universitaria
México, D.F.
Tel. 550 51 50 y 548 33 54

2. M. en I. Augusto Villarreal Aranda
Gerente de Operaciones
Grupo VEA
Asia N° 31
México, D.F.
Tel. 554 45 31 y 554 41 31

3. M. en I. Rubén Téllez Sánchez
Profesor de la Subjefatura del Area
de Ingeniería de Sistemas
División de Estudios de Posgrado
Facultad de Ingeniería,
U N A M
México, D.F.
Tel. 550 52 15 ext. 4482, 4486

PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES

6 de febrero al 16 de marzo

1984

6 al 29 de febrero 18 a 21 h c/d	INTRODUCCION Probabilidad. Estádistica descriptiva e inferencia estadística	Dr. Octavio A. Rascón Chávez M. en I. Rubén Téllez Sánchez
	ESTADISTICA DESCRIPTIVA Obtención de datos muestreo aleatorio simple. Procesamiento de información. Tablas de frecuencia. 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.	
	PROBABILIDAD Eventos. Teoría de conjuntos. Espacio de eventos Probabilidad condicional. Independencia. Teorema de Bayes. Variables aleatorios continuas y discreción de distribución. Momentos y esperanzas. Distribuciones de Bernoulli. Hipergeométrica. Binominal y de Poisson. Proceso de Poisson simple. Distribuciones uniforme, exponencial, normal y extremos. Ejemplos y Aplicaciones.	
2 al 16 de marzo 18 a 21 h c/d	INFERENCIA ESTADISTICA Estimación puntual de los parámetros de una distribución de probabilidades. Estimación por intervalos. Distribuciones muestrales Pruebas de hipótesis que involucren medidas, varianclas o proporciones. Prueba de bondad de ajuste en regresión lineal y en distribuciones de probabilidades. Ejemplos y Aplicaciones	M. en I. Augusto Villarreal Aranda

EVALUACION DEL PERSONAL DOCENTE

CURSO: "PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES"

FECHA: 6 de febrero al 16 de marzo de 1984

	CONFERENCISTA	DOMINIO DEL TEMA	EFICIENCIA EN EL USO DE AYUDAS AUDIOVISUALES	MANTENIMIENTO DEL INTERES. (COMUNICACION CON LOS ASISTENTES, AMENIDAD, FACILIDAD DE EXPRESION).	PUNTUALIDAD	
1.	DR. OCTAVIO A. RASCON CHAVEZ					
2.	M. EN I. RUBEN TELLEZ SANCHEZ					
3.	M. EN I. AUGUSTO VILLARREAL ARANDA					
4.						
5.						
6.						
7.						
8.						
9.						

ESCALA DE EVALUACION: 1 a 10

EVALUACION DE LA ENSEÑANZA

SU EVALUACION SINCERA NOS AYUDARA A MEJORAR LOS PROGRAMAS POSTERIORES QUE DISEÑAREMOS PARA USTED.

"PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES"

6 de febrero al 16 de marzo de 1984

TEMA	ORGANIZACION Y DESARROLLO DEL TEMA	GRADO DE PROFUNDIDAD LOGRADO EN EL TEMA	GRADO DE ACTUALIZACION LOGRADO EN EL TEMA	UTILIDAD PRACTICA DEL TEMA	
INTRODUCCION					
ESTADISTICA DESCRIPTIVA					
PROBABILIDAD					
INFERENCIA ESTADISTICA					

ESCALA DE EVALUACION: 1 a 10

EVALUACION DEL CURSO

	CONCEPTO	EVALUACION
1.	APLICACION INMEDIATA DE LOS CONCEPTOS EXPUESTOS	
2.	CLARIDAD CON QUE SE EXPUSIERON LOS TEMAS	
3.	GRADO DE ACTUALIZACION LOGRADO CON EL CURSO	
4.	CUMPLIMIENTO DE LOS OBJETIVOS DEL CURSO	
5.	CONTINUIDAD EN LOS TEMAS DEL CURSO	
6.	CALIDAD DE LAS NOTAS DEL CURSO	
7.	GRADO DE MOTIVACION LOGRADO CON EL CURSO	

ESCALA DE EVALUACION DE 1 A 10

1. ¿Qué le pareció el ambiente en la División de Educación Continua?

MUY AGRADABLE	AGRADABLE	DESAGRADABLE

2. Medio de comunicación por el que se enteró del curso:

PERIODICO EXCELSIOR ANUNCIO TITULADO DIVISION DE EDUCACION CONTINUA	PERIODICO NOVEDADES ANUNCIO TITULADO DIVISION DE EDUCACION CONTINUA	FOLLETO DEL CURSO

CARTEL MENSUAL	RADIO UNIVERSIDAD	COMUNICACION CARTA, TELEFONO, VERBAL, ETC.

REVISTAS TECNICAS	FOLLETO ANUAL	CARTELERA UNAM "LOS UNIVERSITARIOS HOY"	GACETA UNAM

3. Medio de transporte utilizado para venir al Palacio de Minería:

AUTOMOVIL PARTICULAR	METRO	OTRO MEDIO

4. ¿Qué cambios haría usted en el programa para tratar de perfeccionar el curso?

____

____

____

5. ¿Recomendaría el curso a otras personas?

SI	NO

6. ¿Qué cursos le gustaría que ofreciera la División de Educación Continua?

7. La coordinación académica fue:

EXCELENTE	BUENA	REGULAR	MALA

8. Si está interesado en tomar algún curso intensivo ¿Cuál es el horario más conveniente para usted?

LUNES A VIERNES DE 9 A 13 H. Y DE 14 A 18 H. (CON COMIDAS)	LUNES A VIERNES DE 17 A 21 H.	LUNES, MIERCOLES Y VIERNES DE 18 A 21 H.	MARTES Y JUEVES DE 18 A 21 H.

VIERNES DE 17 A 21 H. SABADOS DE 9 A 14 H.	VIERNES DE 17 A 21 H. SABADOS DE 9 A 13 Y DE 14 a 18 H.	O T R O

9. ¿Qué servicios adicionales desearía que tuviese la División de Educación Continua, para los asistentes?

10. Otras sugerencias:

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

PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES

PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA EN REGRESION LINEAL

DR. OCTAVIO A. RASCON CHAVEZ

FEBRERO, 1984

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

$$\bar{Y} = mx + b$$

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:

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

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

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

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

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

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:

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

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: $b \pm t_c s_b$.

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

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: $m \pm t_c s_m$

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

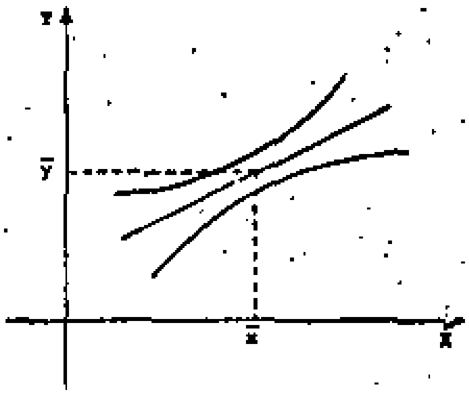
c. PARA LA PREDICCION, $Y_i \in \hat{Y}_i \pm t_c \sigma_{\hat{Y}}$

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

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	20.3	23.1	23.2	23.6	25.8	26.3

SI SE AJUSTA UNA RECTA POR MINIMOS CUADRADOS SE OBTIENE

$$\bar{Y} = 0.225\,x + 13.01$$

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

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.	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$x_i-\bar{x}$,	$(x_i-\bar{x})^2$
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
$\Sigma=285$				$\Sigma=2.40$		$\Sigma=437.2$

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

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

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

INTERVALOS DE CONFIANZA:

a) PARA B :

$$13.01 \pm t_c\, s_{y|x}\sqrt{\frac{\bar{x}^2}{n s_x^2} + \frac{1}{n}}$$

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

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

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

B) PARA M:

$$0.225 \pm t_c \frac{s_{y|x}}{\sqrt{n s_x^2}} = 0.225 \pm 2.776\, \frac{0.77}{\sqrt{6\,(72.9)}} =$$

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

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_I, ESTE ULTIMO PARA UN $X = x_I$ 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 $\dfrac{B - b_0}{s_{y|x}\sqrt{\dfrac{\overline{x^2}}{nS_x^2}}} = \dfrac{B - b_0}{\dfrac{s_{y|x}}{s_x}\sqrt{\dfrac{\overline{x^2}}{n}}} = T$

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

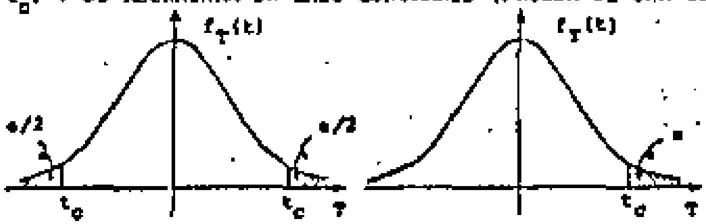
SI SE DESEA PROBAR LA HIPOTESIS

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

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

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

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

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

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: $t = \dfrac{m - m_0}{\dfrac{s_{y|x}}{s_x\sqrt{n}}}$

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

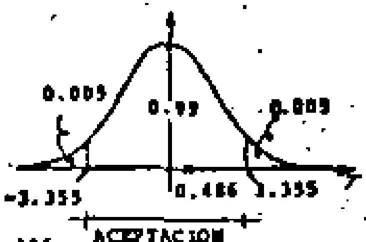
$m = 0.093$, $b = 0.033$, $s^2_{y|x} = 0.01258$

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

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

b. PROBAR LA HIPOTESIS DE QUE $M = 0.1$

CON $\alpha = 0.01$ Y $\sigma_{y|x}$ DESCONOCIDA.



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

$$t = \frac{b - b_0}{s_{y|x}\sqrt{\dfrac{\overline{x^2}}{nS_x^2}}} = \frac{0.033 - 0}{\sqrt{0.01258}\sqrt{\dfrac{28.5}{10 \times 8.25}}} = 0.486$$

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

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

$$t = \frac{m - m_0}{\dfrac{s_{y|x}}{s_x \sqrt{n}}} = \frac{0.093 - 0.1}{\dfrac{\sqrt{0.01258}}{\sqrt{8.25 \times 10}}} = 0.567 < 1.355$$

SE ACEPTA H_0 CON 95% DE NIVEL DE CONFIANZA.

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

PRUEBA

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

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 $\alpha = 0.05$.

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

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

$t_0 = t_{0.975,\ 28} = 2.048 < 13.448$

$\therefore$ 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 FUNDAMENTOS Y APLICACIONES

E J E R C I C I O S

DR. OCTAVIO A. RASCON CHAVEZ

FEBRERO, 1984

1

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.

2

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 Nº	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

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

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

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

$$p_4 = P(X > 4) = \int_3^\infty (0.45)e^{-0.45t}\,dt = 0.259;\quad 50p_4 = 12.95$$

$$\Sigma = 0.998 \approx 1;\quad \Sigma = 49.90 \approx 50$$

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

$\chi^2_{0.95,3} = 7.81 < 8.71$

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 $\alpha = 0.01$.

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

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	108.2	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

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

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

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 DESCRIPTIVA*, MULTIPLICANDO x 10^{-5} CADA NUMERO Y ELIMINANDO LOS 3 ULTIMOS DIGITOS SE OBTIENE LA SIGUIENTE MUESTRA:

0.16 - 0.01 - 0.04 - 0.53 - 0.79 - 0.21 - 0.03 - 0.92 - 0.36 - 0.31
0.59 - 0.73 - 0.47 - 0.47 - 0.87 - 0.99 - 0.00 - 0.80 - 0.71 - 0.18
0.20 - 0.33 - 0.30 - 0.03 - 0.23 - 0.14 - 0.13 - 0.45 - 0.22 - 0.19
0.09 - 0.74 - 0.60 - 0.96 - 0.20 - 0.42 - 0.78 - 0.05 - 0.22 - 0.34
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
					$\Sigma = 10.0$

9

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

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

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

[illegible], O. A., "INTRODUCCION A LA ESTADISTICA DESCRIPTIVA", VOL. I, 3a. ED. (1977), PAG. 21.

10

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.

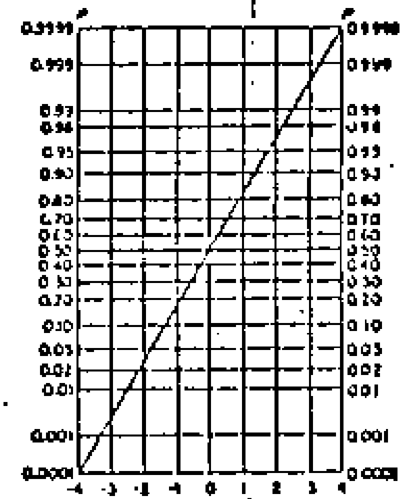


Figure 1-29

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 1-18, 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 1-2 and Figure 1-30. Table 1-2 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 × cum f_i/N. For example, the fifth number in the last column, 25.3, equals 100 × 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

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

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

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

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	20126	20516	20530	-6	142

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

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

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

$S^2(X) = \frac{20126}{17} - 1180.1245 = 15.5225$; $S(X) = 3.9396604$

$S^2(Y) = \frac{20516}{17} - 1204.4982 = 19.9723$; $S(Y) = 4.4690379$

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

$$r_{XY} = \frac{[20530/17] - (34.352941)(34.705882)}{(3.9396604)(4.4690375)} = 0.7742881$$

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

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

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

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

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	19667	19072	-21	175

$\bar{x} = \frac{545}{16} = 35.75$; $\bar{x}^2 = 1139.0625$; $\bar{d} = -1.3125$; $\bar{d}^2 = 1.7226562$

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

$S_x^2(x) = \frac{18614}{16} - 1139.0625 = 24.3125$; $S(x) = 4.9307707$

$S^2(Y) = \frac{19667}{16} - 1229.3789 = 12.3066$; $S(Y) = 3.5063614$

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

$$r_{xy} = \frac{(19072/16)-(35.75)(35.0625)}{(4.9307707)(3.5063614)} = \underline{\underline{0.4994934}}$$

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

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

$$t_d = \left|\frac{(35.75-35.063)\sqrt{15}}{3.036}\right| = 1.67 < 2.13$$

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

AMBAS ESCUELAS JUNTAS

$\Sigma x_i = 1124$, $\Sigma y_i = 1151$, $\Sigma x_i^2 = 38940$, $\Sigma y_i^2 = 39573$, $\Sigma d_i = -27$, $\Sigma d_i^2 = 317$

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

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

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

$S^2(Y) = \frac{40683}{33} - 1216.5297 = 16.2084$; $S(Y) = 4.0358889$

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

$$r_{xy} = \frac{(39573/33)-(34.060606)(34.878787)}{(4.458161)(4.0358889)} = \underline{\underline{0.6201924}}$$

$$t_0 = \left| \frac{(31.061 - 34.879)\sqrt{32}}{2.989} \right| = 1.540 < 2.04$$

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

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

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

<u>RESPUESTAS AL INCISO c</u>

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

$$= 0.364 < 2.04$$

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

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

$$= 0.24 < 2.04$$

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 95% 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 95% DE NIVEL DE CONFIANZA.

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

PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES

E J E R C I C I O S

FEBRERO, 1984

Palacio de Minería Calle de Tacuba 5 primer piso Deleg. Cuauhtemoc 06000 México, D.F. Tel.: 521-40-20 Apdo. Postal M-2285

3.17 Suponga que la vida, en horas, de un cierto tipo de tubo tiene la f.d.p. $f(x) = a/x^2$, $x \geq 500$ y $f(x) = 0$, $x < 500$. Encuentre f.d.a., determine la media y la varianza. ¿Cuál es la probabilidad de que un tubo dure cuando menos 300 horas?

3.18 Un submarino lleva 3 proyectiles dirigidos. Suponga que el único error es en una dirección (por ejemplo, un error de rango pero no lateral) y que un impacto dentro de 64 km del blanco, es un éxito, calcule la probabilidad de una operación exitosa (esto es, una operación en la cual, cuando menos, un impacto es un éxito) si todos los proyectiles son lanzados y el error de la f.d.p. es:

$$
\begin{aligned}
f(x) &= (161 + x)/16100 & -161 < x < 0 \\
&= (161 - x)/16100 & 0 < x < 161 \\
&= 0 & \text{cualquiera}
\end{aligned}
$$

3.19 Respecto al problema anterior, el submarino puede llevar ocho proyectiles de menor tamaño. Sin embargo, en este caso un impacto debe estar dentro de 24 km para que sea un éxito. Suponiendo la misma f.d.p. ¿Se usarían los proyectiles ligeros o los pesados?

3.20 Una estación de servicio es abastecida de gasolina una vez por semana. Su volumen semanal de ventas, en miles de litros, se predice por la f.d.p. $f(x) = 5(1 - x)^4$ para $0 < x < 1$. Determine la capacidad del tanque subterráneo, si la probabilidad de que el suministro se agote en una semana es: 0.01.

3.21 Demuestre que la correlación entre dos variables aleatorias es cero si son estadísticamente independientes.

3.22 Sea $f_1(x) = 1$ la densidad marginal de X para $-\frac{1}{2} < x < \frac{1}{2}$. Para valores dados de X la densidad condicional de Y es:

$$
\begin{aligned}
f(y|x) &= 1;\ x < y < x + 1,\ -\tfrac{1}{2} < x < 0 \\
&= 1;\ -x < y < 1 - x,\ 0 < x < \tfrac{1}{2} \\
&= 0;\ \text{cualquiera}
\end{aligned}
$$

Encuentre la correlación entre X y Y. Discuta la relación entre correlación e independencia estadística.

3.23 En un proceso se producen partes que en promedio son defectuosas en uno por ciento. Se toman al azar diez partes del proceso, teniendo en cuenta que el proceso se suspende si una o más de las diez partes son defectuosas. Se pregunta la probabilidad de que el proceso se suspenda.

3.24 En la inspección de 1,000 juntas soldadas efectuadas por un cierto soldador, se encontraron 150 defectuosas. Si el soldador va a soldar cinco juntas ¿Cuál es la probabilidad de que ninguna sea defectuosa? ¿De que una lo sea? ¿De que dos lo sean? ¿De que dos o más sean defectuosas? Discuta cualquier suposición que usted haga en la solución del problema.

3.25 En el montaje de un avión se usa una gran cantidad de remaches. Se ha determinado que la distribución de probabilidad para el número de remaches defectuosos es la de Poisson para $\lambda = 2$. Encuentre la probabilidad de que el número de remaches defectuosos en un avión no sea más de dos.

3.26 Supóngase que hay un promedio de 1 error tipográfico por cada 10 páginas de un cierto libro. ¿Cuál es la probabilidad de que un capítulo de 30 páginas no contenga errores?

3.27 Un conmutador telefónico opera un promedio de 600 llamadas durante la hora de mayor tráfico. El conmutador puede hacer, como máximo, 20 conexiones por minuto. ¿Cuál es la probabilidad de que el conmutador esté sobrecargado en un minuto dado de la hora de mayor tráfico?

3.28 Suponiendo una distribución normal, encuentre:

(a) $P(-3 < Y \leq -1)$; dados $\mu = 0, \sigma = 1$.
(b) $P(-3 \leq Y < 0.5)$; dados $\mu = 0, \sigma = 1$.
(c) $P(-8 < Y < 0)$; dados $\mu = 2, \sigma^2 = 4$
(d) $P(4 < Y \leq 50)$; dados $\mu = -0.1, \sigma^2 = 4$
(e) $P(Y \geq 3)$; dados $\mu = 0, \sigma^2 = 1$
(f) $P(Y < -3)$; dados $\mu = 2, \sigma^2 = 4$.

3.29 Suponiendo una distribución ji cuadrada, encuentre:

(a) $P(\chi^2 \geq 26.1)$ para $\nu = 14$
(b) $P(\chi^2 < 26.1)$ para $\nu = 15$
(c) $P(23.3 < \chi^2 \leq 36.4)$ para $\nu = 24$
(d) $P(\chi^2 \leq 0.70)$ para $\nu = 6$.

3.30 Suponiendo una distribución t, encuentre:

(a) $P(|t| > 2.015)$ para $\nu = 5$
(b) $P(t > 2.015)$ para $\nu = 5$
(c) $P(-1.341 < t < 2.131)$ para $\nu = 15$
(d) $P(t \leq 1.3)$ para $\nu = 20$.

3.31 Suponiendo una distribución F, encuentre:

(a) $P(F > 7.79)$ para $\nu_1 = 11, \nu_2 = 6$
(b) $P(F > 4.03)$ para $\nu_1 = 11, \nu_2 = 6$
(c) $P(F > 7.79)$ para $\nu_1 = 6, \nu_2 = 11$
(d) $P(0.221 < F \leq 2.62)$ para $\nu_1 = 3, \nu_2 = 24$.

3.32 El diámetro de acabado de un cable eléctrico armado está distribuido normalmente con una media de 19.558 mm y una desviación de 0.254 mm. ¿Cuál es la probabilidad de que el diámetro del cable exceda 20.193 mm. Si las especificaciones de la ingeniería son 19.812 ± 0.508 mm. ¿Cuál es la probabilidad de una pieza de cable defectuosa?

3.33 Si la f.d.p. para la vida de un cierto tipo de componente es $f(x) = (1/100) \exp(-x/100)$ para $x > 0$. ¿Cuál es la probabilidad de que una componente elegida al azar dure 400 horas? ¿de que dure 400 horas si ya ha sobrevivido 200 horas? Si un ensamble usa tres de esas componentes en serie, ¿cuál es la probabilidad de que un ensamble de tres piezas elegidas al azar no falle a causa de la falla de las componentes?

3.34 La relación al azar es definida como $f(x)/[1 - F(x)]$. Si $f(x) = (1/8) \exp(x/8)$ para $x > 0$. ¿Qué es la relación al azar?

Referencias y otras obras de consulta

1. Anderson, R. L., and Bancroft, T. A. *Statistical Theory in Research*. McGraw-Hill Book Company, Inc., New York, 1952.

se seleccionaron valores de d_1 y d_2.

Problemas

5.1 ¿Qué tan grande deberá tomarse una muestra si queremos un 95 por ciento de seguridad de que $\bar{X}$ no se alejará de μ más de $\sigma/2$?

5.2 Un libro de 400 páginas contiene 400 errores de impresión. Calcule la probabilidad de que una página contenga cuando menos tres errores.

5.3 Un lote contiene 1,000 artículos y de él se saca una muestra de 400. Si en la muestra no aparecen más de 2 elementos defectuosos, se acepta el lote. Determine la probabilidad de que el lote sea aceptado, suponiendo que es 1 por ciento defectuoso.

5.4 El ancho de la ranura en un forjado se distribuye normalmente con una media de 2.286 cm y desviación normal de 0.00762 cm. Las especificaciones al respecto son 2.286 ± 0.0127 cm. ¿Qué porcentaje de forjados es defectuoso?

5.5 Respecto al problema 5.4, diariamente se obtienen muestras de tamaño 5 y se calculan sus medias. ¿Qué porcentaje de esos promedios de muestra estarán fuera de especificación?

5.6 Los diámetros de algunas flechas y los diámetros de algunas chumaceras, cada uno de los cuales, se distribuye normalmente con una desviación estándar igual a 0.00254 cm. Si la flecha tiene un diámetro medio de 1.27 cm y la chumacera tiene un diámetro medio de 1.2776 cm, ¿cuál es la probabilidad de interferencia?

5.7 Tres resistencias se conectan en serie. Los valores nominales de dichas resistencias son 10, 15 y 20 ohmios, respectivamente. ¿Si se sabe que las resistencias están normalmente distribuidas con respecto a los valores nominales, y que cada resistencia tiene una desviación estándar de 0.5 ohmios, cuál es la probabilidad de que una conexión tenga una resistencia mayor de 46.5 ohmios?

5.8 Resuelva el problema 5.7, suponiendo que la desviación, en cada caso, es el 5 por ciento del valor nominal respectivo.

5.9 Una caja de ½ kg de dulces se empaca a máquina, y contiene 32 piezas de dulce. Si los pesos de los dulces están distribuidos normalmente con una media de 15.675 g y una desviación estándar de 1.5625 g, ¿Cuáles son las probabilidades de que un cliente reciba:
a) menos de ½ kg, b) menos de 468.75 g, c) más de ½ kg, d) más de 508.25 g, e) exactamente ½ kg?

5.10 Con respecto al problema 5.9 y suponiendo que la desviación estándar permanece constante, ¿cómo cambiaría la media del proceso para que solamente 1 de 100 clientes reciba menos del peso estipulado?

5.11 Una fábrica ensambla estufas a razón de 500 por día. Al hacer la inspección después del ensamblado final, se encuentra que el promedio de estufas defectuosas es de 3 por ciento. ¿Cuál es la probabilidad de que la producción de la próxima semana contenga menos de 20 estufas defectuosas?

5.12 Revise la parte del texto que considera las distribuciones de Poisson, estándar, ji-cuadrada, t y F, y cerciórese de que usted puede utilizar las tablas de los apéndices 2 a 8.

Referencias y otras obras de consulta

1. Burr, I. W. *Engineering Statistics and Quality Control*. McGraw-Hill Book Company, Inc., New York, 1953.
2. Lieberman, G. J., and Owen, D. B. *Tables of the Hypergeometric Distribution*. Stanford University Press, Stanford, Calif., 1961.
3. Mood, A. M. *Introduction to the Theory of Statistics*. McGraw-Hill Book Company, Inc., New York, 1950.
4. National Bureau of Standards. *Tables of the Binomial Probability Distribution*. Applied Mathematics Series 6, U.S. Govt. Print. Off., Washington, D.C., 1949.
5. Pearson, E. S. The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, 24:404-17, Nov., 1932.
6. ———, and Hartley, H. O. The probability integral of the range in samples of *n* observations from a normal population. *Biometrika*, 32:301-10, April, 1942.
7. Robertson, W. H. *Tables of the Binomial Distribution Function for Small Values of p*. Sandia Corporation Monograph SCR-143, Albuquerque, N. Mex., Jan., 1960.
8. Romig, H. G. *50-100 Binomial Tables*. John Wiley and Sons, Inc., New York, 1953.

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

PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES

A N E X O S

FEBRERO, 1984

Palacio de Minería Calle de Tacuba 5 primer piso Deleg. Cuauhtemoc 06000 México, D.F. Tel.: 521-40-20 Apdo. Postal M-2285

PROBABILIDAD Y ESTADISTICA PARA INGENIEROS - I. MILLER, J. FREUND

CAP. II . EDITORIAL REVERTE

4. Utilizando las 20 lecturas de temperatura dadas en el problema 4 de la página 114, estimar la media de la población a partir de la cuartila media.
5. Empleando la distribución de pesos de recubrimientos de estaño obtenidos en el ejercicio 3 de la página 106, determinar la mediana y la cuartila media y comparar con la media obtenida en el problema 10 de la página 114.
6. Usar la distribución de velocidades obtenida en el problema 10 de la página 115 para calcular la cuartila media y estimar μ. Comparar este resultado con $\bar{x}$ y con una estimación de μ obtenida dibujando la gráfica en papel de probabilidades.
7. Utilizar los dos estimadores de cuantilas de la página 197 para dar una estimación de σ para los ingresos no agrupados del problema 6 de la página 105. Calcular la desviación típica muestral y comparar las estimaciones. Estimar, también, σ tomando un cuarto de la diferencia entre la media del 5% superior y el 5% inferior de los datos.
8. Con respecto a las lecturas de temperatura del ejercicio 4 de la página 114, estimar la desviación típica de la población utilizando (a) la desviación típica muestral, (b) las dos fórmulas dadas en la página 197, y (c) una gráfica de probabilidad. Comparar los resultados obtenidos.
9. Emplear la distribución de pesos de recubrimiento de estaño obtenida en el problema 3 de la página 106 para calcular los dos estimadores de cuantilas de σ dados en la página 197. También, comparar con el valor de la desviación típica de la muestra obtenida en el problema 11 de la página 114.
10. Comparar el valor estimado de σ obtenido de la gráfica de probabilidad del ejercicio 6 con los valores estimados basados en las dos fórmulas de la página 197. Calcular las cuantilas necesarias a partir de los datos agrupados.

11.3 Tests de los signos

En esta sección describiremos tests no paramétricos basados en clasificar los datos de acuerdo con dos tributos, representados convenientemente por *signos más* y *signos menos*. Por ejemplo, si queremos contrastar la hipótesis nula $H_0: \mu = \mu_0$ sobre la base de una muestra de azar de tamaño n, podemos cambiar cada observación que exceda de μ_0 por un signo más y, cada observación menor que μ_0, por un signo menos. Si la población de la que obtenemos las muestras es *continua* y *simétrica* la probabilidad de que una observación sea cambiada por un signo más es igual a 1/2 cuando H_0 es cierta. En consecuencia, la prueba de la hipótesis nula $\mu = \mu_0$ es equivalente a una prueba de la hipótesis nula $p = 1/2$, donde p es el parámetro de una distribución binómica. La alternativa bilatera $\mu \neq \mu_0$ es equivalente, ahora, a $p \neq 1/2$, y las alternativas unilaterales $\mu < \mu_0$ y $\mu > \mu_0$ son equivalentes a $p < 1/2$ y $p > 1/2$, respectivamente, siendo p la probabilidad de encontrar un signo más, es decir, una observación mayor que μ_0.

Para ilustrar el *test de los signos de una sola muestra* que acabamos de describir, vamos a contrastar la hipótesis de que la temperatura media a la que opera un termostato es 28° C., utilizando los resultados siguientes obtenidos de 20 termostatos:

29.9	29.2	32.0	30.5	29.3	30.1	27.7	31.4	28.6	27.9
+	+	+	+	+	+	−	+	+	−
26.8	30.3	29.0	28.8	28.0	31.4	32.1	27.5	31.7	29.2
−	+	+	+		+	+	−	+	+

Notemos que hay 15 observaciones mayores que 28.0, 4 observaciones menores que 28.0 y una observación igual a 28.0. Aunque la probabilidad de que una observación, en una población continua, sea exactamente igual a 28.0, es nula, los números anteriores están redondeados y, como no sabemos si el número 28.0 representa un valor mayor o menor que 28, descartamos esta observación. Así pues, debemos determinar si los 15 signos más y los 4 signos menos, o sea, 15 "casos favorables" en 19 pruebas, comprueban la hipótesis de que $p = 1/2$. Aplicando el criterio del test exacto dado en la página 180 con $\alpha = 0.05$, encontramos en la tabla de probabilidades binómicas que $k'_{0.025} = 4$ y $k'_{0.975} = 15$; como hubo 15 signos más y 4 signos menos, se deduce que la hipótesis nula debe ser rechaza. Nótese que, si el tamaño de la muestra es suficientemente grande, podemos usar la curva normal como aproximación de la distribución binómica y los tests dados en la tabla de la página 164.

Utilizando muestras apareadas, podemos extender inmediatamente el test de los signos a test de diferencias entre dos medias de población. En este caso, el test de los signos se puede emplear como una alternativa no paramétrica del test t para muestras apareadas introducida en la página 157. Dadas n observaciones apareadas, con la primera observación proviniente de la población uno y la segunda de la población dos, usamos un *signo más* para reemplazar cada par para el cual la observación de la primera población excede a la de la segunda, y un *signo menos* para substituir cada par en el que la observación de la segunda población excede al de la primera. En el caso en que dos observaciones pareadas sean iguales, se omite esta pareja y la prueba se desarrolla como en el caso de una sola muestra antes descrita.

Para ilustrar el *test de los signos de muestras apareadas*, vamos a comparar dos métodos para anodizar aluminio atendiendo a la apariencia de las piezas anodizadas (brillo, color, etc.). Aunque resulta difícil asignar valores numéricos a estas cualidades, no es difícil comparar piezas apareadas y decidir cuál tiene el aspecto más agradable. Supongamos que 40 unidades apareadas se juzgan, dando un signo más o un signo menos a cada par, de acuerdo con que el anodizado del primer método o el del segundo se considere superior. Dado que hubo 24 signos más, 11 signos menos y 5 empates, queremos probar si el primer método es realmente superior. Empleando la curva normal (ver página 180) y $\alpha = 0.05$ calculamos primero

$$z = \frac{24 - 17.5}{\sqrt{35 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = 2.20$$

y como este valor excede a 1.645, valor crítico para un test unilatero, con un nivel de significación de 0.05, concluimos que la hipótesis nula debe ser rechazada. En otras palabras, concluimos que el primer método de anodización es mejor.

La eficacia del test de los signos es bastante alta para muestras pequeñas, 95% para $n = 6$, pero disminuye a medida que el tamaño de la muestra aumenta hasta llegar a una eficacia límite de 63%. Las hipótesis necesarias para aplicar el test de los signos, lo mismo en el caso de una sola muestra que en el de muestras apareadas, son que las poblaciones consideradas sean continuas y simétricas. Si la población no fuese continua, podría haber una probabilidad positiva de que una observación fuera realmente igual a μ_0 en el caso de una sola muestra, o que las observaciones apareadas fueran exactamente iguales en el caso de muestras apareadas. Entonces, dejará de ser válida la hipótesis de ser $p = 1/2$, a menos que impongamos mayores restricciones. Si la población (o poblaciones) no fueran simétricas, la probabilidad de que una observación fuera mayor que la media, o de que la diferencia entre dos observaciones apareadas fuera mayor que cero, no iguala necesariamente un medio, bajo la hipótesis nula $\mu = \mu_0$ en el caso de una sola muestra, o $\mu_1 = \mu_2$ en el caso de muestras apareadas. Sin embargo, es posible modificar el test de los signos para eliminar la hipótesis de simetría. Para concluir esto, sólo tenemos que considerar las hipótesis concernientes a las medianas de la población en lugar de las concernientes a las medias.

11.4 Tests por suma de números de orden

El test de los signos de muestras apareadas es uno de los métodos no paramétricos para contrastar la hipótesis nula de que dos muestras provienen de poblaciones continuas idénticas, frente a la alternativa de que las poblaciones tienen medias diferentes. Una clase altamente eficaz de tests no paramétricos de esta hipótesis, y otras similares, se basa en la *suma de los números de orden;* esto es, se dan números de orden a las observaciones de acuerdo con su magnitud y los tests se realizan sobre la base de ciertas sumas de estos número de orden. En esta sección, introduciremos tres tests basados en sumas de números de orden. *El test U de Mann-Whitney* se presenta como un substituto del test t de dos muestras, y tiene una eficacia límite de 95.5% cuando las hipótesis necesarias para el test t correspondiente quedan satisfechas. Un test similar al test U, que se puede emplear cuando la hipótesis alternativa especifica que las dos poblaciones tienen *dispersiones* diferentes, se considerará a continuación. Finalmente, introduciremos el *test H de Kruskal-Wallis H* para contrastar si k muestras provienen de poblaciones idénticas, frente a la alternativa de que las poblaciones tengan medias diferentes. Como el test U, el test H tiene, también, una eficacia de 95.5% cuando se compara con el procedimiento "normal" correspondiente, que se analizará en el capítulo 13.

Vamos a describir, en primer lugar, el test U de Mann-Whitney por medio del ejemplo siguiente. Supongamos que queremos comparar dos instrumentos de control o "registradores" diferentes utilizados para la determinación del contenido de humedad dentro de un semiconductor, a partir de las siguientes corrientes medidas en microamperes:

Registrador A:	1.2	0.9	0.8	0.2	0.4	0.6	0.1	5.1	0.2	
Registrador B:	1.7	3.5	7.8	0.9	0.7	2.6	0.3	1.5	15.3	0.7

Primero, ordenamos *conjuntamente* las 19 observaciones de acuerdo con su tamaño, reteniendo la identidad de la muestra en cada observación. Después, asignamos a estas observaciones los puestos 1, 2, 3, ... y 19, como se indica en la tabla siguiente:

Registrador:	A	A	A	B	A	A	B	B	A	A	B	A
Observación:	0.1	0.2	0.2	0.3	0.4	0.6	0.7	0.7	0.8	0.9	0.9	1.2
Rango:	1	2	3	4	5	6	7.5	7.5	9	10.5	10.5	12

Registrador:	B	B	B	B	A	B	B
Observación:	1.5	1.7	2.6	3.5	5.1	7.8	15.3
Rango:	13	14	15	16	17	18	19

Nótese que, si dos o más observaciones están empatadas en el mismo lugar, damos a cada una de ellas el medio de los lugares que ocupan en conjunto.

Si denotamos los tamaños respectivos de las muestras por n_1 y n_2 y la suma de los números de orden ocupados por la primera muestra por R_1, se puede demostrar que la media y la varianza de la distribución muestral del estadístico

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

están dadas por

$$\mu_U = \frac{n_1 n_2}{2} \quad \text{y} \quad \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

(Si hay empates en la ordenación, estas fórmulas son correctas sólo aproximadamente). Si n_1 y n_2 son, ambas, mayores que 8, la distribución del estadístico U se puede aproximar con bastante exactitud por una distribución normal y, por lo tanto, el test se puede basar en el estadístico

$$z = \frac{U - \mu_U}{\sigma_U}$$

y en la tabla III. Existen, también, tablas en las que se pueden basar tests exactos cuando n_1 y n_2 son pequeños (ver la tabla de D. B. Owen mencionada en la bibliografía). Observemos que no tiene consecuencias saber qué muestra se considera "primera", por lo que podemos trabajar con cualquier suma de números de orden, escogiendo la que sea más fácil de obtener.

Volviendo ahora a nuestro ejemplo, tenemos $n_1 = 9$, $n_2 = 10$, $R_1 = 66.5$, y, por consiguiente,

$$U = 9 \cdot 10 + \frac{9 \cdot 10}{2} - 66.5 = 68.5$$

$$\mu_U = \frac{9 \cdot 10}{2} = 45.0$$

$$\sigma_U^2 = \frac{9 \cdot 10 \cdot 20}{12} = 150.0$$

Así,

$$z = \frac{68.5 - 45.0}{\sqrt{150.0}} = 1.92$$

y como este valor se encuentra entre -1.96 y 1.96, valores críticos de una alternativa bilatera con $\alpha = 0.05$, llegamos a la conclusión de que la hipótesis nula de poblaciones idénticas no se puede rechazar.

Si se asignan los números de orden de alguna otra forma diferente, se puede emplear el mismo estadístico U para contrastar la hipótesis nula de poblaciones idénticas frente a la alternativa de que las poblaciones tienen *dispersiones distintas*. Los números de orden se asignan "desde ambos extremos hacia el medio", dando el número 1 a la menor observación; los números 2 y 3, a la mayor y segunda mayor observaciones; los números 4 y 5, a la segunda y tercera menores; los 6 y 7, a la tercera y cuarta mayores, y así sucesivamente. Todos los demás aspectos de este test con dispersiones diferentes son idénticos a los del test U de Mann-Whitney.

El test Kruskal-Wallis para decidir si k muestras independientes provienen de poblaciones idénticas, se desarrolla en una forma similar al test U. Como antes, las observaciones se tratan *en conjunto* para darles el lugar de orden, y si R_i es la suma de los números de orden ocupados por las n_i observaciones de la i-ésima muestra, el test se basa en el estadístico

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

donde $n = n_1 + n_2 + \ldots + n_k$. Cuando $n_i > 5$ para todas las i y la hipótesis nula es válida, la distribución del estadístico H se puede aproximar bastante por la distribución X-cuadrado, con $k-1$ grados de libertad. En la tabla de D. B. Owen, mencionada en la bibliografía, se encuentran tablas especiales para aplicar con valores pequeños seleccionados de las n_i y k.

Para ilustrar la prueba de Kruskal-Wallis H, supondremos que el experimento descrito en la página 200, se amplía para incluir cuatro registradores diferentes, con los resultados mostrados en la tabla siguiente. (Nótese que a las observaciones empatadas se les asignan nuevamente, el medio de los puestos que ocupan en conjunto.)

Registrador A:	0.2	0.3	0.4	0.6	1.7	1.9	2.0
Registrador B:	0.8	1.1	1.3	1.9	2.5	7.8	
Registrador C:	0.7	0.9	8.3	12.0	12.1	15.3	
Registrador D:	0.1	0.1	0.3	0.6	2.9	13.6	

Las observaciones son, otra vez, corrientes de retorno en microamperes. Como se puede verificar fácilmente, las observaciones de la primera muestra ocupan los puestos 3, 4.5, 6, 7.5, 14, 15.5 y 17, por lo que $R_1 = 67.5$. Similarmente, las observaciones de la segunda muestra ocupan los puestos 10, 12, 13, 15.5, 18 y 20, por lo que $R_2 = 88.5$; las observaciones de la tercera muestra ocupan los puestos 9, 11, 21, 22, 23 y 25, por lo que $R_3 = 111.0$; y las observaciones de la cuarta muestra ocupan los puestos 1.5, 1.5, 4.5, 7.5, 19 y 24, por lo que $R_4 = 58.0$. Substituyendo en la fórmula de H, encontramos

$$H = \frac{12}{25 \cdot 26}\left(\frac{67.5^2}{7} + \frac{88.5^2}{6} + \frac{111.0^2}{6} + \frac{58.0^2}{6}\right) - 3(26) = 0.4$$

y si comparamos este valor con 7.815, valor de $\chi^2_{.05}$ con 3 grados de libertad, vemos que no se puede rechazar la hipótesis nula. En otras palabras, no podemos rechazar la hipótesis nula de que las muestras provienen de poblaciones idénticas frente a la alternativa de que las medias de las poblaciones no son iguales.

EJERCICIOS

1. Con respecto a las 100 medidas de los pesos de planchas de estaño de galvanizado electrolítico del problema 3 de la página 106, utilizar el test de los signos con $\alpha = 0.05$ para contrastar la hipótesis nula $\mu = 0.33$ frente a la alternativa $\mu < 0.33$, donde μ es la media de la población de pesos de la que se obtuvo la muestra.
2. Empleando el test de los signos de una sola muestra, contrastar la hipótesis nula de que el octanaje medio de la gasolina de la que se tomaron las 16 muestras siguientes es 100, frente a la hipótesis alternativa de que es mayor de cien.

101.8	98.2	104.5	99.0	102.8	105.4	107.7	99.4
100.3	100.0	102.5	97.1	103.6	101.0	98.7	101.0

Emplear un nivel de significación de 0.05.

3. Utilizar el test de los signos y un nivel de significación de 0.10 para decidir si hay una diferencia sistemática entre las lecturas obtenidas de los dos instrumentos del ejercicio 19 de la página 161.
4. Se sierran vigas de acero por dos métodos, consistente el primero en aserrarlas cuando aún están calientes, y el segundo cuando se han enfriado. Las longitudes finales resultantes (en pies) una vez que todas las vigas se han enfriado hasta la temperatura ambiente, son las siguientes:

Vigas serradas en caliente:	31.8	30.5	31.1	29.7	27.9	30.2
	30.8	31.8	32.6	28.8	29.6	28.5
	28.8	29.9	31.6	30.7	30.3	31.5
Vigas serradas en frío:	30.1	31.0	29.9	29.8	30.0	30.8
	30.8	30.2	31.1	29.4	29.7	29.0
	31.3	30.6	30.1	30.0	30.8	30.3

Apareando al azar las 18 observaciones de las dos muestras, usar el test de los signos de dos muestras, con un nivel de significación de 0.05, para determinar si hay alguna diferencia significante en las longitudes finales medias.

5. En pruebas repetidas, un motor experimental operó, respectivamente, durante 20, 19, 22, 17, 18, 20, 23, 19, 20, 15, 24, 21, 18, 20, 24, 23, 20, 17, 25 y 28 minutos, con un galón de cierta clase de combustible. Empleando el test de los signos y un nivel de significación de 0.01 contrastar la hipótesis nula $\mu = 20$ frente a la alternativa $\mu \neq 20$.
6. Para comparar una bebida con la de una marca de la competencia, 50 personas probaron una bebida y luego la otra y después se les pidió que indicaran su preferida. (El orden de las marcas se escogió al azar para cada persona.) Si 27 prefirieron la marca dada, 18 prefirieron la de la competencia y 5 no encontraron diferencia en el sabor, contrastar, con un nivel de significación de 0.01, si la marca dada es superior en sabor a la de la competencia.
7. Un experimento para comparar la resistencia a la tensión de dos clases de hilos dio los resultados siguientes (en libras):

Hilo A:	143.8	144.8	145.2	144.8	145.8	146.0
	143.0	147.4	144.0	145.6	145.5	144.8
Hilo B:	146.6	147.8	144.4	140.8	143.0	146.6
	153.0	142.4	146.8	143.2	140.9	150.6

Usar el test U y un nivel de significación de 0.05 para contrastar la hipótesis nula de que las dos muestras provienen de poblaciones idénticas, frente a la alternativa de que las dos poblaciones tienen medias diferentes.

8. Repetir el ejercicio 4, usando el test U de Mann-Whitney. También contrastar si las dispersiones de las dos poblaciones son iguales, utilizando el test de suma de números de orden mencionado en la página 211.
9. En el ejercicio 7 utilizar $\alpha = 0.05$ para contrastar la hipótesis nula de que las muestras provienen de poblaciones idénticas, frente a la alternativa de que las dos poblaciones tienen dispersiones diferentes.
10. Los tests denominados de Franklin se establecieron para determinar las propiedades de aislamiento de acero al silicio de granos orientados que fueron recocidos en cinco atmósferas diferentes, con los resultados siguientes:

Atmósfera	*Resultados del ensayo (amperios)*							
1	0.58	0.61	0.69	0.79	0.61	0.89		
2	0.37	0.27	0.58	0.40	0.28	0.44	0.35	
3	0.29	0.19	0.34	0.17	0.70	0.18		
4	0.81	0.89	0.75	0.72	0.68	0.85	0.57	0.77
5	0.26	0.34	0.29	0.47	0.30	0.42		

Emplear el test H de Kruskal-Wallis H y un nivel de significación de 0.05 para decidir si se puede aceptar que estas cinco muestras proceden de poblaciones idénticas.

11. Para investigar tres medidas preventivas contra la corrosión, se probaron muestras al azar de 10 piezas de alambre para cada una de las tres medidas preventivas, dando los siguientes resultados para las profundidades máximas de las partes erosionadas (en milésimas de pulgada):

A:	45	53	40	46	57	62	49	53	53	52
B:	62	58	47	59	63	48	66	52	60	49
C:	47	45	60	64	57	63	48	59	62	60

Contrastar, con un nivel de significación de 0.05, si hay alguna diferencia en la eficacia de las tres medidas preventivas contra la corrosión.

11.5 Tests de las series de términos iguales

Al discutir las muestras aleatorias en el capítulo 7, presentamos varios métodos que daban por anticipado cierta seguridad de que una muestra fuera de azar. Sin embargo, es útil tener una técnica para contrastar si una muestra se puede considerar como aleatoria *después de haber sido obtenida*. Una de estas técnicas está basada en el orden en que se obtuvieron los valores de las muestras; más concisamente, se basa en el número de series de términos iguales mostradas en los resultados de las muestras.

Dada una sucesión de dos símbolos, tales como H y T (que pueden representar, por ejemplo, las caras y las cruces en tiradas sucesivas), una "serie de iguales" es una sucesión de símbolos idénticos comprendidos entre dos símbolos diferentes o sin estos últimos. Es decir, la serie

$$\underline{T\ T}\ \underline{H\ H}\ \underline{T\ T}\ \underline{H\ H\ H}\ \underline{T}\ \underline{H\ H\ H}\ \underline{T\ T\ T\ T}\ \underline{H\ H\ H}$$

contiene 8 series de iguales, como indican los subrayados. El número total de series de iguales en una sucesión de n ensayos da una indicación de si la sucesión se puede considerar como de azar. Entonces, si sólo ha habido dos series de iguales, consistentes en diez caras, seguidas por diez cruces, se puede suponer que la probabilidad de un caso favorable no ha permanecido constante de una prueba a la siguiente. Por otra parte, si la sucesión consta de veinte tiradas formadas alternativamente por caras y cruces, se puede suponer que los ensayos no han sido independientes. En cualquier caso, hay razones para suponer que no se trata de un azar. Notemos que nuestra suposición no procede del *número* de caras y cruces, si no del *orden* en que aparecen.

Si una sucesión contiene n_1 símbolos de una clase y n_2 de otra (y ni n_1 ni n_2 son muy pequeños), la distribución muestral del número total de series de iguales se puede representar muy aproximadamente por una distribución normal de media

$$\mu_u = \frac{2n_1n_2}{n_1 + n_2} + 1$$

y desviación típica

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

donde u denota el número total de series de iguales. Entonces, el test de la hipótesis nula de que la ordenación de los símbolos (y, por consiguiente, de la muestra) sea aleatoria, se puede basar en el estadístico

$$z = \frac{u - \mu_u}{\sigma_u}$$

y en la tabla III. Este test da una excelente aproximación cuando ni n_1 ni n_2 son menores de 10. Se pueden encontrar tablas especiales para hacer tests exactos cuando n_1 o n_2 son pequeñas, en las tablas de D. B. Owen, citadas en la bibliografía.

Para ilustrar este test, examinaremos la siguiente sucesión de 32 vuelos de prueba de un cohete, donde S y F marcan, sucesivamente, los éxitos y los fallos:

$$\underline{F\ F\ F}\ \underline{S\ S}\ \underline{F\ F}\ \underline{S\ S\ S}\ \underline{F}\ \underline{S}\ \underline{F}\ \underline{S\ S\ S\ S}\ \underline{F}\ \underline{S\ S\ S}\ \underline{F}\ \underline{S\ S\ S\ S\ S}\ \underline{F}\ \underline{S\ S\ S\ S}$$

Como hay 22 éxitos, 10 fallos y 14 series de iguales, substituimos $n_1 = 22$, $n_2 = 10$, $u = 14$, y obtenemos

$$\mu_u = \frac{2 \cdot 22 \cdot 10}{32} + 1 = 14.75$$

$$\sigma_u = \sqrt{\frac{2 \cdot 22 \cdot 10(2 \cdot 22 \cdot 10 - 22 - 10)}{(22 + 10)^2(22 + 10 - 1)}} = 2.38$$

y
$$z = \frac{14 - 14.75}{2.38} = -0.31$$

Como este valor queda entre -1.96 y 1.96, no podemos rechazar (con un nivel de significación de 0.05) la hipótesis nula de que la ordenación sea aleatoria. Evidentemente, no hay razones suficientes para concluir que hay una fiabilidad real.

Se puede emplear también el test de series de iguales para contrastar la casualidad que hay en las muestras formadas por datos numéricos, contando las series de iguales a partir de la mediana *por encima y por debajo de ésta*. Si denotamos una observación mayor que la mediana de la muestra por la letra a y una observación menor que la mediana por la letra b, podemos utilizar la sucesión resultante de a y b para contrastar la casualidad, siguiendo el método indicado antes. Una aplicación frecuente de este método es en el control de calidad, donde las medias de muestras pequeñas sucesivas se representan en una gráfica en orden cronológico. El test de series de iguales se puede usar entonces para comprobar si hay alguna tendencia en los datos, que nos indique que es necesario ajustar una máquina o hacer algún otro proceso antes de que ocurra algún daño grave.

Para ilustrar un test de series de iguales por encima y por debajo de la mediana, supongamos que un ingeniero se interesa en la posibilidad de que se hayan hecho demasiados cambios en el ajuste de un torno automático. Para contrastar esta hipótesis, se obtuvieron los siguientes diámetros medios (en pulgadas) de 40 ejes torneados sucesivamente en el torno:

0.261	0.258	0.249	0.251	0.247	0.250	0.250	0.247	0.253	0.249
0.252	0.250	0.253	0.247	0.251	0.243	0.253	0.251	0.245	0.250
0.248	0.252	0.254	0.250	0.247	0.253	0.251	0.248	0.249	0.252
0.247	0.250	0.253	0.247	0.249	0.253	0.248	0.251	0.249	0.253

La mediana de estas medidas es 0.250 y, cambiando cada una de ellas por una letra a si excede de 0.250, por una b si es menor que 0.250, y omitiendo las cinco que son iguales a 0.250, obtenemos la sucesión

aababababaababaabbaabaabbababbababa

que tiene 27 series de iguales. Entonces, $n_1 = 19$, $n_2 = 16$, $u = 27$, tendremos:

$$\mu_u = \frac{2 \cdot 19 \cdot 16}{35} + 1 = 18.37$$

$$\sigma_u = \sqrt{\frac{2 \cdot 19 \cdot 16(2 \cdot 19 \cdot 16 - 19 - 16)}{(19 + 16)^2(19 + 16 - 1)}} = 2.89$$

y
$$z = \frac{27 - 18.37}{2.89} = 2.98$$

Como este valor excede a 1.96, podemos rechazar la hipótesis nula de que la sucesión de medidas sea aleatoria. Como el número de series de iguales es mayor que el que se podría esperar, debido al azar, es razonable suponer que el torno se ha ajustado demasiado; es probable que se haya hecho un ajuste después de tornear cada pieza, tratando, con ello, de compensar cualquier discrepancia que se haya observado con respecto al diámetro nominal de 0.250 pulgadas.

11.6 Tests de Kolmogorov-Smirnov

Los tests de Kolmogorov-Smirnov son test no paramétricos para diferencias entre dos distribuciones totales o acumulativas. El *test uni-muestral* se refiere a la concordancia entre una distribución acumulativa observada de valores de una muestra y una función de distribución continua especificada; es decir, se trata de una prueba de bondad de ajuste. El *test bi-muestral* se refiere a la concordancia entre dos distribuciones acumulativas observadas; se contrasta la hipótesis de si dos muestras independientes provienen de distribuciones continuas idénticas, y es sensible a las diferencias de población en lo que se refiere a la localización, dispersión, o disimetría.

El test uni-muestral de Kolmogorov-Smirnov es, en general, más eficaz que el test X-cuadrado para la bondad de ajuste de muestras pequeñas y puede usarse con muestras muy pequeñas en las que el test X-cuadrado no es aplicable. Debemos recordar, sin embargo, que el test X-cuadrado de la sección 10.5 se puede usar con distribuciones discretas, mientras que el test de Kolmogorov-Smirnov no puede usarse.

El test uni-muestral se basa en la diferencia absoluta máxima D entre los valores de la distribución acumulativa de una muestra aleatoria de tamaño n y una distribución teórica especificada. Como ilustración de este test, se quiere comprobar si los agujeros para clavijas en una placa de hojalata electrolítica están distribuidos uniformemente para lo cual se han tomado medidas de las siguientes distancias (en pulgadas) de 10 agujeros a partir de un extremo de una tira grande de placa de hojalata de 30 pulgadas de ancho:

4.8 14.8 28.2 23.1 4.4 28.7 19.5 2.4 25.0 6.2

Bajo la hipótesis nula de que los agujeros están uniformemente repartidos, la distribución acumulativa teórica con la que queremos comparar la distribución acumulativa observada está dada por

$$F(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ x/30 & \text{para } 0 < x < 30 \\ 1 & \text{para } x \geq 30 \end{cases}$$

La gráfica de esta distribución acumulativa teórica se muestra, junto con la de la distribución acumulativa observada, en la figura 11.1. Como se indica en esta figura, la diferencia máxima entre las dos distribuciones acumulativas es 0.193.

Para determinar si esta diferencia es mayor que lo que se puede esperar razonablemente, encontramos el valor crítico de D en la tabla IX. Para $n = 10$ $\alpha = 0.05$, el valor crítico es $D_{.05} = 0.410$, y de aquí que la hipótesis nula (que los agujeros están uniformemente distribuidos) no se puede rechazar.

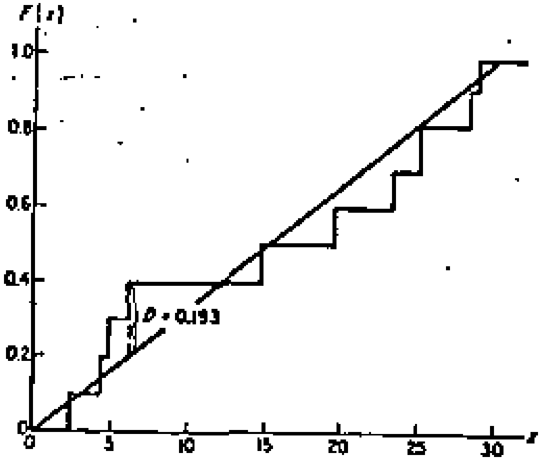


Fig. 11.1 Prueba Kolmogorov-Smirnov

El test bimuestral de Kolmogorov-Smirnov se basa en la diferencia absoluta máxima entre los valores de las dos distribuciones acumulativas observadas. En principio, es muy similar al test uni-muestral, y los valores críticos necesarios se pueden obtener de tablas especiales (por ejemplo, las de D. B. Owen, citada en la bibliografía).

EJERCICIOS

1. Para comprobar si cierta señal de radio contiene un mensaje, se puede subdividir un intervalo de tiempo en cierto número de intervalos muy cortos y determinar después si la fuerza de la señal excede cierto nivel (ruido de fondo) en cada corto intervalo. Supongamos que lo siguiente es parte de una observación de este tipo, donde H indica una señal fuerte y L que la señal no excede cierto nivel de ruido.

   $L\ L\ H\ L\ H\ L\ H\ L\ H\ H\ H\ L\ H\ H\ H\ L\ H\ H\ H\ L\ H\ L\ H\ L\ H\ L\ L$
   $L\ L\ H\ L\ H\ L\ H\ L\ H\ H\ H\ L\ H\ H\ H\ L\ H\ H\ H\ L\ H\ L\ H\ L\ H\ L\ L$

   Verificar si esta sucesión se debe al azar (usando un nivel de significación de 0.05) y comprobar si es razonable suponer que la señal contiene un mensaje.
2. La siguiente es una lista que nos da, leyendo las filas sucesivas de izquierda a derecha, el orden en que una máquina produjo piezas defectuosas (D) y no defectuosas (N) durante cierto período:

   $N\ N\ N\ N\ N\ N\ N\ N\ N\ D\ D\ D\ N\ N\ N\ N\ N\ N\ N\ N$
   $N\ N\ N\ D\ N\ N\ N\ N\ N\ N\ N\ N\ D\ D\ N\ N\ D\ D\ N\ N$
   $N\ N\ N\ N\ N\ N\ N\ N\ N\ N\ N\ N\ N\ D\ D\ D\ N\ N\ D\ N$
   $D\ N\ N\ D\ N\ N\ N\ N\ N\ N\ N\ N\ D\ D\ N\ N\ N\ D\ N\ N$
   $N\ N\ N\ N\ N\ N\ N\ N\ D\ D\ D\ N\ N\ N\ N\ N\ N\ N\ N\ N$

   Comprobar, con un nivel de significación de 0.05, si este arreglo es debido al azar.
3. En la página 92 se describió un método para generar dígitos al seudoazar, en el cual un número de 4 dígitos se elevaba al cuadrado, los 4 números de la parte media del número resultante se elevaban al cuadrado, y así sucesivamente. Comenzando con el número 3571, usar una tabla de cuadrados o una calculadora para continuar este proceso hasta obtener una sucesión de 48 dígitos. Contrastar si la sucesión resultante, para ver si es de azar, utilizando series de términos iguales por encima y por debajo de la mediana y un nivel de significación de 0.05.
4. En una fábrica, el tiempo que no trabaja una máquina durante las horas de trabajo, debido a dificultades tales como roturas o fallos, se llama "tiempo muerto". La tabla que sigue corresponde a 50 tiempos muertos (en minutos) consecutivos observados por un ingeniero de control durante cierto período (léanse las filas sucesivas de izquierda a derecha):

18	25	28	31	29	30	34	30	25	31
22	29	30	34	35	37	30	37	30	25
30	21	28	33	36	35	31	30	28	39
40	30	38	34	35	38	42	30	35	41
38	35	40	46	34	37	39	42	48	41

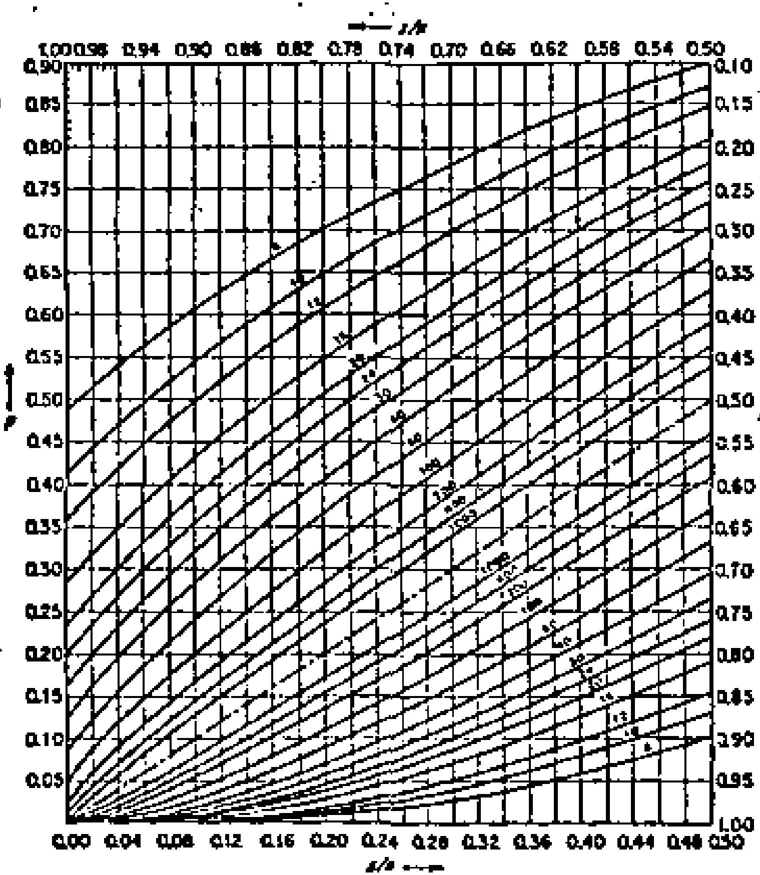
   Usar el test de las series de iguales por encima y por debajo de la mediana, a un nivel de significación de 0.05, para contrastar la hipótesis de que los datos marcan una tendencia.
5. Las temperaturas horarias de un horno (en grados centígrados) tomadas durante un período de 24 horas son las siguientes:

   260, 265, 271, 265, 270, 266, 273, 271, 275, 260, 271, 273
   275, 265, 276, 270, 273, 266, 270, 265, 272, 271, 278, 267

   Contrastar si esta disposición es al azar, con un nivel de significación de 0.01, para investigar si el horno está trabajando cíclicamente en intervalos de dos horas.
6. El problema 13 de la página 107 contiene el número de imperfecciones en muestras tomadas de 50 piezas de tela. Suponiendo que el orden de las muestras es el mismo que el orden en que se han producido las piezas de tela, contrastar la hipótesis de que la presencia de muestras sin imperfecciones se deba al azar, frente a la alternativa de que existe un agrupamiento. (Se toma $\alpha = 0.05$).
7. Utilizar el test de Kolmogorov-Smirnov, con $\alpha = 0.01$, para decidir si las resistencias a la compresión del ejercicio 9 de la página 107 se pueden suponer provenientes de una distribución normal con media de 30,000 libras por pulgada cuadrada y desviación típica de 10.000 libras por pulgada cuadrada. [*Sugerencia:* utilizar papel gráfico de probabilidades.]
8. En un estudio de vibraciones, se sometieron las componentes de un avión a fuertes vibraciones, hasta que se originaron fallos estructurales. Los siguientes son los tiempos obtenidos (en minutos) 4.1, 0.8, 5.3, 5.0, 8.3, 1.7, 2.5, 6.2, 7.3, 9.0, 1.2, 3.7, 9.5, 10.5. Contrastar si se pueden considerar estos datos como una muestra proveniente de una población exponencial con una media de 5 minutos. (Se toma $\alpha = 0.05$.)

Tabla VIII(b)

0.99 INTERVALOS DE CONFIANZA PARA PROPORCIONES*



* Reproducción de la tabla 41 de *Biometrika Tables for Statisticians*, Vol. I (New York: Cambridge University Press, 1954) con autorización de *Biometrika* fideicomiso.

Tabla IX

VALORES CRÍTICOS DE D*

Número de muestra n	$D_{.10}$	$D_{.05}$	$D_{.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29

* Adaptado de F. J. Massey, Jr., "The Kolgomorov-Smirnov test for goodness of fit" *J. Amer. Statist. Ass.*, Vol. 46 (1951), p. 70 con autorización de autor y editores.

Canavos: Probabilidad y Estadística: Fundamentos y Aplicaciones

MUESTRA DEL PESO DE 50 SOBRE DE CORREO AEREO

0.098	0.103	0.104	0.100	0.108	0.099	0.105	0.101	0.110	0.095
0.106	0.105	0.096	0.107	0.101	0.115	0.100	0.103	0.097	0.103
0.103	0.098	0.107	0.099	0.105	0.092	0.104	0.106	0.102	0.111
0.108	0.097	0.109	0.110	0.096	0.102	0.101	0.094	0.104	0.100
0.101	0.104	0.094	0.102	0.112	0.101	0.099	0.107	0.098	0.108

1. Calcular promedio aritmético, el modo y la mediana, la variancia, desviación estándar y coeficiente de variación
   a) con datos básicos
   b) con datos agrupados por valores
   c) con datos agrupados por intervalos.

2. Dibujar el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas