$$s = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = 3\sqrt{\frac{97}{100} - \left(\frac{15}{100}\right)^2} = 3\sqrt{0.9475} = 2.92 \text{ in}$$

4.18. Por métodos de compilación, hallar (*a*) la media y (*b*) la desviación típica para la distribución de salarios del Problema 2.3.

Solución

La tarea es sencilla, como ilustra la Tabla 4.6.

Tabla 4.6

	X	u	f	fu	fu^2
	\$255.00	−2	8	−16	32
	265.00	−1	10	−10	10
$A \longrightarrow$	275.00	0	16	0	0
	285.00	1	14	14	14
	295.00	2	10	20	40
	305.00	3	5	15	45
	315.00	4	2	8	32
			$N = \sum f = 65$	$\sum fu = 31$	$\sum fu^2 = 173$

(*a*) $$\bar{X} = A + c\bar{u} = A + c\frac{\sum fu}{N} = \$275.00 + (\$10.00)\left(\frac{31}{65}\right) = \$279.77$$

(*b*) $$s = \sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = (\$10.00)\sqrt{\frac{173}{65} - \left(\frac{31}{65}\right)^2} = (\$10.00)\sqrt{2.4341} = \$15.60$$

4.19. La Tabla 4.7 muestra los IQ (cocientes de inteligencia) de 480 niños de una escuela elemental. Mediante el método de compilación, hallar (*a*) la media y (*b*) la desviación típica.

Tabla 4.7

Marca de clase (X)	70	74	78	82	86	90	94	98	102	106	110	114	118	122	126
Frecuencia (f)	4	9	16	28	45	66	85	72	54	38	27	18	11	5	2

Solución

El cociente de inteligencia es

$$\text{IQ} = \frac{\text{edad mental}}{\text{edad cronológica}}$$

expresado como porcentaje. Por ejemplo, un niño de 8 años que (de acuerdo con ciertos procedimientos pedagógicos) tiene una mentalidad equivalente a uno de 10 años, tendría un IQ de 10/8 = 12.5 = 125%, o sencillamente 125, quedando sobreentendido el símbolo %.

Para hallar la media y la desviación típica de los IQ de la Tabla 4.7, podemos hacer lo que indica la Tabla 4.8.

(*a*) $$\bar{X} = A + c\bar{u} = A + c\frac{\sum fu}{N} = 94 + 4\left(\frac{236}{480}\right) = 95.97$$

(*b*) $$s = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = 4\sqrt{\frac{3404}{480} - \left(\frac{236}{480}\right)^2} = 4\sqrt{6.8499} = 10.47$$

Tabla 4.8

X	u	f	fu	fu^2
70	−6	4	−24	144
74	−5	9	−45	225
78	−4	16	−64	256
82	−3	28	−84	252
86	−2	45	−90	180
90	−1	66	−66	66
$A \longrightarrow$ 94	0	85	0	0
98	1	72	72	72
102	2	54	108	216
106	3	38	114	342
110	4	27	108	432
114	5	18	90	450
118	6	11	66	396
122	7	5	35	245
126	8	2	16	128
		$N = \sum f = 480$	$\sum fu = 236$	$\sum fu^2 = 3404$

COMPROBACION DE CHARLIER

4.20. Usar la comprobación de Charlier para verificar los cálculos de (*a*) la media y (*b*) la desviación típica, efectuados en el Problema 4.19.

Solución

Para aplicar esa comprobación hay que sumar las columnas de la Tabla 4.9 a las de la 4.8 (excepto la columna 2, que se repite en la Tabla 4.9 por conveniencia).

(*a*) De la Tabla 4.9, $\sum f(u + 1) = 716$; de la Tabla 4.8, $\sum fu + N = 236 + 480 = 716$. Eso da la requerida comprobación sobre la media.

(*b*) De la Tabla 4.9, $\sum f(u + 1)^2 = 4356$; de la Tabla 4.8, $\sum fu^2 + 2\sum fu + N = 3404 + 2(236) + 480 = 4356$. Lo cual proporciona la comprobación pedida sobre la desviación típica.

CORRECCIONES DE SHEPPARD PARA LA VARIANZA

4.21. Aplicar la corrección de Sheppard para determinar la desviación típica de los datos del (*a*) Problema 4.17, (*b*) Problema 4.18 y (*c*) Problema 4.19.

Solución

(*a*) $s^2 = 8.5275$ y $c = 3$. Varianza corregida $= s^2 - c^2/12 = 8.5275 - 3^2/12 = 7.7775$. Desviación típica corregida $= \sqrt{\text{varianza correcta}} = \sqrt{7.7775} = 2.79$ in.

(*b*) $s^2 = 243.41$ y $c = 10$. Varianza corregida $= s^2 - c^2/12 = 243.41 - 10^2/12 = 235.08$. Desviación típica corregida $= \sqrt{235.08} = \$15.33$.

(*c*) $s^2 = 109.60$ y $c = 4$. Varianza corregida $= s^2 - c^2/12 = 109.60 - 4^2/12 = 108.27$. Desviación típica corregida $= \sqrt{108.27} = 10.41$.

Tabla 4.9

$u + 1$	f	$f(u + 1)$	$f(u + 1)^2$
-5	4	-20	100
-4	9	-36	144
-3	16	-48	144
-2	28	-56	112
-1	45	-45	45
0	66	0	0
1	85	85	85
2	72	144	288
3	54	162	486
4	38	152	608
5	27	135	675
6	18	108	648
7	11	77	539
8	5	40	320
9	2	18	162
	$N = \sum f = 480$	$\sum f(u+1) = 716$	$\sum f(u + 1)^2 = 4356$

4.22. Hallar, para la segunda distribución de frecuencias del Problema 2.8, (*a*) la media, (*b*) la desviación típica, (*c*) la desviación típica usando la corrección de Sheppard y (*d*) la verdadera desviación típica para los datos sin agrupar.

Solución

El trabajo lo resume la Tabla 4.10.

Tabla 4.10

X	u	f	fu	fu^2
122	-3	3	-9	27
131	-2	5	-10	20
140	-1	9	-9	9
$A \longrightarrow$ 149	0	12	0	0
158	1	5	5	5
167	2	4	8	16
176	3	2	6	18
		$N = \sum f = 40$	$\sum fu = -9$	$\sum fu^2 = 95$

(*a*) $$\bar{X} = A + c\bar{u} = A + c\frac{\sum fu}{N} = 149 + 9\left(\frac{-9}{40}\right) = 147.0 \text{ lb}$$

(*b*) $$s = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = 9\sqrt{\frac{95}{40} - \left(\frac{-9}{40}\right)^2} = 9\sqrt{2.324375} = 13.7 \text{ lb}$$

(*c*) Varianza corregida $= s^2 - c^2/12 = 188.27 - 9^2/12 = 181.52$. Desviación corregida típica = 13.5 lb.

(*d*) Para calcular la desviación típica de los propios datos originales, conviene restar primero un número adecuado, digamos $A = 150$ lb, de cada peso y usar entonces el método del Problema 4.15. Las desviaciones $d = X - A = X - 150$ son las que figuran en la siguiente tabla:

−12	14	0	−18	−6	−25	−1	7
−4	8	−10	−3	−14	−2	2	−6
18	−24	−12	26	13	−31	4	15
−4	23	−8	−3	−15	3	−10	−15
11	−5	−15	−8	0	6	−5	−22

de donde deducimos que $\sum d = -128$ y $\sum d^2 = 7052$. Entonces

$$s = \sqrt{\overline{d^2} - \bar{d}^2} = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2} = \sqrt{\frac{7052}{40} - \left(\frac{-128}{40}\right)^2} = \sqrt{166.06} = 12.9 \text{ lb}$$

De modo que la corrección de Sheppard produce una cierta mejora en este caso.

RELACIONES EMPIRICAS ENTRE MEDIDAS DE DISPERSION

4.23. Para la distribución de alturas de la Universidad XYZ, discutir la validez de las fórmulas empíricas (*a*) desviación media = $\frac{4}{5}$(desviación típica) y (*b*) rango semi-intercuartil = $\frac{2}{3}$(desviación típica).

Solución

(*a*) De los Problemas 4.4 y 4.11, desviación media ÷ desviación típica = $2.26/2.92 = 0.77$, que está cerca de 4/5.

(*b*) De los Problemas 4.6 y 4.11, rango semi-intercuartil ÷ desviación típica = $1.98/2.92 = 0.68$, que es próximo a 2/3.

Luego las fórmulas empíricas son válidas en este caso.

Notemos que en lo anterior no hemos usado la desviación típica con corrección Sheppard para el agrupamiento, pues no se ha hecho corrección correspondiente para la desviación media o el rango semi-intercuartil.

PROPIEDADES DE LA DESVIACION TIPICA

4.24. Determinar el porcentaje de los IQ del Problema 4.19 que caen en los rangos (*a*) $\bar{X} \pm s$, (*b*) $\bar{X} \pm 2s$ y (*c*) $\bar{X} \pm 3s$.

Solución

(*a*) El rango de IQ desde 85.5 a 106.4 es $\bar{X} \pm s = 95.97 \pm 10.47$. El número de IQ en el rango $\bar{X} \pm s$ es

$$\left(\frac{88 - 85.5}{4}\right)(45) + 66 + 85 + 72 + 54 + \left(\frac{106.4 - 104}{4}\right)(38) = 339$$

El porcentaje de IQ en el rango $\bar{X} \pm s$ es $339/480 = 70.6\%$.

(*b*) El rango de IQ desde 75.0 a 116.9 es $\bar{X} \pm 2s = 95.97 \pm 2(10.47)$. El número de IQ en el rango $\bar{X} \pm 2s$ es

$$\left(\frac{76-75.0}{4}\right)(9) + 16 + 28 + 45 + 66 + 85 + 72 + 54 + 38 + 27 + 18 + \left(\frac{116.9-116}{4}\right)(11) = 451$$

El porcentaje de IQ en el rango $\bar{X} \pm 2s$ es $451/480 = 94.0\%$.

(*c*) El rango de IQ desde 64.6 a 127.4 es $\bar{X} \pm 3s = 95.97 \pm 3(10.47)$. El número de IQ en el rango $\bar{X} \pm 3s$ es

$$480 - \left(\frac{128 - 127.4}{4}\right)(2) = 479.7 \qquad \text{o sea} \qquad 480$$

El porcentaje de IQ en el rango $\bar{X} \pm 3s$ es $479.7/480 = 99.9\%$, es decir, prácticamente el 100 por 100.

Los porcentajes en las partes (*a*), (*b*) y (*c*) están en buen acuerdo con los esperados para una distribución normal: 68.27%, 95.45% y 99.73%, respectivamente.

Nótese que no hemos usado la corrección de Sheppard para la desviación típica. Si se usa, los resultados en este caso coinciden casi con lo obtenido aquí. Por cierto, que éstos pueden también obtenerse de la Tabla 4.11 del Problema 4.32.

4.25. Dados los conjuntos de números 2, 5, 8, 11, 14 y 2, 8, 14, hallar (*a*) la media de cada uno, (*b*) la varianza de cada uno, (*c*) la media combinada y (*d*) la varianza combinada

Solución

(*a*) Media del primer conjunto $= \frac{1}{5}(2 + 5 + 8 + 11 + 14) = 8$. Media del segundo conjunto $= \frac{1}{3}(2 + 8 + 14) = 8$.

(*b*) Varianza del primer conjunto $= s_1^2 = \frac{1}{5}[(2-8)^2 + (5-8)^2 + (8-8)^2 + (11-8)^2 + (14-8)^2] = 18$.
Varianza del segundo conjunto $= s_2^2 = \frac{1}{3}[(2-8)^2 + (8-8)^2 + (14-8)^2] = 24$.

(*c*) La media de ambos conjuntos es

$$\frac{2 + 5 + 8 + 11 + 14 + 2 + 8 + 14}{5 + 3} = 8$$

(*d*) La varianza del conjunto total es

$$s^2 = \frac{(2-8)^2 + (5-8)^2 + (8-8)^2 + (11-8)^2 + (14-8)^2 + (2-8)^2 + (8-8)^2 + (14-8)^2}{5+3} = 20.25$$

Otro método (por fórmula)

$$s^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2} = \frac{(5)(18) + (3)(24)}{5 + 3} = 20.25$$

4.26. Resolver el Problema 4.25 para los conjuntos 2, 5, 8, 11, 14 y 10, 16, 22.

Solución

Aquí las medias de los dos conjuntos son 8 y 16, mientras que las varianzas son las mismas que las de los conjuntos del problema anterior, es decir, $s_1^2 = 18$ y $s_2^2 = 24$.

$$\text{Media de ambos conjuntos} = \frac{2 + 5 + 8 + 11 + 14 + 10 + 16 + 22}{5 + 3} = 11$$

$$s^2 = \frac{(2-11)^2 + (5-11)^2 + (8-11)^2 + (11-11)^2 + (14-11)^2 + (10-11)^2 + (16-11)^2 + (22-11)^2}{5+3} = 35.25$$

Nótese que la fórmula

$$s^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2}$$

que da el valor 20.25, *no* es aplicable en este caso porque las medias de los dos conjuntos *no* son iguales.

4.27. (*a*) Probar que $w^2 + pw + q$, donde p y q son constantes dadas, es un mínimo si y sólo si $w = -\frac{1}{2}p$.

(*b*) Usando la parte (*a*), probar que

$$\frac{\sum_{j=1}^{N}(X_j - a)^2}{N} \quad \text{o brevemente} \quad \frac{\sum (X - a)^2}{N}$$

es un mínimo si y sólo si $a = \bar{X}$.

Solución

(*a*) Tenemos $w^2 + pw + q = (w + \frac{1}{2}p)^2 + q - \frac{1}{4}p^2$. Como $(q - \frac{1}{4}p^2)$ es una constante, la expresión tiene el valor mínimo si y sólo si $w + \frac{1}{2}p = 0$ (i.e., $w = -\frac{1}{2}p$).

(*b*) $$\frac{\sum (X - a)^2}{N} = \frac{\sum (X^2 - 2aX + a^2)}{N} = \frac{\sum X^2 - 2a\sum X + Na^2}{N} = a^2 - 2a\frac{\sum X}{N} + \frac{\sum X^2}{N}$$

Comparando esta última expresión con $(w^2 + pw + q)$, se obtiene

$$w = a \qquad p = -2\frac{\sum X}{N} \qquad q = \frac{\sum X^2}{N}$$

Así pues, la expresión es mínima cuando $a = -\frac{1}{2}p = (\sum X)/N = \bar{X}$, usando el resultado en (*a*).

DISPERSION ABSOLUTA Y RELATIVA: COEFICIENTE DE VARIACION

4.28. Un fabricante de tubos de televisión produce dos tipos de tubos, A y B, que tienen vidas medias respectivas $\bar{X}_A = 1495$ horas y $\bar{X}_B = 1875$ horas, y desviación típica de $s_A = 280$ horas y $s_B = 310$ horas. ¿Qué tubo tiene (*a*) mayor dispersión absoluta y (*b*) mayor dispersión relativa?

Solución

(*a*) La dispersión absoluta de A es $s_A = 280$ horas y la de B es $s_B = 310$ horas. Luego el tubo B tiene mayor dispersión absoluta.

(*b*) Los coeficientes de variación son

$$A = \frac{s_A}{\bar{X}_A} = \frac{280}{1495} = 18.7\% \qquad B = \frac{s_B}{\bar{X}_B} = \frac{310}{1875} = 16.5\%$$

Luego tiene más dispersión relativa el *A*.

4.29. Hallar los coeficientes de variación V para los datos del (*a*) Problema 4.14 y (*b*) Problema 4.18, usando tanto desviaciones típicas corregidas como no corregidas.

Solución

(*a*)
$$V(\text{sin corregir}) = \frac{s(\text{sin corregir})}{\bar{X}} = \frac{2.92}{67.45} = 0.0433 = 4.3\%$$

$$V(\text{corregido}) = \frac{s(\text{corregido})}{\bar{X}} = \frac{2.79}{67.45} = 0.0413 = 4.1\% \qquad \text{por el Problema 4.21}(a)$$

(*b*)
$$V(\text{sin corregir}) = \frac{s(\text{sin corregir})}{\bar{X}} = \frac{15.60}{79.77} = 0.196 = 19.6\%$$

$$V(\text{corregido}) = \frac{s(\text{corregido})}{\bar{X}} = \frac{15.33}{79.77} = 0.192 = 19.2\% \qquad \text{por el Problema 4.21}(b)$$

4.30. (*a*) Definir una medida de la dispersión relativa que pueda utilizarse para un conjunto de datos cuyos cuartiles son conocidos.

(*b*) Ilustrar el cálculo de la medida definida en (*a*) mediante los datos del Problema 4.6.

Solución

(*a*) Si Q_1 y Q_3 son conocidos para un conjunto de números, entonces $\frac{1}{2}(Q_1 + Q_3)$ es una medida de tendencia central de esos datos, o promedio, mientras que $Q = \frac{1}{2}(Q_3 - Q_1)$, el rango semi-intercuartil, es una medida de su dispersión. Podemos, pues, definir una medida de dispersión relativa como

$$V_Q = \frac{\frac{1}{2}(Q_3 - Q_1)}{\frac{1}{2}(Q_1 + Q_3)} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

que llamaremos el *coeficiente de variación cuartil*, o *coeficiente cuartil de dispersión relativa*.

(*b*)
$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{69.61 - 65.64}{69.61 + 65.64} = \frac{3.97}{135.25} = 0.0293 = 2.9\%$$

VARIABLES TIPIFICADAS: UNIDADES ESTANDAR

4.31. Un estudiante obtuvo 84 puntos en el examen final de Matemáticas, en el que la nota media fue 76, y la desviación típica 10. En el examen final de Física obtuvo 90 puntos, siendo la media 82 y la desviación típica 16. ¿En qué examen sobresalió más?

Solución

La variable tipificada $z = (X - \bar{X})/s$ mide la desviación de X respecto de la media $\bar{X}$ en términos de la desviación típica s. En Matemáticas, $z = (84 - 76)/10 = 0.8$; para física, $z = (90 - 82)/16 = 0.5$. Luego su puntuación estaba 0.8 desviaciones típicas sobre la media en matemáticas y sólo 0.5 desviaciones típicas en física. Sobresalió más en matemáticas.

La variable $z = (X - \bar{X})/s$ se usa a menudo en niveles de enseñanza, donde se conoce como una *puntuación o recuento estándar*.

4.32. (*a*) Convertir los IQ del Problema 4.19 en un recuento estándar y (*b*) construir una gráfica de frecuencias relativas versus recuento estándar.

Solución

(*a*) La Tabla 4.11 resume el proceso de conversión. Añadidas a la tabla para su uso en la parte (*b*) están las marcas de clase de IQ 66 y 130, que tienen frecuencia cero. Asimismo, la corrección de Sheppard para la desviación típica no ha sido utilizada; las correcciones en este caso serian casi despreciables.

(*b*) El polígono de frecuencias relativas se muestra en la Figura 4.2. El eje horizontal se mide en términos de la desviación típica s como la unidad. Nótese que la distribución es poco asimétrica y algo sesgada a la derecha.

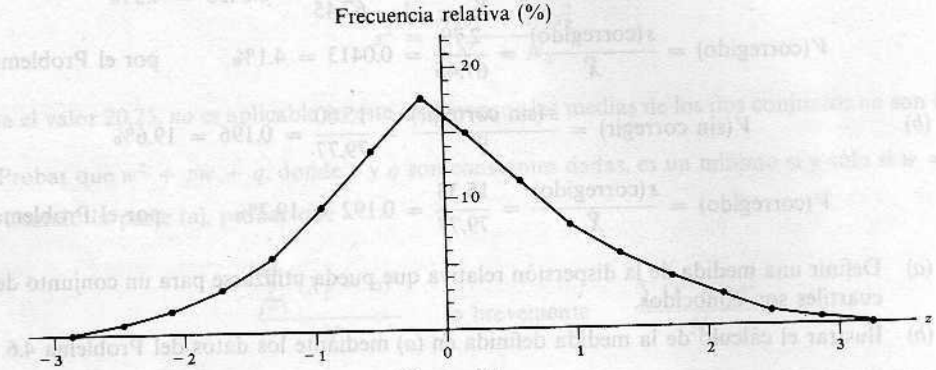


Figura 4.2.

Tabla 4.11. $\bar{X} = 96.0$, $s = 10.5$

IQ (X)	$X - \bar{X}$	$z = \dfrac{X - \bar{X}}{s}$	Frecuencia (f)	Frecuencia relativa (f)/N (%)
66	−30.0	−2.86	0	0.0
70	−26.0	−2.48	4	0.8
74	−22.0	−2.10	9	1.9
78	−18.0	−1.71	16	3.3
82	−14.0	−1.33	28	5.8
86	−10.0	−0.95	45	9.4
90	−6.0	−0.57	66	13.8
94	−2.0	−0.19	85	17.7
98	2.0	0.19	72	15.0
102	6.0	0.57	54	11.2
106	10.0	0.95	38	7.9
110	14.0	1.33	27	5.6
114	18.0	1.71	18	3.8
118	22.0	2.10	11	2.3
122	26.0	2.48	5	1.0
126	30.0	2.86	2	0.4
130	34.0	3.24	0	0.0
			480	100

PROBLEMAS SUPLEMENTARIOS

EL RANGO

4.33. Hallar el rango de los conjuntos de números (*a*) 5, 3, 8, 4, 7, 6, 12, 4, 3 y (*b*) 8.772, 6.453, 10.624, 8.628, 9.434, 6.351.

4.34. Hallar el rango de las cargas máximas del Problema 3.59, Tabla 3.8.

4.35. Hallar el rango de los diámetros de remaches del Problema 3.61, Tabla 3.10.

4.36. La mayor de 50 medidas es 8.34 kilogramos (kg). Si el rango es 0.46 kg, hallar la menor de esas medidas.

4.37. Determinar el rango de los datos en (*a*) Problema 3.62, (*b*) Problema 3.73 y (*c*) Problema 2.20.

LA DESVIACION MEDIA

4.38. Hallar los valores absolutos de (*a*) -18.2, (*b*) $+3.58$, (*c*) 6.21, (*d*) 0, (*e*) $-\sqrt{2}$ y (*f*) $4.00 - 2.36 - 3.52$.

4.39. Hallar la desviación media del conjunto (*a*) 3, 7, 9, 5 y (*b*) 2.4, 1.6, 3.8, 4.1, 3.4.

4.40. Hallar la desviación media de los conjuntos de números del Problema 4.33.

4.41. Hallar la desviación media de las cargas máximas del Problema 3.59, Tabla 3.8.

4.42. (*a*) Hallar la desviación media de los diámetros del Problema 3.61, Tabla 3.10.

(*b*) ¿Qué porcentaje de ellos está entre ($\bar{X} \pm$ $\pm$ MD), ($\bar{X} \pm 2$ MD) y ($\bar{X} \pm 3$ MD)?

4.43. Para el conjunto de números 8, 10, 9, 12, 4, 8, 2, hallar la desviación media respecto de (*a*) la media y (*b*) la mediana. Verificar que la desviación media de la mediana no es mayor que la de la media.

4.44. Para la distribución de la Tabla 3.9, Problema 3.60, hallar la desviación media respecto de (*a*) la media y (*b*) la mediana. Usar los resultados de los Problemas 3.60 y 3.70.

4.45. Para la distribución de la Tabla 3.11, Problema 3.62, hallar la desviación media respecto de (*a*) la media y (*b*) la mediana. Usar los resultados de los Problemas 3.62 y 3.72.

4.46. Explicar por qué la desviación media es o no una buena medida de dispersión para la distribución de la Tabla 3.12 del Problema 3.73.

4.47. Deducir fórmulas de compilación para calcular la desviación media respecto de (*a*) la media y (*b*) la mediana, de una distribución de frecuencias. Aplicar estas fórmulas a la verificación de los resultados de los Problemas 4.44 y 4.45.

EL RANGO SEMI-INTERCUARTIL

4.48. Hallar el rango semi-intercuartil para la distribución del (*a*) Problema 3.59, (*b*) Problema 3.60 y (*c*) Problema 3.107. Interpretar los resultados claramente en cada caso.

4.49. Hallar el rango semi-intercuartil para la distribución de (*a*) Problema 2.31 y (*b*) Problema 3.73, interpretando los resultados en cada caso. Comparando con otras medidas de dispersión, explicar las ventajas del rango semi-intercuartil para este tipo de distribuciones.

4.50. Probar que para cualquier distribución de frecuencias el porcentaje total de casos que caen en el intervalo $\frac{1}{2}(Q_1 + Q_3) \pm \frac{1}{2}(Q_3 - Q_1)$ es 50%. ¿Es eso cierto para el intervalo $Q_2 \pm \frac{1}{2}(Q_3 - Q_1)$? Explicar la respuesta.

4.51. (*a*) ¿Cómo representaría el rango semi-intercuartil de una distribución de frecuencias dada?

(*b*) ¿Cuál es la relación del rango semi-intercuartil con la ojiva de la distribución?

EL RANGO PERCENTIL 10-90

4.52. Hallar el rango percentil 10-90 para las distribuciones de (*a*) Problema 3.59 y (*b*) Problema 3.107. Interpretar cada resultado.

4.53. Hallar el rango percentil 10-90 para las distribuciones de (*a*) Problema 2.31 y (*b*) Problema 3.73. Interpretar los resultados. ¿Qué ventajas y desventajas ofrece el rango percentil 10-90 frente a otras medidas de dispersión?

4.54. ¿Qué ventajas y desventajas tendría un rango percentil 20-80 comparado con el rango percentil 10-90?

4.55. Resolver el Problema 4.51 con referencia al (*a*) rango percentil 10-90, (*b*) rango percentil 20-80 y (*c*) rango percentil 25-75. ¿Cuál es la relación entre (*c*) y el rango semi-intercuartil?

LA DESVIACION TIPICA

4.56. Hallar la desviación típica de los conjuntos de números (*a*) 3, 6, 2, 1, 7, 5; (*b*) 3.2, 4.6, 2.8, 5.2, 4.4 y (*c*) 0, 0, 0, 0, 0, 1, 1, 1.

4.57. (*a*) Sumando 5 a cada número del conjunto 3, 6, 2, 1, 7, 5, obtenemos 8, 11, 7, 6, 12, 10. Probar que ambos conjuntos de números tienen la misma desviación típica pero diferentes medias. ¿Cómo están relacionadas las medias?

(*b*) Multiplicando cada número en 3, 6, 2, 1, 7, y 5 por 2 y sumando entonces 5, obtenemos el conjunto 11, 17, 9, 7, 19, 15. ¿Cuál es la relación entre la desviación típica y las medias de ambos conjuntos?

(*c*) ¿Qué propiedades de la media y de la desviación típica quedan ilustradas por los conjuntos particulares elegidos en las partes (*a*) y (*b*)?

4.58. Hallar la desviación típica del conjunto de números de la progresión aritmética 4, 10, 16, 22, ..., 154.

4.59. Hallar la desviación típica para las distribuciones de (*a*) Problema 3.59, (*b*) Problema 3.60 y (*c*) Problema 3.107.

4.60. Ilustrar el uso de la comprobación de Charlier en cada parte del Problema 4.59.

4.61. Hallar (*a*) la media y (*b*) la desviación típica para la distribución del Problema 2.17, y explicar la relevancia de los resultados obtenidos.

4.62. (*a*) Explicar por qué la desviación típica no es una medida apropiada de dispersión para la distribución del Problema 2.31.

(*b*) ¿Qué medida de dispersión debe utilizarse en su lugar? Ilustrar la respuesta.

4.63. (*a*) Hallar la desviación típica s de los diámetros de remaches de la Tabla 3.10.

(*b*) ¿Qué porcentajes de ellos cae entre $\bar{X} \pm s$, $\bar{X} \pm 2s$ y $\bar{X} \pm 3s$?

(*c*) Comparar los porcentajes de la parte (*b*) con los esperados teóricamente si la distribución fuese normal, y razonar la discrepancia.

4.64. Aplicar la corrección de Sheppard a cada desviación típica del Problema 4.59, y discutir en cada caso si tal aplicación está o no justificada.

4.65. ¿Qué modificaciones se producen en el Problema 4.63 al aplicar la corrección de Sheppard?

4.66. (*a*) Hallar la media y la desviación típica para los datos del Problema 2.8.

(*b*) Construir una distribución de frecuencias para los datos y hallar su desviación típica.

(*c*) Comparar los resultados de (*a*) y (*b*). Determinar si la aplicación de la corrección de Sheppard mejora los resultados.

4.67. Repetir el Problema 4.66 con los datos del Problema 2.27.

4.68. (*a*) De un total de N números, la fracción p son unos, y la fracción $q = 1 - p$ son ceros. Probar que la desviación típica de ese conjunto de números es $\sqrt{pq}$.

(*b*) Aplicar el resultado de (*a*) al Problema 4.56(*c*).

4.69. (*a*) Probar que la varianza de un conjunto de números a, $a + d$, $a + 2d$, ...

$a + (n - 1)d$ (o sea, una progresión aritmética con primer término a y razón d) viene dada por $\frac{1}{12}(n^2 - 1)d^2$.

(*b*) Usar (*a*) del Problema 4.58. [*Ayuda:* Use $1 + 2 + 3 + \cdots + (n-1) = \frac{1}{2}n(n-1)$, $1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 = \frac{1}{6}n(n-1)(2n-1)$.]

4.70. Generalizar y probar la Propiedad 3 de este capítulo (pág. 95).

RELACIONES EMPIRICAS ENTRE MEDIDAS DE DISPERSION

4.71. Comparando las desviaciones típicas obtenidas en el Problema 4.59 con las correspondientes desviaciones medias de los Problemas 4.41, 4.42 y 4.44, determinar si es válida la siguiente relación empírica: Desviación media $= \frac{4}{5}$(desviación típica). Explicar las posibles discrepancias.

4.72. Comparando las desviaciones típicas obtenidas en el Problema 4.59 con los correspondientes rangos semi-intercuartiles del Problema 4.48, determinar si es válida la relación empírica: rango semi-intercuartil $= \frac{2}{3}$(desviación típica). Explicar las posibles discrepancias.

4.73. ¿Qué relación empírica esperaría entre el rango semi-intercuartil y la desviación media de una distribución de frecuencias en forma de campana algo sesgada?

4.74. Una distribución de frecuencias que es casi normal tiene un rango semi-intercuartil igual a 10. ¿Qué valores esperaría para (*a*) la desviación típica y (*b*) la desviación media?

DISPERSION ABSOLUTA Y RELATIVA: COEFICIENTE DE VARIACION

4.75. En un examen final de Estadística, la puntuación media de 150 estudiantes fue de 78, y la desviación típica 8.0. En Algebra, la media fue 73 y la desviación típica 7.6. ¿En qué materia fue mayor (*a*) la dispersión absoluta y (*b*) la dispersión relativa?

4.76. Hallar el coeficiente de variación para los datos de (*a*) Problema 3.59 y (*b*) Problema 3.107.

4.77. (*a*) ¿Por qué no es posible calcular el coeficiente de variación para la distribución del Problema 2.31?

(*b*) Calcular el coeficiente cuartil de dispersión relativa para esta distribución. [Véanse Probs. 3.10(*c*) y 4.30.]

4.78. (*b*) Ilustrar el cálculo de tal medida con los datos del Problema 3.73.

VARIABLES TIPIFICADAS: UNIDADES ESTANDAR

4.79. En los exámenes a que se refiere el Problema 4.75, un alumno tuvo 75 en Estadística y 71 en Algebra. ¿En qué examen sobresalió más?

4.80. Convertir el conjunto 6, 2, 8, 7, 5 en un recuento estándar.

4.81. Probar que la media y la desviación típica de un recuento estándar son 0 y 1, respectivamente. Ilustrar esto mediante el Problema 4.80.

4.82. (*a*) Convertir las puntuaciones del Problema 3.107 en un recuento estándar y (*b*) construir un gráfico de frecuencias relativas versus ese recuento estándar.

CAPITULO 5

Momentos, sesgo y curtosis

MOMENTOS

Si $X_1, X_2, \ldots, X_N$ son los N valores de la variable X, definimos la cantidad

$$\overline{X^r} = \frac{X_1^r + X_2^r + \cdots + X_N^r}{N} = \frac{\sum_{j=1}^{N} X_j^r}{N} = \frac{\sum X^r}{N} \tag{1}$$

llamada r-ésimo *momento*. El primer momento, con $r = 1$, es la media aritmética $\bar{X}$.

El r-ésimo *momento respecto de la media* $\bar{X}$ se define como

$$m_r = \frac{\sum_{j=1}^{N} (X_j - \bar{X})^r}{N} = \frac{\sum (X - \bar{X})^r}{N} = \overline{(X - \bar{X})^r} \tag{2}$$

Si $r = 1$, entonces $m_1 = 0$ (véase Prob. 3.16). Si $r = 2$, entonces $m_2 = s^2$, la varianza.

El r-ésimo *momento respecto de cualquier origen A* se define como

$$m'_r = \frac{\sum_{j=1}^{N} (X_j - A)^r}{N} = \frac{\sum (X - A)^r}{N} = \frac{\sum d^r}{N} = \overline{(X - A)^r} \tag{3}$$

donde $d = X - A$ son las desviaciones de X respecto de A. Si $A = 0$, la ecuación (3) se reduce a la (1). Por esa razón, se suele llamar a (1) el r-ésimo *momento respecto de cero.*

MOMENTOS PARA DATOS AGRUPADOS

Si $X_1, X_2, \ldots, X_K$ ocurren con frecuencias $f_1, f_2, \ldots, f_K$, respectivamente, los momentos anteriores vienen dados por

$$\overline{X^r} = \frac{f_1 X_1^r + f_2 X_2^r + \cdots + f_K X_K^r}{N} = \frac{\sum_{j=1}^{K} f_j X_j^r}{N} = \frac{\sum f X^r}{N} \tag{4}$$

$$m_r = \frac{\sum_{j=1}^{K} f_j(X_j - \bar{X})^r}{N} = \frac{\sum f(X - \bar{X})^r}{N} = \overline{(X - \bar{X})^r} \tag{5}$$

$$m'_r = \frac{\sum_{j=1}^{K} f_j(X_j - A)^r}{N} = \frac{\sum f(X - A)^r}{N} = \overline{(X - A)^r} \tag{6}$$

donde $N = \sum_{j=1}^{K} f_j = \sum f$. Las fórmulas son adecuadas para calcular momentos en datos agrupados.

RELACIONES ENTRE MOMENTOS

Existen las siguientes relaciones entre momentos respecto de la media m_r y momentos respecto de un origen arbitrario m'_r:

$$\begin{aligned} m_2 &= m'_2 - m'^2_1 \\ m_3 &= m'_3 - 3m'_1m'_2 + 2m'^3_1 \\ m_4 &= m'_4 - 4m'_1m'_3 + 6m'^2_1m'_2 - 3m'^4_1 \end{aligned} \tag{7}$$

etcétera (véase Problema 5.5). Nótese que $m'_1 = \bar{X} - A$.

CALCULO DE MOMENTOS PARA DATOS AGRUPADOS

El método de compilación visto en capítulos precedentes para el cálculo de la media y de la desviación típica, puede usarse también como método breve para calcular momentos. Este método se apoya en que $X_j = A + cu_j$ (o más brevemente, $X = A + cu$), así que de la ecuación (6) tenemos

$$m'_r = c^r \frac{\sum fu^r}{N} = c^r\overline{u^r} \tag{8}$$

que puede utilizarse para hallar m_r aplicando las ecuaciones (7).

COMPROBACION DE CHARLIER Y CORRECCIONES DE SHEPPARD

La comprobación de Charlier para calcular momentos por compilación usa las identidades

$$\begin{aligned} \sum f(u+1) &= \sum fu + N \\ \sum f(u+1)^2 &= \sum fu^2 + 2\sum fu + N \\ \sum f(u+1)^3 &= \sum fu^3 + 3\sum fu^2 + 3\sum fu + N \\ \sum f(u+1)^4 &= \sum fu^4 + 4\sum fu^3 + 6\sum fu^2 + 4\sum fu + N \end{aligned} \tag{9}$$

Las correcciones de Sheppard para los momentos son como siguen:

$$m_2 \text{ corregido} = m_2 - \tfrac{1}{12}c^2 \qquad m_4 \text{ corregido} = m_4 - \tfrac{1}{2}c^2 m_2 + \tfrac{7}{240}c^4$$

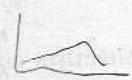
Los momentos m_1 y m_3 no requieren corrección.

MOMENTOS ADIMENSIONALES

Para evitar unidades particulares, podemos definir los *momentos adimensionales* respecto de la media como

$$a_r = \frac{m_r}{s^r} = \frac{m_r}{(\sqrt{m_2})^r} = \frac{m_r}{\sqrt{m_2^r}} \tag{10}$$

donde $s = \sqrt{m_2}$ es la desviación típica. Ya que $m_1 = 0$ y $m_2 = s^2$, se tiene $a_1 = 0$ y $a_2 = 1$.

 sesgo(+)  sesgo (-)

SESGO

Se conoce como *sesgo* el grado de asimetría de una distribución, es decir, cuánto se aparta de la simetría. Si la curva de frecuencias (polígono de frecuencias suavizado) de una distribución tiene a la derecha una cola más larga que a la izquierda, se dice *sesgada a la derecha*, o de *sesgo positivo*. En caso contrario, *sesgada a la izquierda*, o de *sesgo negativo*.

Para distribuciones sesgadas, la media tiende a estar del mismo lado de la moda que la cola larga (véanse Figs. 3.1 y 3.2). Luego una medida de la asimetría viene dada por la diferencia: media − moda, que puede hacerse adimensional dividiéndola por una medida de dispersión, tal como la desviación típica, lo que lleva a la definición

$$\text{Sesgo} = \frac{\text{media} - \text{moda}}{\text{desviación típica}} = \frac{\bar{X} - \text{moda}}{s} \tag{11}$$

Para evitar el uso de la moda, podemos recurrir a la fórmula empírica (10) del Capítulo 3 y definir

$$\text{Sesgo} = \frac{3(\text{media} - \text{mediana})}{\text{desviación típica}} = \frac{3(\bar{X} - \text{mediana})}{s} \tag{12}$$

Las ecuaciones (11) y (12) se llaman, respectivamente, *primer y segundo coeficientes de sesgo de Pearson*.

Otras medidas del sesgo, en términos de cuartiles y percentiles, son

$$\text{Coeficiente cuartil de sesgo} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \tag{13}$$

$$\text{Coeficiente percentil 10-90 de sesgo} = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} \tag{14}$$

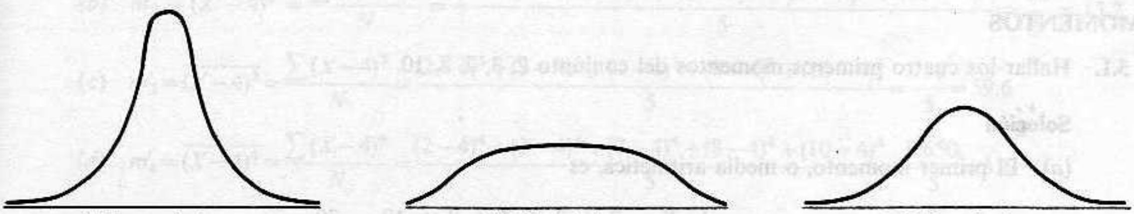
Una importante medida del sesgo usa el tercer momento respecto de la media expresado en forma adimensional y viene dado por

$$\text{Coeficiente momento de sesgo} = a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{m_3}{\sqrt{m_2^3}} \tag{15}$$

Otra usada a veces es $b_1 = a_3^2$. Para curvas perfectamente simétricas, como la curva normal, a_3 y b_1 son cero.

CURTOSIS

La *curtosis* mide cuán puntiaguda es una distribución, en general por referencia a la normal. Si tiene un pico alto, como la de la Figura 5.1(a), se dice *leptocúrtica*, mientras si es aplastada, como la de la Figura 5.1(b), se dice *platicúrtica*. La distribución normal, mostrada en la Figura 5.1(c), que no es ni muy puntiaguda ni muy aplastada, se llama *mesocúrtica*.



(a) Leptocúrtica (b) Platicúrtica (c) Mesocúrtica.

$K>0$ $K<0$ $K=0$

Figura 5.1.

Una medida de la curtosis utiliza el cuarto momento respecto de la media en forma adimensional y viene dada por

$$\text{Coeficiente momento de curtosis} = a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} \tag{16}$$

que se suele denotar por b_2. Para la distribución normal, $b_2 = a_4 = 3$. De ahí que se defina a veces la curtosis como $(b_2 - 3)$, que es positivo para una distribución leptocúrtica, negativo para una platicúrtica y cero para la normal.

Otra medida de curtosis se basa en cuartiles y percentiles, y viene dada por

$$\kappa = \frac{Q}{P_{90} - P_{10}} \tag{17}$$

donde $Q = \frac{1}{2}(Q_3 - Q_1)$ es el rango semi-intercuartil. Nos referiremos a κ (letra griega minúscula *kappa*) como el *coeficiente percentil de curtosis;* para la distribución normal, κ vale 0.263 (véase Problema 5.14).

MOMENTOS, SESGO Y CURTOSIS DE UNA POBLACION

Cuando es necesario distinguir entre los momentos, medidas de sesgo y medidas de curtosis de una población y los de una muestra suya, se suelen usar símbolos latinos para los primeros y griegos para los segundos. Así, si los momentos de la muestra se denotan por m_r y m'_r, los correspondientes símbolos griegos serán μ_r y μ'_r (μ es la letra griega *mu*). Los subíndices serán siempre símbolos latinos.

Análogamente, si las medidas de sesgo y curtosis de la muestra se denotan a_3 y a_4, respectivamente, las de la población serán α_3 y α_4 (α es la letra griega *alfa*).

Ya sabemos, Capítulo 4, que la desviación típica de una muestra y de una población se denotan, respectivamente, por s y σ.

PROBLEMAS RESUELTOS

MOMENTOS

5.1. Hallar los cuatro primeros momentos del conjunto 2, 3, 7, 8, 10.

Solución

(*a*) El primer momento, o media aritmética, es

$$\bar{X} = \frac{\sum X}{N} = \frac{2 + 3 + 7 + 8 + 10}{5} = \frac{30}{5} = 6$$

(*b*) El segundo momento es

$$\overline{X^2} = \frac{\sum X^2}{N} = \frac{2^2 + 3^2 + 7^2 + 8^2 + 10^2}{5} = \frac{226}{5} = 45.2$$

(*c*) El tercer momento es

$$\overline{X^3} = \frac{\sum X^3}{N} = \frac{2^3 + 3^3 + 7^3 + 8^3 + 10^3}{5} = \frac{1890}{5} = 378$$

(*d*) El cuarto momento es

$$\overline{X^4} = \frac{\sum X^4}{N} = \frac{2^4 + 3^4 + 7^4 + 8^4 + 10^4}{5} = \frac{16,594}{5} = 3318.8$$

5.2. Hallar los cuatro primeros momentos respecto de la media para el conjunto de números del Problema 5.1.

Solución

(*a*) $$m_1 = \overline{(X - \bar{X})} = \frac{\sum (X - \bar{X})}{N} = \frac{(2-6) + (3-6) + (7-6) + (8-6) + (10-6)}{5} = \frac{0}{5} = 0$$

m_1 es siempre cero ya que $\overline{X - \bar{X}} = \bar{X} - \bar{X} = 0$ (véase Probl. 3.16).

(*b*) $m_2 = \overline{(X-\bar{X})^2} = \dfrac{\sum (X-\bar{X})^2}{N} = \dfrac{(2-6)^2+(3-6)^2+(7-6)^2+(8-6)^2+(10-6)^2}{6} = \dfrac{46}{5} = 9.2$

Nótese que m_2 es la varianza s^2.

(*c*) $m_3 = \overline{(X-\bar{X})^3} = \dfrac{\sum (X-\bar{X})^3}{N} = \dfrac{(2-6)^3+(3-6)^3+(7-6)^3+(8-6)^3+(10-6)^3}{5} = \dfrac{-18}{5} = -3.6$

(*d*) $m_4 = \overline{(X-\bar{X})^4} = \dfrac{\sum (X-\bar{X})^4}{N} = \dfrac{(2-6)^4+(3-6)^4+(7-6)^4+(8-6)^4+(10-6)^4}{5} = \dfrac{610}{5} = 122$

5.3. Hallar los cuatro primeros momentos respecto del origen para el conjunto de números del Problema 5.1.

Solución

(*a*) $m'_1 = \overline{(X-4)} = \dfrac{\sum (X-4)}{N} = \dfrac{(2-4)+(3-4)+(7-4)+(8-4)+(10-4)}{5} = 2$

(*b*) $m'_2 = \overline{(X-4)^2} = \dfrac{\sum (X-4)^2}{N} = \dfrac{(2-4)^2+(3-4)^2+(7-4)^2+(8-4)^2+(10-4)^2}{5} = \dfrac{66}{5} = 13.2$

(*c*) $m'_3 = \overline{(X-4)^3} = \dfrac{\sum (X-4)^3}{N} = \dfrac{(2-4)^3+(3-4)^3+(7-4)^3+(8-4)^3+(10-4)^3}{5} = \dfrac{298}{5} = 59.6$

(*d*) $m'_4 = \overline{(X-4)^4} = \dfrac{\sum (X-4)^4}{N} = \dfrac{(2-4)^4+(3-4)^4+(7-4)^4+(8-4)^4+(10-4)^4}{5} = \dfrac{1650}{5} = 330$

5.4. Usando los resultados de los Problemas 5.2 y 5.3, verificar las relaciones entre momentos: (*a*) $m_2 = m'_2 - m'^2_1$, (*b*) $m_3 = m'_3 - 3m'_1m'_2 + 2m'^3_1$ y (*c*) $m_4 = m'_4 - 4m'_1m'_3 + 6m'^2_1m'_2 - 3m'^4_1$.

Solución

Por el Problema 5.3 tenemos $m'_1 = 2$, $m'_2 = 13.2$, $m'_3 = 59.6$ y $m'_4 = 330$. Por tanto:

(*a*) $m_2 = m'_2 - m'^2_1 = 13.2 - (2)^2 = 13.2 - 4 = 9.2.$

(*b*) $m_3 = m'_3 - 3m'_1m'_2 + 2m'^3_1 = 59.6 - (3)(2)(13.2) + 2(2)^3 = 59.6 - 79.2 + 16 = -3.6$

(*c*) $m_4 = m'_4 - 4m'_1m'_3 + 6m'^2_1m'_2 - 3m'^4_1 = 330 - 4(2)(59.6) + 6(2)^2(13.2) - 3(2)^4 = 122$

de acuerdo con el Problema 5.2.

5.5. Probar que: (*a*) $m_2 = m'_2 - m'^2_1$, (*b*) $m_3 = m'_3 - 3m'_1m'_2 + 2m'^3_1$ y (*c*) $m_4 = m'_4 - 4m'_1m'_3 + 6m'^2_1m'_2 - 3m'^4_1$.

Solución

Si $d = X - A$, entonces $X = A + d$, $\bar{X} = A + \bar{d}$ y $X - \bar{X} = d - \bar{d}$. Luego:

(*a*)

$$\begin{aligned} m_2 &= \overline{(X-\bar{X})^2} = \overline{(d-\bar{d})^2} = \overline{d^2 - 2\bar{d}d + \bar{d}^2} \\ &= \overline{d^2} - 2\bar{d}^2 + \bar{d}^2 = \overline{d^2} - \bar{d}^2 = m'_2 - m'^2_1 \end{aligned}$$

(*b*)

$$\begin{aligned} m_3 &= \overline{(X-\bar{X})^3} = \overline{(d-\bar{d})^3} = \overline{(d^3 - 3d^2\bar{d} + 3d\bar{d}^2 - \bar{d}^3)} \\ &= \overline{d^3} - 3\bar{d}\overline{d^2} + 3\bar{d}^3 - \bar{d}^3 = \overline{d^3} - 3\bar{d}\overline{d^2} + 2\bar{d}^3 = m'_3 - 3m'_1m'_2 + 2m'^3_1 \end{aligned}$$

(c)
$$m_4 = \overline{(X - \bar{X})^4} = \overline{(d - \bar{d})^4} = \overline{(d^4 - 4d^3\bar{d} + 6d^2\bar{d}^2 - 4d\bar{d}^3 + \bar{d}^4)}$$
$$= \overline{d^4} - 4\bar{d}\overline{d^3} + 6\bar{d}^2\overline{d^2} - 4\bar{d}^4 + \bar{d}^4 = \overline{d^4} - 4\bar{d}\overline{d^3} + 6\bar{d}^2\overline{d^2} - 3\bar{d}^4$$
$$= m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4$$

Por extensión de este método, se pueden deducir resultados similares para m_5, m_6, etc.

CALCULO DE MOMENTOS PARA DATOS AGRUPADOS

5.6. Hallar los cuatro primeros momentos respecto de la media para la distribución de alturas del Problema 3.22.

Solución

Tabla 5.1

X	u	f	fu	fu^2	fu^3	fu^4
61	−2	5	−10	20	−40	80
64	−1	18	−18	18	−18	18
67	0	42	0	0	0	0
70	1	27	27	27	27	27
73	2	8	16	32	64	128
		$N=\sum f=100$	$\sum fu=15$	$\sum fu^2=97$	$\sum fu^3=33$	$\sum fu^4=253$

El trabajo lo resume la Tabla 5.1, de la que vemos que

$$m_1' = c\frac{\sum fu}{N} = (3)\left(\frac{15}{100}\right) = 0.45 \qquad m_3' = c^3\frac{\sum fu^3}{N} = (3)^3\left(\frac{33}{100}\right) = 8.91$$

$$m_2' = c^2\frac{\sum fu^2}{N} = (3)^2\left(\frac{97}{100}\right) = 8.73 \qquad m_4' = c^4\frac{\sum fu^4}{N} = (3)^4\left(\frac{253}{100}\right) = 204.93$$

Así pues,

$$m_1 = 0$$
$$m_2 = m_2' - m_1'^2 = 8.73 - (0.45)^2 = 8.5275$$
$$m_3 = m_3' - 3m_1'm_2' + m_1'^3 = 8.91 - 3(0.45)(8.73) + 2(0.45)^3 = -2.6932$$
$$m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4$$
$$= 204.93 - 4(0.45)(8.91) + 6(0.45)^2(8.73) - 3(0.45)^4 = 199.3759$$

5.7. Calcular: (a) m_1', (b) m_2', (c) m_3', (d) m_4', (e) m_1, (f) m_2, (g) m_3, (h) m_4, (i) $\bar{X}$, (j) s, (k) $\overline{X^2}$ y (l) $\overline{X^3}$ para la distribución de la Tabla 4.7 del Problema 4.19.

Solución

Procédase como indica la Tabla 5.2.

Tabla 5.2

X	u	f	fu	fu^2	fu^3	fu^4
70	−6	4	−24	144	−864	5184
74	−5	9	−45	225	−1125	5625
78	−4	16	−64	256	−1024	4096
82	−3	28	−84	252	−756	2268
86	−2	45	−90	180	−360	720
90	−1	66	−66	66	−66	66
$A \longrightarrow$ 94	0	85	0	0	0	0
98	1	72	72	72	72	72
102	2	54	108	216	432	864
106	3	38	114	342	1026	3078
110	4	27	108	432	1728	6912
114	5	18	90	450	2250	11250
118	6	11	66	396	2376	14256
122	7	5	35	245	1715	12005
126	8	2	16	128	1024	8192
		$N=\sum f=480$	$\sum fu=236$	$\sum fu^2=3404$	$\sum fu^3=6428$	$\sum fu^4=74{,}588$

(*a*) $m_1' = c\dfrac{\sum fu}{N} = (4)\left(\dfrac{236}{480}\right) = 1.9667$

(*b*) $m_2' = c^2\dfrac{\sum fu^2}{N} = (4)^2\left(\dfrac{3404}{480}\right) = 113.4667$

(*c*) $m_3' = c^3\dfrac{\sum fu^3}{N} = (4)^3\left(\dfrac{6428}{480}\right) = 857.0667$

(*d*) $m_4' = c^4\dfrac{\sum fu^4}{N} = (4)^4\left(\dfrac{74{,}588}{480}\right) = 39{,}780.2667$

(*e*) $m_1 = 0$

(*f*) $m_2 = m_2' - m_1'^2 = 113.4667 - (1.9667)^2 = 109.5988$

(*g*) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3 = 857.0667 - 3(1.9667)(113.4667) + 2(1.9667)^3 = 202.8158$

(*h*) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4 = 35{,}627.2853$

(*i*) $\bar{X} = \overline{(A + d)} = A + m_1' = A + c\dfrac{\sum fu}{N} = 94 + 1.9667 = 95.97$

(*j*) $s = \sqrt{m_2} = \sqrt{109.5988} = 10.47$

(*k*) $\overline{X^2} = \overline{(A + d)^2} = \overline{(A^2 + 2Ad + d^2)} = A^2 + 2A\bar{d} + \overline{d^2} = A^2 + 2Am_1' + m_2'$
$= (94)^2 + 2(94)(1.9667) + 113.4667 = 9319.2063$, o sea 9319 con cuatro cifras significativas

(*l*) $\overline{X^3} = \overline{(A + d)^3} = \overline{(A^3 + 3A^2d + 3Ad^2 + d^3)} = A^3 + 3A^2\bar{d} + 3A\overline{d^2} + \overline{d^3}$
$= A^3 + 3A^2m_1' + 3Am_2' + m_3' = 915{,}571.9597$, o sea 915,600 con cuatro cifras significativas

COMPROBACION DE CHARLIER

5.8. Ilustrar el uso de la comprobación de Charlier en los cálculos del Problema 5.7.

Solución

Para ello, sumamos a la Tabla 5.2 las columnas de la Tabla 5.3 (excepto la columna 2 que se repite en la Tabla 5.3 por conveniencia).

En cada uno de los siguientes agrupamientos, el primero está sacado de la Tabla 5.3 y el segundo de la Tabla 5.2. La igualdad de los resultados en cada grupo proporciona la deseada comprobación.

$\sum f(u+1)=716$

$\sum fu+N=236+480=716$

$\sum f(u+1)^2=4356$

$\sum fu^2+2\sum fu+N=3404+2(236)+480=4356$

$\sum f(u+1)^3=17{,}828$

$\sum fu^3+3\sum fu^2+3\sum fu+N=6428+3(3404)+3(236)+480=17{,}828$

$\sum f(u+1)^4=122{,}148$

$\sum fu^4+4\sum fu^3+6\sum fu^2+4\sum fu+N=74{,}588+4(6428)+6(3404)+4(236)+480=122{,}148$

Tabla 5.3

$u+1$	f	$f(u+1)$	$f(u+1)^2$	$f(u+1)^3$	$f(u+1)^4$
−5	4	−20	100	−500	2500
−4	9	−36	144	−576	2304
−3	16	−48	144	−432	1296
−2	28	−56	112	−224	448
−1	45	−45	45	−45	45
0	66	0	0	0	0
1	85	85	85	85	85
2	72	144	288	576	1152
3	54	162	486	1458	4374
4	38	152	608	2432	9728
5	27	135	675	3375	16875
6	18	108	648	3888	23328
7	11	77	539	3773	26411
8	5	40	320	2560	20480
9	2	18	162	1458	13122
	$N=\sum f=$ $=480$	$\sum f(u+1)=$ $=716$	$\sum f(u+1)^2=$ $=4356$	$\sum f(u+1)^3=$ $=17828$	$\sum f(u+1)^4=$ $=122148$

CORRECCIONES DE SHEPPARD PARA LOS MOMENTOS

5.9. Aplicar las correcciones de Sheppard para determinar los momentos respecto de la media para los datos en: (*a*) Problema 5.6 y (*b*) Problema 5.7.

Solución

(*a*) m_2 corregido $= m_2 - c^2/12 = 8.5275 - 3^2/12 = 7.7775$

m_4 corregido $= m_4 - \frac{1}{2}c^2m_2 + \frac{7}{240}c^4$

$= 199.3759 - \frac{1}{2}(3)^2(8.5275) + \frac{7}{240}(3)^4$

$= 163.3646$

m_1 y m_3 no necesitan corrección

(*b*) m_2 corregido $= m_2 - c^2/12 = 109.5988 - 4^2/12 = 108.2655$

m_4 corregido $= m_4 - \frac{1}{2}c^2m_2 + \frac{7}{240}c^4$

$= 35{,}627.2853 - \frac{1}{2}(4)^2(109.5988) + \frac{7}{240}(4)^4$

$= 34{,}757.9616$

SESGO

5.10. Hallar el (*a*) primero y (*b*) segundo coeficientes de Pearson de sesgo para la distribución salarial de los 65 empleados de la empresa P&R (véanse Probs. 3.44 y 4.18).

Solución

Media = \$279.76, mediana = \$279.06, moda = \$277.50 y la desviación típica s = \$15.60. Así pues:

(*a*) $\text{Primer coeficiente de sesgo} = \dfrac{\text{media} - \text{moda}}{s} = \dfrac{\$279.76 - \$277.50}{\$15.60} = 0.1448$, o sea 0.14

(*b*) $\text{Segundo coeficiente de sesgo} = \dfrac{3(\text{media} - \text{mediana})}{s} = \dfrac{3(\$279.76 - \$279.06)}{\$15.60} = 0.1346$, o sea 0.13

Si se usa la desviación típica corregida [véase Prob. 4.21(*b*)], estos coeficientes pasan a ser, respectivamente:

(*a*) $\dfrac{\text{media} - \text{moda}}{s \text{ corregida}} = \dfrac{\$279.76 - \$277.50}{\$15.33} = 0.1474$, o sea 0,15

(*b*) $\dfrac{3(\text{media} - \text{mediana})}{s \text{ corregida}} = \dfrac{3(\$279.76 - \$279.06)}{\$15.33} = 0.1370$, o sea 0.14

Como los coeficientes son positivos, la distribución tiene sesgo positivo (o sea, a la derecha).

5.11. Hallar el coeficiente: (*a*) cuartil y (*b*) percentil de sesgo para la distribución del Problema 5.10 (véase Problema 3.44).

Solución

Q_1 = \$268.25, $Q_2 = P_{50}$ = \$279.06, Q_3 = \$290.75, $P_{10} = D_1$ = \$258.12 y $P_{90} = D_9$ = \$301.00. Luego:

(*a*) $\text{Coeficiente cuartil de sesgo} = \dfrac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \dfrac{\$290.75 - 2(\$279.06) + \$268.25}{\$290.75 - \$268.25} = 0.0391$

(*b*) $\text{Coeficiente percentil de sesgo} = \dfrac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} = \dfrac{\$301.00 - 2(\$279.06) + \$258.12}{\$301.00 - \$258.12} = 0.0233$

5.12. Hallar el coeficiente momento de sesgo a_3 para: (*a*) la distribución de alturas de estudiantes universitarios del Problema 5.6 y (*b*) los IQ de alumnos de escuela elemental del Problema 5.7.

Solución

(*a*) $m_2 = s^2 = 8.5275$ y $m_3 = -2.6932$. Luego

$$a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{-2.6932}{(\sqrt{8.5275})^3} = -0.1081 \quad \text{o sea} \quad -0.11$$

Si se usan correcciones de Sheppard para agrupar [véase Prob. 5.9(*a*)], entonces

$$a_3 \text{ corregido} = \frac{m_3}{(\sqrt{m_2 \text{ corregido}})^3} = \frac{-2.6932}{(\sqrt{7.7775})^3} = -0.1242 \quad \text{o sea} \quad -0.12$$

(*b*)

$$a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{202.8158}{(\sqrt{109.5988})^3} = 0.1768 \quad \text{o sea} \quad 0.18$$

Si se usan correcciones de Sheppard para agrupar [véase Prob. 5.9(*a*)], entonces

$$a_3 \text{ corregido} = \frac{m_3}{(\sqrt{m_2 \text{ corregido}})^3} = \frac{202.8158}{(\sqrt{108.2655})^3} = 0.1800 \quad \text{o sea} \quad 0.18$$

Nótese que ambas distribuciones son poco sesgadas, la (*a*) a la izquierda (negativamente) y la (*b*) a la derecha (positivamente). La (*b*) es más sesgada que la (*a*); esto es, (*a*) es más simétrica que (*b*), como queda patente por el hecho de que el valor numérico (o valor absoluto) del coeficiente de sesgo es mayor para (*b*) que para (*a*).

CURTOSIS

5.13. Hallar el coeficiente momento de curtosis a_4 para los datos de: (*a*) Problema 5.6 y (*b*) Problema 5.7.

Solución

(*a*)

$$a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} = \frac{199.3759}{(8.5275)^2} = 2.7418 \quad \text{o sea} \quad 2.74$$

Si se usan correcciones de Sheppard [véase Prob. 5.9(*a*)], entonces

$$a_4 \text{ corregido} = \frac{m_4 \text{ corregido}}{(m_2 \text{ corregido})^2} = \frac{163.3646}{(7.7775)^2} = 2.7007 \quad \text{o sea} \quad 2.70$$

(*b*)

$$a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} = \frac{35{,}627.2853}{(109.5988)^2} = 2.9660 \quad \text{o sea} \quad 2.97$$

Si se usan correcciones de Sheppard [véase Prob. 5.9(*b*)], entonces

$$a_4 \text{ corregido} = \frac{m_4 \text{ corregido}}{(m_2 \text{ corregido})} = \frac{34{,}757.9616}{(108.2655)^2} = 2.9653 \quad \text{o sea} \quad 2.97$$

Como para una distribución normal $a_4 = 3$, se sigue que ambas distribuciones, (*a*) y (*b*), son platicúrticas con respecto a la normal (o sea, más aplastadas que la distribución normal).

En lo referente a aplastamiento, la distribución (*b*) se aproxima a la normal mucho más que la (*a*). Sin embargo, sabemos del Problema 5.12 que en lo concerniente a la simetría, la (*a*) se aproxima más a la normal.

5.14. (*a*) Calcular el coeficiente percentil de curtosis $\kappa = Q/(P_{90} - P_{10})$, para la distribución del Problema 5.11.
(*b*) ¿Se aproximaría bien por una distribución normal?

Solución

(*a*) $Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(\$290.75 - \$268.25) = \$11.25$, $P_{90} - P_{10} = \$301.00 - \$258.12 = \$42.88$. Por tanto $\kappa = Q/(P_{90} - P_{10}) = 0.262$.

(*b*) Como para la distribución normal κ vale 0.263, se sigue que la distribución dada es mesocúrtica (o sea de aplastamiento más o menos normal). Así pues, la curtosis es la misma que para una distribución normal y nos lleva a creer que sería bien aproximada por ella, al menos en lo referente a curtosis.

PROBLEMAS SUPLEMENTARIOS

MOMENTOS

5.15. Hallar los cuatro primeros momentos del conjunto 4, 7, 5, 9, 8, 3, 6.

5.16. Hallar los cuatro primeros momentos respecto de la media para el conjunto de números del Problema 5.15.

5.17. Hallar los cuatro primeros momentos respecto del número 7 para el conjunto de números del Problema 5.15.

5.18. Usando los resultados de los Problemas 5.16 y 5.17, verificar las relaciones entre momentos: (*a*) $m_2 = m'_2 - m'^2_1$, (*b*) $m_3 = m'_3 - 3m'_1m'_2 + 2m'^3_1$ y (*c*) $m_4 = m'_4 - 4m'_1m'_3 + 6m'^2_1m'_2 - 3m'^4_1$.

5.19. Hallar los cuatro primeros momentos respecto de la media para el conjunto de números de la progresión aritmética 2, 5, 8, 11, 14, 17.

5.20. Probar que: (*a*) $m'_2 = m_2 + h^2$, (*b*) $m'_3 = m_3 + 3hm_2 + h^3$ y (*c*) $m'_4 = m_4 + 4hm_3 + 6h^2m_2 + h^4$, donde $h = m'_1$.

5.21. Si el primer momento respecto del número 2 es 5, ¿cuál es la media?

5.22. Si los primeros cuatro momentos de un conjunto de números respecto del número 3 son -2, 10, -25 y 50, determinar los correspondientes momentos respecto de: (*a*) la media, (*b*) el número 5 y (*c*) el cero.

5.23. Hallar los cuatro primeros momentos respecto de la media para el conjunto de números 0, 0, 0, 1, 1, 1, 1 y 1.

5.24. (*a*) Probar que $m_5 = m'_5 - 5m'_1m'_4 + 10m'^2_1m'_3 - 10m'^3_1m'_2 + 4m'^5_1$.
(*b*) Deducir una fórmula similar para m_6.

5.25. De un total de N números, la fracción p son unos y la fracción $q = 1 - p$ son ceros. Hallar: (*a*) m_1, (*b*) m_2, (*c*) m_3 y (*d*) m_4.

5.26. Probar que los primeros cuatro momentos respecto de la media de la progresión aritmética $a, a + d, a + 2d, \ldots, a + (n - 1)d$ son $m_1 = 0$, $m_2 = \frac{1}{12}(n^2 - 1)d^2$, $m_3 = 0$ y $m_4 = \frac{1}{240}(n^2 - 1)(3n^2 - 7)d^4$. Comparar con el Problema 5.19 (véase también el Problema 4.69). [*Ayuda:* $1^4 + 2^4 + 3^4 + \cdots + (n-1)^4 = \frac{1}{30}n(n-1)(2n-1)(3n^2-3n-1)$.]

MOMENTOS PARA DATOS AGRUPADOS

5.27. Calcular los primeros cuatro momentos respecto de la media para la distribución de la Tabla 5.4.

Tabla 5.4

X	f
12	1
14	4
16	6
18	10
20	7
22	2
Total	30

5.28. Ilustrar el uso de la comprobación de Charlier para los cálculos del Problema 5.27.

5.29. Aplicar las correcciones de Sheppard a los momentos obtenidos en el Problema 5.27.

5.30. Hallar los cuatro primeros momentos respecto de la media para la distribución del Problema 3.59: (*a*) sin correcciones de Sheppard y (*b*) con correcciones de Sheppard.

5.31. Hallar: (*a*) m_1, (*b*) m_2, (*c*) m_3, (*d*) m_4, (*e*) $\bar{X}$, (*f*) s, (*g*) $\overline{X^2}$, (*h*) $\overline{X^3}$, (*i*) $\overline{X^4}$ y (*j*) $\overline{(X+1)^3}$ para la distribución del Problema 3.62.

SESGO

5.32. Hallar el coeficiente momento de sesgo a_3 para la distribución del Problema 5.27: (*a*) sin y (*b*) con correcciones de Sheppard.

5.33. Hallar el coeficiente momento de sesgo a_3 para la distribución del Problema 3.59 (véase Problema 5.30).

5.34. Los segundos momentos respecto de la media de dos distribuciones son 9 y 16, mientras que los terceros momentos respecto de la media son -8.1 y -12.8, respectivamente. ¿Qué distribución es más sesgada a la izquierda?

5.35. Hallar los coeficientes de Pearson: (*a*) primero y (*b*) segundo, para la distribución del Problema 3.59, y explicar la diferencia.

5.36. Hallar el coeficiente de sesgo: (*a*) cuartil y (*b*) percentil, para la distribución del Problema 3.59. Comparar los resultados con los del Problema 5.35 y explicar lo que se aprecie.

5.37. (*a*) Explicar por qué los coeficientes de sesgo de Pearson no son apropiados para la distribución del Problema 2.31.
(*b*) Hallar el coeficiente cuartil de sesgo para ella e interpretar el resultado.

CURTOSIS

5.38. Hallar el coeficiente momento de curtosis a_4 para la distribución del Problema 5.27: (*a*) sin y (*b*) con correcciones de Sheppard.

5.39. Hallar el coeficiente momento de curtosis para la distribución del Problema 3.59: (*a*) sin y (*b*) con correcciones de Sheppard (véase Problema 5.30).

5.40. Los cuartos momentos respecto de la media de las distribuciones del Problema 5.34 son 230 y 780, respectivamente. ¿Qué distribución se aproxima más a la normal desde el punto de vista de: (*a*) aplastamiento y (*b*) sesgo?

5.41. ¿Cuál de las distribuciones del Problema 5.40 es: (*a*) leptocúrtica, (*b*) mesocúrtica y (*c*) platicúrtica?

5.42. La desviación típica de una distribución simétrica es 5. ¿Cuál debe ser el valor del cuarto momento respecto de la media para que la distribución sea: (*a*) leptocúrtica, (*b*) mesocúrtica y (*c*) platicúrtica?

5.43. (*a*) Calcular el coeficiente percentil de curtosis para la distribución del Problema 3.59.
(*b*) Comparar el resultado con el valor teórico 0.263 para la normal e interpretar.
(*c*) ¿Cómo se puede reconciliar este resultado con el del Problema 5.39?

CAPITULO 6

Teoría elemental de probabilidades

DEFINICIONES DE PROBABILIDAD

Definición clásica

Supongamos que un suceso E tiene h posibilidades de ocurrir entre un total de n posibilidades, cada una de las cuales tiene la misma oportunidad de ocurrir que las demás. Entonces, la probabilidad de que ocurra E (o sea un *éxito*) se denota por

$$p = \Pr\{E\} = \frac{h}{n}$$

La probabilidad de que no ocurra E (o sea, un *fracaso*) se denota por

$$q = \Pr\{\text{no } E\} = \frac{n-h}{n} = 1 - \frac{h}{n} = 1 - p = 1 - \Pr\{E\}$$

Así pues, $p + q = 1$, es decir, $\Pr\{E\} + \Pr\{\text{no } E\} = 1$. El suceso «no E» se denotará por $\bar{E}$, $\tilde{E}$ o $\sim E$.

EJEMPLO 1. Sea E el suceso de que al tirar un dado una vez salga un 3 o un 4. Hay seis formas de caer el dado, dando 1, 2, 3, 4, 5 ó 6; y si el dado es bueno (no trucado), como se supondrá en todo lo que sigue salvo mención explícita, podemos suponer que las seis tienen la misma oportunidad de salir. Como E puede ocurrir de dos formas, tenemos $p = \Pr\{E\} = \frac{2}{6} = \frac{1}{3}$.

La probabilidad de que no salga ni 3 ni 4 (o sea, de que salga 1, 2, 5 ó 6) es $q = \Pr\{\bar{E}\} = 1 - \frac{1}{3} = \frac{2}{3}$.

Nótese que la probabilidad de un suceso es un número entre 0 y 1. Si un suceso es imposible, su probabilidad es 0. Si un suceso debe ocurrir necesariamente (suceso seguro) su probabilidad es 1.

Si p es la probabilidad de que ocurra un suceso, las *apuestas* a su favor están $p:q$ (léase «p a q»). Luego las apuestas en su contra están $q:p$. Así, las apuestas contra la aparición de un 3 o un 4 al lanzar un dado bueno son $q:p = \frac{2}{3}:\frac{1}{3} = 2:1$ (o sea, 2 a 1).

Definición como frecuencia relativa

La definición clásica de probabilidad tiene la pega de que las palabras «misma oportunidad» aparecen como sinónimas de «equiprobables», lo cual produce un círculo vicioso. Por ello, algunos defienden una definición estadística de la probabilidad. Para ellos, la probabilidad estimada, o *probabilidad empírica*, de un suceso se toma como la *frecuencia relativa* de ocurrencia del suceso cuando el número de observaciones es muy grande. La probabilidad misma es el *límite* de esa frecuencia relativa cuando el número de observaciones crece indefinidamente.

EJEMPLO 2. Si en 1000 tiradas de una moneda salen 529 caras, la frecuencia relativa de caras es 529/1000 = = 0.529. Si en otros 1000 lanzamientos salen 493 caras, la frecuencia relativa en el total de 2000 tiradas es (529 + + 493)/2000 = 0.511. De acuerdo con la definición estadística, continuando de este modo nos iremos acercando más y más a un número que representa la probabilidad de que salga cara en una sola tirada. De los resultados presentados, éste sería 0.5, con un dígito significativo. Para obtener más dígitos habría que hacer más tiradas.

La definición estadística, si bien útil en la práctica, tiene una desventaja matemática en el hecho de que un límite puede no existir. Por esa razón, la moderna teoría de la probabilidad es *axiomática* y deja el concepto de probabilidad sin definir, al igual que sucede en geometría con el punto y la recta.

PROBABILIDAD CONDICIONAL; SUCESOS INDEPENDIENTES Y SUCESOS DEPENDIENTES

Si E_1 y E_2 son dos sucesos, la probabilidad de que E_2 ocurra dado que haya ocurrido E_1 se denota por $\Pr\{E_2 \mid E_1\}$, o $\Pr\{E_2 \text{ dado } E_1\}$, y se llama la *probabilidad condicional* de E_2 dado E_1.

Si la ocurrencia o no de E_1 no afecta para nada la probabilidad de ocurrencia de E_2, entonces $\Pr\{E_2 \mid E_1\} = \Pr\{E_2\}$, y diremos que E_1 y E_2 son *sucesos independientes;* en caso contrario, se dirá que son *sucesos dependientes*.

Si denotamos por E_1E_2 el suceso de que «ambos E_1 y E_2 ocurran», llamado un suceso compuesto, entonces

$$\Pr\{E_1E_2\} = \Pr\{E_1\}\ \Pr\{E_2 \mid E_1\} \tag{1}$$

En particular,

$$\Pr\{E_1E_2\} = \Pr\{E_1\}\ \Pr\{E_2\} \quad \text{para sucesos independientes} \tag{2}$$

Para tres sucesos E_1, E_2 y E_3, tenemos

$$\Pr\{E_1E_2E_3\} = \Pr\{E_1\}\ \Pr\{E_2 \mid E_1\}\ \Pr\{E_3 \mid E_1E_2\} \tag{3}$$

Esto es, la probabilidad de que ocurran E_1, E_2 y E_3 es igual a (la probabilidad de E_1) × (la probabilidad de E_2 dado E_1) × (la probabilidad de E_3 dados E_1 y E_2). En particular,

$$\Pr\{E_1E_2E_3\} = \Pr\{E_1\}\ \Pr\{E_2\}\ \Pr\{E_3\} \quad \text{para sucesos independientes} \tag{4}$$

En general, si $E_1, E_2, E_3, ..., E_n$ son n sucesos independientes con probabilidades respectivas $p_1, p_2, p_3, ..., p_n$, entonces la probabilidad de que ocurran E_1 y E_2 y E_3 y $\cdots$ E_n es $p_1p_2p_3 \cdots p_n$.

EJEMPLO 3. Sean E_1 y E_2 los sucesos «cara en el quinto lanzamiento» y «cara en el sexto lanzamiento» de una moneda, respectivamente. Entonces, E_1 y E_2 son sucesos independientes y, por tanto, la probabilidad de que salga cara en ambos intentos (supuesta la moneda no trucada, aquí y en lo que sigue) es

$$\Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2\} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

EJEMPLO 4. Si las probabilidades de A y B de estar vivos dentro de 20 años son 0.7 y 0.5, respectivamente, entonces la probabilidad de que ambos lo estén es $(0.7)(0.5) = 0.35$.

EJEMPLO 5. Una caja contiene 3 bolas blancas y 2 bolas negras. Sea E_1 el suceso «la primera bola extraída es negra» y E_2 el suceso «la segunda bola extraída es negra». Las bolas extraídas no se devuelven a la caja. E_1 y E_2 son sucesos dependientes.

La probabilidad de que la primera bola sea negra es $\Pr\{E_1\} = 2/(3+2) = \frac{2}{5}$. La probabilidad de que la segunda sea negra, dado que ya lo haya sido la primera, es $\Pr\{E_2 \mid E_1\} = 1/(3+1) = \frac{1}{4}$. Luego la probabilidad de que ambas sean negras es

$$\Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2 \mid E_1\} = \frac{2}{5}\cdot\frac{1}{4} = \frac{1}{10}$$

SUCESOS MUTUAMENTE EXCLUYENTES Unión

Dos o más sucesos se llaman sucesos *mutuamente excluyentes* si la ocurrencia de cualquiera de ellos excluye la de los otros. De modo que si E_1 y E_2 son sucesos mutuamente excluyentes, entonces $\Pr\{E_1E_2\} = 0$.

Si $E_1 + E_2$ denota el suceso de que «ocurra E_1 o bien E_2 o ambos a la vez», entonces

$$\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1E_2\} \tag{5}$$

En particular,

$$\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} \quad \text{para sucesos mutuamente excluyentes} \tag{6}$$

Como extensión de esto, si $E_1, E_2, \ldots, E_n$ son n sucesos mutuamente excluyentes con probabilidades respectivas E_1 o E_2 o $\cdots$ E_n es $p_1 + p_2 + \cdots + p_n$.

El resultado (5) se puede generalizar a tres o más sucesos mutuamente excluyentes (véase Problema 6.38).

EJEMPLO 6. Sean E_1 el suceso «sacar un as de una baraja» y E_2 «sacar un rey». Entonces $\Pr\{E_1\} = \frac{4}{52} = \frac{1}{13}$ y $\Pr\{E_2\} = \frac{4}{52} = \frac{1}{13}$. La probabilidad de sacar o un as o un rey en un solo ensayo es

$$\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$$

pues no es posible sacar ambos a la vez, y son, por tanto, sucesos mutuamente excluyentes.

EJEMPLO 7. Sean E_1 el suceso «sacar un as» de una baraja y E_2 «sacar una espada». Entonces E_1 y E_2 no son sucesos mutuamente excluyentes, porque puede sacarse el as de espadas. Luego la probabilidad de sacar un as o una espada o ambos es

$$\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1E_2\} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

DISTRIBUCIONES DE PROBABILIDAD

Discretas

Si una variable X puede tomar un conjunto discreto de valores $X_1, X_2, ..., X_K$, con probabilidades respectivas $p_1, p_2, ..., p_K$, donde $p_1 + p_2 + \cdots + p_K = 1$, decimos que tenemos definida una *distribución de probabilidad discreta* para X. La función $p(X)$, que tiene valores $p_1, p_2, ..., p_K$ para $X = X_1, X_2, ..., X_K$, se llama *función de probabilidad* o una *función de frecuencia* de X. Como X puede tomar ciertos valores con ciertas probabilidades, se le llama una *variable aleatoria discreta*. Una variable aleatoria se conoce también como *variable estocástica*.

EJEMPLO 8. Sea X la suma de puntos obtenida al lanzar dos dados. La distribución de probabilidad se muestra en la Tabla 6.1. Por ejemplo, la probabilidad de obtener suma 5 es $\frac{4}{36} = \frac{1}{9}$; así que en 900 tiradas se esperan 100 tiradas con suma 5.

Tabla 6.1

X	2	3	4	5	6	7	8	9	10	11	12
$p(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Nótese que esto es análogo a una distribución de frecuencias relativa, con probabilidad en lugar de frecuencia relativa. De manera que podemos pensar en las distribuciones de probabilidad como formas teóricas o ideales en el límite, de distribuciones de frecuencia relativa cuando el número de observaciones es muy grande. Por eso podemos pensar en las distribuciones de probabilidad como si fueran distribuciones de *poblaciones*, mientras que las distribuciones de frecuencia relativa son distribuciones de *muestras* de esa población.

La distribución de probabilidad se puede representar gráficamente dibujando $p(X)$ versus X, igual que para las distribuciones de frecuencia relativa (véase Prob. 6.11).

Acumulando probabilidades, obtenemos *distribuciones de probabilidad acumulada*, análogas a las distribuciones de frecuencia relativa acumulada. La función asociada con esa distribución se llama una *función de distribución*.

Continuas

Las ideas anteriores se extienden a variables X que pueden tomar un conjunto continuo de valores. El polígono de frecuencias relativas de una muestra se convierte, en el caso teórico o límite de una población, en una curva continua (como la de la Fig. 6.1) de ecuación $Y = p(X)$. El área total bajo esa curva y sobre el eje X es 1, y el área entre $X = a$ y $X = b$ (sombreada en la figura) da la probabilidad de que X esté entre a y b, que se denota por $\Pr\{a < X < b\}$.

Llamamos a $p(x)$ una *función densidad de probabilidad*, o brevemente una *función densidad*, y cuando tal función es dada decimos que se ha definido una *distribución de probabilidad continua* para X. La variable X se llama entonces una *variable aleatoria continua*.

Como en el caso discreto, podemos definir distribuciones de probabilidad acumulada y las asociadas funciones de distribución.

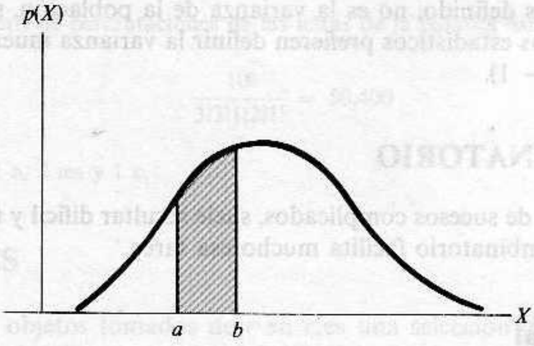


Figura 6.1.

ESPERANZA MATEMATICA

Si p es la probabilidad de que una persona reciba una cantidad S de dinero, la *esperanza matemática* (o simplemente *esperanza*) se define como pS.

EJEMPLO 9. Si la probabilidad de que un hombre gane un premio de \$10 es 1/5, su esperanza matemática es $\frac{1}{5}(\$10) = \$2$.

El concepto de esperanza matemática se extiende fácilmente. Si X denota una variable aleatoria discreta que puede tomar los valores $X_1, X_2, ..., X_K$ con probabilidades $p_1, p_2, ..., p_K$, donde $p_1 + p_2 + + \cdots + p_K = 1$, la *esperanza matemática* de X (o simplemente *esperanza* de X), denotada $E(X)$, y se define como

$$E(X) = p_1X_1 + p_2X_2 + \cdots + p_KX_K = \sum_{j=1}^{K} p_jX_j = \sum pX \tag{7}$$

Si las probabilidades p_j en esa expresión se sustituyen por las frecuencias relativas f_j/N, donde $N = \sum f_j$, la esperanza matemática se reduce a $(\sum fX)/N$, que es la media aritmética $\bar{X}$ de una muestra de tamaño N en la que $X_1, X_2, ..., X_K$ aparecen con estas frecuencias relativas. Al crecer N más y más, las frecuencias relativas se acercan a las probabilidades p_j. Así que nos vemos abocados a interpretar $E(X)$ como la media de la población cuyo muestreo se consideraba. Si llamamos m a la media muestral, podemos denotar la media poblacional por la correspondiente letra griega μ (mu).

Puede definirse, asimismo, la esperanza matemática para variables aleatorias continuas, pero requiere el cálculo.

RELACION ENTRE POBLACION, MEDIA MUESTRAL Y VARIANZA

Si seleccionamos una muestra de tamaño N al azar de una población (o sea, suponemos que todas las posibles muestras son igualmente probables), entonces es posible mostrar que *el valor esperado de la media muestral m es la media poblacional μ.*

No se deduce, sin embargo, que el valor esperado de cualquier cantidad calculada sobre una muestra sea la cantidad correspondiente de la población. Así, el valor esperado de la varianza muestral, como la hemos definido, no es la varianza de la población, sino $(N - 1)/N$ veces dicha varianza. Por eso algunos estadísticos prefieren definir la varianza muestral como nuestra varianza multiplicada por $N/(N - 1)$.

ANALISIS COMBINATORIO

Al hallar probabilidades de sucesos complicados, suele resultar difícil y tediosa una enumeración de los casos. El análisis combinatorio facilita mucho esa tarea.

Principio fundamental

Si un suceso puede ocurrir de n_1 maneras, y si cuando éste ha ocurrido otro suceso puede ocurrir de n_2 maneras, entonces el número de maneras en que ambos pueden ocurrir en el orden especificado es $n_1 n_2$.

EJEMPLO 10. Si hay 3 candidatos para gobernador y 5 para alcalde, los dos cargos pueden ocuparse de $3 \cdot 5 = 15$ formas.

Factorial de *n*

La factorial de n, denotada por $n!$, se define como

$$n! = n(n-1)(n-2)\cdots 1 \tag{8}$$

Así, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, y $4!3! = (4 \cdot 3 \cdot 2 \cdot 1)(3 \cdot 2 \cdot 1) = 144$. Conviene definir $0! = 1$.

Permutaciones

Una permutación de n objetos tomados de r en r es una *elección ordenada* de r objetos de entre n. El número de permutaciones de n objetos tomados de r en r se denota por ${}_nP_r$, $P(n, r)$, o $P_{n,r}$ y viene dado por

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \tag{9}$$

En particular, el número de permutaciones de n objetos tomados de n en n es

$${}_n^nP = n(n-1)(n-2)\cdots 1 = n!$$

EJEMPLO 11. El números de permutaciones que se pueden dar de las letras a, b y c tomadas de dos en dos es ${}_3P_2 = 3 \cdot 2 = 6$. Son ab, ba, ac, ca, bc y cb.

El número de permutaciones de n objetos, de los que n_1 son iguales, n_2 son iguales, ... es

$$\frac{n!}{n_1!n_2!\cdots} \quad \text{donde } n = n_1 + n_2 + \cdots \tag{10}$$

10

EJEMPLO 12. El número de permutaciones de las letras de la palabra «statistics» es

$$\frac{10!}{3!3!1!2!1!} = 50{,}400$$

porque hay 3 eses, 3 tes, 1 a, 2 ies y 1 c.

COMBINACIONES

Una combinación de n objetos tomados de r en r es una selección de r de ellos, sin importar el orden de los r escogidos. El número de combinaciones de n objetos, tomados de r en r se denota por $\binom{n}{r}$ y viene dado por

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \boxed{\frac{n!}{r!(n-r)!}} \tag{11}$$

EJEMPLO 13. El número de combinaciones de las letras a, b y c tomadas de dos en dos es

$$\binom{3}{2} = \frac{3 \cdot 2}{2!} = 3$$

Son ab, ac y bc. Nótese que ab es la misma combinación que ba, pero no la misma permutación.

APROXIMACION DE STIRLING A $n!$

Cuando n es grande, la evaluación directa de $n!$ es horrible. En tal caso, se usa una fórmula aproximada debida a James Stirling:

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n} \tag{12}$$

donde $e = 2.71828 \cdots$ es la base natural de logaritmos (véase Prob. 6.31).

RELACION DE LA PROBABILIDAD CON LA TEORIA DE CONJUNTOS

En la moderna teoría de probabilidad, se piensa en los posibles resultados de un ensayo, experimento, etc., como puntos de un espacio (que puede ser de 1, 2, 3, ..., dimensiones), llamado *espacio muestral S*. Si S contiene sólo un número finito de puntos, a cada punto está asociado un número no negativo, llamado *probabilidad*, tal que la suma de todos ellos es 1. Un suceso es un *conjunto* (o *colección*) de puntos de S, tal como E_1 o E_2 en la Figura 6.2; esa figura se llama un *diagrama de Euler* o de *Venn*.

El suceso $E_1 + E_2$ es el conjunto de puntos que *están en* E_1 *o en* E_2 *o en ambos*, y el suceso E_1E_2 es el conjunto de puntos *comunes a* E_1 *y a* E_2. Así que la probabilidad de un suceso tal como E_1 es la suma de las probabilidades asociadas a todos sus puntos. Análogamente, la probabilidad de $E_1 + E_2$, denotada $\Pr\{E_1 + E_2\}$, es la suma de las probabilidades asociadas a todos los puntos

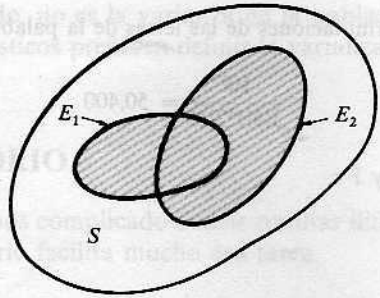


Figura 6.2.

contenidos en el conjunto $E_1 + E_2$. Si E_1 y E_2 no tienen puntos en común (o sea, si son sucesos mutuamente excluyentes), entonces $\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\}$. Si tienen puntos en común, entonces $\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1E_2\}$.

El conjunto $E_1 + E_2$ se denota a veces por $E_1 \cup E_2$ y se llama conjunto *unión* de los dos conjuntos. El conjunto E_1E_2 se suele denotar $E_1 \cap E_2$ y se llama *intersección* de los dos conjuntos. Cabe extender eso a más de dos conjuntos; así, en vez de $E_1 + E_2 + E_3$ y $E_1E_2E_3$, podríamos usar las notaciones $E_1 \cup E_2 \cup E_3$ y $E_1 \cap E_2 \cap E_3$, respectivamente.

El símbolo ϕ (letra griega *phi*) se usa para denotar el *conjunto vacío*, que no contiene punto alguno. La probabilidad asociada con un suceso correspondiente a este conjunto es cero (o sea, $\Pr\{\phi\} = 0$). Si E_1 y E_2 no tienen puntos en común, podemos escribir $E_1E_2 = \phi$, que significa que los correspondientes sucesos son sucesos mutuamente excluyentes, de donde $\Pr\{E_1E_2\} = 0$.

Con este enfoque moderno, una variable aleatoria es una función definida en cada punto del espacio muestral. Por ejemplo, en el Problema 6.36 la variable aleatoria es la suma de las coordenadas de cada punto.

En el caso de que S tenga infinitos puntos, lo anterior se extiende usando nociones del Cálculo.

PROBLEMAS RESUELTOS

REGLAS FUNDAMENTALES DE LA PROBABILIDAD

6.1. Determinar, o estimar, la probabilidad p de los siguientes sucesos:

(*a*) Una tirada de un dado resulte impar. $\frac{3}{6} = \frac{1}{2}$

(*b*) Al menos una cara en dos tiradas de una moneda.

(*c*) Un as, el 10 de diamantes o el 2 de picas aparezca al sacar una sola carta de una baraja francesa de 52 naipes.

(*d*) La suma de dos dados sea 7.

(*e*) Que aparezca una cruz en la próxima tirada de una moneda si han salido 56 caras de 100 tiradas previas.

Solución

(*a*) De los 6 casos equiprobables, tres (si salen 1, 3 ó 5) son favorables al suceso. Luego $p = \frac{3}{6} = \frac{1}{2}$.

(*b*) Si H denota cara y T cruz, pueden salir HH, HT, TH, y TT, con igual probabilidad. Sólo los tres primeros son favorables, luego $p = \frac{3}{4}$.

(*c*) El suceso puede ocurrir de 6 maneras (los 4 ases, el 10 de diamantes y el 2 de picas) de los 52 casos posibles. Luego $p = \frac{6}{52} = \frac{3}{26}$.

(*d*) Emparejando de todos los modos posibles las puntuaciones de lòs dos dados, hay $6 \cdot 6 = 36$ posibles casos. Pueden denotarse (1, 1), (2, 1), (3, 1), ..., (6, 6).

Las seis formas de que sumen 7 son (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) y (6, 1) [véase Prob. 6.37(*a*)]. Luego $p = \frac{6}{36} = \frac{1}{6}$.

(*e*) Como salieron $100 - 56 = 44$ cruces en 100 tiradas, la probabilidad estimada (o empírica) de una cruz es la frecuencia relativa $44/100 = 0.44$.

6.2. Un experimento consiste en tirar un dado y una moneda. Si E_1 es el suceso «cara» al tirar la moneda, y E_2 es el suceso «3 ó 6» al tirar el dado, enunciar en palabras el significado de:

(*a*) $\bar{E}_1$ (*c*) E_1E_2 (*e*) $\Pr\{E_1 \mid E_2\}$

(*b*) $\bar{E}_2$ (*d*) $\Pr\{E_1\bar{E}_2\}$ (*f*) $\Pr\{\bar{E}_1 + \bar{E}_2\}$.

Solución

(*a*) Cruz en la moneda y cualquier cosa en el dado.

(*b*) 1, 2, 4 ó 5 en el dado y cualquier cosa en la moneda.

(*c*) Cara en la moneda y 3 ó 6 en el dado.

(*d*) La probabilidad de cara en la moneda y 1, 2, 4 ó 5 en el dado.

(*e*) La probabilidad de cara en la monedá, dado que en el dado sale 3 ó 6.

(*f*) La probabilidad de cruz en la moneda o 1, 2, 4 ó 5 en el dado, o ambos.

6.3. Se saca al azar una bola de una caja que contiene 6 bolas rojas, 4 blancas y 5 azules. Hallar la probabilidad de que la bola extraída sea: (*a*) roja, (*b*) blanca, (*c*) azul, (*d*) no roja y (*e*) roja o blanca.

Solución

Denotemos *R*, *W* y *B* los sucesos de sacar una bola roja, blanca y azul, respectivamente. Entonces:

(*a*)

$$\Pr\{R\} = \frac{\text{formas de coger una bola roja}}{\text{formas totales de coger una bola}} = \frac{6}{6+4+5} = \frac{6}{15} = \frac{2}{5}$$

(*b*)

$$\Pr\{W\} = \frac{4}{6+4+5} = \frac{4}{15}$$

(*c*)

$$\Pr\{B\} = \frac{5}{6+4+5} = \frac{5}{15} = \frac{1}{3}$$

(*d*)

$$\Pr\{\bar{R}\} = 1 - \Pr\{R\} = 1 - \frac{2}{5} = \frac{3}{5} \quad \text{por la parte } (a)$$

(*e*)

$$\Pr\{R + W\} = \frac{\text{formas de coger una bola roja o una blanca}}{\text{formas totales de coger una bola}} = \frac{6+4}{6+4+5} = \frac{10}{15} = \frac{2}{3}$$

Otro método

$$\Pr\{R + W\} = \Pr\{\bar{B}\} = 1 - \Pr\{B\} = 1 - \frac{1}{3} = \frac{2}{3} \quad \text{por la parte } (c)$$

Nótese que $\Pr\{R + W\} = \Pr\{R\} + \Pr\{W\}$ (es decir, $\frac{2}{3} = \frac{2}{5} + \frac{4}{15}$). Esto ilustra la regla general $\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\}$ válida para sucesos mutuamente excluyentes E_1 y E_2.

6.4. Un dado se lanza dos veces. Hallar la probabilidad de obtener 4, 5 ó 6 en la primera tirada y 1, 2, 3 ó 4 en la segunda.

$P(E_1 E_2) = P(E_2/E_1)P(E_1)$

Solución

Sea E_1 = suceso «4, 5 ó 6» en la primera tirada, y E_2 = suceso «1, 2, 3 ó 4» en la segunda. Los diversos resultados de las dos tiradas se emparejan de $6 \times 6 = 36$ formas posibles, todas equiprobables. Las tres formas de salir el resultado apetecido en la primera y las cuatro de la segunda se emparejan de $3 \times 4 = 12$ formas, los casos favorables en que E_1 y E_2 ocurren ambos, es decir E_1E_2. Luego $\Pr\{E_1E_2\} = 12/36 = 1/3$.

Notemos que $\Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2\}$ (es decir, $\frac{1}{3} = \frac{3}{6} \cdot \frac{4}{6}$) es válida para los *sucesos independientes* E_1 y E_2.

6.5. De una baraja de 52 naipes, mezclados al azar, se sacan dos naipes. Hallar la probabilidad de que ambos sean ases si la primera extraída: (*a*) se devuelve a la baraja y (*b*) si no se devuelve.

Solución

Sea E_1 = suceso «as» en la primera extracción, y E_2 = suceso «as» en la segunda.

(*a*) Si se repone, E_1 y E_2 son sucesos independientes. Así pues, $\Pr\{\text{ambas sean ases}\} = \Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2\} = (\frac{4}{52})(\frac{4}{52}) = \frac{1}{169}$.

(*b*) Si no se repone, la primera carta se saca de entre 52 y la segunda de entre 51, luego ambas pueden sacarse de 52×51 formas, todas equiprobables.

Hay 4 casos favorables a E_1 y 3 a E_2, de modo que ambos, E_1 y E_2, o sea E_1E_2, pueden ocurrir de 4×3 formas. Luego $\Pr\{E_1E_2\} = (4 \cdot 3)/(52 \cdot 51) = \frac{1}{221}$.

Nótese que $\Pr\{E_2 \mid E_1\} = \Pr\{\text{la segunda es un as dado que la primera era un as}\} = \frac{3}{51}$. Por tanto, nuestro resultado ilustra la regla general de que $\Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2 \mid E_1\}$ cuando E_1 y E_2 son sucesos dependientes.

6.6. Se sacan sucesivamente 3 bolas de la caja del Problema 6.3. Hallar la probabilidad de que salgan en el orden roja, blanca, azul si cada bola: (*a*) se repone y (*b*) no se repone.

Solución

Sea R = suceso «roja» en la primera extracción, W = suceso «blanca» en la segunda y B = suceso «azul» en la tercera. Se pide $\Pr\{RWB\}$.

(*a*) Con reposición, R, W y B son sucesos independientes, luego

$$\Pr\{RWB\} = \Pr\{R\}\Pr\{W\}\Pr\{B\} = \left(\frac{6}{6+4+5}\right)\left(\frac{4}{6+4+5}\right)\left(\frac{5}{6+4+5}\right) = \left(\frac{6}{15}\right)\left(\frac{4}{15}\right)\left(\frac{5}{15}\right) = \frac{8}{225}$$

(*b*) Sin reposición, R, W y B son sucesos dependientes y

$$\Pr\{RWB\} = \Pr\{R\}\Pr\{W \mid R\}\Pr\{B \mid WR\} = \left(\frac{6}{6+4+5}\right)\left(\frac{4}{5+4+5}\right)\left(\frac{5}{5+3+5}\right)$$

$$= \left(\frac{6}{15}\right)\left(\frac{4}{14}\right)\left(\frac{5}{13}\right) = \frac{4}{91}$$

donde $\Pr\{B \mid WR\}$ es la probabilidad condicional de sacar una azul si ya han salido una blanca y una roja.

6.7. Hallar la probabilidad de que salga al menos un 4 en dos tiradas de un dado.

Solución

Sea E_1 = suceso «4» en la primera tirada, E_2 = suceso «4» en la segunda y $E_1 + E_2$ = suceso «4» en la primera o «4» en la segunda o en ambas = suceso de que salga al menos un 4. Se pide $\Pr\{E_1 + E_2\}$.

Primer método

El número de formas en que pueden salir los dos dados es $6 \times 6 = 36$. Además,

Número de formas de que salga E_1 pero no $E_2 = 5$
Número de formas de que salga E_2 pero no $E_1 = 5$
Número de formas de que salgan ambos E_1 y $E_2 = 1$

Luego el número de formas en que al menos uno de ellos sale es $5 + 5 + 1 = 11$ y, por tanto, $\Pr\{E_1 + E_2\} = \frac{11}{36}$.

Segundo método

Como E_1 y E_2 no son sucesos mutuamente excluyentes, $\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1E_2\}$. Además, como E_1 y E_2 son sucesos independientes, $\Pr\{E_1E_2\} = \Pr\{E_1\}\Pr\{E_2\}$. Entonces $\Pr\{E_1 + E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1\}\Pr\{E_2\} = \frac{1}{6} + \frac{1}{6} - (\frac{1}{6})(\frac{1}{6}) = \frac{11}{36}$.

Tercer método

$$\Pr\{\text{salir al menos un 4}\} + \Pr\{\text{no salga ningún 4}\} = 1$$

Por tanto

$$\Pr\{\text{al menos un 4}\} = 1 - \Pr\{\text{ningún 4}\}$$
$$= 1 - \Pr\{\text{ni 4 en la primera ni 4 en la segunda}\}$$
$$= 1 - \Pr\{\bar{E}_1\bar{E}_2\} = 1 - \Pr\{\bar{E}_1\}\Pr\{\bar{E}_2\}$$
$$= 1 - \left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \frac{11}{36}$$

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

$\left(\frac{1}{6} + \frac{1}{6}\right) - \left(\frac{1}{6}\cdot\frac{1}{6}\right)$

$\frac{4}{6}\cdot\frac{4}{6} + \frac{1}{36} + \frac{1}{36}$

6.8. Una bolsa contiene 4 bolas blancas y 2 bolas negras; otra contiene 3 bolas blancas y 5 bolas negras. Si se saca una bola de cada bolsa, hallar la probabilidad de que: (*a*) ambas sean blancas, (*b*) ambas sean negras y (*c*) una sea blanca y la otra negra.

Solución

$\frac{4}{6} + \frac{3}{8} = \frac{1}{4}$

Sea W_1 = suceso «bola blanca» de la primera bolsa y W_2 = suceso «bola blanca» de la segunda.

(*a*)

$$\Pr\{W_1W_2\} = \Pr\{W_1\}\Pr\{W_2\} = \left(\frac{4}{4+2}\right)\left(\frac{3}{3+5}\right) = \frac{1}{4}$$

(*b*)

$$\Pr\{\bar{W}_1\bar{W}_2\} = \Pr\{\bar{W}_1\}\Pr\{\bar{W}_2\} = \left(\frac{2}{4+2}\right)\left(\frac{5}{3+5}\right) = \frac{5}{24}$$

$\left(\frac{4}{6}\right)\cdot\left(\frac{5}{8}\right)+$

(*c*) El suceso «una es blanca y la otra negra» es el mismo que «o la primera es blanca, o la segunda es negra o la primera negra y la segunda blanca»; esto es, $W_1\bar{W}_2 + \bar{W}_1 W_2$. Como $W_1\bar{W}_2$ y $\bar{W}_1 W_2$ son sucesos mutuamente excluyentes, tenemos

b N N b + N

$$\begin{aligned}\Pr\{W_1\bar{W}_2 + \bar{W}_1 W_2\} &= \Pr\{W_1\bar{W}_2\} + \Pr\{\bar{W}_1 W_2\}\\ &= \Pr\{W_1\}\Pr\{\bar{W}_2\} + \Pr\{\bar{W}_1\}\Pr\{W_2\}\\ &= \left(\frac{4}{4+2}\right)\left(\frac{5}{3+5}\right) + \left(\frac{2}{4+2}\right)\left(\frac{3}{3+5}\right) = \frac{13}{24}\end{aligned}$$

Otro método

La probabilidad pedida es $1 - \Pr\{W_1 W_2\} - \Pr\{\bar{W}_1\bar{W}_2\} = 1 - \frac{1}{4} - \frac{5}{24} = \frac{13}{24}$.

6.9. *A* y *B* juegan 12 partidas de ajedrez. *A* gana 6, *B* gana 4 y en 2 hacen tablas. Acuerdan jugar un torneo de 3 partidas. Hallar la probabilidad de que: (*a*) *A* gane las 3, (*b*) hagan tablas en 2, (*c*) *A* y *B* ganen alternadamente y (*d*) *B* gane al menos 1 partida.

Solución

Denotemos por A_1, A_2 y A_3 los sucesos «*A* gana» en la primera, segunda y tercera partidas, respectivamente; y por B_1, B_2 y B_3 lo análogo para *B*. Sean T_1, T_2 y T_3 los sucesos «tablas» en las tres partidas sucesivas.

Sobre la base de su experiencia pasada (probabilidad empírica), supondremos que $\Pr\{A$ gana cualquier partida$\} = \frac{6}{12} = \frac{1}{2}$, que $\Pr\{B$ gana cualquier partida$\} = \frac{4}{12} = \frac{1}{3}$, y que $\Pr\{$tablas en cualquier partida$\} = \frac{2}{12} = \frac{1}{6}$.

(*a*) $\Pr\{A \text{ gane los 3 juegos}\} = \Pr\{A_1A_2A_3\} = \Pr\{A_1\}\Pr\{A_2\}\Pr\{A_3\} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$

suponiendo que los resultados de cada partida sean independientes, lo cual parece justificable (a menos que los jugadores se dejen *influir psicológicamente* por las derrotas).

(*b*) $\Pr\{\text{tablas en 2 partidas}\} = \Pr\{1.^a \text{ y } 2.^a \text{ en tablas, o } 1.^a \text{ y } 3.^a \text{ en tablas, o } 2.^a \text{ y } 3.^a \text{ en tablas}\}$

$$\begin{aligned}&= \Pr\{T_1T_2\bar{T}_3\} + \Pr\{T_1\bar{T}_2T_3\} + \Pr\{\bar{T}_1T_2T_3\}\\ &= \Pr\{T_1\}\Pr\{T_2\}\Pr\{\bar{T}_3\} + \Pr\{T_1\}\Pr\{\bar{T}_2\}\Pr\{T_3\} + \Pr\{\bar{T}_1\}\Pr\{T_2\}\Pr\{T_3\}\\ &= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{15}{216} = \frac{5}{72}\end{aligned}$$

(*c*) $\Pr\{A \text{ y } B \text{ ganan alternadamente}\} = \Pr\{\text{ganan } ABA \text{ o ganan } BAB\}$

$$\begin{aligned}&= \Pr\{A_1B_2A_3 + B_1A_2B_3\} = \Pr\{A_1B_2A_3\} + \Pr\{B_1A_2B_3\}\\ &= \Pr\{A_1\}\Pr\{B_2\}\Pr\{A_3\} + \Pr\{B_1\}\Pr\{A_2\}\Pr\{B_2\}\\ &= \left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{5}{36}\end{aligned}$$

(*d*) $\Pr\{B \text{ gana al menos 1 partida}\} = 1 - \Pr\{B \text{ pierde las tres}\}$

$$\begin{aligned}&= 1 - \Pr\{\bar{B}_1\bar{B}_2\bar{B}_3\} = 1 - \Pr\{\bar{B}_1\}\Pr\{\bar{B}_2\}\Pr\{\bar{B}_3\}\\ &= 1 - \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{19}{27}\end{aligned}$$

DISTRIBUCIONES DE PROBABILIDAD

6.10. Hallar la probabilidad de cada reparto en chicos y chicas en familias con 3 hijos, supuesta igual probabilidad para ambos.

Solución

Sea B = suceso «chico» y G = suceso «chica». De acuerdo con la hipótesis de igual probabilidad, $\Pr\{B\} = \Pr\{G\} = \frac{1}{2}$. En familias de 3 hijos, pueden ocurrir los siguientes sucesos mutuamente excluyentes con las probabilidades indicadas:

(*a*) Tres chicos (*BBB*):

$$\Pr\{BBB\} = \Pr\{B\}\Pr\{B\}\Pr\{B\} = \frac{1}{8}$$

Aquí suponemos que el nacimiento de cada hijo es independiente de los demás nacimientos.

(*b*) Tres chicas (*GGG*): Como en la parte (*a*) por simetría,

$$\Pr\{GGG\} = \frac{1}{8}$$

(*c*) Dos chicos y una chica ($BBG + BGB + GBB$): $\left(\frac{1}{2}\right)^2 \cdot \frac{1}{2}$ $\frac{1}{}$

$$\Pr\{BBG + BGB + GBB\} = \Pr\{BBG\} + \Pr\{BGB\} + \Pr\{GBB\}$$

$$= \Pr\{B\}\Pr\{B\}\Pr\{G\} + \Pr\{B\}\Pr\{G\}\Pr\{B\} + \Pr\{G\}\Pr\{B\}\Pr\{B\}$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

(*d*) Dos chicas y un chico ($GGB + GBG + BGG$): Como en la parte (*c*) o por simetría, la probabilidad es $\frac{3}{8}$.

Si llamamos X a la *variable aleatoria* que indica el número de chicos en cada familia de 3 hijos, su distribución de probabilidad se muestra en la Tabla 6.2.

Tabla 6.2

Número de chicos X	0	1	2	3
Probabilidad $p(X)$	1/8	3/8	3/8	1/8

6.11. Representar la distribución del Problema 6.10.

Solución

El gráfico puede representarse como en la Figura 6.3 o como en la 6.4. La suma de las áreas de los rectángulos de la Figura 6.4 es 1; en ella, llamada un *histograma de probabilidad*, estamos considerando a X como una variable continua aunque es discreta en verdad, un procedimiento que resulta útil a menudo. La Figura 6.3, por su lado, se usa cuando uno no quiere tratar la variable como continua.

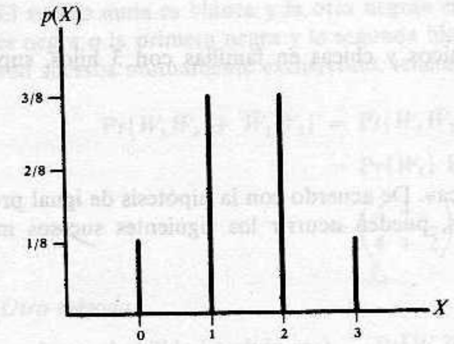


Figura 6.3.

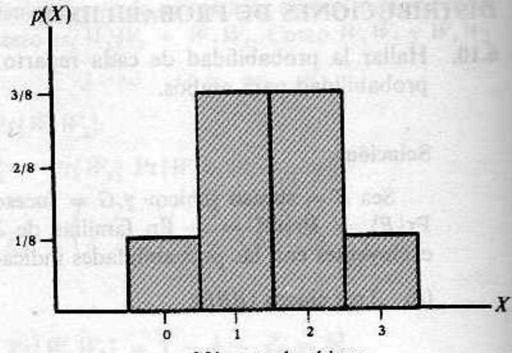


Figura 6.4.

6.12. Una variable aleatoria continua X con valores entre 0 y 4 tiene una función densidad dada por $p(X) = \frac{1}{2} - aX$, donde a es una constante.

(*a*) Calcular a.
(*b*) Hallar $\Pr\{1 < X < 2\}$.

Solución

(*a*) El gráfico de $p(X) = \frac{1}{2} - aX$ es una recta, como muestra la Figura 6.5. Para hallar a, debemos constatar primero que el área total bajo la recta entre $X = 0$ y $X = 4$, y sobre el eje X, ha de ser 1: en $X = 0$, $p(X) = \frac{1}{2}$, y en $X = 4$, $p(X) = \frac{1}{2} - 4a$. Entonces debemos elegir a de modo que el área del trapecio $= 1$. Area del trapecio $=$ (altura) $\times$ (suma de las bases/2 $= \frac{1}{2}(4)(\frac{1}{2} + \frac{1}{2} - 4a) = 2(1 - 4a) = 1$, de donde $(1 - 4a) = \frac{1}{2}$, $4a = \frac{1}{2}$ y $a = \frac{1}{8}$. Luego $(\frac{1}{2} - 4a)$ es realmente igual a cero y, por tanto, la gráfica correcta se muestra en la Figura 6.6.

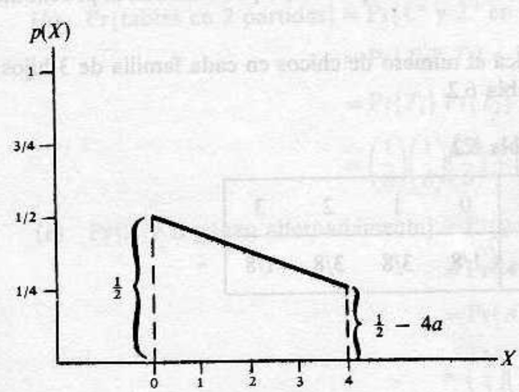


Figura 6.5.

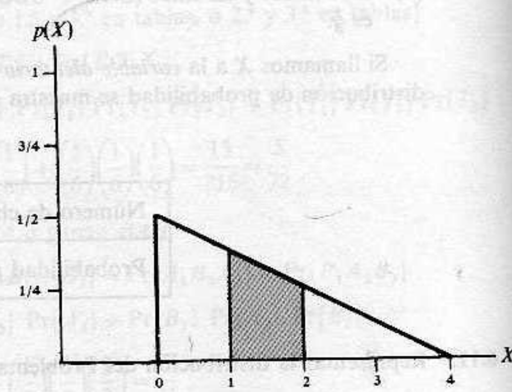


Figura 6.6.

(*b*) La requerida probabilidad es el área entre $X = 1$ y $X = 2$, sombreada en la Figura 6.6. De la parte (*a*), $p(X) = \frac{1}{2} - \frac{1}{8}X$; así que $p(1) = \frac{3}{8}$ y $p(2) = \frac{1}{4}$ son las ordenadas en $X = 1$ y $X = 2$, respectivamente. El área del trapecio pedida es $\frac{1}{2}(1)(\frac{3}{8} + \frac{1}{4}) = \frac{5}{16}$, que es la probabilidad deseada.

$P(1 < X < 2) = F(2) - F(1) - P(X = 2)$

$\frac{1}{4} - \frac{3}{80} - F(2^+) + F(2)$

ESPERANZA MATEMATICA

6.13. Un boleto de una rifa ofrece dos premios, uno de \$5000 y otro de \$2000, con probabilidades 0.001 y 0.003. ¿Cuál sería el precio justo a pagar por él?

Solución

Su esperanza matemática es (\$5000)(0.001) + (\$2000)(0.003) = \$5 + \$6 = \$11, que es el precio justo.

6.14. En un negocio aventurado, una señora puede ganar \$300 con probabilidad 0.6 o perder \$100 con probabilidad 0.4. Hallar su esperanza matemática.

Solución

Su esperanza matemática es (\$300)(0.6) + (−\$100)(0.4) = \$180 − \$40 = \$140.

6.15. Hallar: (*a*) $E(X)$, (*b*) $E(X^2)$ y (*c*) $E[(X - \bar{X})^2]$ para la distribución de probabilidad que muestra la Tabla 6.3.

Tabla 6.3

X	8	12	16	20	24
$p(X)$	1/8	1/6	3/8	1/4	1/12

Solución

(*a*) $E(X) = \sum Xp(X) = (8)(\frac{1}{8}) + (12)(\frac{1}{6}) + (16)(\frac{3}{8}) + (20)(\frac{1}{4}) + (24)(\frac{1}{12}) = 16$; esto representa la *media* de la distribución.

(*b*) $E(X^2) = \sum X^2p(X) = (8)^2(\frac{1}{8}) + (12)^2(\frac{1}{6}) + (16)^2(\frac{3}{8}) + (20)^2(\frac{1}{4}) + (24)^2(\frac{1}{12}) = 276$; esto representa el segundo momento respecto del origen cero.

(*c*) $E[(X - \bar{X})^2] = \sum (X - \bar{X})^2p(X) = (8 - 16)^2(\frac{1}{8}) + (12 - 16)^2(\frac{1}{6}) + (16 - 16)^2(\frac{3}{8}) + (20 - 16)^2(\frac{1}{4}) + (24 - 16)^2(\frac{1}{12}) = 20$; esto representa la *varianza* de la distribución.

6.16. Una bolsa contiene 2 bolas blancas y 3 bolas negras. Cada una de cuatro personas, A, B, C y D, en ese orden, saca una bola y no la repone. El primero que la saque blanca recibe \$10. Determinar las esperanzas matemáticas de A, B, C y D.

Solución

Como sólo hay 3 bolas blancas, alguien ganará en su primer intento. Sean A, B, C y D los sucesos «A gana», «B gana», «C gana» y «D gana», respectivamente.

$$\Pr\{A \text{ gana}\} = \Pr\{A\} = \frac{2}{3 + 2} = \frac{2}{5}$$

La esperanza matemática de $A = \frac{2}{5}(\$10) = \$4$.

$$\Pr\{A \text{ pierde y } B \text{ gana}\} = \Pr\{\bar{A}B\} = \Pr\{\bar{A}\}\Pr\{B \mid \bar{A}\} = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{3}{10}$$

Así que la esperanza matemática de B = \$3.

$$\Pr\{A \text{ y } B \text{ pierden y } C \text{ gana}\} = \Pr\{\bar{A}\bar{B}C\} = \Pr\{\bar{A}\}\Pr\{\bar{B}\mid\bar{A}\}\Pr\{C\bar{A}\bar{B}\} = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = \frac{1}{5}$$

Luego la esperanza matemática de C = \$2.

$$\Pr\{A, B \text{ y } C \text{ pierden y } D \text{ gana}\} = \Pr\{\bar{A}\bar{B}\bar{C}D\}$$
$$= \Pr\{\bar{A}\} \quad \Pr\{\bar{B}\mid\bar{A}\} \quad \Pr\{\bar{C}\mid\bar{A}\bar{B}\}$$
$$\Pr\{D\mid\bar{A}\bar{B}\bar{C}\}$$
$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)\left(\frac{1}{1}\right) = \frac{1}{10}$$

Y la de D = \$1.

Comprobación: \$4 + \$3 + \$2 + \$1 = \$10 y $\frac{2}{5} + \frac{3}{10} + \frac{1}{5} + \frac{1}{10} = 1$

PERMUTACIONES

6.17. ¿De cuántas maneras se pueden poner en fila 5 fichas de colores distintos?

Solución

Debemos colocarlas en cinco posiciones: — — — — —. La primera posición puede ser ocupada por cualquier ficha (o sea, hay 5 formas de ocupar esa posición). Una vez ocupada ella, hay 4 maneras de ocupar la siguiente, y entonces 3 de ocupar la tercera, 2 de ocupar la cuarta y sólo una de ocupar la quinta y última. En consecuencia:

$$\text{Número de ordenaciones de 5 fichas en fila} = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 5! = 120$$

En general,

$$\text{Número de ordenaciones de } n \text{ objetos distintos en fila} = n(n-1)(n-2)\cdots 1 = n!$$

Eso se llama el número de *permutaciones* de n objetos distintos tomados de n en n, y se denota por $_nP_n$

6.18. ¿De cuántas maneras se pueden sentar 10 personas en un banco si hay 4 sitios disponibles?

Solución

El primer sitio se puede ocupar de 10 formas, y una vez ocupado, el segundo se puede ocupar de 9 maneras, el tercero de 8 y el cuarto de 7. Por tanto,

$$\text{Número de colocaciones de 10 personas tomadas de 4 en 4} = 10\cdot 9\cdot 8\cdot 7 = 5040$$

En general,

$$\text{Número de colocaciones de } n \text{ objetos distintos de } r \text{ en } r = n(n-1)\cdots(n-r+1)$$

Esto se llama el número de *permutaciones* de n objetos distintos tomados de n en n y se denota por $_nP_r$, $P(n, r)$ o $P_{n,r}$. Nótese que cuando $r = n$, $_nP_n = n!$, como en el Problema 6.17.

6.19. Evaluar: (*a*) ${}_8P_3$, (*b*) ${}_6P_4$, (*c*) ${}_{15}P_1$ y ${}_3P_3$.

Solución

(*a*) ${}_8P_3 = 8 \cdot 7 \cdot 6 = 336$, (*b*) ${}_6P_4 = 6 \cdot 5 \cdot 4 \cdot 3 = 360$, (*c*) ${}_{15}P_1 = 15$ y (*d*) ${}_3P_3 = 3 \cdot 2 \cdot 1 = 6$.

6.20. Hay que colocar a 5 hombres y 4 mujeres en una fila de modo que las mujeres ocupen los lugares pares. ¿De cuántas maneras puede hacerse?

Solución

Los hombres se pueden colocar de ${}_5P_5$ maneras y las mujeres de ${}_4P_4$ maneras; cada colocación de ellos se puede asociar con una de ellas, luego el número pedido es ${}_5P_5 \cdot {}_4P_4 = 5!4! = (120)(24) = 2880$.

6.21. ¿Cuántos números de 4 dígitos se pueden formar con las cifras 0, 1, 2, 3, ..., 9: (*a*) permitiendo repeticiones, (*b*) sin repeticiones y (*c*) si el último dígito ha de ser cero y no se permiten repeticiones?

Solución

(*a*) El primero de los dígitos puede ser cualquiera de los 9 no nulos (el cero no se permite en esta posición, pues daría lugar a un número de 3 cifras). El segundo, tercero y cuarto dígitos pueden ya ser cualquiera de los 10. Luego se pueden formar $9 \cdot 10 \cdot 10 \cdot 10 = 9000$ números.

(*b*) El primer dígito puede ser cualquiera salvo el 0. El segundo cualquiera de los 9 que quedan al suprimir el ya empleado. El tercero uno de los 8 que aún no se han colocado y el cuarto cualquiera de los 7 no utilizados todavía. Así que se pueden formar $9 \cdot 9 \cdot 8 \cdot 7 = 4536$ números.

Otro método

El primero de los dígitos puede ser elegido entre 9, y los tres restantes de ${}_9P_3$ maneras. Por tanto, hay $9 \cdot {}_9P_3 = 9 \cdot 9 \cdot 8 \cdot 7 = 4536$ números.

(*c*) El primer dígito se puede elegir de 9 formas, el segundo de 8 y el tercero de 7. Luego se podrán formar $9 \cdot 8 \cdot 7 = 504$ números.

Otro método

El primero de los dígitos se puede tomar de 9 maneras y los otros dos de ${}_8P_2$ maneras. Luego $9 \cdot {}_8P_2 = 9 \cdot 8 \cdot 7 = 504$ números se pueden formar.

6.22. Cuatro libros diferentes de matemáticas, 6 de física y 2 de química han de ser colocados en una estantería. ¿Cuántas colocaciones distintas admiten si: (*a*) los libros de cada materia han de estar juntos y (*b*) sólo los de matemáticas tienen que estar juntos?

Solución

(*a*) Los de matemáticas se pueden colocar entre sí de ${}_4P_4 = 4!$ formas, los de física en ${}_6P_6 = 6!$, los de química de ${}_2P_2 = 2!$ y los tres grupos de ${}_3P_3 = 3!$ maneras entre sí. Luego el número requerido es $= 4!6!2!3! = 207{,}360$.

(*b*) Consideremos los 4 de matemáticas como una sola obra. Entonces tenemos 9 libros, que se pueden colocar de ${}_9P_9 = 6!$ maneras. En cada una de ellas, los 4 de matemáticas están juntos. Pero estos 4 se pueden colocar entre sí de ${}_4P_4 = 4!$ maneras. Luego la solución es $9!4! = 8{,}709{,}120$.

6.23. Cinco fichas rojas, 2 blancas y 3 azules se colocan en fila. Las de un color no son distinguibles entre sí. ¿Cuántas colocaciones distintas son posibles?

Solución

Sea P el número de colocaciones. Multiplicando P por el número de colocaciones de: (*a*) las 5 rojas entre sí, (*b*) las 2 blancas entre sí y (*c*) las 3 azules entre sí (o sea, multiplicando por P es 5!2!3!), obtendremos el número de colocaciones de 10 fichas distinguibles (o sea 10!). Luego

$$(5!2!3!)P = 10! \quad \text{y} \quad P = \frac{10!}{5!2!3!}$$

En general, el número de colocaciones diferentes de n objetos, de los que n_1 son iguales, n_2 son iguales, ..., n_k son iguales, es

$$\frac{n!}{n_1!n_2! \cdots n_k!}$$

donde $n_1 + n_2 + \cdots + n_k = n$.

6.24. ¿De cuántas formas se pueden sentar 7 personas en torno a una mesa redonda si: (*a*) son libres de elegir el asiento que deseen y (*b*) 2 personas particulares no pueden sentarse juntas?

Solución

(*a*) Sentemos a una en una silla. Entonces, los 6 restantes se pueden sentar de $6! = 720$ formas, que es el número total pedido.

(*b*) Consideremos a esas dos especiales como una sola persona. Entonces habría 6 personas, que se pueden sentar de 5! formas. Pero las 2 especiales se pueden colocar entre sí de 2! maneras, luego el número de formas en que se pueden situar 6 personas en una mesa redonda estando dos prefijadas juntas es $= 5!2! = 240$.

Usando la parte (*a*), la solución a (*b*) no es otra que $= 720 - 240 = 480$ maneras de sentarse con las condiciones impuestas.

COMBINACIONES

6.25. ¿De cuántas formas se pueden repartir 10 objetos en dos grupos de 4 y 6 objetos, respectivamente?

Solución

Es el mismo que el número de colocaciones de 10 objetos de los que 4 son iguales y los otros 6 son iguales. Por el Problema 6.23, es

$$\frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210$$

El problema equivale a hallar el número de selecciones de 4 entre 10 objetos (o 6 entre 10), siendo irrelevante el orden de selección.

En general, el número de selecciones de r entre n objetos, llamado el número de *combinaciones* de n objetos tomados de r en r, se denota por $\binom{n}{r}$ y viene dado por

combinaci $\binom{8}{2} = \frac{8!}{2!(6)!} =$ $$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{{}_nP_r}{r!}$$

$\frac{8!}{2!\,6!}$ $= \frac{8!}{6!(2!)}$ $\frac{10!}{4!(6)!} = 210$ $\frac{210!}{4!\,6!}$ n = iguales

6.26. Calcular: (*a*) $\binom{7}{4}$, (*b*) $\binom{6}{5}$ y (*c*) $\binom{4}{4}$.

$\frac{7!}{4!\,3!}$ = $\frac{4!}{4!\,0!}$ = 1

Solución

(*a*)

$$\binom{7}{4} = \frac{7!}{4!3!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

(*b*)

$$\binom{6}{5} = \frac{6!}{5!1!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{5!} = 6 \quad \text{o} \quad \binom{6}{5} = \binom{6}{1} = 6$$

(*c*) $\binom{4}{4}$ es el número de selecciones de 4 objetos tomados todos de golpe, y hay una sola selección, así que $\binom{4}{4} = 1$. Nótese que formalmente

$$\binom{4}{4} = \frac{4!}{4!0!} = 1$$

si definimos $0! = 1$.

6.27. ¿De cuántas maneras se puede formar con 9 personas una comisión de 5 miembros?

Solución

$$\binom{9}{5} = \frac{9!}{5!4!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{5!} = 126$$

6.28. De entre 5 matemáticos y 7 físicos hay que constituir una comisión de 2 matemáticos y 3 físicos. ¿De cuántas formas podrá hacerse si: (*a*) todos son elegibles, (*b*) un físico particular ha de estar en esa comisión y (*c*) dos matemáticos concretos tienen prohibido pertenecer a la comisión?

Solución

(*a*) Dos matemáticos entre 5 se pueden escoger de $\binom{5}{2}$ maneras, y 3 físicos de entre 7, de $\binom{7}{3}$ maneras. El número total de posibles selecciones es

$$\binom{5}{2} \cdot \binom{7}{3} = 10 \cdot 35 = 350$$

(*b*) Dos matemáticos entre 5 se pueden escoger de $\binom{5}{2}$ maneras, y los 2 físicos adicionales de entre 6 de $\binom{6}{2}$ formas. El número total de selecciones posibles es

$$\binom{5}{2} \cdot \binom{6}{2} = 10 \cdot 15 = 150$$

(*c*) Dos matemáticos entre 3 son elegibles de $\binom{3}{2}$ maneras, y 3 físicos de entre 7, de $\binom{7}{3}$ maneras. Luego el número total de selecciones posibles es

$$\binom{3}{2} \cdot \binom{7}{3} = 3 \cdot 35 = 105$$

6.29. ¿Cuántos ramilletes distintos se pueden formar con 5 flores de variedades diferentes?

$\binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} =$

Solución

Cada flor puede elegirse o no. Esas dos posibilidades ocurren para cada flor, luego en total 2^5. Pero de estas 2^5 opciones hay que excluir la consistente en no escoger ninguna. Luego el número de ramilletes es $= 2^5 - 1 = 31$.

Otro método

Podemos elegir 1 de las 5, o 2 de las 5, ..., o las 5 flores. De modo que el número pedido es

$$\binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 5 + 10 + 10 + 5 + 1 = 31$$

En general, para todo entero n positivo,

$$\binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - 1$$

6.30. Con 7 consonantes y 5 vocales, ¿cuántas palabras se pueden formar que tengan 4 consonantes distintas y 3 vocales distintas? Se admiten palabras sin significado.

Solución

Las 4 consonantes se pueden escoger de $\binom{7}{4}$ maneras, las 3 vocales de $\binom{5}{3}$ maneras y las 7 letras ya elegidas se pueden colocar entre sí de ${}_7P_7 = 7!$ maneras. Así que el número requerido es

$$\binom{7}{4} \cdot \binom{5}{3} \cdot 7! = 35 \cdot 10 \cdot 5040 = 1{,}764{,}000$$

APROXIMACION DE STIRLING A $n!$

6.31. Calcular aproximadamente 50!

Solución

Para n grande, tenemos $n! \approx \sqrt{2\pi n}\, n^n e^{-n}$; así que

$$50! \approx \sqrt{2\pi(50)}50^{50} e^{-50} = S$$

Para evaluar S usemos logaritmos en base 10. Tendremos

$$\begin{aligned} \log S = \log(\sqrt{100\pi 50^{50} e^{-50}}) &= \tfrac{1}{2}\log 100 + \tfrac{1}{2}\log \pi + 50 \log 50 - 50 \log e \\ &= \tfrac{1}{2}\log 100 + \tfrac{1}{2}\log 3.142 + 50 \log 50 - 50 \log 2.718 \\ &= \tfrac{1}{2}(2) + \tfrac{1}{2}(0.4972) + 50(1.6990) - 50(0.4343) = 64.4846 \end{aligned}$$

de donde $S = 3.05 \times 10^{64}$,

PROBABILIDAD Y ANALISIS COMBINATORIO

6.32. Una caja contiene 8 bolas rojas, 3 blancas y 9 azules. Si se sacan 3 bolas al azar, determinar la probabilidad de que: (*a*) las 3 sean rojas, (*b*) las 3 sean blancas, (*c*) 2 sean rojas y 1 blanca, (*d*) al menos 1 sea blanca, (*e*) sean una de cada color y (*f*) salgan en el orden roja, blanca, azul.

Solución

(*a*) *Primer método*

Denotemos por R_1, R_2 y R_3 los sucesos «la primera bola es roja», «la segunda bola es roja» y «la tercera bola es roja», respectivamente. Entonces $R_1 R_2 R_3$ denota el suceso de que las 3 sean rojas.

$$\Pr\{R_1 R_2 R_3\} = \Pr\{R_1\}\ \Pr\{R_2 \mid R_1\}\ \Pr\{R_3 \mid R_1 R_2\} = \left(\frac{8}{20}\right)\left(\frac{7}{19}\right)\left(\frac{6}{18}\right) = \frac{14}{285}$$

Segundo método

$$\text{Probabilidad requerida} = \frac{\text{número selecciones de 3 entre 8}}{\text{número selecciones de 3 entre 20}} = \frac{\binom{8}{3}}{\binom{20}{3}} = \frac{14}{285}$$

(*b*) Usando el segundo método de la parte (*a*),

$$\Pr\{\text{las 3 son blancas}\} = \frac{\binom{3}{3}}{\binom{20}{3}} = \frac{1}{1140}$$

Podía usarse también el primer método de (*a*).

(*c*)

$$\Pr\{\text{2 son rojas y 1 blanca}\} = \frac{\begin{pmatrix}\text{selecciones de 2 entre}\\ \text{8 bolas rojas}\end{pmatrix}\begin{pmatrix}\text{selecciones de 1 entre}\\ \text{3 bolas blancas}\end{pmatrix}}{\text{número de selecciones de 3 entre 20 bolas}} = \frac{\binom{8}{2}\binom{3}{1}}{\binom{20}{3}} = \frac{7}{95}$$

(*d*)

$$\Pr\{\text{ninguna es blanca}\} = \frac{\binom{17}{3}}{\binom{20}{3}} = \frac{34}{57} \qquad \text{o} \qquad \Pr\{\text{al menos 1 es blanca}\} = 1 - \frac{34}{57} = \frac{23}{57}$$

(*e*)

$$\Pr\{\text{sacar 1 de cada color}\} = \frac{\binom{8}{1}\binom{3}{1}\binom{9}{1}}{\binom{20}{3}} = \frac{18}{95}$$

(*f*) Usando la parte (*e*),

$$\Pr\{\text{bolas en orden roja, blanca, azul}\} = \frac{1}{3!}\Pr\{\text{1 de cada color}\} = \frac{1}{6}\left(\frac{18}{95}\right) = \frac{3}{95}$$

Otro método

$$\Pr\{R_1 W_2 B_2\} = \Pr\{R_1\}\ \Pr\{W_2 \mid R_1\}\ \Pr\{B_3 \mid R_1 W_2\} = \left(\frac{8}{20}\right)\left(\frac{3}{19}\right)\left(\frac{9}{18}\right) = \frac{3}{95}$$

6.33. De una baraja de 52 naipes bien mezclada se sacan 5 naipes. Hallar la probabilidad de que: (*a*) 4 sean ases, (*b*) 4 sean ases y 1 rey, (*c*) 3 sean dieces y 2 sotas, (*d*) salgan nueve, diez, sota, caballo y rey en cualquier orden, (*e*) 3 son de un palo y 2 de otro y (*f*) al menos uno sea un as.

Solución

(*a*)

$$\Pr\{4 \text{ ases}\} = \binom{4}{4}\cdot\frac{\binom{48}{1}}{\binom{52}{5}} = \frac{1}{54{,}145}$$

(*b*)

$$\Pr\{4 \text{ ases y 1 rey}\} = \binom{4}{4}\cdot\frac{\binom{4}{1}}{\binom{52}{5}} = \frac{1}{649{,}740}$$

(*c*)

$$\Pr\{3 \text{ son dieces y 2 son sotas}\} = \binom{4}{3}\cdot\frac{\binom{4}{2}}{\binom{52}{5}} = \frac{1}{108{,}290}$$

(*d*)

$$\Pr\{\text{nueve, diez, sota, caballo, rey en cualquier orden}\} = \frac{\binom{4}{1}\cdot\binom{4}{1}\cdot\binom{4}{1}\cdot\binom{4}{1}\cdot\binom{4}{1}}{\binom{52}{5}} = \frac{64}{162{,}435}$$

(*e*) Como hay cuatro formas de escoger el primer palo y tres de elegir el segundo,

$$\Pr\{3 \text{ de cualquier figura, 2 de otra}\} = \frac{4\binom{13}{3}\cdot 3\binom{13}{2}}{\binom{52}{5}} = \frac{429}{4165}$$

(*f*)

$$\Pr\{\text{ningún as}\} = \frac{\binom{48}{5}}{\binom{52}{5}} = \frac{35{,}673}{54{,}145} \quad \text{y} \quad \Pr\{\text{al menos 1 as}\} = 1 - \frac{35{,}673}{54{,}145} = \frac{18{,}482}{54{,}145}$$

6.34. Determinar la probabilidad de sacar 3 seises en 5 tiradas de un dado.

Solución

Representemos las 5 tiradas por 5 espacios — — — — —. En cada espacio tendremos los sucesos 6 o no 6 ($\bar{6}$); por ejemplo, tres 6 y dos no 6 pueden ocurrir como 6 6 $\bar{6}$ 6 $\bar{6}$ o como 6 $\bar{6}$ 6 $\bar{6}$ 6, etc.

Ahora bien, la probabilidad de un suceso tal como 6 6 $\bar{6}$ 6 $\bar{6}$ es

$$\Pr\{6\ 6\ \bar{6}\ 6\ \bar{6}\} = \Pr\{6\}\ \Pr\{6\}\ \Pr\{\bar{6}\}\ \Pr\{6\}\ \Pr\{\bar{6}\} = \frac{1}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{5}{6} = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

Similar $\Pr\{6\ \bar{6}\ 6\ \bar{6}\ 6\} = (\frac{1}{6})^3(\frac{5}{6})^2$, etc., para todos los sucesos en los que salen tres 6 y dos no 6. Pero hay $\binom{5}{3} = 10$ de tales sucesos, y esos sucesos son sucesos mutuamente excluyentes; por tanto, la probabilidad requerida es

$$\Pr\{6\ 6\ \bar{6}\ 6\ \bar{6}\ \text{ó}\ 6\ \bar{6}\ 6\ \bar{6}\ 6\ \text{o etc.}\} = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{125}{3888}$$

En general, si $p = \Pr\{E\}$ y $q = \Pr\{\bar{E}\}$, entonces usando el mismo argumento que antes, la probabilidad de obtener exactamente X veces E en N intentos es $\binom{N}{X}p^X q^{N-X}$.

6.35. Una factoría observa que, en promedio, el 20% de las tuercas producidas por una máquina son defectuosas. Si se toman 10 tuercas al azar, hallar la probabilidad de que: (*a*) exactamente 2 sean defectuosas, (*b*) 2 o más sean defectuosas y (*c*) más de 5 sean defectuosas.

Solución

(*a*) Por un razonamiento similar al del Problema 6.34,

$$\Pr\{2 \text{ tuercas defectuosas}\} = \binom{10}{2}(0.2)^2(0.8)^8 = 45(0.04)(0.1678) = 0.3020$$

(*b*)

$$\begin{aligned}\Pr\{2 \text{ o más tuercas defectuosas}\} &= 1 - \Pr\{0 \text{ tuercas defectuosas}\} - \Pr\{1 \text{ tuercas defectuosas}\}\\ &= 1 - \binom{10}{0}(0.2)^0(0.8)^{10} - \binom{10}{1}(0.2)^1(0.8)^9\\ &= 1 - (0.8)^{10} - 10(0.2)(0.8)^9\\ &= 1 - 0.1074 - 0.2684 = 0.6242\end{aligned}$$

(*c*)

$$\begin{aligned}\Pr\{\text{más de 5 tuercas defectuosas}\} &= \Pr\{6 \text{ tuercas defectuosas}\} + \Pr\{7 \text{ tuercas defectuosas}\}\\ &\quad + \Pr\{8 \text{ tuercas defectuosas}\} + \Pr\{9 \text{ tuercas defectuosas}\}\\ &\quad + \Pr\{10 \text{ tuercas defectuosas}\}\\ &= \binom{10}{6}(0.2)^6(0.8)^4 + \binom{10}{7}(0.2)^7(0.8)^3 + \binom{10}{8}(0.2)^8(0.8)^2\\ &\quad + \binom{10}{9}(0.2)^9(0.8) + \binom{10}{10}(0.2)^{10}\\ &= 0.00637\end{aligned}$$

6.36. Si se tomaran 1000 muestras de 10 tuercas cada una en el Problema 6.35, ¿de cuántas de ellas cabría esperar que tuvieran: (*a*) exactamente 2 defectuosas, (*b*) 2 o más defectuosas y (*c*) más de 5 defectuosas?

Solución

(*a*) Número esperado = (1000)(0.3020) = 302, por el Problema 6.35(*a*).
(*b*) Número esperado = (1000)(0.6242) = 624, por el Problema 6.35(*b*).
(*c*) Número esperado = (1000)(0.00637) = 6, por el Problema 6.35(*c*).

ESPACIO MUESTRAL Y DIAGRAMAS DE EULER

6.37. (*a*) Describir un espacio muestral para una tirada de un par de dados.
(*b*) Determinar a partir de él la probabilidad de que la suma de los dados sea 7 u 11.

Solución

(*a*) El espacio muestral consta de los puntos de la Figura 6.7, cuyas primeras coordenadas son las puntuaciones del primer dado y las segundas coordenadas son las puntuaciones del segundo dado. Hay 36 puntos, y a cada uno le asignamos una probabilidad de $\frac{1}{36}$. La suma de todas esas probabilidades es 1.

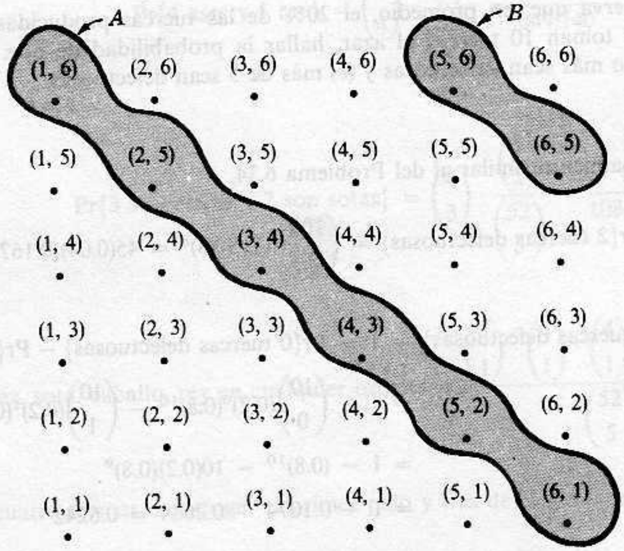


Figura 6.7.

(*b*) Los conjuntos de puntos correspondientes a los sucesos «suma 7» y «suma 11» se indican por A y B, respectivamente.

$\Pr\{A\}$ = suma de probabilidades asociadas con cada punto de $A = \frac{6}{36}$
$\Pr\{B\}$ = suma de probabilidades asociadas con cada punto de $B = \frac{2}{36}$
$\Pr\{A + B\}$ = suma de probabilidades de los puntos en A, en B o en ambos

Nótese que en este caso $\Pr\{A + B\} = \Pr\{A\} + \Pr\{B\}$. Ello ocurre porque A y B no tienen puntos en común (es decir, son sucesos mutuamente excluyentes).

6.38. Usando un espacio muestral, probar que:

(*a*) $\Pr\{A + B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}$

(*b*) $\Pr\{A + B + C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} - \Pr\{AB\} - \Pr\{BC\} - \Pr\{AC\} + \Pr\{ABC\}$

Solución

(*a*) Sean A y B dos conjuntos de puntos con puntos comunes denotados por AB, como en la Figura 6.8. A consta de $A\bar{B}$ y de AB, mientras B está compuesto por $B\bar{A}$ y AB. La totalidad de puntos en $A + B$ (o bien A, o B o ambos) = totalidad de puntos en A + totalidad de puntos en B − totalidad de puntos en AB. Como la probabilidad de un suceso conjunto es la suma de las probabilidades asociadas a sus puntos, tenemos

$$\Pr\{A + B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}$$

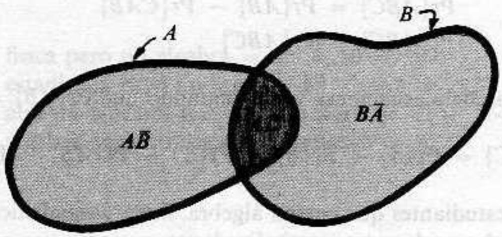


Figura 6.8.

Otro método

Denotemos por $A - AB$ el conjunto de puntos que están en A, pero no en B (es lo mismo que $A\bar{B}$); entonces $A - AB$ y B son mutuamente excluyentes (o sea, sin puntos en común). Además, $\Pr\{A - AB\} = \Pr\{A\} - \Pr\{AB\}$. Luego

$$\Pr\{A + B\} = \Pr\{A - AB\} + \Pr\{B\} = \Pr\{A\} - \Pr\{AB\} + \Pr\{B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}$$

(*b*) Sean A, B y C tres conjuntos de puntos, como indica la Figura 6.9. El símbolo $AB\bar{C}$ significa el conjunto de puntos en A y B que no están en C, y los otros símbolos son análogos.

Podemos considerar puntos que están en A o B o C como incluidos en los 7 conjuntos mutuamente excluyentes de la Figura 6.9, cuatro de los cuales están sombreados y tres sin sombrear. La probabilidad pedida viene dada por

$$\Pr\{A+B+C\} = \Pr\{A\bar{B}\bar{C}\} + \Pr\{B\bar{C}\bar{A}\} + \Pr\{C\bar{A}\bar{B}\} + \Pr\{AB\bar{C}\} + \Pr\{BC\bar{A}\} + \Pr\{CA\bar{B}\} + \Pr\{ABC\}$$

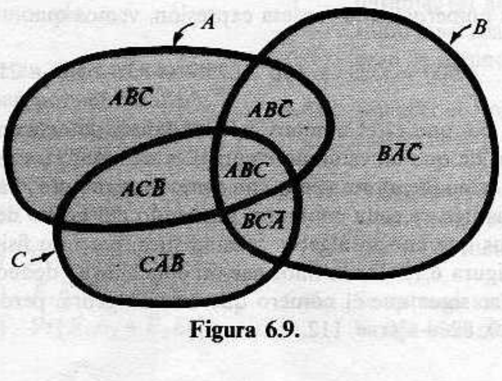


Figura 6.9.

Para obtener ahora $A\bar{B}\bar{C}$, por ejemplo, eliminamos los puntos comunes a A, B y a A, C; pero al hacerlo, hemos quitado los puntos comunes a A, B, C dos veces. Por tanto, $A\bar{B}\bar{C} = A - AB - AC + ABC$, y

$$\Pr\{A\bar{B}\bar{C}\} = \Pr\{A\} - \Pr\{AB\} - \Pr\{AC\} + \Pr\{ABC\}$$

Análogamente, se encuentra

$$\begin{aligned}
\Pr\{B\bar{C}\bar{A}\} &= \Pr\{B\} - \Pr\{BC\} - \Pr\{BA\} + \Pr\{BCA\}\\
\Pr\{C\bar{A}\bar{B}\} &= \Pr\{C\} - \Pr\{CA\} - \Pr\{CB\} + \Pr\{CAB\}\\
\Pr\{BC\bar{A}\} &= \Pr\{BC\} - \Pr\{ABC\}\\
\Pr\{CA\bar{B}\} &= \Pr\{CA\} - \Pr\{BCA\}\\
\Pr\{AB\bar{C}\} &= \Pr\{AB\} - \Pr\{CAB\}\\
\Pr\{ABC\} &= \Pr\{ABC\}
\end{aligned}$$

Sumando esas siete ecuaciones y considerando que $\Pr\{AB\} = \Pr\{BA\}$, etc., obtenemos

$$\Pr\{A + B + C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} - \Pr\{AB\} - \Pr\{BC\} - \Pr\{AC\} + \Pr\{ABC\}$$

6.39. Un recuento de 500 estudiantes que cursan álgebra, física y estadística reveló los siguintes números de estudiantes matriculados en las materias indicadas:

Algebra	329	Algebra y física	83
Física	186	Algebra y estadística	217
Estadística	295	Física y estadística	63

¿Cuántos estudiantes están matriculados en: (*a*) las tres, (*b*) álgebra pero no estadística, (*c*) física pero no álgebra, *d*) estadística pero no física, (*e*) álgebra o estadística pero no física y (*f*) álgebra pero no física ni estadística?

Solución

Sea A el conjunto de estudiantes matriculados en álgebra y (A) el número de ellos. Lo mismo con B, (B) para la física, y con C, (C) para la estadística. Entonces $(A + B + C)$ denota el número de estudiantes matriculados bien en álgebra o en física o en estadística o combinaciones de ellas, (AB) el de los matriculados en ambas, álgebra y física, etc. Como en el Problema 6.38, se sigue que

$$(A + B + C) = (A) + (B) + (C) - (AB) - (BC) - (AC) + (ABC)$$

(*a*) Sustituyendo los números dados en esa expresión, vemos que

$$500 = 329 + 186 + 295 - 83 - 63 - 217 + (ABC)$$

o sea $(ABC) = 53$, que es el número de estudiantes que cursan las tres. Nótese que la probabilidad (empírica) de que un estudiante curse las tres materias es $\frac{53}{500}$.

(*b*) Para obtener la deseada información, conviene construir un diagrama de Euler que muestre el número de estudiantes en cada conjunto. Partiendo del hecho de que 53 de ellos cursan las tres, deducimos que los que cursan álgebra y estadística, pero no física, son $217 - 53 = 164$, como se indica en la Figura 6.10. De la información conocida se deducen los otros números.

De los datos se sigue que el número que cursa álgebra, pero no estadística $= 329 - 217$; y por la Figura 6.10, $82 + 30 = 112$.

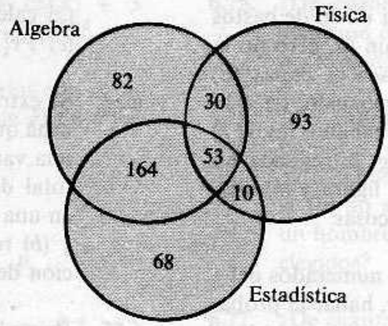


Figura 6.10.

(*c*) Número que cursa física pero no álgebra = 93 + 10 = 103
(*d*) Número que cursa estadística pero no física = 68 + 164 = 232
(*e*) Número que cursa álgebra o estadística pero no física = 82 + 164 + 68 = 314
(*f*) Número que cursa álgebra pero no física ni estadística = 82

PROBLEMAS SUPLEMENTARIOS

REGLAS FUNDAMENTALES DE LA PROBABILIDAD

6.40. Determinar la probabilidad *p*, o estimarla, para los sucesos:

(*a*) Al extraer una carta de una baraja bien mezclada se saca as, rey o la sota de bastos o el caballo de oros.
(*b*) Al lanzar un par de dados salga suma 8.
(*c*) Encontrar una tuerca defectuosa si entre 600 ya examinadas había 12 defectuosas.
(*d*) Sumar 7 u 11 en una tirada de un par de dados.
(*e*) Sacar al menos una cara en tres lanzamientos de una moneda.

6.41. Un experimento consiste en sacar tres cartas sucesivamente de una baraja bien mezclada. Sea E_1 el suceso «rey» en la primera extracción, E_2 el suceso «rey» en la segunda y E_3 el suceso «rey» en la tercera. Expresar en palabras el significado de:

(*a*) $\Pr\{E_1\bar{E}_2\}$	(*d*) $\Pr\{E_3 \mid E_1\bar{E}_2\}$
(*b*) $\Pr\{E_1 + E_2\}$	(*e*) $E_1E_2E_3$
(*c*) $\bar{E}_1 + \bar{E}_2$	(*f*) $\Pr\{E_1E_2 + \bar{E}_2E_3\}$

6.42. Se saca al azar una bola de una caja que contiene 10 rojas, 30 blancas, 20 azules y 15 naranja. Hallar la probabilidad de que la bola extraída sea: (*a*) roja o naranja, (*b*) ni roja ni azul, (*c*) no azul, (*d*) blanca y (*e*) roja, blanca o azul.

6.43. De la caja del Problema 6.42 se saca una bola, se repone y se hace una nueva extracción. Hallar la probabilidad de que: (*a*) ambas sean blancas, (*b*) la primera sea roja y la segunda blanca, (*c*) ninguna sea naranja, (*d*) ambas son rojas, o blancas o una de cada, (*e*) la segunda no sea azul, (*f*) la primera sea naranja, (*g*) al menos una sea azul, (*h*) a lo sumo una sea roja, (*i*) la primera sea azul, pero la segunda no y (*j*) sólo una sea roja.

6.44. Rehacer el Problema 6.43 sin reponer tras la extracción.

6.45. Hallar la probabilidad de obtener un total de 7 puntos en dos tiradas de un dado: (*a*) una vez, (*b*) al menos una vez y (*c*) dos veces.

6.46. Se extraen sucesivamente dos cartas de una baraja bien mezclada. Hallar la probabilidad de que: (*a*) la primera no sea un 10 de bastos o un as, (*b*) la primera sea un as, pero no la segunda, (*c*) al menos una sea de copas, (*d*) las cartas no sean del mismo palo, (*e*) a lo sumo una sea figura (sota, caballo, rey), (*f*) la segunda no sea figura, (*g*) la segunda no sea figura si la primera era figura y (*h*) sean figuras o espadas o ambas cosas.

6.47. Una caja contiene 9 tickets numerados del 1 al 9. Si se extraen 3 a la vez, hallar la probabilidad de que sean: (*a*) impar, par, impar, o (*b*) par, impar, par.

6.48. Las apuestas a favor de que A gane una partida de ajedrez contra B están 3 : 2. Si se disputan 3 partidas, ¿cuáles son las apuestas: (*a*) a favor de que A gane al menos dos y (*b*) en contra de que A pierda las dos primeras?

6.49. Un bolso contiene 2 monedas de plata y 4 de cobre, y otro contiene 4 de plata y 3 de cobre. Si se coge al azar de uno de los bolsos una moneda, ¿cuál es la probabilidad de que sea de plata?

6.50. La probabilidad de que un hombre siga vivo dentro de 25 años es $\frac{3}{5}$, y la de que su esposa lo esté es de $\frac{2}{3}$. Hallar la probabilidad de que en ese momento: (*a*) ambos estén vivos, (*b*) sólo el hombre viva, (*c*) sólo viva la esposa y (*d*) al menos uno esté vivo.

6.51. De entre 800 familias con 4 hijos cada una, ¿qué porcentaje es de esperar que tenga: (*a*) 2 chicos y dos chicas, (*b*) al menos un chico, (*c*) ninguna chica y (*d*) a lo sumo 2 chicas? Se supone igual probabilidad para chicos y chicas.

DISTRIBUCIONES DE PROBABILIDAD

6.52. Si X es la variable aleatoria que da el número de chicos en familias de 4 hijos (véase Prob. 6.51): (*a*) construir una tabla que muestre su distribución de probabilidad y (*b*) representar la distribución de probabilidad de la parte (*a*) gráficamente.

6.53. Una variable aleatoria continua X que toma valores entre 2 y 8 inclusive, tiene una función densidad dada por $a(X + 3)$, con a constante: (*a*) calcular a; hallar: (*b*) $\Pr\{3 < X < 5\}$, (*c*) $\Pr\{X \geqslant 4\}$ y (*d*) $\Pr\{|X - 5| < 0.5\}$.

6.54. Se extraen, sin reposición, tres fichas de una urna que contiene 4 rojas y 6 blancas. Si X es una variable aleatoria que denota el número total de fichas rojas extraídas: (*a*) construir en una tabla su distribución de probabilidad y (*b*) representar gráficamente esa distribución de probabilidad.

6.55. Para el Problema 6.54, hallar: (*a*) $\Pr\{X = 2\}$, y (*b*) $(\Pr\{1 \leqslant X \leqslant 3\}$, e interpretar los resultados.

ESPERANZA MATEMATICA

6.56. ¿Cuál es el precio justo para participar en un juego en el que se ganan \$25 con probabilidad 0.2 y \$10 con probabilidad 0.4?

6.57. Si llueve, un vendedor de paraguas gana \$30 al día, y si no llueve pierde \$6 al día. ¿Cuál es su esperanza matemática si la probabilidad de lluvia es 0.3?

6.58. A y B juegan a tirar una moneda tres veces. Gana el primero que saque cara. Si A lanza primero y el montante de la apuesta es \$20, ¿cuánto debe poner cada uno para que el juego sea justo?

6.59. Hallar: (*a*) $E(X)$, (*b*) $E(X^2)$, (*c*) $E[(X - \bar{X})^2]$ y (*d*) $E(X^3)$ para la distribución de probabilidad de la Tabla 6.4.

Tabla 6.4

X	-10	-20	30
$p(X)$	1/5	3/10	1/2

6.60. Refiriéndonos al Problema 6.54, hallar: (*a*) la media, (*b*) la varianza y (*c*) la desviación típica de la distribución de X, e interpretar los resultados.

6.61. Una variable aleatoria toma el valor 1 con probabilidad p y el 0 con probabilidad $q = 1 - p$. Probar que: (*a*) $E(X) = p$ y (*b*) $E[(X - \bar{X})^2] = pq$.

6.62. Probar que: (*a*) $E(2X + 3) = 2E(X) + 3$ y (*b*) $E[(X - \bar{X})^2] = E(X^2) - [E(X)]^2$.

6.63. Sea X e Y dos variables aleatorias con idéntica distribución. Demostrar que $E(X+Y) = E(X) + E(Y)$.

PERMUTACIONES

6.64. Evaluar: (*a*) ${}_4P_2$, (*b*) ${}_7P_5$ y (*c*) ${}_{10}P_3$.

6.65. ¿Para qué valor de n es ${}_{n+1}P_3 = {}_nP_4$?

6.66. ¿De cuántas maneras pueden sentarse 5 personas en un sofá de 3 plazas?

6.67. ¿De cuántas maneras pueden colocarse 7 libros en una estantería si: (*a*) cualquier colocación es admitida, (*b*) 3 libros particulares han de estar juntos y (*c*) 2 libros particulares deben ocupar los extremos?

6.68. ¿Cuántos números de 5 cifras diferentes se pueden formar con los dígitos 1, 2, 3, ..., 9 si: (*a*) cada número ha de ser impar y (*b*) los dos primeros dígitos han de ser pares?

6.69. Resolver el Problema 6.68 permitiendo repeticiones de dígitos.

6.70. ¿Cuántos números de tres dígitos se pueden formar con 3 cuatros, 4 doses y 2 treses?

6.71. ¿De cuántas maneras pueden sentarse 3 hombres y 3 mujeres en una mesa redonda si: (*a*) no se imponen restricciones, (*b*) 2 mujeres particulares no pueden sentarse juntas y (*c*) cada mujer ha de estar entre dos hombres?

COMBINACIONES

6.72. Evaluar: (*a*) $\binom{7}{3}$, (*b*) $\binom{8}{4}$ y (*c*) $\binom{10}{8}$.

6.73. ¿Para qué valor de n es $3\binom{n+1}{3} = 7\binom{n}{2}$?

6.74. ¿De cuántas maneras pueden seleccionarse 6 cuestiones de entre un total de 10?

6.75. ¿De cuántas maneras puede formarse una comisión de 3 hombres y 4 mujeres de entre un total de 8 hombres y 6 mujeres?

6.76. ¿De cuántas maneras pueden escogerse 2 hombres, 4 mujeres, 3 niños y 3 niñas de entre 6 hombres, 8 mujeres, 4 niños y 5 niñas si: (*a*) no se impone restricción alguna y (*b*) un hombre y una mujer concretos deben ser elegidos?

6.77. ¿De cuántas maneras puede dividirse un grupo de 10 personas en dos grupos de 7 y 3 personas?

6.78. ¿De cuántas maneras puede elegirse una comisión de 3 estadísticos y 2 economistas de entre 5 estadísticos y 6 economistas si: (*a*) no se imponen restricciones, (*b*) 2 estadísticos particulares han de figurar en ella y (*c*) un economista concreto tiene vetado el figurar en ella?

6.79. Hallar el número de: (*a*) combinaciones y (*b*) permutaciones de 4 letras que pueden formarse con las letras de la palabra *Tennessee*.

6.80. Demostrar que $1 - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n\binom{n}{n} = 0$.

APROXIMACION DE STIRLING A $n!$

6.81. ¿De cuántas maneras pueden seleccionarse 30 individuos de entre 100?

6.82. Probar que $\binom{2n}{n} = 2^{2n}/\sqrt{\pi n}$, aproximadamente, para grandes valores de n.

PROBLEMAS DIVERSOS

6.83. Se sacan 3 cartas de una baraja de 52 cartas. Hallar la probabilidad de que: (*a*) dos sean sotas y una rey, (*b*) todas sean del mismo palo, (*c*) sean de palos diferentes y (*d*) al menos dos sean ases.

6.84. Hallar la probabilidad de al menos dos sietes en 4 tiradas de un par de dados.

6.85. Si el 10% de los remaches producidos por una máquina son defectuosos, ¿cuál es la probabilidad de que entre 5 elegidos al azar: (*a*) ninguno sea defectuoso, (*b*) haya uno defectuoso y (*c*) al menos dos lo sean?

6.86. (*a*) Describir un espacio muestral para los resultados de dos lanzamientos de una moneda, usando 1 para representar «cara» y 0 para «cruz».
(*b*) Con tal espacio muestral, determinar la probabilidad de al menos una cara.
(*c*) ¿Puede dar un espacio muestral para los resultados de lanzar 3 veces una moneda? En caso afirmativo, determine con su ayuda la probabilidad de al menos 2 caras.

6.87. Un muestreo de 200 votantes revela la siguiente información referente a tres candidatos A, B y C de un cierto partido que se disputaban tres cargos diferentes:

28 a favor de ambos A y B
98 a favor de A o B pero no C
42 a favor de B pero no A o C
122 a favor de B o C pero no A
64 a favor de C pero no A o B
14 a favor de A y C pero no B

¿Cuántos de los votantes están a favor de: (*a*) los tres candidatos, (*b*) de A e indiferentes a B y C, (*c*) de B e indiferentes a A y C, (*d*) de C e indiferentes a A y B, (*e*) de A y B, pero no de C y (*f*) sólo de uno de los candidatos?

6.88. (*a*) Probar que para cualesquiera sucesos E_1 y E_2, $\Pr\{E_1 + E_2\} \leqslant \Pr\{E_1\} + \Pr\{E_2\}$.
(*b*) Generalizar el resultado de la parte (*a*).

6.89. Sean E_1, E_2 y E_3 tres sucesos diferentes, al menos uno de los cuales se sabe que ha ocurrido. Si todas las probabilidades $\Pr\{E_1\}$, $\Pr\{E_2\}$, $\Pr\{E_3\}$ y $\Pr\{A \mid E_1\}$, $\Pr\{A \mid E_2\}$, $\Pr\{A \mid E_3\}$ se suponen conocidas, probar que

$$\Pr\{E_1 \mid A\} = \frac{\Pr\{E_1\}\Pr\{A \mid E_1\}}{\sum_1^3 j \Pr\{E_j\} \Pr\{A|E_j\}}$$

con resultados similares para $\Pr\{E_2 \mid A\}$ y $\Pr\{E_3 \mid A\}$. Esto se conoce como *regla* o *teorema* de Bayes. Es útil al calcular probabilidades de varias hipótesis que han resultado en el suceso A. El resultado es generalizable.

6.90. Tres joyeros idénticos tienen cada uno dos cajones. Cada cajón del primero contiene un reloj de oro, y cada uno del segundo un reloj de plata. En un cajón del tercero hay uno de oro y en el otro uno de plata. Si seleccionamos un joyero al azar, abrimos uno de sus cajones y en él hay un reloj de plata, ¿cuál es la probabilidad de que en el otro cajón haya un reloj de oro? [*Ayuda:* Aplicar el Problema 6.89.]

6.91. Hallar la probabilidad de acertar una loto en la que se deben marcar 6 números de entre 1, 2, 3, ..., 40 en cualquier orden.

6.92. Rehacer el Problema 6.91 si se marcan: (*a*) 5, (*b*) 4 y (*c*) 3 de los números.

6.93. En el póquer se dan a cada jugador 5 cartas de una baraja de 52 cartas. Determinar las apuestas en contra de que un jugador reciba:
(*a*) Escalera de color máxima (10, J, Q, K y as del mismo palo).
(*b*) Escalera de color (cinco cartas sucesivas del mismo palo, por ejemplo, 3, 4, 5, 6 y 7 de tréboles).
(*c*) Un póquer (cuatro cartas iguales, por ejemplo, cuatro sietes).
(*d*) Un «full» (un trío y una pareja, por ejemplo, tres reyes y dos cincos).

6.94. A y B deciden encontrarse entre las 3 y las 4 de la tarde, pero acuerdan que cada uno no espera más de 10 minutos al otro. Hallar la probabilidad de que se encuentren.

6.95. Se escogen al azar dos puntos en un segmento recto de longitud $a > 0$. Hallar la probabilidad de que los tres segmentos así formados puedan ser los lados de un triángulo.

CAPITULO 7

Las distribuciones binomial, normal y de Poisson

LA DISTRIBUCION BINOMIAL

Si p es la probabilidad de que ocurra un suceso en un solo intento (llamada probabilidad de *éxito*) y $q = 1 - p$ es la probabilidad de que no ocurra en un solo intento (llamada probabilidad de *fracaso*), entonces la probabilidad de que el suceso ocurra exactamente X veces en N intentos (o sea, X éxitos y $N - X$ fracasos) viene dada por

$$p(X) = \binom{N}{X} p^X q^{N-X} = \frac{N!}{X!(N-X)!} p^X q^{N-X} \tag{1}$$

donde $X = 0, 1, 2, \ldots, N$; $N! = N(N-1)(N-2)\cdots 1$; y $0! = 1$ por definición (véase Prob. 6.34).

EJEMPLO 1. La probabilidad de obtener exactamente 2 caras en 6 tiradas de una moneda es

$$\binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{6-2} = \frac{6}{2!4!}\left(\frac{1}{2}\right)^6 = \frac{15}{64}$$

$\frac{6!}{2!(6-2)!}$

usando la fórmula (1) con $N = 6$, $X = 2$ y $p = q = \frac{1}{2}$.

EJEMPLO 2. La probabilidad de obtener al menos 4 caras en 6 tiradas de una moneda es

$$\binom{6}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{6-4} + \binom{6}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{6-5} + \binom{6}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^{6-6} = \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{11}{32}$$

La distribución de probabilidad discreta (1) se llama *distribución binomial* porque para $X = 0, 1, 2, \ldots, N$ corresponde a términos sucesivos de la *fórmula binomial*, o *desarrollo del binomio*,

$$(q + p)^N = q^N + \binom{N}{1} q^{N-1} p + \binom{N}{2} q^{N-2} p^2 + \cdots + p^N \tag{2}$$

donde $1, \binom{N}{1}, \binom{N}{2}, \ldots$ se llaman *coeficientes binomiales*.

EJEMPLO 3.

$$(q+p)^4 = q^4 + \binom{4}{1}q^3p + \binom{4}{2}q^2p^2 + \binom{4}{3}qp^3 + p^4$$
$$= q^4 + 4q^3p + 6q^2p^2 + 4qp^3 + p^4$$

La distribución (1) se llama también *distribución de Bernoulli*, en honor de James Bernoulli, quien la descubrió a finales del siglo XVII. Algunas propiedades de la distribución binomial se recogen en la Tabla 7.1.

EJEMPLO 4. En 100 tiradas de una moneda el número medio de caras es $\mu = Np = (100)(\frac{1}{2}) = 50$; este es el número esperado de caras en 100 lanzamientos. La desviación típica es $\sigma = \sqrt{Npq} = \sqrt{(100)(\frac{1}{2})(\frac{1}{2})} = 5$.

LA DISTRIBUCION NORMAL

Uno de los más importantes ejemplos de una distribución de probabilidad continua es la *distribución normal, curva normal* o *distribución gaussiana*, definida por la ecuación

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2/\sigma^2} \tag{3}$$

Tabla 7.1. Distribución binomial

Media	$\mu = Np$
Varianza	$\sigma^2 = Npq$
Desviación típica	$\sigma = \sqrt{Npq}$
Coeficiente de sesgo	$\alpha_3 = \dfrac{q-p}{\sqrt{Npq}}$
Coeficiente de curtosis	$\alpha_4 = 3 + \dfrac{1-6pq}{Npq}$

donde μ = media, σ = desviación típica, $\pi = 3.14159\cdots$ y $e = 2.71828\cdots$. El área total limitada por la curva (3) y el eje X es 1; por tanto, el área bajo la curva entre $X = a$ y $X = b$, con $a < b$, representa la probabilidad de que X esté entre a y b. Esta probabilidad se denota por $\Pr\{a < X < b\}$.

Cuando se expresa la variable X en unidades estándar $[z = (X - \mu)/\sigma]$, la ecuación (3) es reemplazada por la llamada *forma canónica*

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{4}$$

En tal caso, decimos que z está *normalmente distribuida con media 0 y varianza* 1. La Figura 7.1 es un gráfico de esta forma canónica. Muestra que las áreas comprendidas entre $z = -1$ y $+1$, $z = -2$ y $+2$, y $z = -3$ y $+3$ son iguales, respectivamente, a 68.27%, 95.45% y 99.73% del área total, que es 1. La tabla del Apéndice II muestra las áreas bajo esta curva acotadas por las ordenadas $z = 0$ y cualquier valor positivo de z. De esa tabla se puede deducir el área entre todo par de coordenadas usando la simetría de la curva respecto de $z = 0$.

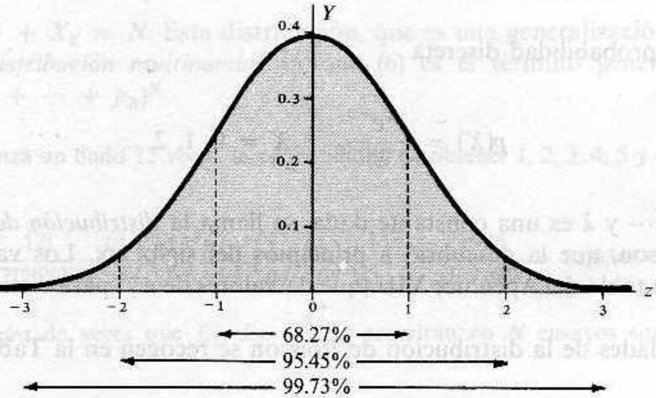


Figura 7.1.

Algunas propiedades de la distribución normal (3) se listan en la Tabla 7.2.

Tabla 7.2. Distribución normal

Media	μ
Varianza	σ^2
Desviación típica	σ
Coeficiente de sesgo	$\alpha_3 = 0$
Coeficiente de curtosis	$\alpha_4 = 3$
Desviación media	$\sigma\sqrt{2/\pi} = 0.7979\sigma$

RELACION ENTRE LA DISTRIBUCION BINOMIAL Y LA DISTRIBUCION NORMAL

Si N es grande y si ni p ni q son muy próximos a cero, la distribución binomial puede aproximarse estrechamente por una distribución normal con variable canónica dada por

$$z = \frac{X - Np}{\sqrt{Npq}}$$

La aproximación mejora al crecer N, y en el límite es exacta; esto se muestra en las Tablas 7.1 y 7.2, donde es claro que al crecer N, el sesgo y la curtosis de la distribución binomial se aproximan a los de la distribución normal. En la práctica, la aproximación es muy buena si tanto Np como Nq son mayores que 5.

LA DISTRIBUCION DE POISSON

La distribución de probabilidad discreta

$$p(X) = \frac{\lambda^X e^{-\lambda}}{X!} \qquad X = 0, 1, 2, \ldots \tag{5}$$

donde $e = 2.71828\cdots$ y λ es una constante dada, se llama la *distribución de Poisson* en honor de Siméon-Denis Poisson, que la descubrió a principios del siglo XIX. Los valores de $p(X)$ pueden calcularse usando la tabla del Apéndice VIII (que da valores de $e^{-\lambda}$ para distintos λ) o por medio de logaritmos.

Algunas propiedades de la distribución de Poisson se recogen en la Tabla 7.3.

Tabla 7.3. Distribución de Poisson

Media	$\mu = \lambda$
Varianza	$\sigma^2 = \lambda$
Desviación típica	$\sigma = \sqrt{\lambda}$
Coeficiente de sesgo	$\alpha_3 = 1/\sqrt{\lambda}$
Coeficiente de curtosis	$\alpha_4 = 3 + 1/\lambda$

RELACION ENTRE LA DISTRIBUCION BINOMIAL Y LA DISTRIBUCION DE POISSON

En la distribución binomial (1), si N es grande y la probabilidad p de ocurrencia de un suceso es muy pequeña, de modo que $q = 1 - p$ es casi 1, el suceso se llama un *suceso raro*. En la práctica, un suceso se considera raro si el número de ensayos es al menos 50 ($N \geqslant 50$) mientras Np es menor que 5. En tal caso, la distribución binomial queda aproximada muy estrechamente por la distribución de Poisson (5) con $\lambda = Np$. Esto se comprueba comparando las Tablas 7.1 y 7.3, pues al poner $\lambda = Np$, $q \approx 1$ y $p \approx 0$ en la Tabla 7.1 obtenemos los resultados de la Tabla 7.3.

Como hay una relación entre la distribución binomial y la distribución normal, se sigue que también están relacionadas la distribución de Poisson y la distribución normal. De hecho, puede probarse que la distribución de Poisson tiende a una distribución normal con variable canónica $(X - \lambda)/\sqrt{\lambda}$ cuando λ crece indefinidamente.

LA DISTRIBUCION MULTINOMIAL

Si los sucesos $E_1, E_2, ..., E_K$ pueden ocurrir con frecuencias $p_1, p_2, ..., p_K$, respectivamente, entonces la probabilidad de que $E_1, E_2, ..., E_K$ ocurran $X_1, X_2, ..., X_K$ veces, respectivamente, es

$$\frac{N!}{X_1!X_2!\cdots X_K!}\, p_1^{X_1}p_2^{X_2}\cdots p_K^{X_K} \tag{6}$$

donde $X_1 + X_2 + \cdots + X_K = N$. Esta distribución, que es una generalización de la distribución binomial, se llama *distribución multinomial* ya que (6) es el término general en el *desarrollo multinomial* $(p_1 + p_2 + \cdots + p_K)^N$.

EJEMPLO 5. Si se lanza un dado 12 veces, la probabilidad de obtener 1, 2, 3, 4, 5 y 6 puntos exactamente dos veces cada uno es

$$\frac{12!}{2!2!2!2!2!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2 = \frac{1925}{559{,}872} = 0.00344$$

Los números *esperados* de veces que $E_1, E_2, ..., E_K$ ocurrirán en N ensayos son $Np_1, Np_2, ..., Np_K$, respectivamente.

AJUSTE DE DISTRIBUCIONES DE FRECUENCIAS MUESTRALES MEDIANTE DISTRIBUCIONES TEORICAS

Cuando se tiene una cierta indicación sobre la distribución de una población por argumentos probabilísticos o de otra índole, suele ser posible ajustar esa distribución teórica (llamada también distribución *esperada* o *modelo*) a distribuciones de frecuencias obtenidas de una muestra de esa población. El método usado consiste en emplear la media y la desviación típica de la muestra para estimar las de la población (véanse Probs. 7.31, 7.33 y 7.34).

Para comprobar la *bondad* del ajuste de las distribuciones teóricas, usamos el *test ji-cuadrado* (Cap. 12). Al intentar determinar si una distribución normal representa un buen ajuste para datos dados, es conveniente usar *papel gráfico de curva normal*, o *papel gráfico de probabilidad* como se le llama a veces (véase Prob. 7.32).

PROBLEMAS RESUELTOS

LA DISTRIBUCION BINOMIAL

7.1. Calcular:

(a) $5!$ (c) $\binom{8}{3}$ (e) $\binom{4}{4}$

(b) $\frac{6!}{2!4!}$ (d) $\binom{7}{5}$ (f) $\binom{4}{0}$

Solución

(*a*) $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

(*b*) $\dfrac{6!}{2!4!} = \dfrac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(4 \cdot 3 \cdot 2 \cdot 1)} = \dfrac{6 \cdot 5}{2 \cdot 1} = 15$

(*c*) $\dbinom{8}{3} = \dfrac{8!}{3!(8-3)!} = \dfrac{8!}{3!5!} = \dfrac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = \dfrac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56$

(*d*) $\dbinom{7}{5} = \dfrac{7!}{5!2!} = \dfrac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)(2 \cdot 1)} = \dfrac{7 \cdot 6}{2 \cdot 1} = 21$

(*e*) $\dbinom{4}{4} = \dfrac{4!}{4!0!} = 1$ porque $0! = 1$ por definición

(*f*) $\dbinom{4}{0} = \dfrac{4!}{0!4!} = 1$

7.2. Hallar la probabilidad de que al lanzar una moneda tres veces, aparezcan: (*a*) 3 caras, (*b*) 2 caras y una cruz, (*c*) 2 caras y una cara y (*d*) 3 cruces.

Solución

Primer método

Denotemos «cara» por H y «cruz» por T, y supongamos que designamos por HTH el que ocurra cara en el primer lanzamiento, cruz en el segundo y cara en el tercero. Como las posibilidades cara y cruz pueden aparecer en cada tirada, hay $(2)(2)(2) = 8$ posibles resultados. Son

$$HHH \quad HHT \quad HTH \quad HTT \quad TTH \quad THH \quad THT \quad TTT$$

Cada una de esas posibilidades es igualmente probable, con probabilidad $\frac{1}{8}$.

(*a*) 3 caras (HHH) sólo ocurren una vez; luego su probabilidad es $\frac{1}{8}$.
(*b*) 2 caras y 1 cruz ocurren tres veces (HHT, HTH y THH); luego $\Pr\{2 \text{ caras y una cruz}\} = \frac{3}{8}$.
(*c*) 1 cara y dos cruces ocurren tres veces (HTT, TTH y THT); luego $\Pr\{1 \text{ cara y 2 cruces}\} = \frac{3}{8}$.
(*d*) 3 cruces (TTT) ocurren sólo una vez; luego $\Pr\{TTT\} = \Pr\{3 \text{ cruces}\} = \frac{1}{8}$.

Segundo método [usando la fórmula (1)]

(*a*) $\Pr\{3 \text{ caras}\} = \dbinom{3}{3}\left(\dfrac{1}{2}\right)^3\left(\dfrac{1}{2}\right)^0 = (1)\left(\dfrac{1}{8}\right)(1) = \dfrac{1}{8}$

(*b*) $\Pr\{2 \text{ caras y 1 cruz}\} = \dbinom{3}{2}\left(\dfrac{1}{2}\right)^2\left(\dfrac{1}{2}\right) = (3)\left(\dfrac{1}{4}\right)\left(\dfrac{1}{2}\right) = \dfrac{3}{8}$

(*c*) $\Pr\{1 \text{ cara y 2 cruces}\} = \dbinom{3}{1}\left(\dfrac{1}{2}\right)^1\left(\dfrac{1}{2}\right)^2 = (3)\left(\dfrac{1}{2}\right)\left(\dfrac{1}{4}\right) = \dfrac{3}{8}$

(*d*) $\Pr\{3 \text{ cruces}\} = \dbinom{3}{0}\left(\dfrac{1}{2}\right)^0\left(\dfrac{1}{2}\right)^3 = (1)(1)\left(\dfrac{1}{2}\right) = \dfrac{1}{8}$

Podría procederse, asimismo, como en el Problema 6.10.

7.3. Hallar la probabilidad de que en 5 tiradas de un dado aparezca el 3: (*a*) ninguna vez, (*b*) 1 vez, (*c*) 2 veces, (*d*) 3 veces, (*e*) 4 veces y (*f*) 5 veces.

Solución

La probabilidad del 3 en una sola tirada $= p = \frac{1}{6}$, y la de no sacar $3 = q = 1 - p = \frac{5}{6}$; luego:

(*a*) $\Pr\{3 \text{ ocurra cero veces}\} = \binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 = (1)(1)\left(\frac{5}{6}\right)^5 = \frac{3125}{7776}$

(*b*) $\Pr\{3 \text{ ocurra una vez}\} = \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 = (5)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 = \frac{3125}{7776}$

(*c*) $\Pr\{3 \text{ ocurra dos veces}\} = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = (10)\left(\frac{1}{36}\right)\left(\frac{125}{216}\right) = \frac{625}{3888}$

(*d*) $\Pr\{3 \text{ ocurra tres veces}\} = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = (10)\left(\frac{1}{216}\right)\left(\frac{25}{36}\right) = \frac{125}{3888}$

(*e*) $\Pr\{3 \text{ ocurra cuatro veces}\} = \binom{5}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^1 = (5)\left(\frac{1}{1296}\right)\left(\frac{5}{6}\right) = \frac{25}{7776}$

(*f*) $\Pr\{3 \text{ ocurra cinco veces}\} = \binom{5}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^0 = (1)\left(\frac{1}{7776}\right)(1) = \frac{1}{7776}$

Nótese que estas probabilidades representan los términos del desarrollo binomial

$$\left(\frac{5}{6}+\frac{1}{6}\right)^5 = \left(\frac{5}{6}\right)^5 + \binom{5}{1}\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) + \binom{5}{2}\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)^2 + \binom{5}{3}\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^3 + \binom{5}{4}\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^4 + \left(\frac{1}{6}\right)^5 = 1$$

7.4. Escribir el desarrollo binomial para: (*a*) $(q + p)^4$ y (*b*) $(q + p)^6$.

Solución

(*a*)

$$(q + p)^4 = q^4 + \binom{4}{1}q^3p + \binom{4}{2}q^2p^2 + \binom{4}{3}qp^3 + p^4$$

$$= q^4 + 4q^3p + 6q^2p^2 + 4qp^3 + p^4$$

(*b*)

$$(q + p)^6 = q^6 + \binom{6}{1}q^5p + \binom{6}{2}q^4p^2 + \binom{6}{3}q^3p^3 + \binom{6}{4}q^2p^4 + \binom{6}{5}qp^5 + p^6$$

$$= q^6 + 6q^5p + 15q^4p^2 + 20q^3p^3 + 15q^2p^4 + 6qp^5 + p^6$$

Los coeficientes 1, 4, 6, 4, 1 y 1, 6, 15, 20, 15, 6, 1 se llaman *coeficientes binomiales* correspondientes a $N = 4$ y $N = 6$, respectivamente. Escribiendo estos coeficientes para $N = 0, 1, 2, 3, \ldots$, como muestra la disposición triangular adjunta, obtenemos el llamado *triángulo de Pascal*. Notemos que el primero y el último de los números de cada fila son 1 y que todo otro número se obtiene sumando sus dos vecinos de la fila de encima.

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

7.5. Hallar la probabilidad de que en una familia con 5 hijos haya: (*a*) al menos un chico y (*b*) al menos un chico y una chica. Suponemos que la probabilidad de que nazca chico es $\frac{1}{2}$.

Solución

(*a*)

$$\Pr\{1 \text{ chico}\} = \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4} \qquad \Pr\{3 \text{ chicos}\} = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = \frac{1}{4}$$

$$\Pr\{2 \text{ chicos}\} = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8} \qquad \Pr\{4 \text{ chicos}\} = \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$$

Por tanto

$$\Pr\{\text{al menos 1 chico}\} = \Pr\{1 \text{ chico}\} + \Pr\{2 \text{ chicos}\} + \Pr\{3 \text{ chicos}\} + \Pr\{4 \text{ chicos}\}$$

$$= \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{15}{16}$$

Otro método

$$\Pr\{\text{al menos 1 chico}\} = 1 - \Pr\{\text{ningún chico}\} = 1 - \left(\frac{1}{2}\right)^4 = 1 - \frac{1}{16} = \frac{15}{16}$$

(*b*)

$$\Pr\{\text{al menos 1 chico y 1 chica}\} = 1 - \Pr\{\text{ningún chico}\} - \Pr\{\text{ninguna chica}\} = 1 - \frac{1}{16} - \frac{1}{16} = \frac{7}{8}$$

7.6. De entre 2000 familias con 4 hijos, ¿cuántas cabe esperar que tengan: (*a*) al menos 1 chico, (*b*) 2 chicos, (*c*) 1 ó 2 chicas y (*d*) ninguna chica? Véase el Problema 7.5(*a*).

Solución

(*a*) Número esperado de familias con al menos 1 chico $= 2000(\frac{15}{16}) = 1875$

(*b*) Número esperado de familias con 2 chicos $= 2000 \cdot \Pr\{2 \text{ chicos}\} = 2000(\frac{3}{8}) = 750$

(*c*) $\Pr\{1 \text{ ó } 2 \text{ chicas}\} = \Pr\{1 \text{ chica}\} + \Pr\{2 \text{ chicas}\} = \Pr\{1 \text{ chico}\} + \Pr\{2 \text{ chicos}\} = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$.
Número esperado de familias con 1 ó 2 chicas $= 2000(\frac{5}{8}) = 1250$

(*d*) Número esperado de familias con ninguna chica $= 2000(\frac{1}{16}) = 125$

7.7. Si el 20% de los pernos producidos por una máquina son defectuosos, determinar la probabilidad de que, entre 4 pernos elegidos al azar: (*a*) 1, (*b*) 0 y (*c*) a lo sumo 2 sean defectuosos.

Solución

La probabilidad de un perno defectuoso es $p = 0.2$ y la de uno no defectuoso es $q = 1 - p = 0.8$.

(*a*)

$$\Pr\{1 \text{ defectuoso entre } 4\} = \binom{4}{1}(0.2)^1(0.8)^3 = 0.4096$$

(*b*)

$$\Pr\{0 \text{ defectuosos}\} = \binom{4}{0}(0.2)^0(0.8)^4 = 0.4096$$

(*c*)

$$\Pr\{2 \text{ defectuosos}\} = \binom{4}{2}(0.2)^2(0.8)^2 = 0.1536$$

Entonces

$$\begin{aligned}\Pr\{\text{de más de 2 pernos defectuosos}\} &= \Pr\{0 \text{ defectuosos}\} + \Pr\{1 \text{ defectuoso}\} + \{2 \text{ defectuosos}\}\\ &= 0.4096 + 0.4096 + 0.1536 = 0.9728\end{aligned}$$

7.8 La probabilidad de que un estudiante que ingresa en la Universidad se licencie es 0.4. Hallar la probabilidad de que entre 5 estudiantes elegidos al azar: (*a*) ninguno, (*b*) 1, (*c*) al menos 1 y (*d*) todos, se licencien.

Solución

(*a*) $\Pr\{\text{ninguno se licencie}\} = \binom{5}{0}(0.4)^0(0.6)^5 = 0.07776$ o sea, aproximadamente 0.08

(*b*) $\Pr\{1 \text{ se licencie}\} = \binom{5}{1}(0.4)^1(0.6)^4 = 0.2592$ o sea, aproximadamente 0.26

(*c*) $\Pr\{\text{al menos 1 se licencie}\} = 1 - \Pr\{\text{ninguno se licencie}\} = 0.92224$ o sea, aprox. 0.92

(*d*) $\Pr\{\text{todos se licenciarán}\} = \binom{5}{5}(0.4)^5(0.6)^0 = 0.01024$ o sea, aproximadamente 0.01

7.9 ¿Cuál es la probabilidad de obtener un total de 9: (*a*) dos veces y (*b*) al menos dos veces, en 6 tiradas de un par de dados?

Solución

Asociando los 6 posibles resultados del primer dado con los 6 del segundo, resulta un total de $6 \cdot 6 = 36$ posibles formas de caer los dados. Son: 1 en el primero y 1 en el segundo, 1 en el primero y 2 en el segundo, etc., denotadas por (1, 1), (1, 2), etc.

De esas 36 posibilidades equiprobables, la suma 9 ocurre en cuatro de ellas: (3, 6), (4, 5), (5, 4) y (6, 3). Luego la probabilidad de sacar 9 en una tirada es $p = \frac{4}{36} = \frac{1}{9}$, y la de no sacar 9 es $q = 1 - p = \frac{8}{9}$.

(*a*) $\Pr\{2 \text{ nueves en 6 tiradas}\} = \binom{6}{2}\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)^{6-2} = \frac{61{,}440}{531{,}441}$

(*b*) $\Pr\{\text{al menos 2 nueves}\} = \Pr\{2 \text{ nueves}\} + \Pr\{3 \text{ nueves}\} + \Pr\{4 \text{ nueves}\} + \Pr\{5 \text{ nueves}\} + \Pr\{6 \text{ nueves}\}$

$$= \binom{6}{2}\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)^4 + \binom{6}{3}\left(\frac{1}{9}\right)^3\left(\frac{8}{9}\right)^3 + \binom{6}{4}\left(\frac{1}{9}\right)^4\left(\frac{8}{9}\right)^2 + \binom{6}{5}\left(\frac{1}{9}\right)^5\left(\frac{8}{9}\right)^1 + \binom{6}{6}\left(\frac{1}{9}\right)^6\left(\frac{8}{9}\right)^0$$

$$= \frac{61{,}440}{531{,}441} + \frac{10{,}240}{531{,}441} + \frac{960}{531{,}441} + \frac{48}{531{,}441} + \frac{1}{531{,}441} = \frac{72{,}689}{531{,}441}$$

Otro método

$$\Pr\{\text{al menos 2 nueves}\} = 1 - \Pr\{0 \text{ nueves}\} - \Pr\{1 \text{ nueve}\}$$

$$= 1 - \binom{6}{0}\left(\frac{1}{9}\right)^0\left(\frac{8}{9}\right)^6 - \binom{6}{1}\left(\frac{1}{9}\right)^1\left(\frac{8}{9}\right)^5 = \frac{72{,}689}{531{,}441}$$

7.10. Evaluar: (*a*) $\sum_{X=0}^{N} Xp(X)$ y (*b*) $\sum_{X=0}^{N} X^2 p(X)$, donde $p(X) = \binom{N}{X} p^X q^{N-X}$.

Solución

(*a*) Como $q + p = 1$,

$$\sum_{X=0}^{N} Xp(X) = \sum_{X=1}^{N} X \frac{N!}{X!(N-X)!} p^X q^{N-X} = Np \sum_{X=1}^{N} \frac{(N-1)!}{(X-1)!(N-X)!} p^{X-1} q^{N-X}$$
$$= Np(q+p)^{N-1} = Np$$

(*b*)
$$\sum_{X=0}^{N} X^2 p(X) = \sum_{X=1}^{N} \frac{N!}{X!(N-X)!} p^X q^{N-X} = \sum_{X=1}^{N} [X(X-1) + X] \frac{N!}{X!(N-X)!} p^X q^{N-X}$$
$$= \sum_{X=2}^{N} X(X-1) \frac{N!}{X!(N-X)!} p^X q^{N-X} + \sum_{X=1}^{N} X \frac{N!}{X!(N-X)!} p^X q^{N-X}$$
$$= N(N-1)p^2 \sum_{X=2}^{N} \frac{(N-2)!}{(X-2)!(N-X)!} p^{X-2} q^{N-X} + Np = N(N-1)p^2(q+p)^{N-2} + Np$$
$$= N(N-1)p^2 + Np$$

Nota: Los resultados en las partes (*a*) y (*b*) son las esperanzas de X y X^2, denotadas por $E(X)$ y $E(X^2)$, respectivamente (véase Cap. 6).

7.11. Si una variable está normalmente distribuida, determinar: (*a*) su media μ y (*b*) su varianza σ^2.

Solución

(*a*) Por el Problema 7.10(*a*),

$$\mu = \text{valor esperado de la variable} = \sum_{X=0}^{N} Xp(X) = Np$$

(*b*) Usando $\mu = Np$ y los resultados del Problema 7.10,

$$\sigma^2 = \sum_{X=0}^{N} (X-\mu)^2 p(X) = \sum_{X=0}^{N} (X^2 - 2\mu X + \mu^2)p(X) = \sum_{X=0}^{N} X^2 p(X) - 2\mu \sum_{X=0}^{N} Xp(X) + \mu^2 \sum_{X=0}^{N} p(X)$$
$$= N(N-1)p^2 + Np - 2(Np)(Np) + (Np)^2(1) = Np - Np^2 = Np(1-p) = Npq$$

Se desprende que la desviación típica de una variable normalmente distribuida es $\sigma = \sqrt{Npq}$.

Otro método

Por el Problema 6.62(*b*),

$$E[(X - \bar{X})]^2 = E(X^2) - [E(X)]^2 = N(N-1)p^2 + Np - N^2p^2 = Np - Np^2 = Npq$$

7.12. Si la probabilidad de un perno defectuoso es 0.1, hallar: (*a*) la media y (*b*) la desviación típica, para la distribución de pernos defectuosos en un total de 400.

Solución

(*a*) La media es $Np = 400(0.1) = 40$; esto es, esperamos 40 pernos defectuosos.

(*b*) La varianza es $Npq = 400(0.1)(0.9) = 36$. Por tanto, la desviación típica es $\sqrt{36} = 6$.

7.13. Hallar los coeficientes momento de: (*a*) sesgo y (*b*) curtosis de la distribución del Problema 7.12.

Solución

(*a*)

$$\text{Coeficiente momento de sesgo} = \frac{q-p}{\sqrt{Npq}} = \frac{0.9 - 0.1}{6} = 0.133$$

Como es positivo, la distribución es sesgada a la derecha.

(*b*)

$$\text{Coeficiente momento de curtosis} = 3 + \frac{1 - 6pq}{Npq} = 3 + \frac{1 - 6(0.1)(0.9)}{36} = 3.01$$

La distribución es un poco *leptocúrtica* con respecto a la distribución normal (o sea, algo más puntiaguda; véase Cap. 5).

LA DISTRIBUCION NORMAL

7.14. En un examen de matemáticas, la calificación media fue 72 y la desviación típica 15. Determinar en unidades estándar las puntuaciones de los alumnos que obtuvieron: (*a*) 60, (*b*) 93 y (*c*) 72.

Solución

(*a*) $z = \dfrac{X - \bar{X}}{s} = \dfrac{60 - 72}{15} = -0.8$ (*c*) $z = \dfrac{X - \bar{X}}{s} = \dfrac{72 - 72}{15} = 0$

(*b*) $z = \dfrac{X - \bar{X}}{s} = \dfrac{93 - 72}{15} = 1.4$

7.15. Con referencia al Problema 7.14, hallar las puntuaciones correspondientes a las puntuaciones estándar: (*a*) -1 y (*b*) 1.6.

Solución

(*a*) $X = \bar{X} + zs = 72 + (-1)(15) = 57$ (*b*) $X = \bar{X} + zs = 72 + (1.6)(15) = 96$

7.16. Se informó a dos estudiantes que habían recibido puntuaciones estándar de 0.8 y -0.4, respectivamente, en una prueba de inglés. Si sus puntuaciones fueron 88 y 64, respectivamente, hallar la media y la desviación típica de las puntuaciones de esa prueba.

Solución

Usando la ecuación $X = \bar{X} + zs$, tenemos $88 = \bar{X} + 0.8s$ para el primer estudiante y $64 = \bar{X} - 0.4s$ para el segundo. Resolviendo esas ecuaciones se obtiene $\bar{X} = 72$ y $s = 20$.

7.17. Hallar el área bajo la curva normal en cada uno de los casos siguientes: (*a*) a (*g*), que corresponden a las Figuras 7.2(*a*) a 7.2(*g*), respectivamente. Usar el Apéndice II.

(*a*) Entre $z = 0$ y $z = 1.2$
(*b*) Entre $z = -0.68$ y $z = 0$
(*c*) Entre $z = -0.46$ y $z = 2.21$
(*d*) Entre $z = 0.81$ y $z = 1.94$
(*e*) A la izquierda de $z = -0.6$
(*f*) A la derecha de $z = -1.28$
(*g*) A la derecha de $z = 2.05$, y a la izquierda de $z = -1.44$

Solución

(*a*) En el Apéndice II miramos en la columna marcada z hasta ver la entrada 1.2; entonces nos desplazamos a la derecha a la columna marcada 0. El resultado, 0.3849, es el área pedida y representa la probabilidad de que z esté entre 0 y 1.2, denotada $\Pr\{0 \leqslant z \leqslant 1.2\}$.

(*b*) Por simetría, el área solicitada es la que hay entre $z = 0$ y $z = 0.68$. Para hallarla, buscamos en la columna marcada z en el Apéndice II hasta localizar 0.6; entonces a la derecha hasta la columna 8. El resultado, 0.2517, es el área buscada y representa la probabilidad de que z esté entre -0.68 y 0, denotada $\Pr\{-0.68 \leqslant z \leqslant 0\}$.

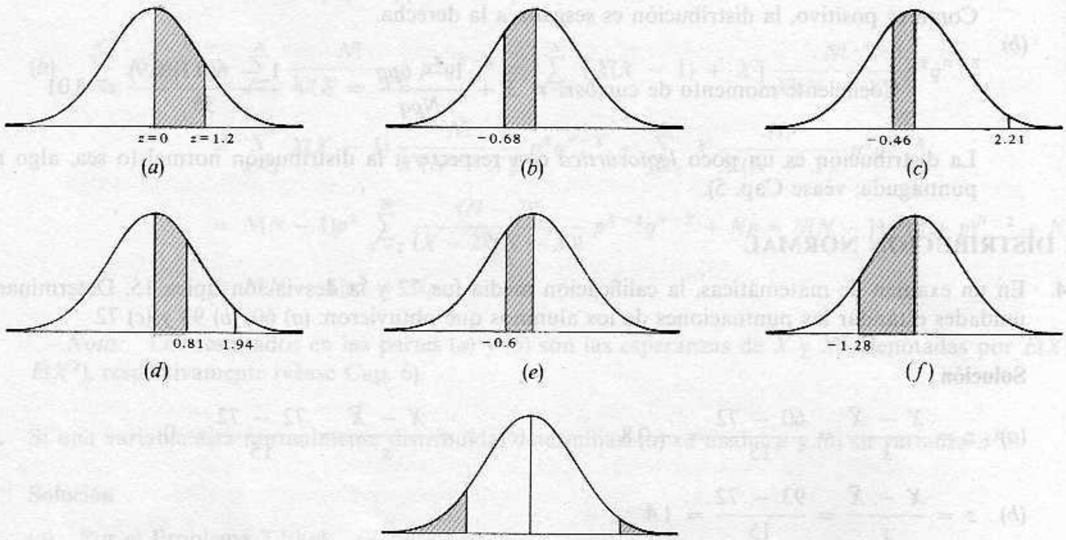


Figura 7.2.

(c) Area pedida $=$ (área entre $z = -0.46$ y $z = 0$) $+$ (área entre $z = 0$ y $z = 2.21$)
$=$ (área entre $z = 0$ y $z = 0.46$) $+$ (área entre $z = 0$ y $z = 2.21$)
$= 0.1772 + 0.4864 = 0.6636$

(d) Area pedida $=$ (área entre $z = 0$ y $z = 1.94$) $-$ (área entre $z = 0$ y $z = 0.81$)
$= 0.4738 - 0.2910 = 0.1828$

(e) Area pedida $=$ (área a la izquierda de $z = 0$) $-$ (área entre $z = -0.6$ y $z = 0$)
$=$ (área a la izquierda de $z = 0$) $-$ (área entre $z = 0$ y $z = 0.6$)
$= 0.5 - 0.2258 = 0.2742$

(f) Area pedida $=$ (área entre $z = -1.28$ y $z = 0$) $+$ (área a la derecha de $z = 0$)
$= 0.3997 + 0.5 = 0.8997$

(g) Area pedida $=$ área total $-$ (área entre $z = -1.44$ y $z = 0$) $-$ (área entre $z = 0$ y $z = 2.05$)
$= 1 - 0.4251 - 0.4798 = 1 - 0.9049 = 0.0951$

7.18. Determinar el valor o valores de z en los casos: (a), (b) y (c), que corresponden a las Figuras 7.3(a) a 7.3(c), respectivamente. La palabra «área» se refiere al área bajo la curva normal.

(a) El área entre 0 y z es 0.3770.
(b) El área a la izquierda de z es 0.8621.
(c) El área entre -1.5 y z es 0.0217.

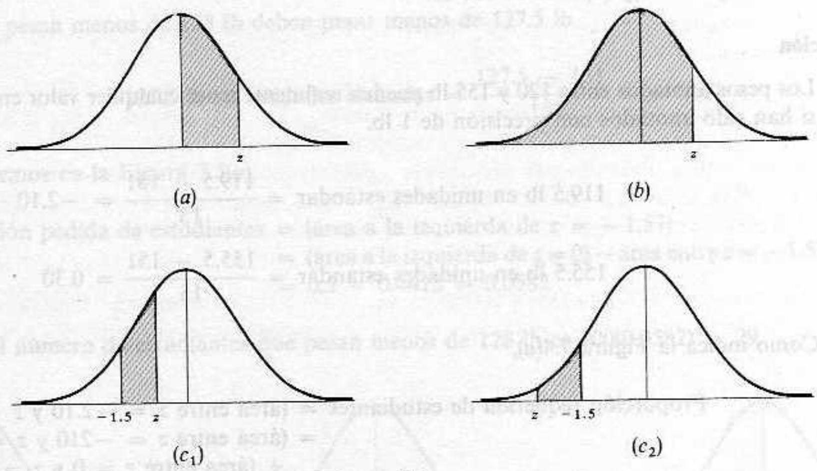


Figura 7.3.

Solución

(a) En el Apéndice II la entrada 0.3770 está a la derecha de la fila marcada 1.1 y bajo la columna 6; así pues, el z pedido es $z = 1.16$. Por simetría, $z = -1.16$ es otro valor solución de z, con lo que $z = \pm 1.16$.

(b) Como el área es mayor que 0.5, z debe ser positivo. El área entre 0 y $z = 0.8621 - 0.5 = 0.3621$, de donde $z = 1.09$.

(c) Si z fuera positivo, el área sería mayor que el área entre -1.5 y 0, que es 0.4332; luego z es negativo.

Caso 1 [z negativo, pero a la derecha de -1.5; véase Fig. 7.3(c_1)]

El área entre -1.5 y z = (área entre -1.5 y 0) $-$ (área entre 0 y z), y $0.0217 = 0.4332 -$ (área entre 0 y z). Así pues, el área entre 0 y $z = 0.4332 - 0.0217 = 0.4115$, de donde $z = -1.35$.

Caso 2 [z negativo, pero a la izquierda de -1.5; véase Fig. 7.3(c_2)]

El área entre z y -1.5 = (área entre z y 0) $-$ (área entre -1.5 y 0), y 0.0217 = (área entre 0 y z) $-$ 0.4332. Luego el área entre 0 y $z = 0.0217 + 0.4332 = 0.4549$, y $z = -1.694$ por interpolación lineal; o sea, con menos precisión, $z = -1.69$.

7.19. Hallar las ordenadas de la curva normal en: (*a*) $z = 0.84$, (*b*) $z = -1.27$ y (*c*) $z = -0.05$.

Solución

(*a*) En el Apéndice I, buscamos la entrada 0.8 en la columna de z y luego nos movemos a la derecha hasta la columna 4. La entrada 0.2803 es la ordenada pedida.
(*b*) Por simetría: (ordenada en $z = -1.27$) = (ordenada en $z = 1.27$) = 0.1781.
(*c*) (Ordenada en $z = -0.05$) = (ordenada en $z = 0.05$) = 0.3984.

7.20. El peso medio de 500 estudiantes varones de cierta Universidad es 151 libras (lb), y la desviación típica es 15 lb. Supuesto que los pesos están normalmente distribuidos, hallar cuántos estudiantes pesan: (*a*) entre 120 y 155 lb y (*b*) más de 185 lb.

Solución

(*a*) Los pesos anotados entre 120 y 155 lb pueden realmente tener cualquier valor entre 119.5 a 155.5 lb, si han sido anotados con precisión de 1 lb.

$$119.5 \text{ lb en unidades estándar} = \frac{119.5 - 151}{15} = -2.10$$

$$155.5 \text{ lb en unidades estándar} = \frac{155.5 - 151}{15} = 0.30$$

Como indica la Figura 7.4(*a*),

$$\begin{aligned}\text{Proporción requerida de estudiantes} &= (\text{área entre } z = -2.10 \text{ y } z = 0.30)\\ &= (\text{área entre } z = -210 \text{ y } z = 0)\\ &\quad + (\text{área entre } z = 0 \text{ y } z = 0.30)\\ &= 0.4821 + 0.1179 = 0.6000\end{aligned}$$

Luego el número de estudiantes que pesan entre 120 y 155 lb es 500(0.6000) = 300.

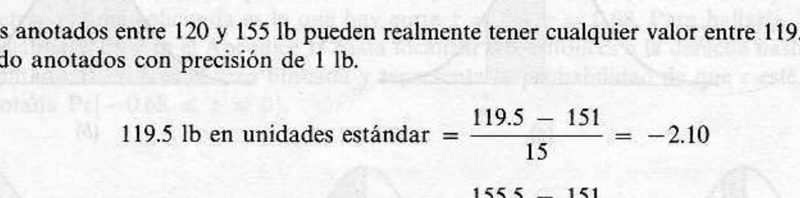


Figura 7.4.

(*b*) Los estudiantes que pesan más de 185 lb han de pesar al menos 185.5 lb.

$$185.5 \text{ lb en unidades estándar} = \frac{185.5 - 151}{15} = 2.30$$

Como se ve en la Figura 7.4(*b*),

$$\begin{aligned}\text{Proporción de estudiantes requerida} &= (\text{área a la derecha de } z = 2.30)\\ &= (\text{área a la derecha de } z = 0) - (\text{área entre } z = 0 \text{ y } z = 2.30)\\ &= 0.5 - 0.4893 = 0.0107\end{aligned}$$

Así que el número de estudiantes que pesan más de 185 lb es 500(0.0107) = 5.

Si W denota el peso de un estudiante al azar, podemos resumir los resultados precedentes en términos de probabilidad escribiendo

$$\Pr\{119.5 \leqslant W \leqslant 155.5\} = 0.6000 \qquad \text{y} \qquad \Pr\{W \geqslant 185.5\} = 0.0107$$

7.21. Determinar cuántos de los 500 estudiantes del problema anterior pesan: (*a*) menos de 128 lb, (*b*) 128 lb, y (*c*) no más de 128 lb.

Solución

(*a*) Los que pesan menos de 128 lb deben pesar menos de 127.5 lb

$$127.5 \text{ lb en unidades estándar} = \frac{127.5 - 151}{15} = -1.57$$

Como vemos en la Figura 7.5(*a*),

$$\begin{aligned}\text{Proporción pedida de estudiantes} &= (\text{área a la izquierda de } z = -1.57)\\ &= (\text{área a la izquierda de } z = 0) - \text{área entre } z = -1.57 \text{ y } z = 0)\\ &= 0.5 - 0.4418 = 0.0582\end{aligned}$$

Luego el número de estudiantes que pesan menos de 128 lb es 500(0.0582) = 29.

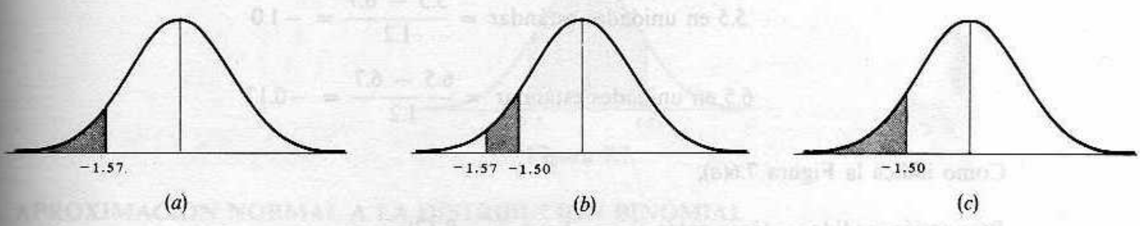


Figura 7.5.

(*b*) Los que pesan 128, en realidad pesan entre 127.5 y 128.5 lb

$$127.5 \text{ lb en unidades estándar} = \frac{127.5 - 151}{15} = -1.57$$

$$128.5 \text{ lb en unidades estándar} = \frac{128.5 - 151}{15} = -1.50$$

Como muestra la Figura 7.5(b),

$$\begin{aligned}\text{Proporción requerida de estudiantes} &= (\text{área entre } z = -1.57 \text{ y } z = -1.50)\\ &= (\text{área entre } z = -1.57 \text{ y } z = 0)\\ &\quad -(\text{área entre } z = -1.50 \text{ y } z = 0)\\ &= 0.4418 - 0.4332 = 0.0086\end{aligned}$$

Por tanto, el número de estudiantes que pesan 128 lb es 500(0.0086) = 4.

(c) Los que no pasan de 128 lb deben pesar 128.5 lb

$$128.5 \text{ lb en unidades estándar} = \frac{128.5 - 151}{15} = -1.50$$

Como muestra la Figura 7.5(c),

$$\begin{aligned}\text{Proporción requerida de estudiantes} &= (\text{área a la izquierda de } z = -1.50)\\ &= (\text{área a la izquierda de } z = 0) - (\text{área entre } z = -1.50 \text{ y } z = 0)\\ &= 0.5 - 0.4332 = 0.0668\end{aligned}$$

Luego el número de estudiantes que no sobrepasan las 128 lb es 500(0.0668) = 33.

Otro método [usando las partes (a) y (b)]

El número de los que no pasan de 128 lb es (los que pesan menos de 128 lb) + (los que pesan 128 lb) = 29 + 4 = 33.

7.22. Las puntuaciones en un test de biología eran 0, 1, 2, ..., 10 puntos, según el número de respuestas correctas de entre las 10 cuestiones. La nota media fue 6.7 y la desviación típica 1.2. Supuesto que las notas estuvieran normalmente distribuidas, determinar: (a) el porcentaje de estudiantes que tuvo 6 puntos, (b) la nota máxima del 10% más bajo y (c) la nota mínima del 10% más alto de la clase.

Solución

(a) Para aplicar la distribución normal a datos discretos es necesario tratar los datos como si fueran continuos. Así que una nota de 6 puntos se considera que está entre 5.5 y 6.5 puntos

$$5.5 \text{ en unidades estándar} = \frac{5.5 - 6.7}{1.2} = -1.0$$

$$6.5 \text{ en unidades estándar} = \frac{6.5 - 6.7}{1.2} = -0.17$$

Como indica la Figura 7.6(a),

$$\begin{aligned}\text{Proporción pedida} &= (\text{área entre } z = -1 \text{ y } z = -0.17)\\ &= (\text{área entre } z = -1 \text{ y } z = 0) - (\text{área entre } z = -0.17 \text{ y } z = 0)\\ &= 0.3413 - 0.0675 = 0.2738 = 27\%\end{aligned}$$

(b) Sea X_1 la nota máxima y z_1 la nota en unidades estándar. De la Figura 7.6(b) se ve que el área a la izquierda de z_1 es 10% = 0.10; por tanto: (área entre z_1 y 0) = 0.40, y $z_1 = -1.28$ (muy aproximadamente). Luego $z_1 = (X_1 - 6.7)/1.2 = -1.28$; y $X_1 = 5.2$, o sea 5 redondeando.

(c) Sea X_2 la nota mínima y z_2 la nota en unidades estándar. De la parte (b), por simetría, $z_2 = 1.28$. Luego $(X_2 = 6.7)/1.2 = 1.28$; y $X_2 = 8.2$, o sea 8 redondeando.

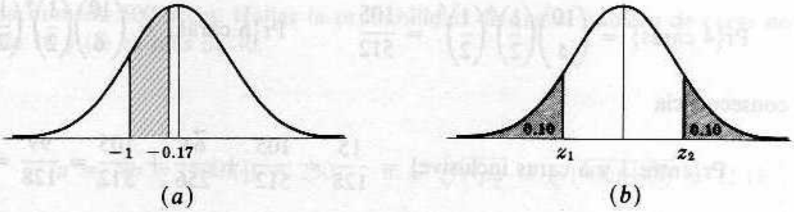


Figura 7.6.

7.23. El diámetro medio interior de una muestra de 200 tubos producidos por una máquina es 0.502 pulgadas (in) y la desviación típica es 0.005 in. El uso de los tubos permitirá una tolerancia en el diámetro de 0.496 a 0.508 in; de otro modo, se considerarán defectuosos. Determinar el porcentaje de tubos defectuosos, supuesto que los tubos producidos por esa máquina están normalmente distribuidos.

Solución

$$0.496 \text{ en unidades estándar} = \frac{0.496 - 0.502}{0.005} = -1.2$$

$$0.508 \text{ en unidades estándar} = \frac{0.508 - 0.502}{0.005} = 1.2$$

Como muestra la Figura 7.7,

$$\begin{aligned} \text{Proporción de tubos defectuosos} &= (\text{área bajo la curva normal entre } z = -1.2 \text{ y } z = 1.2) \\ &= (\text{doble del área entre } z = 0 \text{ y } z = 1.2) \\ &= 2(0.3849) = 0.7698 \quad \text{o sea} \quad 77\% \end{aligned}$$

Luego el porcentaje de tubos defectuosos es 100% − 77% = 23%.

Nótese que si pensamos que el intervalo de 0.496 a 0.508 representa diámetros desde 0.4955 hasta 0.5085 in, el resultado anterior cambia ligeramente. Con dos cifras significativas, sin embargo, el resultado se mantiene.

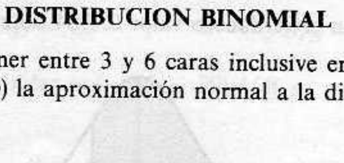


Figura 7.7.

APROXIMACION NORMAL A LA DISTRIBUCION BINOMIAL

7.24. Hallar la probabilidad de obtener entre 3 y 6 caras inclusive en 10 tiradas de una moneda, usando: (*a*) la distribución binomial y (*b*) la aproximación normal a la distribución binomial.

Solución

(*a*)

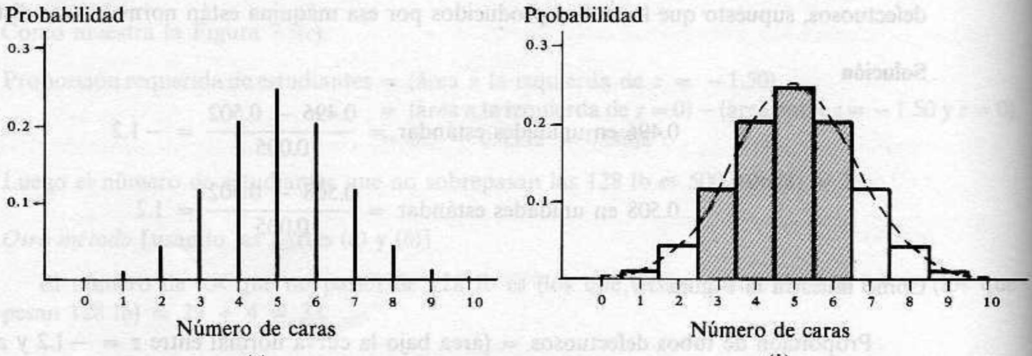
$$\Pr\{3 \text{ caras}\} = \binom{10}{3}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^{15} = \frac{15}{128} \qquad \Pr\{5 \text{ caras}\} = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{63}{256}$$

$$\Pr\{4 \text{ caras}\} = \binom{10}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 = \frac{105}{512} \qquad \Pr\{6 \text{ caras}\} = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = \frac{105}{512}$$

En consecuencia

$$\Pr\{\text{entre 3 y 6 caras inclusive}\} = \frac{15}{128} + \frac{105}{512} + \frac{63}{256} + \frac{105}{512} = \frac{99}{128} = 0.7734$$

(*b*) La distribución de Poisson para el número de caras en 10 tiradas está representada en las Figuras 7.8(*a*) y (*b*), donde esta última trata los datos como si fueran continuos. La probabilidad pedida es la suma de las áreas de los rectángulos sombreados de la Figura 7.8(*b*) y se puede aproximar por el área correspondiente bajo la curva normal, en sombra en la figura.



(*a*) (*b*)

Figura 7.8.

Considerando los datos como continuos, se sigue que 3 a 6 caras es como decir de 2.5 a 6.5 caras. Además, la media y la varianza de la distribución binomial vienen dados por $\mu = Np = 10(\frac{1}{2}) = 5$ y $\sigma = \sqrt{Npq} = \sqrt{(10)(\frac{1}{2})(\frac{1}{2})} = 1.58$

$$2.5 \text{ en unidades estándar} = \frac{2.5 - 5}{1.58} = -1.58$$

$$6.5 \text{ en unidades estándar} = \frac{6.5 - 5}{1.58} = 0.95$$

Como se ve en la Figura 7.9,

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{área entre } z = -1.58 \text{ y } z = 0.95)\\ &= (\text{área entre } z = -1.58 \text{ y } z = 0) + (\text{área entre } z = 0 \text{ y } z = 0.95)\\ &= 0.4429 + 0.3289 = 0.7718\end{aligned}$$

que encaja muy bien con el verdadero valor 0.7734 obtenido en la parte (*a*). La precisión es aún mayor para grandes *N*.

Figura 7.9.

7.25. Se lanza una moneda 500 veces. Hallar la probabilidad de que el número de caras no difiera de 250: (*a*) en más de 10 y (*b*) en más de 30.

Solución

$$\mu = Np = (500)(\tfrac{1}{2}) = 250 \qquad \sigma = \sqrt{Npq} = \sqrt{(500)(\tfrac{1}{2})(\tfrac{1}{2})} = 11.18$$

(*a*) Se nos pide la probabilidad de que el número de caras esté entre 240 y 260, o sea, considerando los datos como continuos, entre 239.5 y 260.5. Como 239.5 en unidades estándar es $(239.5 - 250)/11.18 = -0.94$, y 260.5 en unidades estándar es 0.94, tenemos

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{área bajo la curva normal entre } z = -0.94 \text{ y } z = 0.94)\\ &= (\text{doble del área entre } z = 0 \text{ y } z = 0.94) = 2(0.3264) = 0.6528\end{aligned}$$

(*b*) Se pide la probabilidad de que el número de caras esté entre 220 y 280, o considerados los datos como continuos, entre 219.5 y 280.5. Como 219.5 en unidades estándar es $(219.5 - 250)/11.18 = -2.73$, y 280.5 en unidades estándar es 2.73, tenemos

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{el doble del área bajo la curva normal entre } z = 0 \text{ y } z = -2.73)\\ &= 2(0.4968) = 0.9936\end{aligned}$$

Se sigue que, con gran confianza, el número de caras no diferirá del esperado (250) en más de 30. Así pues, si resultase que el número real de caras fuera 280, tendríamos derecho a sospechar que la moneda estaba trucada o era falsa.

7.26. Se lanza un dado 120 veces. Hallar la probabilidad de que salga el 4: (*a*) 18 veces o menos y (*b*) 14 veces o menos, supuesto como siempre que el dado no está trucado.

Solución

El 4 tiene probabilidad $p = \frac{1}{6}$ de salir y probabilidad $q = \frac{5}{6}$ de no salir.

(*a*) Queremos calcular la probabilidad de que el número de cuatros esté entre 0 y 18, y eso es exactamente

$$\binom{120}{18}\left(\frac{1}{6}\right)^{18}\left(\frac{5}{6}\right)^{102} + \binom{120}{17}\left(\frac{1}{6}\right)^{17}\left(\frac{5}{6}\right)^{103} + \cdots + \binom{120}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{120}$$

pero como la tarea de calcular esto es ímproba, usemos la aproximación normal.

Considerando los datos como continuos, de 0 a 18 significa de -0.5 a 18.5. Además,

$$\mu = Np = 120(\tfrac{1}{6}) = 20 \qquad \text{y} \qquad \sigma = \sqrt{Npq} = \sqrt{(120)(\tfrac{1}{6})(\tfrac{5}{6})} = 4.08$$

Como -0.5 en unidades estándar es $(-0.5 - 20)/4.08 = -5.02$, y 18.5 en unidades estándar es -0.37, se tiene

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{área bajo la curva normal entre } z = -5.02 \text{ y } z = -0.37)\\ &= (\text{área entre } z = 0 \text{ y } z = -5.02)\\ &\quad - (\text{área entre } z = 0 \text{ y } z = -0.37)\\ &= 0.5 - 0.1443 = 0.3557\end{aligned}$$

(b) Procedemos como en (a), sustituyendo 18 por 14. Como -0.5 en unidades estándar es -5.02, y 14.5 en unidades estándar es $(14.5 - 20)/4.08 = -1.35$, tenemos

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{área bajo la curva normal entre } z = -5.02 \text{ y } z = -1.35)\\ &= (\text{área entre } z = 0 \text{ y } z = -5.02)\\ &\quad - (\text{área entre } z = 0 \text{ y } z = -1.35)\\ &= 0.5 - 0.4115 = 0.0885\end{aligned}$$

Se desprende que si tomamos repetidas muestras de 120 lanzamientos de un dado, el 4 saldría 14 veces o menos en aproximadamente un 10% de esas muestras.

LA DISTRIBUCION DE POISSON

7.27. Un 10% de las herramientas producidas en una fábrica son defectuosas. Hallar la probabilidad de que en una muestra de 10 herramientas tomadas al azar exactamente 2 sean defectuosas, usando: (a) la distribución binomial y (b) la aproximación de Poisson a la distribución binomial.

Solución

La probabilidad de una herramienta defectuosa es $p = 0.1$.

(a)
$$\Pr\{2 \text{ objetos defectuosos en } 10\} = \binom{10}{2}(0.1)^2(0.9)^8 = 0.1937 \quad \text{o sea} \quad 0.19$$

(b) Con $\lambda = Np = 10(0.1) = 1$ y usando $e = 2.718$,

$$\Pr\{2 \text{ objetos defectuosos en } 10\} = \frac{\lambda^X e^{-\lambda}}{X!} = \frac{(1)^2 e^{-1}}{2!} = \frac{e^{-1}}{2} = \frac{1}{2e} = 0.1839 \quad \text{o sea} \quad 0.18$$

En general, la aproximación de Poisson es buena si $p \leqslant 0.1$ y $\lambda = Np \leqslant 5$.

7.28. Si la probabilidad de que un individuo sufra una reacción negativa ante una inyección de cierto suero es 0.001, hallar la probabilidad de que entre 2000 individuos: (a) exactamente 3 y (b) más de 2 de ellos reaccionen negativamente.

Solución

$$\Pr\{X \text{ individuos reaccionen negativamente}\} = \frac{\lambda^X e^{-\lambda}}{X!} = \frac{2^X e^{-2}}{X!}$$

donde $\lambda = Np = (2000)(0.001) = 2$.

(a)
$$\Pr\{3 \text{ individuos reaccionen negativamente}\} = \frac{2^3 e^{-2}}{3!} = \frac{4}{3e^2} = 0.180$$

(b)
$$\Pr\{0 \text{ la sufran}\} = \frac{2^0 e^{-2}}{0!} = \frac{1}{e^2} \qquad \Pr\{1 \text{ la sufra}\} = \frac{2^1 e^{-2}}{1!} = \frac{2}{e^2} \qquad \Pr\{2 \text{ la sufran}\} = \frac{2^2 e^{-2}}{2!} = \frac{2}{e^2}$$

$$\begin{aligned}\Pr\{\text{más de 2 la sufran}\} &= 1 - \Pr\{0 \text{ ó } 1 \text{ ó } 2 \text{ la sufran}\}\\ &= 1 - \left(\frac{1}{e^2} + \frac{2}{e^2} + \frac{2}{e^2}\right) = 1 - \frac{5}{e^2} = 0.323\end{aligned}$$

Nótese que de acuerdo con la distribución binomial las probabilidades solicitadas en (*a*) y (*b*) son, respectivamente,

(*a*) $\binom{2000}{3}(0.001)^3(0.999)^{1997}$

(*b*) $1 - \left\{\binom{2000}{0}(0.001)^0(0.999)^{2000} + \binom{2000}{1}(0.001)^1(0.999)^{1999} + \binom{2000}{2}(0.001)^2(0.999)^{1998}\right\}$

mucho más difíciles de evaluar directamente.

7.29. Una distribución de Poisson viene dada por

$$p(X) = \frac{(0.72)^X e^{-0.72}}{X!}$$

Calcular: (*a*) $p(0)$, (*b*) $p(1)$, (*c*) $p(2)$ y (*d*) $p(3)$.

Solución

(*a*)

$$p(0) = \frac{(0.72)^0 e^{-0.72}}{0!} = \frac{(1)e^{-0.72}}{1} = e^{-0.72} = 0.4868 \quad \text{usando el Apéndice VIII}$$

(*b*)

$$p(1) = \frac{(0.72)^1 e^{-0.72}}{1!} = 0.72e^{-0.72} = (0.72)(0.4868) = 0.3505$$

(*c*)

$$p(2) = \frac{(0.72)^2 e^{-0.72}}{2!} = \frac{(0.5184)e^{-0.72}}{2} = (0.2592)(0.4868) = 0.1262$$

Otro método

$$p(2) = \frac{0.72}{2}\,p(1) = (0.36)(0.3505) = 0.1262$$

(*d*)

$$p(3) = \frac{(0.72)^3 e^{-0.72}}{3!} = \frac{0.72}{3}\,p(2) = (0.24)(0.1262) = 0.0303$$

DISTRIBUCION MULTINOMIAL

Una caja contiene 5 bolas rojas, 4 blancas y 3 azules. Se saca al azar una bola de la caja, se anota su color y se vuelve a meter en la caja. Hallar la probabilidad de que entre 6 bolas así seleccionadas, 3 sean rojas, 2 blancas y 1 azul.

Solución

Pr{roja en cualquier extracción} = $\frac{5}{12}$, Pr{blanca en cualquier extracción} = $\frac{4}{12}$, Pr{azul en cualquier extracción} = $\frac{3}{12}$; luego

$$\Pr\{3 \text{ son rojas, } 2 \text{ son blancas, } 1 \text{ es azul}\} = \frac{6!}{3!2!1!}\left(\frac{5}{12}\right)^3\left(\frac{4}{12}\right)^2\left(\frac{3}{12}\right)^1 = \frac{625}{5184}$$

AJUSTE DE DATOS MEDIANTE DISTRIBUCIONES TEORICAS

7.31. Ajustar una distribución binomial a los datos del Problema 2.17.

Solución

$\Pr\{X$ caras en una tirada de 5 monedas$\} = p(X) = \binom{5}{X}p^X q^{5-X}$, donde p y q son las respectivas probabilidades de cara y cruz en una sola tirada. Por el Problema 7.11(*a*), el número medio de caras es $\mu = Np = 5p$. Para la distribución de frecuencias realmente observada, el número medio de caras es

$$\frac{\sum fX}{\sum f} = \frac{(38)(0) + (144)(1) + (342)(2) + (287)(3) + (164)(4) + (25)(5)}{1000} = \frac{2470}{1000} = 2.47$$

Igualando la media teórica con la observada, $5p = 2.47$, o sea $p = 0.494$. Luego la distribución binomial de ajuste viene dada por $p(X) = \binom{5}{X}(0.494)^X(0.506)^{5-X}$.

La Tabla 7.4 recoge las probabilidades así como las frecuencias esperadas (teóricas) y observadas. Se ve que el ajuste es bueno. Su bondad se investigará en el Problema 12.12.

Tabla 7.4

Número de caras (X)	$\Pr\{X$ caras$\}$	Frecuencia esperada	Frecuencia observada
0	0.0332	33.2, o sea 33	38
1	0.1619	161.9, o sea 162	144
2	0.3162	316.2, o sea 316	342
3	0.3087	308.7, o sea 309	287
4	0.1507	150.7, o sea 151	164
5	0.0294	29.4, o sea 29	25

7.32. Usar papel gráfico de probabilidad para determinar si la distribución de frecuencias de la Tabla 2.1 puede aproximarse bien por una distribución normal.

Solución

Primero se convierte la distribución de frecuencias dada en una distribución de frecuencias relativas acumuladas, como indica la Tabla 7.5. Entonces, las frecuencias relativas acumuladas, expresadas en porcentajes, se marcan en el gráfico del papel especial citado (Fig. 7.10). El grado en que tales puntos caen sobre una recta determina la precisión del ajuste de la distribución dada a una distribución normal. De lo anterior vemos que hay una distribución normal que ajusta muy bien los datos (véase el Problema 7.33).

Tabla 7.5

Altura (in)	Frecuencia relativa acumulada (%)
Menor que 62.5	5.0
Menor que 65.5	23.0 (c_1)
Menor que 68.5	65.0
Menor que 71.5	92.0
Menor que 74.5	100.0

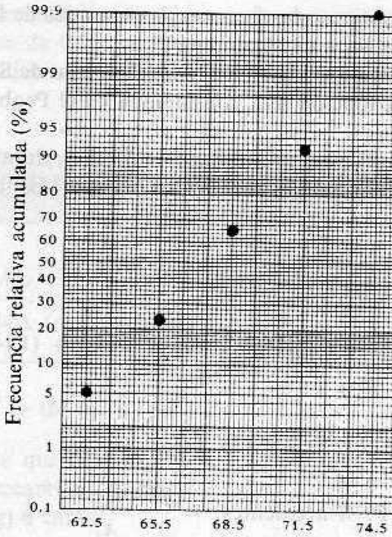


Figura 7.10.

7.33. Ajustar con una curva normal los datos de la Tabla 2.1.

Solución

El método lo esboza la Tabla 7.6. Al calcular z para las fronteras de clase, usamos $z = (X - \bar{X})/s$, donde la media $\bar{X}$ y la desviación típica s se han obtenido, respectivamente, en los Problemas 3.22 y 4.17.

Tabla 7.6

Alturas (in)	Fronteras de clase (X)	z para fronteras de clase	Area bajo la curva normal desde 0 a z	Area para cada clase	Frecuencia esperada	Frecuencia observada
	59.5	−2.72	0.4967			
60-62				0.0413	4.13, o sea 4	5
	62.5	−1.70	0.4554			
63-65				0.2068	20.68, o sea 21	18
	65.5	−0.67	0.2486			
66-68			Suma	→0.3892	38.92, o sea 39	42
	68.5	0.36	0.1406			
69-71				0.2771	27.71, o sea 28	27
	71.5	1.39	0.4177			
72-74				0.0743	7.43, o sea 7	8
	74.5	2.41	0.4920			

Suma $\qquad \bar{X} = 67.45$ in $\qquad s = 2.92$ in

En la columna 4 de la Tabla 7.6, las áreas bajo la curva normal entre 0 y z se han obtenido del Apéndice II. De ahí hallamos las áreas bajo la curva normal entre sucesivos valores de z, como muestra la columna 5. Se obtienen sin más que restar las áreas sucesivas de la columna 4 cuando las correspondientes z tienen el mismo signo, y sumando si son de signo opuesto (lo que ocurre sólo una vez en la tabla).

Multiplicando las entradas de la columna 5 (que representan frecuencias relativas) por la frecuencia total N (en este caso $N = 100$) se obtienen las frecuencias esperadas de la columna 6. Veamos que hay buen acuerdo con las frecuencias observadas (columna 7).

Si se desea, puede emplearse la desviación típica con corrección de Sheppard [véase Prob. 4.21(a)]. La bondad del ajuste de la distribución será considerada en el Problema 12.13.

7.34. La Tabla 7.7 muestra el número f de días, en un plazo de 50 días, durante los cuales se produjeron X accidentes de automóvil en una cierta ciudad. Ajustar los datos mediante una distribución de Poisson.

Solución

El número medio de accidentes es

$$\lambda = \frac{\sum fX}{\sum f} = \frac{(21)(0) + (18)(1) + (7)(2) + (3)(3) + (1)(4)}{50} = \frac{45}{50} = 0.90$$

Luego, de acuerdo con la distribución de Poisson,

$$\Pr\{X \text{ accidentes}\} = \frac{(0.90)^X e^{-0.90}}{X!}$$

Tabla 7.7

Número de accidentes (X)	Número de días (f)
0	21
1	18
2	7
3	3
4	1
	Total 50

La Tabla 7.8 da las probabilidades de 0, 1, 2, 3 y 4 accidentes que predice la distribución de Poisson y el número esperado o teórico en los cuales se producen X accidentes (obtenidos multiplicando las respectivas probabilidades por 50). Para facilitar la comparación, la columna 4 repite el número real de días de la Tabla 7.7.

Tabla 7.8

Número de accidentes (X)	$\Pr\{X \text{ accidentes}\}$	Número esperado de días	Número real de días
0	0.4066	20.33, o sea 20	21
1	0.3659	18.30, o sea 18	18
2	0.1647	8.24, o sea 8	7
3	0.0494	2.47, o sea 2	3
4	0.0111	0.56, o sea 1	1

Nótese que el ajuste es bueno.

Para una verdadera distribución de Poisson, la varianza $\sigma^2 = \lambda$. El cálculo de la varianza de la distribución propuesta nos da 0.97, que se compara favorablemente con el valor 0.90 para λ, lo que añade más evidencia a lo adecuado de la distribución de Poisson como aproximación de nuestros datos.

PROBLEMAS SUPLEMENTARIOS

LA DISTRIBUCION BINOMIAL

7.35. Evaluar: (*a*) $7!$, (*b*) $10!/(6!4!)$, (*c*) $\binom{9}{5}$, (*d*) $\binom{11}{8}$ y (*e*) $\binom{6}{1}$.

7.36. Desarrollar: (*a*) $(q + p)^7$ y (*b*) $(q + p)^{10}$

7.37. Hallar la probabilidad de que al lanzar 6 veces una moneda aparezcan: (*a*) 0, (*b*) 1, (*c*) 2, (*d*) 3, (*e*) 4, (*f*) 5 y (*g*) 6 caras.

7.38. Hallar la probabilidad de: (*a*) 2 o más caras, y (*b*) menos de 4 caras, en una tirada de 6 monedas.

7.39. Si X denota el número de caras en una sola tirada de 4 monedas, hallar: (*a*) $\Pr\{X = 3\}$, (*b*) $\Pr\{X < 2\}$, (*c*) $\Pr\{X \leqslant 2\}$ y (*d*) $\Pr\{1 < X \leqslant 3\}$.

7.40. Entre 800 familias con 5 hijos, ¿cuántas cabe esperar que tengan: (*a*) 3 chicos, (*b*) 5 chicas y (*c*) 2 ó 3 chicos? Se suponen probabilidades iguales para chicos y chicas.

7.41. Hallar la probabilidad de obtener una suma de 11 puntos (*a*) una vez y (*b*) dos veces, en dos lanzamientos de un par de dados.

7.42. ¿Cuál es la probabilidad de sacar 9 exactamente una vez en 3 lanzamientos de un par de dados?

7.43. Hallar la probabilidad de acertar al azar la respuesta de al menos 6 de entre 10 cuestiones en un test verdadero-falso.

7.44. Un agente de seguros contrata 5 pólizas con personas de la misma edad y de buena salud. Según las tablas en uso, la probabilidad de que un hombre de esa edad esté vivo dentro de 30 años es $\frac{2}{3}$. Hallar la probabilidad de que dentro de 30 años vivan: (*a*) los 5, (*b*) al menos 3, (*c*) sólo 2 y (*d*) al menos uno.

7.45. Calcular: (*a*) la media, (*b*) la desviación típica, (*c*) el coeficiente momento de sesgo y (*d*) el coeficiente momento de curtosis, para una distribución binomial en la que $p = 0.7$ y $N = 60$. Interpretar los resultados.

7.46. Probar que si una distribución binomial con $N = 100$ es simétrica, su coeficiente momento de curtosis es 2.98.

7.47. Evaluar: (*a*) $\sum (X - \mu)^3 p(X)$
(*b*) $\sum (X - \mu)^4 p(X)$ para la distribución binomial.

7.48. Probar las fórmulas (1) y (2) del comienzo de este capítulo para los coeficientes momento de sesgo y curtosis.

LA DISTRIBUCION NORMAL

7.49. En un examen de estadística, la media fue 78 y la desviación típica 10.

(*a*) Determinar las puntuaciones estándar de dos estudiantes que obtuvieron 93 y 62 puntos.

(*b*) Hallar las puntuaciones de dos estudiantes cuyas puntuaciones estándar fueron -0.6 y 1.2.

7.50. Hallar: (*a*) la media y (*b*) la desviación típica en un examen en el que las notas 70 y 88 correspondieron a puntuaciones estándar de -0.6 y 1.4, respectivamente.

7.51. Hallar el área bajo la curva normal entre: (*a*) $z = -1.20$ y $z = 2.40$, (*b*) $z = 1.23$ y $z = 1.87$, (*c*) $z = -2.35$ y $z = -0.50$.

7.52. Hallar el área bajo la curva normal: (*a*) a la izquierda de $z = -1.78$, (*b*) a la izquierda de $z = 0.56$, (*c*) a la derecha de $z = -1.45$, (*d*) correspondiente a $z \geqslant 2.16$, (*e*) correspondiente a $-0.80 \leqslant z \leqslant 1.53$ y (*f*) a la izquierda de $z = -2.52$ y a la derecha de $z = 1.83$.

7.53. Si z está normalmente distribuida con media 0 y varianza 1, hallar: (*a*) $\Pr\{z \geqslant -1.64\}$, (*b*) $\Pr\{-1.96 \leqslant z \leqslant 1.96\}$, (*c*) $\Pr\{|z| \geqslant 1\}$.

7.54. Hallar el valor de z tal que: (*a*) el área a su derecha sea 0.2266, (*b*) el área a su izquierda sea 0.0314, (*c*) el área entre -0.23 y z sea 0.5722, (*d*) el área entre 1.15 y z sea 0.0730 y (*e*) el área entre $-z$ y z sea 0.9000.

7.55. Hallar z_1 si $\Pr\{z \geqslant z_1\} = 0.84$, donde z está normalmente distribuida con media 0 y varianza 1.

7.56. Hallar las ordenadas de la curva normal en: (*a*) $z = 2.25$, (*b*) $z = -0.32$ y (*c*) $z = -1.18$.

7.57. Si las alturas de 300 estudiantes están normalmente distribuidas con media 68.0 in y desviación típica 3.0 in, ¿cuántos estudiantes tienen altura: (*a*) mayor que 72 in, (*b*) menor o igual que 64 in, (*c*) entre 65 y 71 in inclusive y (*d*) de 68 in? Se supone que las alturas se han medido con precisión de 1 in.

7.58. Si los diámetros de las bolas de cojinetes están normalmente distribuidas con media 0.6140 in y desviación típica 0.0025 in, determinar el porcentaje de ellas con diámetros: (*a*) entre 0.610 y 0.618 inclusive, (*b*) mayores que 0.617 in, (*c*) menores que 0.608 in y (*d*) iguales a 0.615 in.

7.59. La nota media en un examen es 72 y la desviación típica 9. El 10% del curso recibirá grado *A*. ¿Cuál es la nota mínima para optar a él?

7.60. Si un conjunto de medidas está normalmente distribuida, ¿qué porcentaje de ellas difiere de la media: (*a*) más de 0.5 desviaciones típicas y (*b*) menos de 0.75 desviaciones típicas?

7.61. Si $\bar{X}$ es la media y s la desviación típica de un conjunto de medidas normalmente distribuidas, ¿qué porcentaje de ellas: (*a*) cae en el rango $\bar{X} \pm 2s$, (*b*) fuera del rango $\bar{X} \pm 1.2s$ y (*c*) son mayores que $\bar{X} - 1.5s$?

7.62. En el Problema 7.61, hallar a de manera que el porcentaje de casos: (*a*) en el rango $\bar{X} \pm as$ sea el 75% y (*b*) menor que $\bar{X} - as$ sea 22%.

APROXIMACION NORMAL A LA DISTRIBUCION BINOMIAL

7.63. Hallar la probabilidad de que en 200 lanzamientos de una moneda haya: (*a*) entre 80 y 120 caras inclusive, (*b*) menos de 90 caras, (*c*) menos de 85 o más de 115 caras y (*d*) 100 caras exactamente.

7.64. Hallar la probabilidad de que en un test verdadero-falso un estudiante conjeture acertadamente: (*a*) 12 o más de 20 y (*b*) 24 o más de 40 cuestiones.

7.65. El 10% de las piezas producidas en una máquina son defectuosas. Hallar la probabilidad de que en una muestra aleatoria de 400 piezas sean defectuosas: (*a*) a lo sumo 30, (*b*) entre 30 y 50, (*c*) entre 35 y 45 y (*d*) 55 o más.

7.66. Hallar la probabilidad de obtener más de 25 veces 7 en 100 tiradas de un par de dados.

LA DISTRIBUCION DE POISSON

7.67. Si el 3% de las válvulas manufacturadas por una compañía son defectuosas, hallar la probabilidad de que en una muestra de 100 válvulas: (*a*) 0, (*b*) 1, (*c*) 2, (*d*) 3, (*e*) 4 y (*f*) 5 sean defectuosas.

7.68. En el Problema 7.67, hallar la probabilidad de que sean defectuosas: (*a*) más de 5, (*b*) entre 1 y 3, (*c*) no más de 2 válvulas.

7.69. Una bolsa contiene 1 ficha roja y 7 blancas. Se saca una al azar, se anota su color y se devuelve a la bolsa, tras lo cual se remueven de nuevo. Usando: (*a*) la distribución binomial y (*b*) la aproximación de Poisson a la distribución binomial, hallar la probabilidad de que en 8 de esas extracciones salga la roja 3 veces exactamente.

7.70. De acuerdo con la National Office of Vital Statistics of the U.S. Department of Health, Education, and Welfare, el número medio de ahogados por accidente al año en EE.UU. es 3.0 por cada 100,000 habitantes. Hallar la probabilidad de que en una ciudad de 200,000 habitantes haya: (*a*) 0, (*b*) 2, (*c*) 6, (*d*) 8, (*e*) entre 4 y 8 y (*f*) menos de 3 ahogados por accidente al año.

7.71. Entre las 2 y las 4 P.M., el número medio de llamadas telefónicas por minuto que recibe una centralita es 2.5. Hallar la probabilidad de que durante un minuto concreto se produzcan: (*a*) 0, (*b*) 1, (*c*) 2, (*d*) 3, (*e*) 4 o menos y (*f*) más de 6 llamadas.

LA DISTRIBUCION MULTINOMIAL

7.72. Se lanza un dado 6 veces. Hallar la probabilidad de que: (*a*) salgan 1 uno, 2 doses y 3 treses y (*b*) que salga cada número una vez.

7.73. Una caja contiene un gran número de fichas rojas, blancas, azules y amarillas, en la proporción 4 : 3 : 2 : 1, respectivamente. Hallar la probabilidad de que en 10 extracciones salgan: (*a*) 4 rojas, 3 blancas, 2 azules y 1 amarilla y (*b*) 8 rojas y 2 amarillas.

7.74. Hallar la probilidad de no sacar ni 1, ni 2, ni 3 en cuatro tiradas de un dado.

AJUSTE DE DATOS MEDIANTE DISTRIBUCIONES TEORICAS

7.75. Ajustar una distribución binomial a los datos de la Tabla 7.9.

Tabla 7.9

X	0	1	2	3	4
f	30	62	46	10	2

7.76. Determinar, usando papel gráfico de probabilidad, si los datos del Problema 3.59 se pueden aproximar bien por una distribución normal.

7.77. Ajustar una distribución normal a los datos del Problema 3.59.

7.78. Ajustar una distribución normal a los datos del Problema 3.61.

7.79. Ajustar una distribución de Poisson a los datos del Problema 7.75 y comparar este ajuste con el obtenido mediante la distribución binomial.

7.80. La Tabla 7.10 muestra el número de muertos al año por unidad, a causa de coces de los caballos, entre 10 unidades del ejército prusiano en un periodo de 20 años (1875 a 1894). Ajustar una distribución de Poisson a esos datos.

Tabla 7.10

X	0	1	2	3	4
f	109	65	22	3	1

CAPITULO 8

Teoría elemental del muestreo

TEORIA DEL MUESTREO

La *teoría del muestreo* estudia la relación entre una población y las muestras tomadas de ella. Es de gran utilidad en muchos campos. Por ejemplo, para *estimar* magnitudes desconocidas de una población, tales como media y varianza, llamadas a menudo *parámetros* de la población o simplemente parámetros, a partir del conocimiento de esas magnitudes sobre muestras, que se llaman *estadísticos de la muestra* o simplemente *estadísticos*. Los problemas de estimación se consideran en el Capítulo 9.

La teoría del muestreo es también útil para determinar si las diferencias observadas entre dos muestras son debidas a variaciones fortuitas o si son realmente significativas. Tales cuestiones aparecen, por ejemplo, al probar un nuevo suero como tratamiento de una enfermedad o al decidir si un proceso de producción es mejor que otro. Las respuestas implican el uso de los llamados *contrastes (o tests) de hipótesis y de significación*, importantes en la *teoría de las decisiones*, considerada en el Capítulo 10.

En general, un estudio de las inferencias hechas sobre una población a partir de muestras suyas, con indicación de la precisión de tales inferencias, se llama *inferencia estadística*.

MUESTRAS ALEATORIAS Y NUMEROS ALEATORIOS

Para que las conclusiones de la teoría del muestreo y de la inferencia estadística sean válidas, las muestras deben escogerse *representativas* de la población. El análisis de los métodos de muestreo y problemas relacionados se llama el *diseño del experimento*.

Una forma de obtener una muestra representativa es mediante *muestreo aleatorio*, de acuerdo con el cual, cada miembro de la población tiene la misma probabilidad de ser incluido en la muestra. Un método para lograrlo es asignarles a cada uno un número, escribir cada número en una papeleta, y realizar en una urna un sorteo justo con ellas. Un método alternativo consiste en recurrir a una tabla de *números aleatorios* (véase Apéndice IX) especialmente construida al efecto. Véase Problema 8.6.

MUESTREO CON Y SIN REPOSICION

Si sacamos un número de una urna, podemos volverlo a poner en ella o no, antes de la siguiente extracción. En el primer caso, ese número puede salir de nuevo más veces, mientras que en el segundo sólo puede salir cada número una vez. Estos dos tipos de muestreo se llaman, respectivamente, *muestreo con reposición* y *muestreo sin reposición*.

Las poblaciones son finitas o infinitas. Si, por ejemplo, sacamos 10 bolas sucesivamente, sin reposición, de una urna que contiene 100 bolas, estamos tomando muestra en una población finita; mientras que si lanzamos 50 veces una moneda y contamos el número de caras, estamos ante una muestra de una población infinita.

Una población finita en la que se efectúa muestreo con reposición, puede considerarse infinita teóricamente, ya que se puede tomar cualquier número de muestras sin agotarla. Para muchos efectos prácticos, una población muy grande se puede considerar como si fuera infinita.

DISTRIBUCIONES DE MUESTREO

Consideremos todas las posibles muestras de tamaño N en una población dada (con o sin reposición). Para cada muestra, podemos calcular un estadístico (tal como la media o la desviación típica) que variará de muestra a muestra. De esta manera obtenemos una distribución del estadístico que se llama su *distribución de muestreo.*

Si, por ejemplo, el estadístico utilizado es la media muestral, entonces la distribución se llama la *distribución de muestreo de medias*, o *distribución de muestreo de la media*. Análogamente, podríamos tener distribución de muestreo de la desviación típica, de la varianza, de la mediana, de las proporciones, etcétera.

Para cada distribución de muestreo podemos calcular la media, la desviación típica, etc. Así pues, podremos hablar de la media y la desviación típica de la distribución de muestreo de medias, etcétera.

DISTRIBUCION DE MUESTREO DE MEDIAS

Supongamos que se toman todas las posibles muestras de tamaño N, sin reposición, de una población finita de tamaño $N_p > N$. Si denotamos la media y la desviación típica de la distribución de muestreo de medias por $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$ y las de la población por μ y σ, respectivamente, entonces

$$\mu_{\bar{X}} = \mu \quad \text{y} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} \tag{1}$$

Si la población es infinita o si el muestreo es con reposición, los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu \quad \text{y} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \tag{2}$$

Para valores grandes de N ($N \geqslant 30$), la distribución de muestreo de medias es aproximadamente normal con media $\mu_{\bar{X}}$ y desviación típica $\sigma_{\bar{X}}$, independientemente de la población (en tanto en cuanto la media poblacional y la varianza sean finitas y el tamaño de la población sea al menos doble que el de la muestra). Este resultado para una población infinita es un caso especial del *teorema del límite central* de teoría avanzada de probabilidades, que afirma que la precisión de la aproximación mejora al crecer N. Esto se indica en ocasiones diciendo que la distribución de muestreo es *asintóticamente normal*.

En caso de que la población esté normalmente distribuida, la distribución de muestreo de medias también lo está, incluso para pequeños valores de N (o sea, $N < 30$).

DISTRIBUCION DE MUESTREO DE PROPORCIONES

Supongamos que una población es infinita y que la probabilidad de ocurrencia de un suceso (su éxito) es p, mientras la probabilidad de que no ocurra es $q = 1 - p$. Por ejemplo, la población puede ser la de todas las posibles tiradas de una moneda, en la que la probabilidad del suceso «cara» es $p = \frac{1}{2}$. Consideremos todas las posibles muestras de tamaño N de tal población, y para cada una de ellas determinemos la proporción de éxitos P. En el caso de una moneda, P sería la proporción de caras en N tiradas. Obtenemos así una *distribución de muestreo de proporciones* cuya media μ_P y cuya desviación típica σ_P vienen dadas por

$$\mu_P = p \qquad \text{y} \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-p)}{N}} \tag{3}$$

que se pueden obtener de (2) poniendo $\mu = p$ y $\sigma = \sqrt{pq}$. Para valores grandes de N ($N \geq 30$), la distribución de muestreo está, muy aproximadamente, normalmente distribuida. Nótese que la población está *binomialmente distribuida*.

Las ecuaciones (3) son válidas también para una población finita en la que se hace muestreo con reposición. Para poblaciones finitas en que se haga muestreo sin reposición, las ecuaciones (3) quedan sustituidas por las ecuaciones (1) con $\mu = p$ y $\sigma = \sqrt{pq}$.

Notemos que (3) se deducen fácilmente dividiendo la media y la desviación típica (Np y $\sqrt{Npq}$) de la distribución binomial por N (véase Cap. 7).

DISTRIBUCION DE MUESTREO DE DIFERENCIAS Y SUMAS

Sean dadas dos poblaciones. Para cada muestra de tamaño N_1 de la primera, calculamos un estadístico S_1; eso da una distribución de muestreo para S_1, cuya media y desviación típica denotaremos por μ_{S1} y σ_{S1}. Del mismo modo, para cada muestra de tamaño N_2 de la segunda población, calculamos un estadístico S_2; eso nos da una distribución de muestreo para S_2, cuya media y desviación típica denotaremos por μ_{S2} y σ_{S2}. De todas las posibles combinaciones de estas muestras de las dos poblaciones podemos obtener una distribución de las diferencias, $S_1 - S_2$, que se llama *distribución de muestreo de las diferencias de los estadísticos*. La media y la desviación típica de esta distribución de muestreo, denotadas respectivamente por μ_{S1-S2} y σ_{S1-S2}, vienen dadas por

$$\mu_{S1-S2} = \mu_{S1} - \mu_{S2} \qquad \text{y} \qquad \sigma_{S1-S2} = \sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{4}$$

supuesto que las muestras escogidas no dependan en absoluto una de otra (o sea, que sean *independientes*).

Si S_1 y S_2 son las medias muestrales de ambas poblaciones, cuyas medias denotaremos por $\bar{X}_1$ y $\bar{X}_2$, respectivamente, entonces la distribución de muestreo de las diferencias de medias viene dada para poblaciones infinitas con medias y desviaciones típicas (μ_1, σ_1) y (μ_2, σ_2), respectivamente por

$$\mu_{\bar{X}1-\bar{X}2} = \mu_{\bar{X}1} - \mu_{\bar{X}2} = \mu_1 - \mu_2 \qquad \text{y} \qquad \sigma_{\bar{X}1-\bar{X}2} = \sqrt{\sigma_{\bar{X}1}^2 + \sigma_{\bar{X}2}^2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \tag{5}$$

usando las ecuaciones (2). El resultado es válido también para poblaciones finitas si el muestreo es con reposición. Análogos resultados pueden alcanzarse para poblaciones finitas en que el muestreo sea sin reposición, usando (1).

Resultados correspondientes se pueden obtener para las distribuciones de muestreo de diferencias de proporciones de dos poblaciones binomialmente distribuidas con parámetros (p_1, q_1) y (p_2, q_2), respectivamente. En este caso, S_1 y S_2 corresponden a la proporción de éxitos P_1 y P_2, y las ecuaciones (4) llevan a

$$\mu_{P1-P2} = \mu_{P1} - \mu_{P2} = p_1 - p_2 \qquad \text{y} \qquad \sigma_{P1-P2} = \sqrt{\sigma_{P1}^2 + \sigma_{P2}^2} = \sqrt{\frac{p_1q_1}{N_1} + \frac{p_2q_2}{N_2}} \tag{6}$$

Si N_1 y N_2 son grandes ($N_1, N_2 \geqslant 30$), la distribución de muestreo de diferencias de medias o proporciones están casi normalmente distribuidas.

A veces es útil hablar de la *distribución de muestreo de la suma de estadísticos*. La media y la desviación típica de tal distribución son

$$\mu_{S1+S2} = \mu_{S1} + \mu_{S2} \qquad \text{y} \qquad \sigma_{S1+S2} = \sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{7}$$

supuesto que las muestras sean independientes.

ERRORES TIPICOS

La desviación típica de una distribución de muestreo de un estadístico se suele llamar su *error típico*. La Tabla 8.1 presenta errores típicos de distribución de muestreo para varios estadísticos bajo las condiciones de muestreo aleatorio de una población infinita (o muy grande) o de muestreo con reposición de una finita. También recoge observaciones particulares que garantizan la validez de estos resultados y otras notas pertinentes.

Las cantidades μ, σ, p, μ_r y $\bar{X}$, s, P, m_r denotan, respectivamente, las medias de la población y de la muestra, las desviaciones típicas, proporciones y r-ésimos momentos respecto de la media.

Hay que hacer notar que si el tamaño de la muestra es lo bastante grande, las distribuciones de muestreo son normales o casi normales. Por ello, los métodos se conocen como *métodos de grandes muestras*. Cuando $N < 30$, las muestras se llaman pequeñas. La teoría de *pequeñas muestras* o *teoría exacta del muestreo*, como se le llama a veces, se trata en el Capitulo 11.

Cuando los parámetros de la población, tales como σ, p o μ_r son desconocidos, pueden ser estimados con precisión por sus correspondientes estadísticos muestrales, a saber, s (o sea $\hat{s} = \sqrt{N/(N-1)}s$), P y m_r si las muestras son suficientemente grandes.

Tabla 8.1. Errores típicos para algunas distribuciones de muestreo

Distribución de muestreo	Error típico	Observaciones
Medias	$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{N}}$	Esto es cierto para muestras grandes y pequeñas. La distribución muestral de medias es casi normal para $N \geqslant 30$, incluso cuando la población no es normal. $\mu_{\bar{X}} = \mu$, la media de la población, en todos los casos.
Proporciones	$\sigma_P = \sqrt{\dfrac{p(1-p)}{N}} = \sqrt{\dfrac{pq}{N}}$	La nota precedente para las medias se aplica aquí también. $\mu_P = p$ en todos los casos.
Desviaciones típicas	(1) $\sigma_s = \dfrac{\sigma}{\sqrt{2N}}$ (2) $\sigma_s = \sqrt{\dfrac{\mu_4 - \mu_2^2}{4N\mu_2}}$	Para $N \geqslant 100$, la distribución muestral de s es casi normal. σ_s viene dada por (1) sólo si la población es normal (o aproximadamente normal). Si la población no es normal, se puede usar (2). Nótese que (2) se reduce a (1) cuando $\mu_2 = \sigma^2$ y $\mu_4 = 3\sigma^4$, lo cual es cierto para poblaciones normales. Para $N \geqslant 100$, $\mu_s = \sigma$ muy aproximadamente.
Medianas	$\sigma_{\text{med}} = \sigma\sqrt{\dfrac{\pi}{2N}} = \dfrac{1.2533\sigma}{\sqrt{N}}$	Para $N \geqslant 30$, la distribución de muestreo de la mediana es muy aproximadamente normal. El resultado dado es válido sólo si la población es normal (o casi normal). $\mu_{\text{med}} = \mu$
Primer y tercer cuartiles	$\sigma_{Q1} = \sigma_{Q3} = \dfrac{1.3626\sigma}{\sqrt{N}}$	Los comentarios hechos para las medianas se aplican aquí también. μ_{Q1} y μ_{Q3} son casi iguales al primer y tercer cuartiles de la población. Nótese que $\sigma_{Q2} = \sigma_{\text{med}}$
Deciles	$\sigma_{D1} = \sigma_{D9} = \dfrac{1.7094\sigma}{\sqrt{N}}$ $\sigma_{D2} = \sigma_{D8} = \dfrac{1.4288\sigma}{\sqrt{N}}$ $\sigma_{D3} = \sigma_{D7} = \dfrac{1.3180\sigma}{\sqrt{N}}$ $\sigma_{D4} = \sigma_{D6} = \dfrac{1.2680\sigma}{\sqrt{N}}$	De nuevo son aplicables aquí las observaciones hechas en el caso de las medianas. $\mu_{D1}, \mu_{D2}, \ldots$ son casi iguales al primer, segundo, ... deciles de la población. Nótese que $\sigma_{D5} = \sigma_{\text{med}}$

Tabla 8.1. (*Continuación*)

Distribución de muestreo	Error típico	Observaciones
Rangos semi-intercuartiles	$\sigma_Q = \dfrac{0.7867\sigma}{\sqrt{N}}$	Las observaciones hechas acerca de las medianas se aplican de nuevo aquí. μ_Q es casi igual al rango semi-intercuartil de la población.
Varianzas	(1) $\sigma_{S^2} = \sigma^2\sqrt{\dfrac{2}{N}}$ (2) $\sigma_{S^2} = \sqrt{\dfrac{\mu_4 - \mu_2^2}{N}}$	Las observaciones hechas sobre la desviación típica son aplicables también aquí. Hagamos notar que (2) da (1) en el caso de poblaciones normales. $\mu_{S^2} = \sigma^2(N-1)/N$, que es casi igual a σ^2 para N grandes.
Coeficientes de varianza	$\sigma_V = \dfrac{v}{\sqrt{2N}}\sqrt{1+2v^2}$	Aquí $v = \sigma/\mu$ es el coeficiente de variación de la población. El resultado dado es válido para poblaciones normales (o casi normales) y $N \geqslant 100$.

PROBLEMAS RESUELTOS

DISTRIBUCION DE MUESTREO DE MEDIAS

8.1. Una población consta de los números 2, 3, 6, 8 y 11. Consideremos todas las posibles muestras de tamaño 2 que pueden tomarse con reposición de esa población. Hallar (*a*) la media de la población, (*b*) la desviación típica de la población, (*c*) la media de la distribución de muestreo de medias y (*d*) la desviación típica de la distribución de muestreo de medias (o sea, el error típico de medias).

Solución

(*a*)
$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6.0$$

(*b*)
$$\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5} = \frac{16+9+0+4+25}{5} = 10.8$$

y $\sigma = 3.29$.

(*c*) Hay 5(5) = 25 muestras de tamaño 2 que se pueden tomar, con reposición de la población (porque cualquiera de los 5 números de la primera extracción puede asociarse con uno cualquiera de la segunda). Y son

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

Las correspondientes medias muestrales son

$$\begin{array}{ccccc} 2.0 & 2.5 & 4.0 & 5.0 & 6.5 \\ 2.5 & 3.0 & 4.5 & 5.5 & 7.0 \\ 4.0 & 4.5 & 6.0 & 7.0 & 8.5 \\ 5.0 & 5.5 & 7.0 & 8.0 & 9.5 \\ 6.5 & 7.0 & 8.5 & 9.5 & 11.0 \end{array} \qquad (8)$$

y la media de la distribución de muestreo de medias es

$$\mu_{\bar{X}} = \frac{\text{suma de todas las medias muestrales en (8)}}{25} = \frac{150}{25} = 6.0$$

ilustrando el hecho de que $\mu_{\bar{X}} = \mu$.

(*d*) La varianza $\sigma_{\bar{X}}^2$ de la distribución de muestreo de medias se obtiene restando la media 6 de cada número en (8), elevando al cuadrado el resultado, sumando los 25 números así obtenidos y dividiendo por 25. El resultado final es $\sigma_{\bar{X}}^2 = 135/25 = 5.40$, y por tanto $\sigma_{\bar{X}} = \sqrt{5.40} = 2.32$. Ello ilustra el que para poblaciones finitas y muestreo con reposición (o para poblaciones infinitas), $\sigma_{\bar{X}}^2 = \sigma^2/N$ porque el lado derecho es $10.8/2 = 5.40$, que coincide con el anterior valor.

8.2. Resolver el Problema 8.1 para el caso de muestreo sin reposición.

Solución

Como en las partes (*a*) y (*b*) del Problema 8.1, $\mu = 6$ y $\sigma = 3.29$.

(*c*) Hay $\binom{5}{2}$ muestras de tamaño 2 que se pueden elegir sin reposición (eso significa que sacamos un número y luego otro distinto del anterior) de la población: (2, 3), (2, 6), (2, 8), (2, 11), (3, 6), (3, 8), (3, 11), (6, 8), (6, 11) y (8, 11). La selección (2, 3), por ejemplo, se considera la misma que la (3, 2).

Las correspondientes medias de la muestra son 2.5, 4.0, 5.0, 6.5, 4.5, 5.5, 7.0, 7.0, 8.5 y 9.5, y la media de la distribución de muestreo de medias es

$$\mu_{\bar{X}} = \frac{2.5 + 4.0 + 5.0 + 6.5 + 4.5 + 5.5 + 7.0 + 7.0 + 8.5 + 9.5}{10} = 6.0$$

ilustrando el hecho de que $\mu_{\bar{X}} = \mu$.

(*d*) La varianza de la distribución de muestreo de medias es

$$\sigma_{\bar{X}}^2 = \frac{(2.5 - 6.0)^2 + (4.0 - 6.0)^2 + (5.0 - 6.0)^2 + \cdots + (9.5 - 6.0)^2}{10} = 4.05$$

y $\sigma_{\bar{X}} = 2.01$. Esto ilustra que

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N}\left(\frac{N_p - N}{N_p - 1}\right)$$

ya que el lado derecho es igual a

$$\frac{10.8}{2}\left(\frac{5 - 2}{5 - 1}\right) = 4.05$$

como se había obtenido antes.

8.3. Las alturas de 3000 estudiantes varones de una Universidad están normalmente distribuidas con media 68.0 in y desviación típica 3.0 in. Si se toman 80 muestras de 25 estudiantes cada una, ¿cuáles serán la media y la desviación típica esperadas de la resultante distribución de muestreo de medias, si el muestreo se hizo (*a*) y con (*b*) sin reposición?

Solución

El número de muestras de tamaño 25 que podrían elegirse de un grupo de 3000 estudiantes con y sin reposición son $(3000)^{25}$ y $\binom{3000}{25}$, que son mucho mayores que 80. Por tanto no obtenemos una verdadera distribución de muestreo de medias, sino sólo una distribución de muestreo *experimental*. No obstante, como el número de muestras es grande, debiera haber buen acuerdo entre ambas distribuciones de muestreo. Así que la media y la desviación típica esperadas deben estar próximas a las de la distribución teórica. Por tanto, tenemos

(*a*) $$\mu_{\bar{X}} = \mu = 68.0 \text{ in} \quad \text{y} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{25}} = 0.6 \text{ in}$$

(*b*) $$\mu_{\bar{X}} = 68.0 \text{ in} \quad \text{y} \quad \delta_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{3}{\sqrt{25}}\sqrt{\frac{3000 - 25}{3000 - 1}}$$

que es sólo muy ligeramente menor que 0.6 in y puede ser considerada, a todos los efectos prácticos, la misma que en muestreo con reposición.

Así pues, esperaríamos que la distribución de muestreo experimental de medias esté casi normalmente distribuida con media 68.0 in y desviación típica 0.6 in.

8.4. ¿En cuántas muestras del Problema 8.3 esperaríamos encontrar una media (*a*) entre 66.8 y 68.3 in y (*b*) menor que 66.4 in?

Solución

La media $\bar{X}$ de una muestra en unidades estándar viene dada aquí por

$$z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 68.0}{0.6}$$

(*a*) $$66.8 \text{ en unidades estándar} = \frac{66.8 - 68.0}{0.6} = -2.0$$

$$68.3 \text{ en unidades estándar} = \frac{68.3 - 68.0}{0.6} = 0.5$$

Como muestra la Figura 8.1(*a*),

Proporción de muestras con medias entre 66.8 y 68.3 in =

= (área bajo la curva normal entre $z = -2.0$ y $z = 0.5$)

= (área entre $z = -2$ y $z = 0$) + (área entre $z = 0$ y $z = 0.5$)

= 0.4772 + 0.1915 = 0.6687

Así pues, el número esperado de muestras es (80)(0.6687) = 53.496, o 53.

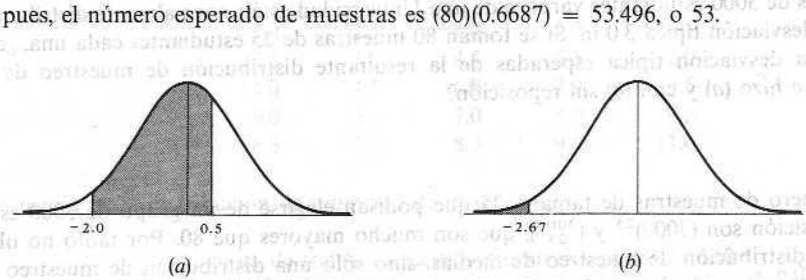


Figura 8.1.

(*b*) $$66.4 \text{ in en unidades estándar} = \frac{66.4 - 68.0}{0.6} = -2.67$$

Como muestra la Figura 8.1(*b*),

Proporción de muestras con media menor que 66.4 in =

= (área bajo la curva normal a la izquierda de $z = -2.67$)

= (área a la izquierda de $z = 0$) − (*área entre* $z = -2.67$ y $z = 0$)

= 0.5 − 0.4962 = 0.0038

Luego el número esperado de muestras es (80)(0.0038) = 0.304, o cero.

8.5. 500 bolas de cojinete tienen un peso medio de 5.02 gramos (g) y una desviación típica de 0.30 g. Hallar la probabilidad de que una muestra al azar de 100 bolas de ese conjunto tengan un peso total (*a*) entre 496 y 500 g y (*b*) más de 510 g.

Solución

Para la distribución de muestreo de medias, $\mu_{\bar{X}} = \mu = 5.02$ g, y

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{0.30}{\sqrt{100}}\sqrt{\frac{500 - 100}{500 - 1}} = 0.027 \text{ g}$$

(*a*) El peso total estaría entre 496 y 500 g si el peso medio de las 100 bolas está entre 4.96 y 5.00 g.

$$4.96 \text{ en unidades estándar} = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$5.00 \text{ en unidades estándar} = \frac{5.00 - 5.02}{0.027} = -0.74$$

Como muestra la Figura 8.2(*a*),

Probabilidad pedida = (área entre $z = -2.22$ y $z = -0.74$)

= (área entre $z = -2.22$ y $z = 0$) − (área entre $z = -0.74$ y $z = 0$)

= 0.4868 − 0.2704 = 0.2164

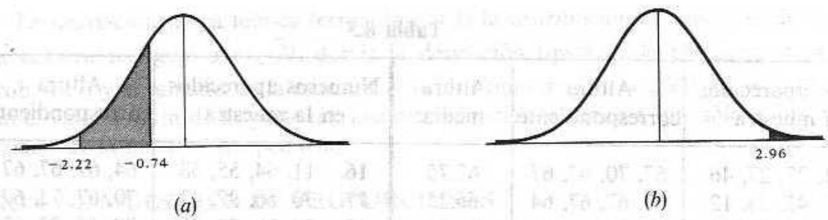


Figura 8.2.

(*b*) El peso total excecerá de 510 g si el peso medio de las 100 bolas excede de 5.10 g.

$$5.10 \text{ en unidades estándar} = \frac{5.10 - 5.02}{0.027} = 2.96$$

Como enseña la Figura 8.2(*b*),

$$\begin{aligned} \text{Probabilidad pedida} &= (\text{área a la derecha de } z = 2.96) \\ &= (\text{área a la derecha de } z = 0) - (\text{área entre } z = 0 \text{ y } z = 2.96) \\ &= 0.5 - 0.4985 = 0.0015 \end{aligned}$$

Luego sólo hay 3 oportunidades en 2000 de tomar una muestra de 100 bolas que supere los 510 g.

8.6. (*a*) Indicar cómo se seleccionarían al azar 30 muestras de 4 estudiantes cada una (con reposición) de la Tabla 2.1, usando números aleatorios.

(*b*) Hallar la media y la desviación típica de la distribución de muestreo de medias en la parte (*a*).

(*c*) Comparar los resultados de (*b*) con los valores teóricos, explicando cualquier discrepancia.

Solución

(*a*) Usamos dos dígitos para numerar a los 100 estudiantes: 00, 01, 02, ..., 99 (véase Tabla 8.2). Así pues, los 5 estudiantes con pesos 60-62 in están numerados 00-04, los 18 con pesos 63-65 con 05-22, etc. Cada número de estudiante es un *número de muestreo*.

Ahora sacamos números de muestreo de la tabla de números aleatorios (Apéndice IX). En la primera línea vemos 51, 77, 27, 46, 40, etc., que tomamos como números aleatorios de muestreo, cada uno de los cuales da la altura de un estudiante particular. Así, 51 corresponde a un estudiante de 66-68 in, que tomamos como 67 in (la marca de clase). Análogamente, 77, 27 y 46 dan alturas 70, 67 y 67 respectivamente.

Por este proceso se obtiene la Tabla 8.3, que recoge los números de muestreo extraídos, las alturas correspondientes y la altura media para cada una de las 30 muestras. Debemos mencionar que aunque hemos entrado en la tabla de números aleatorios por su primera fila, se podía haber entrado de *cualquier* otra forma.

Tabla 8.2

Altura (in)	Frecuencia	Número de muestreo
60-62	5	00-04
63-65	18	05-22
66-68	42	23-64
69-71	27	65-91
72-74	8	92-99

Tabla 8.3

Números aparecidos en la muestra	Altura correspondiente	Altura media	Números aparecidos en la muestra	Altura correspondiente	Altura media
1. 51, 77, 27, 46	67, 70, 67, 67	67.75	**16.** 11, 64, 55, 58	64, 67, 67, 67	66.25
2. 40, 42, 33, 12	67, 67, 67, 64	66.25	**17.** 70, 56, 97, 43	70, 67, 73, 67	69.25
3. 90, 44, 46, 62	70, 67, 67, 67	67.75	**18.** 74, 28, 93, 50	70, 67, 73, 67	69.25
4. 16, 28, 98, 93	64, 67, 73, 73	69.25	**19.** 79, 42, 71, 30	70, 67, 70, 67	68.50
5. 58, 20, 41, 86	67, 64, 67, 70	67.00	**20.** 58, 60, 21, 33	67, 67, 64, 67	66.25
6. 19, 64, 08, 70	64, 67, 64, 70	66.25	**21.** 75, 79, 74, 54	70, 70, 70, 67	69.25
7. 56, 24, 03, 32	67, 67, 61, 67	65.50	**22.** 06, 31, 04, 18	64, 67, 61, 64	64.00
8. 34, 91, 83, 58	67, 70, 70, 67	68.50	**23.** 67, 07, 12, 97	70, 64, 64, 73	67.75
9. 70, 65, 68, 21	70, 70, 70, 64	68.50	**24.** 31, 71, 69, 88	67, 70, 70. 70	69.25
10. 96, 02, 13, 87	73, 61, 64, 70	67.00	**25.** 11, 64, 21, 87	64, 67, 64, 70	66.25
11. 76, 10, 51, 08	70, 64, 67, 64	66.25	**26.** 03, 58, 57, 93	61, 67, 67, 73	67.00
12. 63, 97, 45, 39	67, 73, 67, 67	68.50	**27.** 53, 81, 93, 88	67, 70, 73, 70	70.00
13. 05, 81, 45, 93	64, 70, 67, 73	68.50	**28.** 23, 22, 96, 79	67, 64, 73, 70	68.50
14. 96, 01, 73, 52	73, 61, 70, 67	67.75	**29.** 98, 56, 59, 36	73, 67, 67, 67	68.50
15. 07, 82, 54, 24	64, 70, 67, 67	67.00	**30.** 08, 15, 08, 84	64, 64, 64, 70	65.50

(*b*) La Tabla 8.4 da la distribución de frecuencias de las alturas medias de las muestras en la parte (*a*). Eso es una *distribución de muestreo de medias*. La media y la desviación típica se obtienen como de costumbre por métodos de compilación (Caps. 3 y 4):

$$\text{Media} = A + c\bar{u} = A + \frac{c\sum fu}{N} = 67.00 + \frac{(0.75)(23)}{30} = 67.58 \text{ in}$$

$$\text{Desviación típica} = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = 0.75\sqrt{\frac{123}{30} - \left(\frac{23}{30}\right)^2} = 1.41 \text{ in}$$

Tabla 8.4

Media muestral	Recuento	f	u	fu	fu^2
64.00	/	1	−4	−4	16
64.75		0	−3	0	0
65.50	//	2	−2	−4	8
66.25	~~////~~ /	6	−1	−6	6
$A \rightarrow$ 67.00	////	4	0	0	0
67.75	////	4	1	4	4
68.50	~~////~~ //	7	2	14	28
69.25	~~////~~	5	3	15	45
70.00	/	1	4	4	16
		$\sum f = N = 30$		$\sum fu = 23$	$\sum fu^2 = 123$

(*c*) La media teórica de la distribución de muestreo de medias, dada por $\mu_{\bar{X}}$, debiera ser igual a la media μ de la población, que es 67.45 in (véase Prob. 3.22), de acuerdo con el valor 67.58 in de la parte (*b*).

La desviación típica teórica (error típico) de la distribución de muestreo de medias, dada por $\sigma_{\bar{X}}$, debiera ser igual a $\sigma/\sqrt{N}$, donde la desviación típica de la población $\sigma = 2.92$ in (véase Prob. 4.17) y el tamaño de la muestra $N = 4$. Como $\sigma/\sqrt{N} = 2.92/\sqrt{4} = 1.46$ in, hay acuerdo con el valor 1.41 in de la parte (*b*). Las discrepancias se deben a que sólo había 30 muestras y el tamaño de la muestra era pequeño.

DISTRIBUCION DE MUESTREO DE PROPORCIONES

8.7. Hallar la probabilidad de que en 120 lanzamientos de una moneda (*a*) entre el 40% y 60% sean caras y (*b*) $\frac{5}{8}$ o más sean caras.

Solución

Primer método

Consideremos los 120 lanzamientos como una muestra de la población infinita de todos los posibles lanzamientos de la moneda. En esa población, la probabilidad de cara es $p = \frac{1}{2}$ y la de cruz es $q = 1 - p = \frac{1}{2}$.

(*a*) Se pide la probabilidad de que el número de caras en 120 lanzamientos esté entre (40% de 120) = 48 y (60% de 120) = 72. Procederemos como en el Capítulo 7, usando la aproximación normal a la distribución binomial. Puesto que el número de caras es una variable discreta, nos preguntamos por la probabilidad de que el número de caras esté entre 47.5 y 72.5.

$$\mu = \text{números esperados de caras} = Np = 120(\tfrac{1}{2}) = 60 \qquad \text{y} \qquad \sigma = \sqrt{Npq} = \sqrt{(120)(\tfrac{1}{2})(\tfrac{1}{2})} = 5.48$$

$$47.5 \text{ en unidades estándar} = \frac{47.5 - 60}{5.48} = -2.28$$

$$72.5 \text{ en unidades estándar} = \frac{72.5 - 60}{5.48} = 2.28$$

Como indica la Figura 8.3,

$$\begin{aligned}\text{Probabilidad pedida} &= (\text{área bajo la curva normal entre } z = -2.28 \text{ y } z = 2.28)\\ &= 2(\text{área entre } z = 0 \text{ y } z = 2.28)\\ &= 2(0.4887) = 0.9774\end{aligned}$$

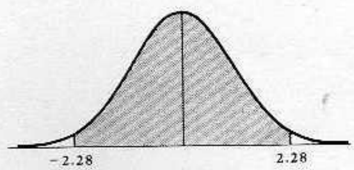


Figura 8.3.

Segundo método

$$\mu_P = p = \tfrac{1}{2} = 0.50 \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{120}} = 0.0456$$

$$40\% \text{ en unidades estándar} = \frac{0.40 - 0.50}{0.0456} = -2.19$$

$$60\% \text{ en unidades estándar} = \frac{0.60 - 0.50}{0.0456} = 2.19$$

$$\text{Probabilidad pedida} = (\text{área bajo la curva normal entre } z = -21.9 \quad \text{y} \quad z = 2.19)$$
$$= 2(0.4857) = 0.9714$$

Aunque este resultado es correcto en dos cifras significativas, no coincide exactamente ya que no hemos hecho uso de que la proporción es en realidad una variable discreta. Para tenerlo en cuenta, restamos $1/2N = 1/2(120)$ de 0.40 y sumamos $1/2N = 1/2(120)$ a 0.60; así pues, como $1/240 = 0.00417$, las proporciones pedidas en unidades estándar son

$$\frac{0.40 - 0.00417 - 0.50}{0.0456} = -2.28 \quad \text{y} \quad \frac{0.60 + 0.00417 - 0.50}{0.0456} = 2.28$$

logrando ya el acuerdo con el primer método.

Nótese que $(0.40 - 0.00417)$ y $(0.60 + 0.00417)$ corresponde a las proporciones 47.5/120 y 72.5/120 en el primer método.

(*b*) Usando el segundo método de la parte (*a*), vemos que como $\frac{5}{8} = 0.6250$,

$$(0.6250 - 0.00417) \text{ en unidades estándar} = \frac{0.6250 - 0.00417 - 0.50}{0.0456} = 2.65$$

$$\text{Probabilidad requerida} = (\text{área bajo la curva normal a la derecha de } z = 2.65)$$
$$= (\text{área a la derecha de } z = 0) - (\text{área entre } z = 0 \text{ y } z = 2.65)$$
$$= 0.5 - 0.4960 = 0.0040$$

8.8. Cada persona de un grupo de 500 lanza una moneda 120 veces. ¿Cuántas personas se espera que (*a*) saquen entre 40% y 60% de caras y (*b*) $\frac{5}{8}$ de sus lanzamientos o más de caras?

Solución

Este problema está muy relacionado con el Problema 8.7. Aquí consideramos 500 muestras de tamaño 120 cada una de una población infinita (todos los posibles lanzamientos de la moneda).

(*a*) La parte (*a*) del Problema 8.7 establece que de todas las posibles muestras, consistentes cada una en 120 lanzamientos, podemos esperar un 97.74% con un porcentaje de caras entre 40% y 60%. Luego en 500 muestras cabe esperar unas (97.74% de 500) = 489 muestras con esa propiedad. Por tanto, unas 489 personas verán aparecer entre un 40% y un 60% de caras.

Es interesante notar que $500 - 489 = 11$ personas se espera que den porcentajes de caras que no caen entre 40% y 60%. Tales personas pueden razonablemente concluir que sus monedas estaban trucadas, aunque fueran buenas. Este tipo de error es un riesgo omnipresente al tratar con probabilidades.

(*b*) Argumentando como en (*a*), deducimos que unas $(500)(0.0040) = 2$ personas verían salir $\frac{5}{8}$ o más de sus lanzamientos con cara.

8.9. Se ha encontrado que el 2% de las piezas fabricadas en una cierta máquina son defectuosas. ¿Cuál es la probabilidad de que en un envío de 400 piezas (*a*) el 3% o más y (*b*) el 2% o menos, sean defectuosas?

Solución

$$\mu_P = p = 0.02 \quad \text{y} \quad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.02)(0.98)}{400}} = \frac{0.14}{20} = 0.007$$

(*a*) *Primer método*

Usando la corrección por variables discretas, $1/2N = 1/800 = 0.00125$, tenemos

$$(0.03 - 0.00125) \text{ en unidades estándar} = \frac{0.03 - 0.00125 - 0.02}{0.007} = 1.25$$

Probabilidad requerida = (área bajo la curva normal a la derecha de $z = 1.25$) = 0.1056

Sin corrección se hubiera llegado al valor 0.0764.

Otro método

(3% de 400) = 12 piezas defectuosas. Sobre base continua, 12 o más significa 11.5 o más.

$$\bar{X} = (2\% \text{ de } 400) = 8 \quad \text{y} \quad \sigma = \sqrt{Npq} = \sqrt{(400)(0.02)(0.98)} = 2.8$$

Entonces, 11.5 en unidades estándar = $(11.5 - 8)/2.8 = 1.25$, y como antes la probabilidad pedida es 0.1056.

(*b*)

$$(0.02 + 0.00125) \text{ en unidades estándar} = \frac{0.02 + 0.00125 - 0.02}{0.007} = 0.18$$

Probabilidad requerida = (área bajo la curva normal a la izquierda de $z = 0.18$)

$= 0.5000 + 0.0714 = 0.5714$

Sin corrección se obtendría 0.5000. El segundo método de la parte (*a*) también es aplicable.

8.10. En unas elecciones uno de los candidatos obtuvo el 46% de los votos. Hallar la probabilidad de qu en un muestreo de (*a*) 200 y (*b*) 1000 votantes elegidos al azar saliera mayoría a su favor.

Solución

(*a*)

$$\mu_P = p = 0.46 \quad \text{y} \quad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.46)(0.54)}{200}} = 0.0352$$

Como $1/2N = 1/400 = 0.0025$, la muestra daría una mayoría si la proporción en favor de tal candidato fuese $0.50 + 0.0025 = 0.5025$ o más. (Esta proporción se puede obtener también recordando que 101 o más es mayoría, pero como variable continua eso es 100.5, y por tanto la proporción es $100.5/200 = 0.5025$.)

$$0.5025 \text{ en unidades estándar} = \frac{0.5025 - 0.46}{0.0352} = 1.21$$

Probabilidad requerida = (área bajo la curva normal a la derecha de $z = 1.21$)

$= 0.5000 - 0.3869 = 0.1131$

(*b*) $$\mu_P = p = 0.46 \quad \text{y} \quad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.46)(0.54)}{1000}} = 0.0158$$

$$0.5025 \text{ en unidades estándar} = \frac{0.5025 - 0.46}{0.0158} = 2.69$$

$$\text{Probabilidad requerida} = (\text{área bajo la curva normal a la derecha de } z = 2.69)$$
$$= 0.5000 - 0.4964 = 0.0036$$

DISTRIBUCION DE MUESTREO DE DIFERENCIAS Y SUMAS

8.11. Sea U_1 una variable que recorre los elementos de la población 3, 7, 8 y U_2 una variable que recorre los de la población 2, 4. Calcular (*a*) μ_{U1}, (*b*) μ_{U2}, (*c*) μ_{U1-U2}, (*d*) σ_{U1}, (*e*) σ_{U2} y (*f*) σ_{U1-U2}.

Solución

(*a*) μ_{U1} = media de la población $U_1 = \frac{1}{3}(3 + 7 + 8) = 6$.
(*b*) μ_{U2} = media de la población $U_2 = \frac{1}{2}(2 + 4) = 3$.
(*c*) La población consistente de las diferencias de cualquier elemento de U_1 y cualquiera de U_2, es

$$\begin{matrix} 3-2 & 7-2 & 8-2 \\ 3-4 & 7-4 & 8-4 \end{matrix} \quad \text{o} \quad \begin{matrix} 1 & 5 & 6 \\ -1 & 3 & 4 \end{matrix}$$

Luego $$\mu_{U1-U2} = \text{media de } (U_1 - U_2) = \frac{1 + 5 + 6 + (-1) + 3 + 4}{6} = 3$$

Eso ilustra el resultado general $\mu_{U1-U2} = \mu_{U1} - \mu_{U2}$, como se ve de las partes (*a*) y (*b*).

(*d*) $$\sigma^2_{U1} = \text{varianza de la población } U_1 = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

es decir $\sigma_{U1} = \sqrt{\frac{14}{3}}$

(*e*) $$\sigma^2_{U2} = \text{varianza de la población } U_2 = \frac{(2-3)^2 + (4-3)^2}{2} = 1 \quad \text{O sea.} \quad \sigma_{U2} = 1$$

(*f*) $$\sigma^2_{U1-U2} = \text{varianza de la población } (U_1 - U_2)$$
$$= \frac{(1-3)^2 + (5-3)^2 + (6-3)^2 + (-1-3) + (3-3)^2 + (4-3)^2}{6} = \frac{17}{3}$$

es decir

$$\sigma_{U1-U2} = \sqrt{\frac{17}{3}}$$

Esto ilustra el resultado general, $\sigma_{U1-U2} = \sqrt{\sigma^2_{U1} + \sigma^2_{U2}}$, para muestras independientes, como se ve de las partes (*d*) y (*e*).

5.12. Las lámparas de un fabricante A tienen vida media de 1400 horas (h) con desviación típica de 200 h, mientras que las de otro fabricante B tienen vida media de 1200 h con desviación típica de 100 h. Si se toma una muestra de 125 lámparas de cada clase, ¿cuál es la probabilidad de que las de A tengan una vida media que sea al menos (*a*) de 160 h y (*b*) 250 h, más que las de B?

Solución

Denotemos por $\bar{X}_A$ y $\bar{X}_B$ las vidas medias de las muestras A y B, respectivamente. Entonces

$$\mu_{\bar{X}_A-\bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 1400 - 1200 = 200 \text{ h}$$

y
$$\sigma_{\bar{X}_A-\bar{X}_B} = \sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}} = \sqrt{\frac{(100)^2}{125} + \frac{(200)^2}{125}} = 20 \text{ h}$$

La variable tipificada para la diferencia en medias es

$$z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{\bar{X}_A-\bar{X}_B})}{\sigma_{\bar{X}_A-\bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 200}{20}$$

y está casi normalmente distribuida.

(*a*) La diferencia 160 h en unidades estándar es $(160 - 200)/20 = -2$. Luego

$$\begin{aligned}\text{Probabilidad requerida} &= (\text{área bajo la curva normal a la derecha de } z = -2)\\ &= 0.5000 + 0.4772 = 0.9772\end{aligned}$$

(*b*) La diferencia 250 h en unidades estándar es $(250 - 200)/20 = 2.50$. Por tanto

$$\begin{aligned}\text{Probabilidad requerida} &= (\text{área bajo la curva normal a la derecha de } z = 2.50)\\ &= 0.5000 - 0.4938 = 0.0062\end{aligned}$$

5.13. Las bolas de rodamientos de cierto fabricante pesan 0.50 g de media, con desviación típica de 0.02 g. ¿Cuál es la probabilidad de que dos lotes de 1000 bolas cada uno difieran en peso en más de 2 g?

Solución

Sean $\bar{X}_1$ y $\bar{X}_2$ los pesos medios de las bolas de ambos lotes. Entonces

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = 0.50 - 0.50 = 0$$

y
$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{(0.02)^2}{1000} + \frac{(0.02)^2}{1000}} = 0.000895$$

La variable tipificada para la diferencia en medias es

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{0.000895}$$

y es casi normalmente distribuida.

Una diferencia de 2 g en los lotes equivale a una diferencia de $2/1000 = 0.002$ g en las medias.

Esto puede suceder si $\bar{X}_1 - \bar{X}_2 \geqslant 0.002$ o $\bar{X}_1 - \bar{X}_2 \leqslant -0.002$; esto es

$$z \geqslant \frac{0.002 - 0}{0.000895} = 2.23 \quad \text{o} \quad z \leqslant \frac{-0.002 - 0}{0.00895} = -2.23$$

Entonces $\Pr\{z \geqslant 2.23 \text{ o } z \leqslant -2.23\} = \Pr\{z \geqslant 2.23\} + \Pr\{z \leqslant -2.23\} = 2(0.5000 - 0.4871) = 0.0258.$

8.14. *A* y *B* juegan a «cara o cruz» tirando 50 monedas. *A* ganará el juego si consigue 5 o más caras que *B*; de lo contrario, es *B* quien gana. Determinar las apuestas en contra de que *A* gane un juego.

Solución

Sean P_A y P_B las proporciones de caras logradas por *A* y *B*. Si suponemos que las monedas son buenas, como siempre, la probabilidad de cara es $p = \frac{1}{2}$. Así que

$$\mu_{P_A - P_B} = \mu_{P_A} - \mu_{P_B} = 0$$

y

$$\sigma_{P_A - P_B} = \sqrt{\sigma^2_{P_A} + \sigma^2_{P_B}} = \sqrt{\frac{pq}{N_A} + \frac{pq}{N_B}} = \sqrt{\frac{2(\frac{1}{2})(\frac{1}{2})}{50}} = 0.10$$

La variable tipificada para la diferencia en proporciones es $z = (P_A - P_B - 0)/0.10$.

Sobre una base continua, 5 o más quiere decir 4.5 o más, de modo que la diferencia en proporciones debería ser $4.5/50 = 0.09$ o más; esto es, z mayor o igual que $(0.09 - 0)/0.10 = 0.9$ (o sea $z \geqslant 0.9$). La probabilidad de esto es el área bajo la curva normal a la derecha de $z = 0.9$, que es $(0.5000 - 0.3159) = 0.1841$.

Por tanto, las apuestas contra *A* están $(1 - 0.1841):0.1841 = 0.8159:0.1841$, o sea 4.43 a 1.

8.15. Dos distancias se han medido como 27.3 cm y 15.6 cm con desviación típica (error típico) de 0.16 cm y 0.08 cm, respectivamente. Hallar la media y la desviación típica de (*a*) la suma y (*b*) la diferencia, de esas distancias.

Solución

Si denotamos las distancias por D_1 y D_2, entonces:

(*a*)
$$\mu_{D1+D2} = \mu_{D1} + \mu_{D2} = 27.3 + 15.6 = 42.9 \text{ cm}$$
$$\sigma_{D1+D2} = \sqrt{\sigma^2_{D1} + \sigma^2_{D2}} = \sqrt{(0.16)^2 + (0.08)^2} = 0.18 \text{ cm}$$

(*b*)
$$\mu_{D1-D2} = \mu_{D1} - \mu_{D2} = 27.3 - 15.6 = 11.7 \text{ cm}$$
$$\sigma_{D1-D2} = \sqrt{\sigma^2_{D1} + \sigma^2_{D2}} = \sqrt{(0.16)^2 + (0.08)^2} = 0.18 \text{ cm}$$

8.16. Un cierto tipo de lámparas tiene una vida media de 1500 h y una desviación típica de 150 h. Se conectan tres de ellas de manera que en cuanto una falle se encenderá otra. Suponiendo que las vidas medias están normalmente distribuidas, ¿cuál es la probabilidad de que den luz durante (*a*) al menos 500 h y (*b*) a lo sumo 4200 h?

Solución

Supongamos que las vidas medias sean L_1, L_2 y L_3. Entonces

$$\mu_{L1+L2+L3} = \mu_{L1} + \mu_{L2} + \mu_{L3} = 1500 + 1500 + 1500 = 4500 \text{ h}$$
$$\sigma_{L1+L2+L3} = \sqrt{\sigma_{L1}^2 + \sigma_{L2}^2 + \sigma_{L3}^2} = \sqrt{3(150)^2} = 260 \text{ h}$$

(*a*) $$5000 \text{ h en unidades estándar} = \frac{5000 - 4500}{260} = 1.92$$

Probabilidad pedida = (área bajo la curva normal a la derecha de $z = 1.92$)
$$= 0.5000 - 0.4726 = 0.0274$$

(*b*) $$4200 \text{ h en unidades estándar} = \frac{4200 - 4500}{260} = -1.15$$

Probabilidad pedida = (área bajo la curva normal a la izquierda de $z = -1.15$)
$$= 0.5000 - 0.3749 = 0.1251$$

PROBLEMAS DIVERSOS

8.17. Con referencia al Problema 8.1, hallar (*a*) la media de la distribución de muestreo de varianzas y (*b*) la desviación típica de la distribución de muestreo de varianza (o sea, el error típico de varianzas).

Solución

(*a*) Las varianzas muestrales correspondientes a cada una de las 25 muestras del Problema 8.1 son

0	0.25	4.00	9.00	20.25
0.25	0	2.25	6.25	16.00
4.00	2.25	0	1.00	6.25
9.00	6.25	1.00	0	2.25
20.25	16.00	6.25	2.25	0

La media de la distribución de muestreo de varianzas es

$$\mu_{s^2} = \frac{\text{suma de todas las varianzas en la tabla anterior}}{25} = \frac{135}{25} = 5.40$$

Eso pone de relieve el hecho de que $\mu_{s^2} = (N-1)(\sigma^2)/N$, ya que para $N = 2$ y $\sigma^2 = 10.8$ [véase Prob. 8.1(*b*)], el lado derecho es $\frac{1}{2}(10.8) = 5.4$.

El resultado dice que es deseable definir una varianza corregida para las muestras como

$$\hat{s}^2 = \frac{N}{N-1} s^2$$

Se seguiría entonces que $\mu_{\hat{s}^2} = \sigma^2$. Debemos hacer constar que las varianzas de la población se definirían igual que antes y que sólo las varianzas muestrales serían corregidas.

(*b*) La varianza de la distribución de muestreo de varianzas $\sigma_{s^2}^2$ se obtiene restando la media 5.40 de cada uno de los 25 números en la tabla anterior, elevando al cuadrado, sumándolos y dividiendo el resultado por 25. Así pues $\sigma_{s^2}^2 = 575.75/25 = 23.03$, o sea $\sigma_{s^2} = 4.80$.

8.18. Rehacer el Problema 8.17 sin reposición.

Solución

(*a*) Hay 10 muestras cuyas varianzas vienen dadas por los números de encima (o debajo) de la diagonal de la tabla del Problema 8.17(*a*). Luego

$$\mu_{s^2} = \frac{0.25 + 4.00 + 9.00 + 20.25 + 2.25 + 6.25 + 16.00 + 1.00 + 6.25 + 2.25}{10} = 6.75$$

Esto es un caso especial del resultado general

$$\mu_{s^2} = \left(\frac{N_p}{N_p - 1}\right)\left(\frac{N-1}{N}\right)\sigma^2$$

como se comprueba poniendo $N_p = 5$, $N = 2$ y $\sigma^2 = 10.8$ en el lado derecho para llegar a que $\mu_{s^2} = (\frac{5}{4})(\frac{1}{2})(10.8) = 6.75$.

(*b*) Restando 6.75 de cada uno de los 10 números sobre la diagonal de ceros de la tabla del Problema 8.17(*a*), elevando al cuadrado, sumando los resultados y dividiendo por 10, se ve que $\sigma_{s^2}^2 = 39.675$, o sea $\sigma_{s^2} = 6.30$.

8.19. La desviación típica de los pesos de una población muy numerosa de estudiantes es 10.0 lb. Se toman muestras de 200 estudiantes de dicha población y se calculan sus desviaciones típicas en altura. Hallar (*a*) la media y (*b*) la desviación típica de la distribución de muestreo de desviación típicas.

Solución

Podemos considerar que el muestreo es o bien de una población infinita o de una finita con reposición. De la Tabla 8.1 se tiene:

(*a*) La media de la distribución de muestreo de desviación típicas es $\mu_s = \sigma = 10.0$ lb.

(*b*) La desviación típica de la distribución de muestreo de desviaciones típicas es $\sigma_s = \sigma/\sqrt{2N} = 10/\sqrt{400} = 0.50$ lb.

8.20. ¿Qué porcentaje de las muestras del Problema 8.19 tendrian desviación típicas (*a*) mayores que 11.0 lb y (*b*) menores que 8.8 lb?

Solución

La distribución de muestreo de desviación típicas está casi normalmente distribuida con media 10.0 lb y desviación típica 0.50 lb.

(*a*) 11.10 lb en unidades estándar es $(11.0 - 10.0)/0.50 = 2.0$. El área bajo la curva normal a la derecha de $z = 2.0$ es $(0.5 - 0.4772) = 0.0228$; por tanto el porcentaje pedido es 2.3%.

(*b*) 8.8 lb en unidades estándar es $(8.8 - 10.0)/0.50 = -2.4$. El área bajo la curva normal a la izquierda de $z = -2.4$ es $(0.5 - 0.4918) = 0.0082$; luego el requerido porcentaje es 0.8%.

PROBLEMAS SUPLEMENTARIOS

DISTRIBUCION DE MUESTREO DE MEDIAS

8.21. Una población consiste en los números 3, 7, 11 y 15. Consideremos todas las posibles muestras de tamaño 2 que se pueden tomar de esa población con reposición. Hallar (*a*) la media de la población, (*b*) la desviación típica de la población, (*c*) la media de la distribución de muestreo de medias y (*d*) la desviación típica de la distribución de muestreo de medias. Verificar las partes (*c*) y (*d*) directamente de (*a*) y (*b*) usando fórmulas adecuadas.

8.22. Resolver el Problema 8.21 si el muestreo se hace con reposición.

8.23. Las masas de 1500 bolas de rodamientos están normalmente distribuidas, con media 22.40 g y desviación típica 0.048 g. Si se toman 300 muestras aleatorias de tamaño 36 en esa población, determinar la media esperada y la desviación típica esperada de la distribución de muestreo de medias, si el muestreo se hace (*a*) con, y (*b*) sin reposición.

8.24. Resolver el Problema 8.23 si la población consiste en 72 bolas.

8.25. ¿Cuántas de las muestras aleatorias del Problema 8.23 tendrían sus medias (*a*) entre 22.39 y 22.41 g, (*b*) mayor que 22.42 g, (*c*) menor que 22.37 g, y (*d*) menor que 22.38 g y más de 22.41 g?

8.26. Las lámparas que fabrica cierta empresa tienen una vida media de 800 h y una desviación típica de 60 h. Hallar la probabilidad de que una muestra aleatoria de 16 lámparas tenga una vida media (*a*) entre 790 y 810 h, (*b*) menor que 785 h, (*c*) más de 820 h y (*d*) entre 770 y 830 h.

8.27. Repetir el Problema 8.26 si se toma una muestra de 64 lámparas. Explicar la diferencia.

8.28. Los paquetes recibidos en un almacén tienen un peso medio de 300 lb y una desviación típica de 50 lb. ¿Cuál es la probabilidad de que 25 de esos paquetes, elegidos al azar y metidos en un montacargas, excedan el límite de carga de éste, que es de 8200 lb?

NUMEROS ALEATORIOS

8.29. Rehacer el Problema 8.6 usando un conjunto diferente de números aleatorios y seleccionando (*a*) 15, (*b*) 30, (*c*) 45 y (*d*) 60 muestras de tamaño 4 con reposición. Comparar en cada caso con los resultados teóricos.

8.30. Repetir el Problema 8.29 seleccionando muestras de tamaño (*a*) 2 y (*b*) 8 con reposición, en lugar de tamaño 4 con reposición.

8.31. Resolver el Problema 8.6 sin reposición. Comparar con los resultados teóricos.

8.32. (*a*) Mostrar cómo seleccionar 30 muestras de tamaño 2 de la distribución del Problema 3.61.
(*b*) Calcular la media y la desviación típica de la distribución de muestreo resultante de medias, y comparar con los resultados teóricos.

8.33. Resolver el Problema 8.32 usando muestras de tamaño 4.

DISTRIBUCION DE MUESTREO DE PROPORCIONES

8.34. Hallar la probabilidad de que en los 200 próximos nacimientos (*a*) menos del 40% sean niños, (*b*) entre 43% y 57% sean niñas y (*c*) más del 54% sean niños. Suponemos probabilidades de nacimiento iguales para niño y niña.

8.35. De 1000 muestras de 200 niños cada una, ¿en cuántas cabe esperar encontrar (*a*) menos del 40% de niños, (*b*) entre 40% y 60% son niñas y (*c*) el 53% o más son niñas?

8.36. Rehacer el Problema 8.34 si se consideran 100 niños en vez de 200, y explicar las diferencias en los resultados.

8.37. En una urna hay 80 fichas, de las que el 60% son rojas y el 40% blancas. De entre 50 muestras de 20 fichas cada una seleccionadas al azar, ¿cuántas es de esperar que tengan (*a*) tantas rojas como blancas, (*b*) 12 rojas y 8 blancas, (*c*) 8 rojas y 12 blancas y (*d*) 10 o más blancas?

8.38. Diseñar un experimento que ilustre los resultados del Problema 8.37. En vez de fichas rojas y blancas, puede usar papeletas en las que se han escrito *R* y *B* en las proporciones adecuadas. ¿Qué errores podrían introducirse al usar dos conjuntos diferentes de piezas?

8.39. Un fabricante envía 1000 lotes de 100 bombillas cada uno. Si el 5% de las bombillas son defectuosas, ¿en cuántos de los lotes se puede esperar que haya (*a*) menos de 90 bombillas buenas y (*b*) 98 o más buenas?

DISTRIBUCION DE MUESTREO DE DIFERENCIAS Y SUMAS

8.40. *A* y *B* producen dos tipos de cables que soportan cargas máximas medias de 4000 lb y 4500 lb, con desviación típica respectivas de 300 lb y 200 lb. Si se analizan 100 cables *A* y 50 cables *B*, ¿cuál es la probabilidad de que la carga máxima que soporta *B* sea (*a*) al menos 600 lb mayor que la de *A* y (*b*) al menos 450 lb mayor que la de *A*.

8.41. ¿Cuáles son las probabilidades en el Problema 8.40 si se analizan 100 cables de cada tipo? Razonar las diferencias.

8.42. La puntuación media en una prueba de aptitud es de 72 puntos con una desviación típica de 8 puntos. ¿Cuál es la probabilidad de que dos grupos de 28 y 36 estudiantes respectivamente, difieran en su puntuación media (*a*) 3 o más puntos, (*b*) 6 o más puntos y (*c*) entre 2 y 5 puntos?

8.43. Una urna contiene 60 piezas rojas y 40 blancas. Se sacan dos conjuntos de 30 con reposición y se anotan sus colores. ¿Cuál es la probabilidad de que los dos conjuntos difieran en 8 o más piezas rojas?

8.44. Resolver el Problema 8.43 sin reposición.

8.45. Un candidato recibe en unas elecciones el 65% de los votos. Hallar la probabilidad de que dos muestras aleatorias de 200 votantes indicasen una diferencia de más del 10% de votos a su favor.

8.46. Si U_1 y U_2 son los conjuntos de números del Problema 8,11, comprobar que (*a*) $\mu_{U1+U2} = \mu_{U1} + \mu_{U2}$ y (*b*) $\sigma_{U1+U2} = \sqrt{\sigma^2_{U1} + \sigma^2_{U2}}$.

8.47. Se han medido tres masas como 20.48, 35.97, y 62.34 g con desviaciones típicas de 0.21, 0.46 y 0.54 g, respectivamente. Hallar (*a*) la media y (*b*) la desviación típica de la suma de las masas.

8.48. El voltaje medio de unas baterías es 15.0 voltios (V) y la desviación típica es 0.2 V. ¿Cuál es la probabilidad de que 4 de ellas, conectadas en serie, tengan un voltaje combinado de 60.8 V o más?

PROBLEMAS DIVERSOS

8.49. Una población de 7 números tiene una media de 40 y una desviación típica de 3. Si se toman muestras de tamaño 5 de esa población, y se calcula la varianza de cada muestra, hallar la media de la distribución de muestreo de varianzas si el muestreo se hace (*a*) con y (*b*) sin reposición.

8.50. Los tubos fabricados en cierta empresa tienen una vida media de 900 h y una desviación típica de 80 h. Se envían 1000 lotes de 100 tubos cada uno. ¿En cuántos de esos lotes se puede esperar que (*a*) la vida media exceda de 900 h y (*b*) la desviación típica de las vidas medias exceda de 95 h? ¿Qué hipótesis hay que hacer?

8.51. Si la mediana de las vidas medias del Problema 8.50 es 900 h, ¿en cuántos lotes cabe esperar que la mediana de las vidas medias sea mayor que 910 h? Comparar la respuesta con el Problema 8.50(*a*) y explicar los resultados.

8.52. En un examen las notas estuvieron normalmente distribuidas con media 72 y desviación típica 8.

(*a*) Hallar la nota mínima del 20% de estudiantes mejores.

(*b*) Hallar la probabilidad de que en una muestra aleatoria de 100 estudiantes, la nota más baja sea inferior a 76.

CAPITULO 9

Teoría de la estimación estadística

ESTIMACION DE PARAMETROS

En el último capítulo vimos cómo se puede emplear la teoría del muestreo para recabar información acerca de muestras aleatorias tomadas de una población conocida. Desde un punto de vista práctico, no obstante, suele resultar más importante ser capaz de inferir información sobre la población a partir de muestras suyas. Con tal situación trata la *inferencia estadística*, que usa los principios de la teoría del muestreo.

Un problema importante de la inferencia estadística es la estimación de *parámetros de la población*, o brevemente *parámetros* (tales como la media o la varianza de la población), de los correspondientes *estadísticos muestrales*, o simplemente *estadísticos* (tales como la media y la varianza de la muestra). Consideramos este problema en el presente capítulo.

ESTIMACIONES SIN SESGO

Si la media de las distribuciones de muestreo de un estadístico es igual que la del correspondiente parámetro de la población, el estadístico se llama un *estimador sin sesgo* del parámetro; si no, se llama un estimador *sesgado*. Los correspondientes valores de tales estadísticos se llaman *estimaciones sin sesgo y sesgadas*, respectivamente.

EJEMPLO 1. La media de las distribuciones de muestreo de medias $\mu_{\bar{X}}$ e μ, la media de la población. Por tanto, la media muestral $\bar{X}$ es una estimación sin sesgo de la media de la población μ.

EJEMPLO 2. La media de las distribuciones de muestreo de varianza es

$$\mu_{s^2} = \frac{N-1}{N}\sigma^2$$

donde σ^2 es la varianza de la población y N es el tamaño de la muestra (véase Tabla 8.1). Así pues, la varianza de la muestra s^2 es una estimación sesgada de la varianza de la población σ^2. Usando la varianza modificada

$$\hat{s}^2 = \frac{N}{N-1}s^2$$

encontramos $\mu_{\hat{s}^2} = \sigma^2$, de manera que $\hat{s}^2$ es una estimación sin sesgo de σ^2. Sin embargo, $\hat{s}$ es una estimación sesgada de σ.

En términos de esperanzas (Cap. 6) podríamos decir que un estadístico es insesgado si su esperanza es igual al correspondiente parámetro de población. Así, $\bar{X}$ y $\hat{s}^2$ son insesgados porque $E\{\bar{X}\} = \mu$ y $E\{\hat{s}^2\} = \sigma^2$.

ESTIMACION EFICIENTE

Si las distribuciones de muestreo de dos estadísticos tienen la misma media (o esperanza), el de menor varianza se llama un *estimador eficiente* de la media, mientras que el otro se llama un *estimador ineficiente*. Los valores correspondientes de los estadísticos se llaman *estimación eficiente* e *estimación ineficiente*, respectivamente.

Si consideramos todos los posibles estadísticos cuyas distribuciones de muestreo tienen la misma media, aquel de varianza mínima se llama a veces el *estimador de máxima eficiencia*, o sea, *el mejor estimador*.

EJEMPLO 3. Las distribuciones de muestreo de media y mediana tienen ambas la misma media, a saber, la media de la población. Sin embargo, la varianza de la distribución de muestreo de medias es menor que la varianza de la distribución de muestreo de medianas (véase Tabla. 8.1). Por tanto, la media muestral da una estimación eficiente de la media de la población, mientras la mediana de la muestra da una estimación ineficiente de ella.

De todos los estadísticos que estiman la media de la población, la media muestral proporciona la mejor (la más eficiente) estimación.

En la práctica, estimaciones ineficientes se usan con frecuencia a causa de la relativa sencillez con que se obtienen algunas de ellas.

ESTIMACIONES DE PUNTO Y ESTIMACIONES DE INTERVALO; SU FIABILIDAD

Una estimación de un parámetro de la población dada por un solo número se llama una *estimación de punto* del parámetro. Una estimación de un parámetro de la población dada por dos números, entre los cuales se puede considerar encajado al parámetro, se llama una *estimación de intervalo* del parámetro.

Las estimaciones de intervalo indican la precisión de una estimación y son por tanto preferibles a las estimaciones de punto.

EJEMPLO 4. Si decimos que una distancia se ha medido como 5.28 metros (m), estamos dando una estimación de punto. Por otra parte, si decimos que la distancia es 5.28 ± 0.03 m (o sea, que está entre 5.25 y 5.31 m), estamos dando una estimación de intervalo.

El margen de error (o la precisión) de una estimación nos informa de su *fiabilidad*.

ESTIMACIONES DE INTERVALO DE CONFIANZA PARA PARAMETROS DE POBLACION

Sean μ_S y σ_S la media y la desviación típica (error típico) de la distribución de muestreo de un estadístico S. Entonces, si la distribución de muestreo de S es aproximadamente normal (que como hemos visto es cierto para muchos estadísticos si el tamaño de la muestra $N \geqslant 30$), podemos esperar hallar un estadístico muestral real S que esté en los intervalos $\mu_S - \sigma_S$ a $\mu_S + \sigma_S$, $\mu_S - 2\sigma_S$ a $\mu_S + 2\sigma_S$, o $\mu_S - 3\sigma_S$ a $\mu_S + 3\sigma_S$ alrededor del 68.27%, 95.45% y 99.73% del tiempo, respectivamente.

Equivalentemente, podemos esperar hallar (o sea, podemos estar *confiados* en encontrar) μ_S en los intervalos $S - \sigma_S$ a $S + \sigma_S$, $S - 2\sigma_S$ a $S + 2\sigma_S$, o $S - 3\sigma_S$ a $S + 3\sigma_S$ alrededor del 68.27%, 95.45% y 99.73% del tiempo, respectivamente. Por esa razón, llamamos a esos respectivos intervalos los *intervalos de confianza* 68.27%, 95.45% y 99.73% para estimar μ_S. Los números extremos de estos intervalos ($S \pm \sigma_S$, $S \pm 2\sigma_S$ y $S \pm 3\sigma_S$) se llaman entonces los *límites de confianza* 68.27%, 95.45% y 99.73% o *límites fiduciales*.

Análogamente, $S \pm 1.96\sigma_S$ y $S \pm 2.58\sigma_S$ son los límites de confianza 95% y 99% (o sea, 0.95 y 0.99) para S. El porcentaje de confianza se suele llamar *nivel de confianza*. Los números 1.96, 2.58, etc. en los límites de confianza se llaman *coeficientes de confianza* o *valores críticos*, y se denotan por z_c. De los niveles de confianza podemos deducir los coeficientes de confianza y viceversa.

La Tabla 9.1 muestra los valores de z_c correspondientes a varios niveles de confianza usados en la práctica. Para niveles de confianza que no aparecen en la tabla, los valores de z_c se pueden encontrar gracias a las tablas de áreas bajo la curva normal (Apéndice II).

Tabla 9.1

Nivel de confianza	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

Intervalos de confianza para las medias

Si el estadístico S es la media $\bar{X}$ de la muestra, entonces los límites de confianza 95% y 99% para estimar la media μ de la población vienen dados por $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\sigma_{\bar{X}}$, respectivamente. Más en general, los límites de confianza para estimar la media de la población μ vienen dados por $\bar{X} \pm z_c\sigma_{\bar{X}}$, donde z_c (que depende del nivel particular de confianza deseado) se puede leer en la Tabla 9.1. Usando los valores de $\sigma_{\bar{X}}$ obtenidos en el Capítulo 8, vemos que los límites de confianza para la media de la población están dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \tag{1}$$

si el muestreo es de una población infinita o de una finita con reposición, y vienen dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} \tag{2}$$

si el muestreo es sin reposición de una población finita de tamaño N_p.

Generalmente, la desviación típica σ de la población no es conocida; así pues, para obtener los anteriores límites de confianza usamos la estimación muestral $\hat{s}$ o s. Esto se verá que es satisfactorio para $N \geqslant 30$. Para $N < 30$, la aproximación es pobre y debe emplearse la teoría de pequeñas muestras (Cap. 11).

Intervalos de confianza para proporciones

Si el estadístico S es la proporción de «éxitos» en una muestra de tamaño N sacada de una población binomial en la que p es la proporción de éxitos (o sea, la probabilidad de éxito), entonces los límites de confianza para p vienen dados por $P \pm z_c\sigma_P$, donde P es la proporción de éxitos en la muestra de tamaño N. Usando los valores de σ_P obtenidos en el Capítulo 8, vemos que los límites de confianza para la proporción en la población vienen dados por

$$P \pm z_c\sqrt{\frac{pq}{N}} = P \pm z_c\sqrt{\frac{p(1-p)}{N}} \tag{3}$$

si el muestreo es de una población infinita o finita con reposición, y por

$$P \pm z_c\sqrt{\frac{pq}{N}}\sqrt{\frac{N_p - N}{N_p - 1}} \tag{4}$$

si el muestreo es de una población finita de tamaño N_p y sin reposición.

Para calcular estos límites de confianza, podemos usar la estimación muestral P para p, que generalmente resultará satisfactoria si $N \geqslant 30$. Un método más exacto para obtener los límites de confianza se presenta en el Problema 9.12.

Intervalos de confianza para diferencias y sumas

Si S_1 y S_2 son dos estadísticos muestrales con distribuciones de muestreo aproximadamente normales, los límites de confianza para la diferencia de los parámetros de población correspondientes a S_1 y S_2 vienen dados por

$$S_1 - S_2 \pm z_c\sigma_{S1-S2} = S_1 - S_2 \pm z_c\sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{5}$$

mientras que los límites de confianza para la suma de los parámetros de población vienen dados por

$$S_1 + S_2 \pm z_c\sigma_{S1+S2} = S_1 + S_2 \pm z_c\sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{6}$$

supuesto que las muestras sean independientes (véase Cap. 8).

Por ejemplo, los límites de confianza para la diferencia de dos medias poblacionales, en el caso de poblaciones infinitas, se calculan como

$$\bar{X}_1 - \bar{X}_2 \pm z_c\sigma_{\bar{X}1-\bar{X}2} = \bar{X}_1 - \bar{X}_2 \pm z_c\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \tag{7}$$

donde $\bar{X}_1$, σ_1, N_1 y $\bar{X}_2$, σ_2, N_2 son las respectivas medias, desviaciones típicas y tamaños de las dos muestras sacadas de las poblaciones.

De forma similar, los límites de confianza para la diferencia de dos proporciones poblacionales, con poblaciones infinitas, están dados por

$$P_1 - P_2 \pm z_c \sigma_{P1-P2} = P_1 - P_2 \pm z_c \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}} \tag{8}$$

donde P_1 y P_2 son las dos proporciones muestrales, N_1 y N_2 los tamaños de las dos muestras, y p_1 y p_2 las proporciones en las dos poblaciones (estimadas por P_1 y P_2).

Intervalos de confianza para desviaciones típicas

Los límites de confianza para la desviación típica σ de una población normalmente distribuida, estimados con una muestra con desviación típica s, vienen dados por

$$s \pm z_c \sigma_s = s \pm z_c \frac{\sigma}{\sqrt{2N}} \tag{9}$$

usando la Tabla 8.1. Al calcular estos límites de confianza, usamos s o $\hat{s}$ para estimar σ.

ERROR PROBABLE

Los límites de confianza 50% de los parámetros de población correspondientes a un estadístico S vienen dados por $S \pm 0.6745\sigma_S$. La cantidad $0.6745\sigma_S$ se conoce como el *error probable de la estimación.*

PROBLEMAS RESUELTOS

ESTIMACIONES SIN SESGO Y EFICIENTES

9.1. Dar un ejemplo de estimadores (o estimaciones) que sean (*a*) sin sesgo y eficiente, (*b*) sin sesgo e ineficiente y (*c*) sesgado e ineficiente.

Solución

(*a*) La media muestral $\bar{X}$ y la varianza muestral modificada

$$\hat{s}^2 = \frac{N}{N-1} s^2$$

son dos ejemplos.

(*b*) La mediana muestral y el estadístico muestral $\frac{1}{2}(Q_1 + Q_3)$, donde Q_1 y Q_3 son los cuartiles muestrales inferior y superior, son dos ejemplos. Ambos son estimaciones sin sesgo de la media de la población, pues la media de sus distribuciones de muestreo es la media de la población.

(c) La desviación típica muestral s, la desviación típica modificada $\hat{s}$, la desviación media y el rango semi-intercuartil son cuatro ejemplos.

9.2. En una muestra de cinco medidas, un científico anotó 6.33, 6.37, 6.36, 6.32 y 6.37 centímetros (cm). Determinar estimaciones insesgadas y eficientes de (a) la verdadera media y (b) la varianza.

Solución

(a) La estimación sin sesgo y eficiente de la media verdadera (o sea, la de la población) es

$$\bar{X} = \frac{\sum X}{N} = \frac{6.33 + 6.37 + 6.36 + 6.32 + 6.37}{5} = 6.35 \text{ cm}$$

(b) La estimación sin sesgo y eficiente de la media verdadera (o sea, la de la población) es

$$\hat{s}^2 = \frac{N}{N-1} s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

$$= \frac{(6.33 - 6.35)^2 + (6.37 - 6.35)^2 + (6.36 - 6.35)^2 + (6.32 + 6.35)^2 + (6.37 - 6.35)^2}{5 - 1} =$$

$$= 0.00055 \text{ cm}^2$$

Nótese que aunque $\hat{s} = \sqrt{0.00055} = 0.023$ cm es una estimación de la verdadera desviación típica, esta estimación no es ni eficiente ni insesgada.

9.3. Supongamos que las alturas de 100 estudiantes varones de la Universidad XYZ representan una muestra aleatoria de las de los 1546 estudiantes de esa Universidad. Determinar estimaciones sin sesgo y eficientes de (a) la media verdadera y (b) la varianza verdadera.

Solución

(a) Por el Problema 3.22, la estimación sin sesgo y eficiente de la verdadera media es $\bar{X} = 67.45$ in.

(b) Del Problema 4.17 se sigue que la estimación sin sesgo y eficiente de la verdadera varianza es

$$\hat{s}^2 = \frac{N}{N-1} s^2 = \frac{100}{99} (8.5275) = 8.6136$$

Así pues, $\hat{s} = \sqrt{8.6136} = 2.93$ in. Notemos que ya que N es grande, no hay diferencia casi entre s^2 y $\hat{s}^2$, o sea entre s y $\hat{s}$.

No hemos usado la corrección de Sheppard para el agrupamiento. Para tener esto en cuenta, usaríamos $s = 2.79$ in (véase Prob. 4.21).

9.4. Dar una estimación sin sesgo e ineficiente para la verdadera media del diámetro de la esfera del Problema 9.2.

Solución

La mediana es un ejemplo. Para las cinco medidas, ordenadas por magnitud, la mediana es 6.36 cm.

donde $\bar{X}_1$, σ_1, N_1 y $\bar{X}_2$, σ_2, N_2 son las respectivas medias, desviaciones típicas y tamaños de las dos muestras sacadas de las poblaciones.

De forma similar, los límites de confianza para la diferencia de dos proporciones poblacionales, con poblaciones infinitas, están dados por

$$P_1 - P_2 \pm z_c\sigma_{P1-P2} = P_1 - P_2 \pm z_c\sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}} \tag{8}$$

donde P_1 y P_2 son las dos proporciones muestrales, N_1 y N_2 los tamaños de las dos muestras, y p_1 y p_2 las proporciones en las dos poblaciones (estimadas por P_1 y P_2).

Intervalos de confianza para desviaciones típicas

Los límites de confianza para la desviación típica σ de una población normalmente distribuida, estimados con una muestra con desviación típica s, vienen dados por

$$s \pm z_c\sigma_s = s \pm z_c\frac{\sigma}{\sqrt{2N}} \tag{9}$$

usando la Tabla 8.1. Al calcular estos límites de confianza, usamos s o $\hat{s}$ para estimar σ.

ERROR PROBABLE

Los límites de confianza 50% de los parámetros de población correspondientes a un estadístico S vienen dados por $S \pm 0.6745\sigma_S$. La cantidad $0.6745\sigma_S$ se conoce como el *error probable de la estimación*.

PROBLEMAS RESUELTOS

ESTIMACIONES SIN SESGO Y EFICIENTES

9.1. Dar un ejemplo de estimadores (o estimaciones) que sean (*a*) sin sesgo y eficiente, (*b*) sin sesgo e ineficiente y (*c*) sesgado e ineficiente.

Solución

(*a*) La media muestral $\bar{X}$ y la varianza muestral modificada

$$\hat{s}^2 = \frac{N}{N-1}s^2$$

son dos ejemplos.

(*b*) La mediana muestral y el estadístico muestral $\frac{1}{2}(Q_1 + Q_3)$, donde Q_1 y Q_3 son los cuartiles muestrales inferior y superior, son dos ejemplos. Ambos son estimaciones sin sesgo de la media de la población, pues la media de sus distribuciones de muestreo es la media de la población.

(c) La desviación típica muestral s, la desviación típica modificada $\hat{s}$, la desviación media y el rango semi-intercuartil son cuatro ejemplos.

9.2. En una muestra de cinco medidas, un científico anotó 6.33, 6.37, 6.36, 6.32 y 6.37 centímetros (cm). Determinar estimaciones insesgadas y eficientes de (a) la verdadera media y (b) la varianza.

Solución

(a) La estimación sin sesgo y eficiente de la media verdadera (o sea, la de la población) es

$$\bar{X} = \frac{\sum X}{N} = \frac{6.33 + 6.37 + 6.36 + 6.32 + 6.37}{5} = 6.35 \text{ cm}$$

(b) La estimación sin sesgo y eficiente de la media verdadera (o sea, la de la población) es

$$\hat{s}^2 = \frac{N}{N-1}s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

$$= \frac{(6.33 - 6.35)^2 + (6.37 - 6.35)^2 + (6.36 - 6.35)^2 + (6.32 + 6.35)^2 + (6.37 - 6.35)^2}{5 - 1} =$$

$$= 0.00055 \text{ cm}^2$$

Nótese que aunque $\hat{s} = \sqrt{0.00055} = 0.023$ cm es una estimación de la verdadera desviación típica, esta estimación no es ni eficiente ni insesgada.

9.3. Supongamos que las alturas de 100 estudiantes varones de la Universidad XYZ representan una muestra aleatoria de las de los 1546 estudiantes de esa Universidad. Determinar estimaciones sin sesgo y eficientes de (a) la media verdadera y (b) la varianza verdadera.

Solución

(a) Por el Problema 3.22, la estimación sin sesgo y eficiente de la verdadera media es $\bar{X} = 67.45$ in.
(b) Del Problema 4.17 se sigue que la estimación sin sesgo y eficiente de la verdadera varianza es

$$\hat{s}^2 = \frac{N}{N-1}s^2 = \frac{100}{99}(8.5275) = 8.6136$$

Así pues, $\hat{s} = \sqrt{8.6136} = 2.93$ in. Notemos que ya que N es grande, no hay diferencia casi entre s^2 y $\hat{s}^2$, o sea entre s y $\hat{s}$.

No hemos usado la corrección de Sheppard para el agrupamiento. Para tener esto en cuenta, usaríamos $s = 2.79$ in (véase Prob. 4.21).

9.4. Dar una estimación sin sesgo e ineficiente para la verdadera media del diámetro de la esfera del Problema 9.2.

Solución

La mediana es un ejemplo. Para las cinco medidas, ordenadas por magnitud, la mediana es 6.36 cm.

INTERVALOS DE CONFIANZA PARA MEDIAS

9.5. Hallar los intervalos de confianza (*a*) 95% y (*b*) 99% para estimar la altura media de los estudiantes del Problema 9.3.

Solución

(*a*) Los límites de confianza 95% son $\bar{X} \pm 1.96\sigma/\sqrt{N}$. Usando $\bar{X} = 67.45$ in y $\hat{s} = 2.93$ in como estimación de σ (véase Prob. 9.3), los límites de confianza son $67.45 \pm 1.96(2.93/\sqrt{100})$, o 67.45 ± 0.57 in. Luego el intervalo de confianza 95% para la media de la población μ es 66.88 a 68.02 in, que denotamos por $66.88 < \mu < 68.02$.

Podemos decir, por tanto, que la probabilidad de que la altura media de la población esté entre 66.88 y 68.02 in es del 95%, o sea 0.95. En símbolos escribimos $\Pr\{66.88 < \mu < 68.02\} = 0.95$. Esto equivale a decir que tenemos 95% de *confianza* de que la media de la población (o media verdadera) esté entre 66.88 y 68.02 in.

(*b*) Los límites de confianza 99% son $\bar{X} \pm 2.58\sigma/\sqrt{N} = \bar{X} \pm 2.58\hat{s}/\sqrt{N} = 67.45 \pm 2.58(2.93/\sqrt{100}) = 67.45 \pm 0.76$ in. Luego el intervalo de confianza 99% para la media de la población μ es 66.69 a 68.21 in, que se denota por $66.69 < \mu < 68.21$.

Al hallar los anteriores intervalos de confianza, hemos supuesto que la población era infinita o tan grande que podíamos considerarla como con reposición. Para poblaciones finitas y muestreo sin reposición, debe usarse

$$\frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} \quad \text{en lugar de} \quad \frac{\sigma}{\sqrt{N}}$$

No obstante, podemos considerar el factor

$$\sqrt{\frac{N_p - N}{N_p - 1}} = \sqrt{\frac{1546 - 100}{1546 - 1}} = 0.967$$

como esencialmente 1.0, y por tanto no es necesario usarlo. Si se usa, los límites de confianza anteriores se convierten en 67.45 ± 0.56 in y 67.45 ± 0.73 in, respectivamente.

9.6. Las medidas de los diámetros de una muestra aleatoria de 200 bolas de rodamientos producidas por una máquina en una semana, dieron una media de 0.824 cm y una desviación típica de 0.042 cm. Hallar los límites de confianza (*a*) 95% y (*b*) 99% para el diámetro medio de todas las bolas.

Solución

(*a*) Los límites de confianza 95% son

$$\bar{X} \pm \frac{1.96\sigma}{\sqrt{N}} = \bar{X} \pm \frac{1.96\hat{s}}{\sqrt{N}} = 0.824 \pm 1.96\,\frac{0.042}{\sqrt{200}} = 0.824 + 0.0058 \text{ cm} \quad \text{o sea} \quad 0.824 \pm 0.006 \text{ cm}$$

(*b*) Los límites de confianza 99% son

$$\bar{X} \pm \frac{2.58\sigma}{\sqrt{N}} = \bar{X} \pm \frac{2.58\hat{s}}{\sqrt{N}} = 0.824 \pm 2.58\,\frac{0.042}{\sqrt{200}} = 0.824 \pm 0.0077 \text{ cm} \quad \text{o sea} \quad 0.824 \pm 0.008 \text{ cm}$$

Nótese que hemos supuesto la desviación típica dada como la desviación típica *modificada* $\hat{s}$. Si la desviación típica hubiera sido s, hubiéramos usado $\hat{s} = \sqrt{N/(N-1)}s = \sqrt{200/199}s$, que puede ser tomada como s a efectos prácticos. En general, para $N \geqslant 30$ podemos suponer que s y $\hat{s}$ son prácticamente iguales.

9.7. Hallar los límites de confianza (*a*) 98%, (*b*) 90% y (*c*) 99.73% para el diámetro medio de las bolas del Problema 9.6.

Solución

(*a*) Sea $z = z_c$ tal que el área bajo la curva normal a su derecha es 1%. Entonces, por simetría, el área a la izquierda de $z = -z_c$ es también 1%, así que el área sombreada es el 98% del total; véase Figura 9.1(*a*). Como el área total bajo la curva es 1, el área desde $z = 0$ hasta $z = z_c$ es 0.49; por tanto, $z_c = 2.33$. Luego los límites de confianza 98% son $\bar{X} \pm 2.33\sigma/\sqrt{N} = 0.824 \pm$ $\pm\ 2.33(0.042/\sqrt{200}) = 0.824 \pm 0.0069$ cm.

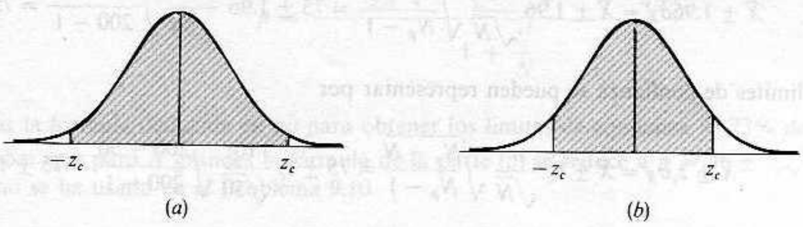


Figura 9.1.

(*b*) Deseamos un z_c tal que el área desde $z = 0$ hasta $z = z_c$ es 0.45, como muestra la Figura 9.1(*b*); entonces $z_c = 1.645$. Así pues, los límites de confianza 90% son $\bar{X} \pm 1.645\sigma/\sqrt{N} = 0.824 \pm$ $\pm\ 1.645(0.042/\sqrt{200}) = 0.824 \pm 0.0049$ cm.

(*c*) Los límites de confianza del 99.73% son

$$\bar{X} \pm 3\sigma/\sqrt{N} = 0.824 \pm 3(0.042/\sqrt{200}) = 0.824 \pm 0.0089 \text{ cm}$$

9.8. Al medir el tiempo de reacción, un psicólogo estima que la desviación típica es 0.05 segundos. ¿De qué tamaño ha de tomarse una muestra de medidas para tener una confianza del (*a*) 95% y (*b*) 99% de que el error de la estimación no supera 0.01 segundos?

Solución

(*a*) Los límites de confianza 95% son $\bar{X} \pm 1.96\sigma/\sqrt{N}$, siendo el error de la estimación $1.96\sigma/\sqrt{N}$. Tomando $\sigma = s = 0.05$ seg, vemos que este error será igual a 0.01 seg si $(1.96)(0.05)/\sqrt{N} = 0.01$; esto es, $\sqrt{N} = (1.96)(0.05)/0.01 = 9.8$, o sea $N = 96.04$. Luego podemos estar confidentes al 95% de que el error de la estimación será menor que 0.01 seg si N es 97 o mayor.

Otro método

$$\frac{(1.96)(0.05)}{\sqrt{N}} \leqslant 0.01 \quad \text{si} \quad \frac{\sqrt{N}}{(1.96)(0.05)} \geqslant \frac{1}{0.01} \quad \text{o sea} \quad \sqrt{N} \geqslant \frac{(1.96)(0.05)}{0.01} = 9.8$$

Entonces $N \geqslant 96.04$, o sea $N \geqslant 97$.

(*b*) Los límites de confianza 99% son $\bar{X} \pm 2.58\sigma/\sqrt{N}$. Entonces $(2.58)(0.05)/\sqrt{N} = 0.01$, es decir

$N = 166.4$. Luego podemos tener confianza al 99% de que el error de la estimación será menor que 0.01 seg si N es 167 o mayor.

9.9. Una muestra al azar de 50 notas de matemáticas de entre un total de 200, revela una media de 75 y una desviación típica de 10.

(*a*) ¿Cuáles son los límites de confianza 95% para estimaciones de la media de las 200 notas?
(*b*) ¿Con qué grado de confianza podríamos decir que la media de las 200 es 75 ± 1?

Solución

(*a*) Como la población no es muy grande comparada con el tamaño de la muestra, debemos tenerlo en cuenta. Por tanto, los límites de confianza 95% son

$$\bar{X} \pm 1.96\sigma_{\bar{X}} = \bar{X} \pm 1.96 \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} = 75 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{200 - 50}{200 - 1}} = 75 \pm 2.4$$

(*b*) Los límites de confianza se pueden representar por

$$\bar{X} \pm z_c\sigma_{\bar{X}} = \bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} = 75 \pm z_c \frac{10}{\sqrt{50}} \sqrt{\frac{200 - 50}{200 - 1}} = 75 \pm 1.23z_c$$

Como esto ha de ser igual a 75 ± 1, tenemos $1.23\, z_c = 1$, o sea $z_c = 0.81$. El área bajo la curva normal entre $z = 0$ y $z = 0.81$ es 0.2910; luego el requerido grado de confianza es $2(0.2910) = 0.582$, o sea 58.2%.

INTERVALOS DE CONFIANZA PARA PROPORCIONES

9.10. Un sondeo de 100 votantes elegidos al azar en un distrito indica que el 55% de ellos estaban a favor de un cierto candidato. Hallar los límites de confianza (*a*) 95%, (*b*) 99% y (*c*) 99.73% para la proporción de todos los votantes favorables a ese candidato.

Solución

(*a*) Los límites de confianza 95% para la población p son $P \pm 1.96\sigma_P = P \pm 1.96\sqrt{p(1-p)/N} = 0.55 \pm 1.96\sqrt{(0.55)(0.45)/100} = 0.55 \pm 0.10$, donde hemos usado la proporción muestral P para estimar p.

(*b*) Los límites de confianza 99% para p son $0.55 \pm 2.58\sqrt{(0.55)(0.45)/100} = 0.55 \pm 0.13$.

(*c*) Los límites de confianza 99.73% para p son $0.55 \pm 3\sqrt{(0.55)(0.45)/100} = 0.55 \pm 0.15$.

9.11. ¿De qué tamaño hay que tomar el sondeo del Problema 9.10 para tener confianza al (*a*) 95% y (*b*) 99.73% de que el candidato saldrá elegido?

Solución

Los límites de confianza para p son $P \pm z_c\sqrt{p(1-p)/N} = 0.55 \pm z_c\sqrt{(0.55)(0.45)/N} = 0.55 \pm 0.50z_c/\sqrt{N}$, donde hemos usado la estimación $P = p = 0.55$ basados en el Problema 9.10. Como el candidato ganará sólo si recibe más del 50% de los votos de la población, exigimos que $0.50z_c/\sqrt{N}$ sea menor que 0.05.

(*a*) Para 95% de confianza, $0.50z_c/\sqrt{N} = 0.50(1.96)/\sqrt{N} = 0.05$ cuando $N = 384.2$. Luego N debe ser al menos 385.

(*b*) Para 99.73 de confianza, $0.50z_c/\sqrt{N} = 0.50(3)/\sqrt{N} = 0.05$ cuando $N = 900$. Luego N debe ser al menos 901.

Otro método

$1.50/\sqrt{N} < 0.05$ cuando $\sqrt{N}/1.50 > 1/0.05$ o sea $\sqrt{N} > 1.50/0.05$. Entonces $\sqrt{N} > 30$, es decir, $N > 900$, así que N ha de ser al menos 901.

9.12. (*a*) Si P es la proporción observada de éxitos en una muestra de tamaño N, probar que los límites de confianza para estimar la proporción de éxitos p de la población en el nivel de confianza determinado por z_c vienen dados por

$$p = \frac{P + \dfrac{z_c^2}{2N} \pm z_c\sqrt{\dfrac{P(1-P)}{N} + \dfrac{z_c^2}{4N^2}}}{1 + \dfrac{z_c^2}{N}}$$

(*b*) Usar la fórmula deducida en (*a*) para obtener los límites de confianza 99.73% del Problema 9.10.

(*c*) Probar que para N grandes la fórmula de la parte (*a*) se reduce a $p = P \pm z_c\sqrt{P(1-P)/N}$, tal como se ha usado en el Problema 9.10.

Solución

(*a*) La proporción muestral P en unidades estándar es

$$\frac{P - p}{\sigma_P} = \frac{P - p}{\sqrt{p(1-p)/N}}$$

Los valores máximo y mínimo de esta variable tipificada son $\pm z_c$, donde z_c determina el valor de confianza. En estos valores extremos debemos tener en consecuencia

$$P - p = \pm z_c\sqrt{\frac{p(1-p)}{N}}$$

Elevando al cuadrado $\quad P^2 - 2pP + p^2 = \dfrac{z_c^2 p(1-p)}{N}$

Multiplicando ambos lados por N y simplificando, encontramos que

$$(N + z_c^2)p^2 - (2NP + z_c^2)p + NP^2 = 0$$

Si $a = N + z_c^2$, $b = -(2NP + z_c^2)$ y $c = NP^2$, esta ecuación pasa a ser $ap^2 + bp + c = 0$ cuya solución para p viene dada por la fórmula cuadrática

$$p = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{2NP + z_c^2 \pm \sqrt{(2NP + z_c^2)^2 - 4(N + z_c^2)(NP^2)}}{2(N + z_c^2)}$$

$$= \frac{2NP + z_c^2 \pm z_c\sqrt{4NP(1-P) + z_c^2}}{2(N + z_c^2)}$$