5

Información de retorno sobre las actividades de estudio independiente

En este capítulo se proporciona la información de retorno referente a las actividades de estudio independiente; se desarrollan diversos procesos de cálculo y se generan las respuestas a las preguntas formuladas en las mencionadas actividades; aquellas que el estudiante o el lector debería haber resuelto una vez hubiese realizado una lectura comprensiva de los temas y revisado los ejemplos indicados a través de cada capítulo. Con el propósito de aclarar algunas dudas y preguntas de investigación, se desarrollan completamente los ejercicios que se propusieron en esas actividades por capítulos.

Capítulo 1

1.1
a) La estadística *descriptiva*

b) La estadística *inferencial*

c) Una *muestra*

d) Se denominan *variables*

e) Variables *cualitativas*

f) Variables *cuantitativas.*

g) Una variable *continua*

h) Una variable *discreta*

1.2
a) cualitativa en escala nominal

b) cualitativa en escala ordinal

c) cuantitativa en escala ordinal

d) cuantitativa en escala de intervalo

e) cuantitativa en escala de razón

1.3
a) n = 30

b) 3/30 = 0.3 = 30 %

c) No es adecuado calcular el promedio, porque son datos correspondientes a una variable cualitativa; para esta se recomienda calcular un porcentaje o una proporción.

d) 63.8914

e) 3.14225

f) 0.0491 = 4.91 %, este es un valor inferior al 8 % e indica que los datos de la variable estatura en esta muestra de estudiantes son homogéneos.

1.4 Del conjunto de datos del ejemplo 1.11 es posible seleccionar 10 muestras aleatorias de tamaño $n = 3$ mediante un muestreo aleatorio sin reemplazo, una de ellas está constituida por los datos 9, 10, 8.

Luego un valor $\overline{x}$ de la variable promedio muestral $\overline{X}$ o estimación de la media poblacional es:

$$\overline{x} = \frac{\sum_{i=1}^{3} x_i}{3} = \frac{9+10+8}{3} = \frac{27}{3} = 9$$

Un valor s^2 de la variable S^2 o estimación de varianza poblacional es:

$$s^2 = \frac{\sum_{i=1}^{3} (x_i - 9)^2}{3} = \frac{(9-9)^2 + (10-9)^2 + (8-9)^2}{3}$$

$$s^2 = \frac{(0)^2 + (1)^2 + (-1)^2}{3} = \frac{0+1+1}{3} = \frac{2}{3} \cong 0.6666$$

Otra estimación de la varianza poblacional es $\hat{s}^2$, valor particular de la variable

$\hat{S}^2$; esta se obtiene así:

$$\hat{s}^2 = \frac{\sum_{i=1}^{3}(x_i - 9)^2}{3-1} = \frac{2}{2} = 1$$

Se observa que el valor $\hat{s}^2$ se obtiene usando la siguiente expresión:

$$\hat{s}^2 = \frac{n}{n-1}s^2 = \frac{3}{3-1}\left(\frac{2}{3}\right) = \frac{6}{6} = 1$$

Un valor s para la variable aleatoria S, denominada *desviación estándar* muestral, es:

$$s = \sqrt{\frac{\sum_{i=1}^{3}(x_i - 9)^2}{3}} = \sqrt{0.6666} \cong 0.8164$$

Un valor $\hat{s}$ para la variable $\hat{S}$ desviación estándar corregida o cuasidesviación estándar es:

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^{3}(x_i - 9)^2}{3-1}} = \sqrt{1} = 1$$

Un valor del coeficiente de variación muestral es:

$$cv = \frac{\hat{s}}{\overline{x}} = \frac{1}{9} \cong 0.1111 = 11.11\ \%$$

El anterior valor se encuentra entre el 8 % y el 18 %, por lo tanto, los datos de esta muestra son casi homogéneos.

Ahora, si x representa el número de empresas que obtuvieron por lo menos 10 millones de pesos por día en esta muestra, entonces la proporción o porcentaje muestral es:

$$\hat{p} = \frac{x}{n} = \frac{1}{3} \cong 0.3333 = 33.33\ \%$$

1.5 Se tiene una población conformada por 100 individuos; en ellos interesa estudiar sus gastos semanales en miles de pesos. Se quiere seleccionar una muestra aleatoria de tamaño 5 usando muestreo aleatorio simple sin reemplazo. El número total de muestras distintas que se obtiene es:

$$\binom{100}{5} = \frac{100!}{6!(100-5)!} = \frac{100!}{5!95!} = 75\ 287\ 520$$

La probabilidad de seleccionar una muestra compuesta por cinco individuos determinados es:

$$\frac{1}{75287520} = 0.0000000132$$

La probabilidad de que un individuo cualquiera de la población pertenezca a la muestra es:

$$\frac{5}{100} = 0.05$$

1.6 Se tiene una población conformada por 20 individuos, de los cuales interesa estudiar la relación peso-talla, a fin de implementar una dieta para disminuir su peso. Se quiere seleccionar una muestra aleatoria de tamaño 3 usando muestreo aleatorio simple con reemplazo. El número total de muestras de tamaño 3 con repetición que se pueden obtener es:

$$20^3 = 8\ 000$$

La probabilidad de que una sucesión cualquiera conformada por 3 individuos sea seleccionada es:

$$\frac{1}{20^3} = \frac{1}{8000} = 0.000125$$

La probabilidad de que un individuo específico sea seleccionado, al menos una vez, para conformar la muestra es:

$$1 - \left(\frac{20-1}{20}\right)^3 = 1 - \left(\frac{19}{20}\right)^3 = 1 - \left(\frac{6859}{8000}\right) = 1 - 0.857375 = 0.142625$$

1.7 Se tiene una población de 5000 individuos dividida en 4 estratos, como se indica en la Figura 5.1; se quiere seleccionar una muestra aleatoria de tamaño 100.

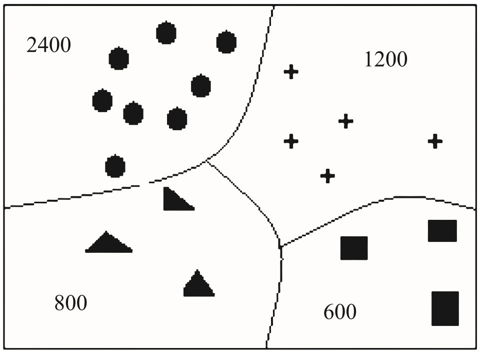

Figura 5.1 Población conformada por cuatro estratos

El tamaño de la población es N =5 000, el tamaño de la muestra es n =100

$$n_1 = \frac{2400}{5000} \star 100 = 48$$

$$n_2 = \frac{1200}{5000} \star 100 = 24$$

$$n_3 = \frac{600}{5000} \star 100 = 12$$

$$n_4 = \frac{800}{5000} \star 100 = 16$$

El tamaño de la muestra, en este caso, se determina de la siguiente forma:

$$n = n_1 + n_2 + n_3 + n_4 = 48+24+12+16=100$$

En este contexto, 48 individuos se han de seleccionar mediante muestreo aleatorio simple del estrato 1; además, 24 individuos del estrato 2, 12 individuos del estrato 3 y 16 individuos del estrato 4.

ii) Ahora, usando muestreo aleatorio no proporcional, o por cuotas iguales, se toma igual número de elementos en cada estrato, usando la expresión:

$$n_i = \frac{n}{k} = \frac{100}{4} = 25$$

Donde k es el número de estratos y n es el tamaño de la muestra.

En este ejemplo se han de seleccionar 25 individuos de cada uno de los cuatro estratos por medio de muestreo aleatorio simple.

1.8 Si se tiene una población con N=2 500 individuos y se desea seleccionar una muestra de n=90 individuos, entonces se ordenan los datos correspondientes a los individuos y se calcula:

$$c = \frac{N}{n} = \frac{2500}{90} = 27.77$$

Ahora el número λ se toma de forma aleatoria como un número menor que 27, y se usará como punto de partida. Si después de realizar un procedimiento aleatorio para elegir el valor de λ, el valor resultante fuera 20, entonces el segundo individuo es aquel en la posición 57, el tercero ocupa la posición 84 y así sucesivamente; el individuo 90 se encuentra en la posición:
20 + (90-1) 27 = 2 423.

Capítulo 2

2.1 Graficar y determinar las siguientes probabilidades utilizando la distribución normal estándar a) $P(Z \leq -1.25)$, b) $P(Z \geq 1.31)$ c) $P(0.29 \leq Z \leq 2.48)$

$a)\ P(Z \leq -1.25) = \Phi(-1.25) = 0.1056 \cong 10.56\ \%$

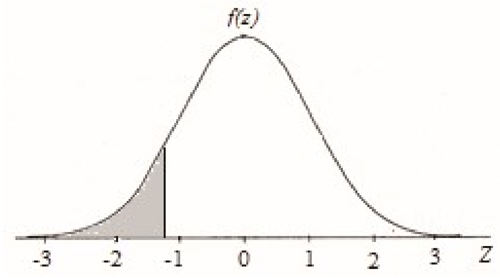

Figura 5.2 Área bajo la curva normal estándar $P(Z \leq -1.25)$

Fuente: los autores con la ayuda del *software* libre R.

La Figura 5.2 muestra el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad.

$b)\ P(Z \geq 1.31) = 1 - P(Z \leq 1.31) = 1 - 0.9049 = 0.0951 \cong 9.51\ \%$

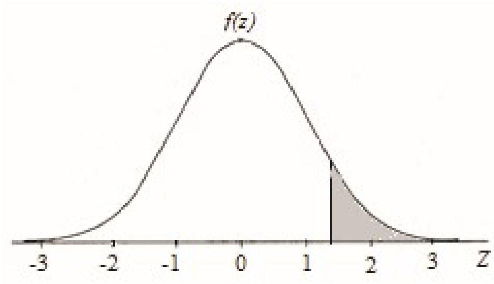

Figura 5.3. Área bajo la curva normal estándar $P(Z \geq 1.31)$

Fuente: los autores con la ayuda del *software* libre R.

La Figura 5.3 muestra el valor del área bajo la curva normal estándar, equivalente al cálculo de la probabilidad.

$$c)\ P(0.29 \leq Z \leq 2.48) = P(Z \leq 2.48) - P(Z \leq 0.29)$$

$$P(0.29 \leq Z \leq 2.48) = 0.9934 - 0.6141 = 0.3793 \cong 37.93\ \%$$

La Figura 5.4 muestra el valor del área bajo la curva normal estándar, equivalente al cálculo de la probabilidad.

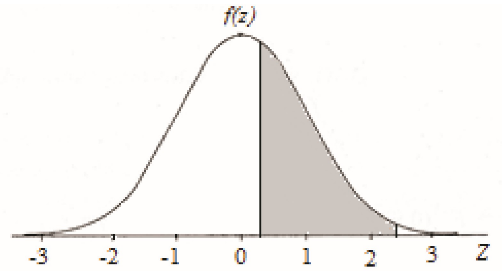

Figura 5.4 Área bajo la curva normal estándar $P(0.29 \leq Z \leq 2.48$

Fuente: los autores con la ayuda del *software* libre R.

2.2 La empresa S&S produce lámparas cuya duración en horas se distribuye normalmente con media de 900 horas y desviación estándar de 50 horas. Si se toma al azar una lámpara de la producción,

a) ¿Cuál es la probabilidad de que dure máximo 940 horas?
b) ¿Cuál es la probabilidad de que dure por lo menos 822 horas?
c) ¿Cuál es la probabilidad de que dure entre 880 y 1020 horas?
d) ¿Cuál es la probabilidad de que dure más de 1200 horas?

$\mu =$ 900 horas, $\sigma =$50 horas

X : duración en horas de una lámpara tomada al azar producida por la empresa S&S.

A) $P(X \leq 940) = P\left(\frac{X-\mu}{\sigma} \leq \frac{940-\mu}{\sigma}\right) = P\left(Z \leq \frac{940-900}{50}\right)$

$$P(X \leq 940) = P\left(Z \leq \frac{40}{50}\right) = P(Z \leq 0.8) = 0.7881 \cong 78.81\ \%$$

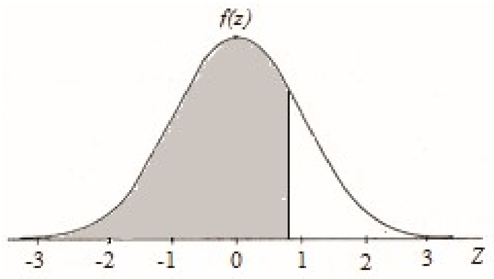

Figura 5.5. Área bajo la curva normal estándar $P(Z \leq 0.8)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que una lámpara escogida al azar de la producción dure máximo 940 horas es del 78.81 %. Una representación de la situación se tiene en la Figura 5.5.

b) $P(X \geq 822) = P\left(\frac{X-\mu}{\sigma} \geq \frac{822-\mu}{\sigma}\right) = P\left(Z \geq \frac{822-900}{50}\right)$

$$P(X \geq 822) = P(Z \geq -1.56) = 1 - P(Z \leq -1.56) = 1 - 0.0594 = 0.9406$$

$$P(X \geq 822) = 0.9406 \cong 94.06\ \%$$

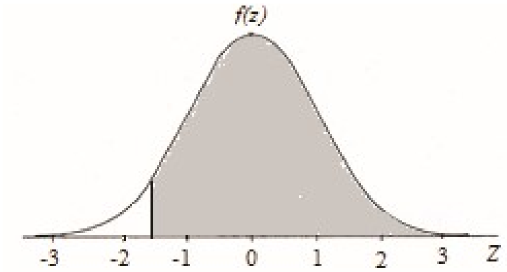

Figura 5.6. Área bajo la curva normal estándar $P(Z \geq -1.56)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que una lámpara escogida al azar de la producción dure por lo menos 822 horas es del 94.06 %. Una representación de la anterior situación se tiene en la Figura 5.6.

c) $P(880 \leq X \leq 1020) = P\left(\dfrac{880-\mu}{\sigma} \leq \dfrac{X-\mu}{\sigma} \leq \dfrac{1020-\mu}{\sigma}\right)$

$$P(880 \leq X \leq 1020) = P\left(\frac{880-900}{50} \leq Z \leq \frac{1020-900}{50}\right) = P(-0.4 \leq Z \leq 2.4)$$

$$P(880 \leq X \leq 1020) = 0.9918 - 0.3446 = 0.6472 \cong 64.72\ \%$$

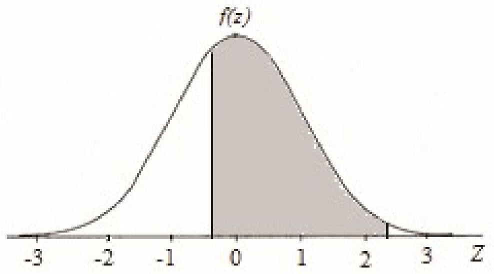

Figura 5.7. Área bajo la curva normal estándar $P(-0.4 \leq Z \leq 2.4)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que una lámpara escogida al azar de la producción dure entre 880 y 1020 horas es de 64.72 %.Ver Figura 5.7.

d) $P(X \geq 1200) = P\left(\dfrac{X-\mu}{\sigma} \geq \dfrac{1200-900}{50}\right) = P(Z \geq 6) \cong 0$

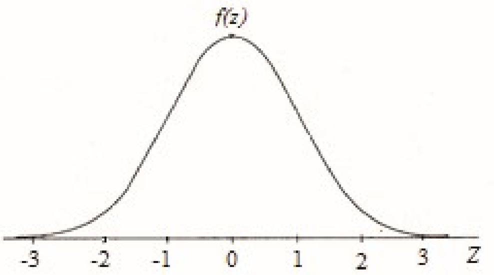

Figura 5.8. Área bajo la curva normal estándar $P(Z \geq 6)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que una lámpara escogida al azar de la producción dure más de 1200 horas es del 0 % aproximadamente.Ver Figura 5.8.

2.3 El peso de los paquetes de arroz empacados por la máquina H&H es una variable aleatoria que se distribuye normalmente con media $\mu = 500$ g, y una desviación estándar $\sigma = 20$ g. Si se escoge aleatoriamente un paquete de arroz empacado por la máquina H&H,

a) ¿Cuál es la probabilidad de que su peso sea de por lo menos 486 gramos?

b) ¿Cuál es la probabilidad de que su peso sea máximo de 480 gramos?

c) ¿Cuál es la probabilidad de que su peso esté entre 476 y 550 gramos?

d) ¿Cuál es la probabilidad de que su peso sea de menos de 400 gramos?

e) ¿Cuál es la probabilidad de que su peso sea de más de 440 gramos?

$\mu = 500$ gramos, $\sigma = 20$ gramos

X : peso del paquete de arroz escogido aleatoriamente.

a) $P(X \geq 486) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{486-500}{20}\right)$

$$P(X \geq 486) = P(Z \geq -0.7) = 1 - P(Z \leq -0.7) = 1 - 0.2420 = 0.758 \cong 75.8\ \%$$

La probabilidad de que el peso del paquete de arroz escogido aleatoriamente sea de por lo menos 486 g es de 75.8 %. Obsérvese la Figura 5.9.

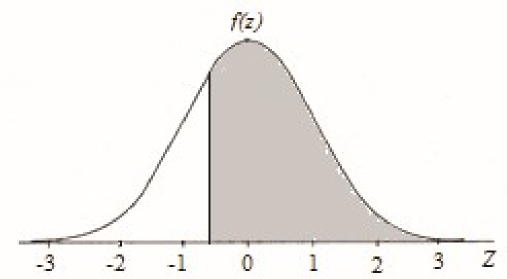

Figura 5.9. Área bajo la curva normal estándar $P(Z \geq -0.7)$

Fuente: los autores con la ayuda del *software* libre R.

b) $P(X \leq 480) = P\left(\frac{X-\mu}{\sigma} \leq \frac{480-500}{20}\right) = P(Z \leq -1) = 0.1587 \cong 15.87\ \%.$

La probabilidad de que el peso del paquete de arroz escogido aleatoriamente sea de máximo 480 g es de 15.87 %.Ver Figura 5.10.

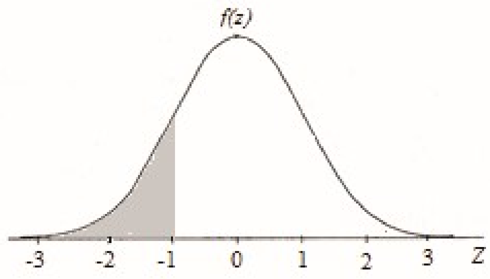

Figura 5.10. Área bajo la curva normal estándar $P(Z \leq -1)$

Fuente: los autores con la ayuda del *software* libre R.

c) $P(476 \leq X \leq 550) = P\left(\frac{476-500}{20} \leq Z \leq \frac{550-500}{20}\right)$

$$P(476 \leq X \leq 550) = P\left(-1.2 \leq Z \leq 2.5\right) = P(Z \leq 2.5) - P(Z \leq -1.2)$$

$$P(476 \leq X \leq 550) = 0.9938 - 0.1151 = 0.8787 \cong 87.87 \text{ \%}.$$

La situación anterior se presenta en la Figura 5.11.

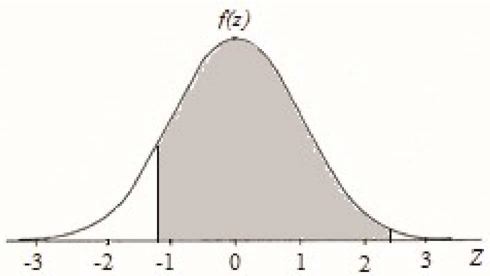

Figura 5.11. Área bajo la curva normal estándar $P(-1.2 \leq Z \leq 2.5)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que el peso del paquete de arroz escogido aleatoriamente esté entre 476 y 550 g es del 87.87 %.

d) $P(X \leq 400) = P\left(\frac{X-\mu}{\sigma} \leq \frac{400-500}{20}\right) = P(Z \leq -5) \cong 0 \text{ \%}$

Esta situación se observa en la Figura 5.12.

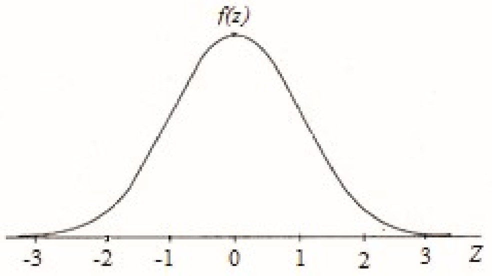

Figura 5.12. Área bajo la curva normal estándar $P(Z \leq -5)$

Fuente: los autores con la ayuda del *software* libre R.

La probabilidad de que el peso del paquete de arroz escogido aleatoriamente sea de menos de 400 g es de 0 %, aproximadamente.

e) $P(X \geq 440) = P\left(\frac{X - \mu}{\sigma} \geq \frac{440 - 500}{20}\right) = P(Z \geq -3) = 1 - P(Z \leq -3)$

$P(X \geq 440) = 1 - 0.0013 = 0.9987 \cong 99.87\ \%$

La probabilidad de que el peso del paquete de arroz escogido aleatoriamente sea de más de 440 g es de 99.87 %.Ver Figura 5.13.

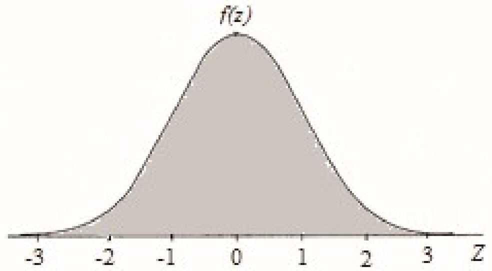

Figura 5.13. Área bajo la curva normal estándar $P(Z \geq -3)$

Fuente: los autores con la ayuda del *software* libre R.

2.4 Determinar los siguientes valores correspondientes a las probabilidades indicadas y elaborar una representación gráfica:

a) $\chi^2_{0.99,17} = ?$

b) $\chi^2_{0.025,12} = ?$

Para la situación *a)* se debe determinar la distancia *chi-cuadrado* correspondiente a una probabilidad de 0.99 con *n*=17 grados de libertad; esta corresponde al valor 33.4087, es decir, $\chi^2_{0.99,17} = 33.4087$; una representación gráfica se indica en la Figura 5.14.

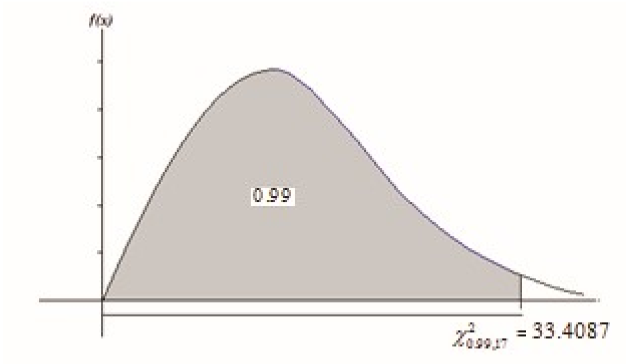

Figura 5.14 Área bajo la curva chi-cuadrado con n=17 grados de libertad

Fuente: los autores con la ayuda del *software* libre R

Para la situación *b)* se debe determinar la distancia *chi-cuadrado* correspondiente a una probabilidad de 0.025 con n=12 grados de libertad; esta corresponde al valor 4.404; es decir,

$\chi^2_{0.025,12} = 4.404$; la representación gráfica se observa en la Figura 5.15.

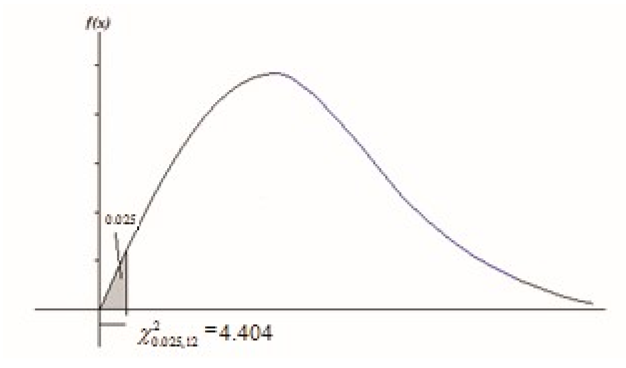

Figura 5.15 Área bajo la curva chi-cuadrado con n = 12 grados de libertad

Fuente: los autores con la ayuda del *software* libre R

2.5 Obtener las probabilidades y graficar:

a) $t_{0.99,16} = ?$

b) $t_{0.05,13} = ?$

Para el caso *a)* se debe determinar un valor en eje horizontal correspondiente a una probabilidad de 0.99 con n=16 grados de libertad, usando la distribución *t-student;* este corresponde a 2.583, es decir, $t_{0.99,16} = 2.583$; una representación gráfica se observa en la Figura 5.16.

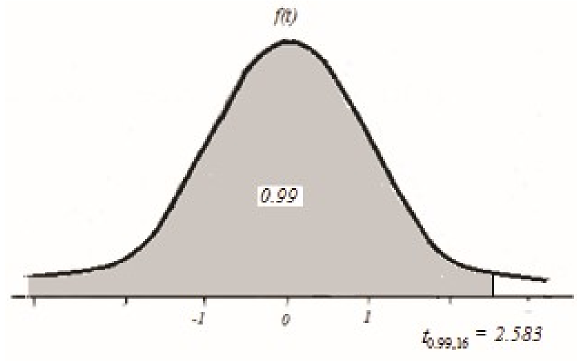

Figura 5.16 Área bajo la curva t-student con n = 16 grados de libertad

Fuente: los autores con la ayuda del software libre R

Para el caso *b)* se debe determinar un valor en eje horizontal correspondiente a una probabilidad de 0.05 con n=13 grados de libertad, usando la distribución *t-student;* este corresponde a 1.771; es decir, $t_{0.05,13} = -1.771$; una representación gráfica se observa en la Figura 5.17.

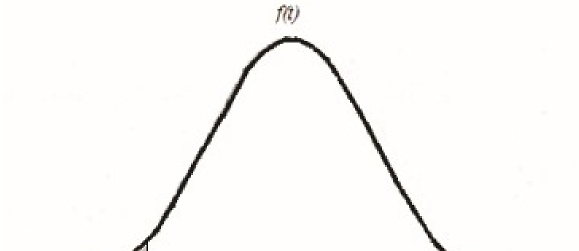

Figura 5.17 Área bajo la curva t-student con n=13 grados de libertad

Fuente: los autores con la ayuda del software libre R

2.6 Leer el valor de probabilidad en la distribución de Fisher y elaborar un gráfico:

a) $F_{0.99,12,10} = ?$

b) $F_{0.05,7,7} = ?$

Para el caso *a)* se debe determinar un valor en eje horizontal correspondiente a una probabilidad de 0.99 con m=12 y n=10 grados de libertad, usando la distribución de *Fisher;* este corresponde a 4.71, es decir, $F_{0.99,12,10} = 4.71$; la representación gráfica se indica en la Figura 5.18.

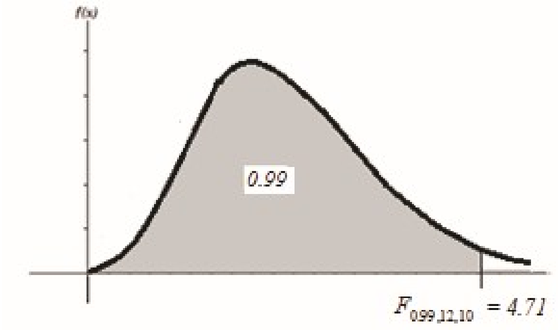

Figura 5.18 Área bajo la curva de Fisher con *m*= 12 y *n*=10 grados de libertad

Fuente: los autores con la ayuda del software libre R.

Para la situación *b)* se debe determinar un valor en eje horizontal correspondiente a una probabilidad de 0.05 con *m*=7 y *n*=7 grados de libertad, usando la distribución de *Fisher;* este corresponde a 0.2638 y se obtiene así:

$F_{0.05,7,7} = \frac{1}{F_{0.95,7,7}} = \frac{1}{3.79} \cong 0.2638$; la representación gráfica se presenta en la Figura 5.19.

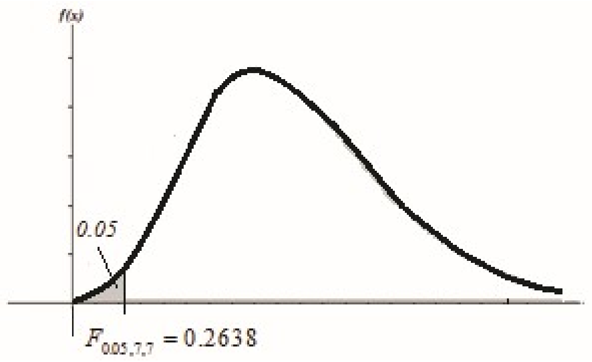

Figura 5.19 Área bajo la curva de Fisher con *m* =7 y *n* = 7 grados de libertad

Fuente: los autores con la ayuda del software libre R.

2.7 Para la variable *X: g*astos en transporte por día (miles de pesos) de los cinco mejores estudiantes de la universidad B, en la ciudad de Tunja en Boyacá-Colombia, en el año 2015, cuyos valores son: 8, 10, 14, 12, 6; *a)* determinar todas las muestras de tamaño *n*=2 que son posibles de seleccionar por medio de un muestreo aleatorio simple sin reemplazo, *b)* calcular el promedio para cada muestra, *c)* construir la distribución de probabilidad para la variable aleatoria "media muestral", *d)* obtener su valor esperado y su varianza, *e)* verificar si se cumplen las igualdades para la esperanza matemática y la varianza presentadas por Freund y Miller (2000).

a) Como se tiene que N=5 y n=2, entonces, en este caso también el número total de muestras distintas por obtener es:

$$\binom{N}{n}=\binom{5}{2}=\frac{5!}{2!(5-2)!}=\frac{5!}{2!3!}=\frac{120}{12}=10$$

Estas muestras son las siguientes:

$$M_1=\{14,12\},M_2=\{14,10\},M_3=\{14,8\},M_4=\{14,6\},M_5=\{12,10\},$$

$$M_6=\{12,8\},M_7=\{12,6\},M_8=\{10,8\},M_9=\{10,6\},M_{10}=\{8,6\}$$

b) Las correspondientes medias muestrales son:

$$\bar{X}_1=\frac{14+12}{2}=13,\bar{X}_2=\frac{14+10}{2}=12,\bar{X}_3=\frac{14+8}{2}=11,\bar{X}_4=\frac{14+6}{2}=10,\bar{X}_5=\frac{12+10}{2}=11$$

$$\bar{X}_6=\frac{12+8}{2}=10,\bar{X}_7=\frac{12+6}{2}=9,\bar{X}_8=\frac{10+8}{2}=9,\bar{X}_9=\frac{10+6}{2}=8,\bar{X}_{10}\frac{8+6}{2}=7$$

Los valores anteriores indican que la media muestral es una variable aleatoria; esta cambia de valor a medida que se cambia de muestra.

c) De acuerdo con lo establecido en el apartado 1.3 del primer capítulo, la función de probabilidad asociada a esta variable aleatoria es:

$$f\left(\bar{X}=13\right)=\frac{1}{10},\ f\left(\bar{X}=12\right)=\frac{1}{10},\ f\left(\bar{X}=11\right)=\frac{2}{10},\ f\left(\bar{X}=10\right)=\frac{2}{10},$$

$$f\left(\bar{X}=9\right)=\frac{2}{10},\ f\left(\bar{X}=8\right)=\frac{1}{10},\ f\left(\bar{X}=7\right)=\frac{1}{10}$$

d) El valor esperado o esperanza matemática de la variable aleatoria es:

$$E\left(\bar{X}\right)=13\left(\frac{1}{10}\right)+12\left(\frac{1}{10}\right)+11\left(\frac{2}{10}\right)+10\left(\frac{2}{10}\right)+9\left(\frac{2}{10}\right)+8\left(\frac{1}{10}\right)+7\left(\frac{1}{10}\right)$$

$$E(\bar{X})=\frac{13+12+22+20+18+8+7}{10}=\frac{100}{10}=10$$

Ahora, para calcular la varianza, en primera instancia se calcula:

$$E(\bar{X}^2)=13^2\left(\frac{1}{10}\right)+12^2\left(\frac{1}{10}\right)+11^2\left(\frac{2}{10}\right)+10^2\left(\frac{2}{10}\right)+9^2\left(\frac{2}{10}\right)+8^2\left(\frac{1}{10}\right)+7^2\left(\frac{1}{10}\right)$$

$$E(\bar{X}^2)=\frac{169+144+2(121)+2(100)+2(81)+64+49}{10}$$

$$E(\bar{X}^2)=\frac{169+144+242+200+162+64+49}{10}=\frac{1030}{10}=103$$

Con los resultados anteriores se calcula la varianza así:

$$Var(\bar{X})=E(\bar{X}^2)-(E(\bar{X}))^2=103-(10)^2=103-100=3$$

Usando los cinco datos de la población se establecen los parámetros:

$$\mu=\frac{14+12+10+8+6}{5}=\frac{50}{5}=10$$

$$\sigma^2=\frac{(14-10)^2+(12-10)^2+(10-10)^2+(8-10)^2+(6-10)}{5}$$

$$\sigma^2=\frac{4^2+2^2+0^2+(-2)^2+(-4)^2}{5}=\frac{16+4+0+4+16}{5}=\frac{40}{5}=8$$

e) Por otro lado, en concordancia con Freund y Miller (2000), se cumple que:

$$E(\bar{X})=10=\mu$$

$$Var(\bar{X})=\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)=\frac{8}{2}\left(\frac{5-2}{5-1}\right)=\frac{24}{8}=3$$

2.8 La duración promedio de cierta marca de teclados es de 900 días, con una desviación estándar de 70 días, siempre que se usen 8 horas por día. Determinar la probabilidad de que una muestra aleatoria de 36 teclados tenga una duración promedio: *a)* comprendida entre 870 y 925 días, *b)* menor o igual a 910 días. En este caso se tiene:

$\mu=900,\ \sigma=70,\ n=36$

Se define la variable:

$\overline{X}$: duración promedio en días de los teclados en la muestra

Así, entonces, para el caso *a)* se ha de usar la distribución de la media muestral, expresión 2.1, porque se conoce la desviación estándar poblacional:

$$P(870 \leq \overline{X} \leq 925) = P\left(\frac{870-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{925-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{870-900}{\frac{70}{\sqrt{36}}} \leq Z \leq \frac{925-900}{\frac{70}{\sqrt{36}}}\right)$$

$$P(870 \leq \overline{X} \leq 925) = P\left(-2.57 \leq Z \leq 2.14\right) = P\left(Z \leq 2.14\right) - \left(Z \leq -2.57\right)$$

$$P(870 \leq \overline{X} \leq 925) = 0.9838 - 0.0051 = 0.9787 \cong 97.87 \ \%$$

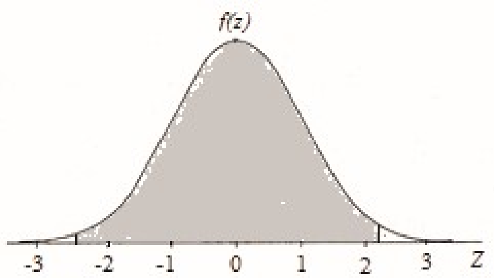

Figura 5.20. Área bajo la curva normal estándar $P\left(-2.57 \leq Z \leq 2.14\right)$

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que en una muestra aleatoria de 36 teclados se obtenga una duración promedio entre 870 y 925 días es el 97.87 %, aproximadamente. Ver Figura 5.20.

Para el caso *b)* se tiene,

$$P(\overline{X} \leq 910) = P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{910-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z \leq \frac{910-900}{\frac{70}{\sqrt{36}}}\right)$$

$$P(\overline{X} \leq 910) = P(Z \leq 0.857) \cong P(Z \leq 0.86) = 0.8051 \cong 80.51 \ \%$$

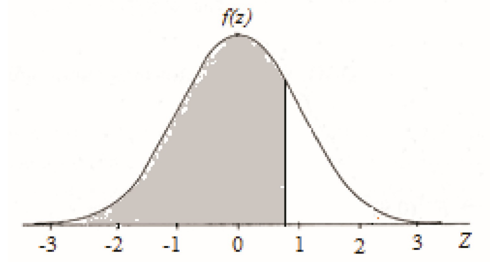

Figura 5.21. Área bajo la curva normal estándar $P(Z \leq 0.857)$

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que una muestra aleatoria de 36 teclados tenga una duración promedio menor o igual a 910 días es del 80.51 %, aproximadamente. Ver Figura 5.21.

2.9 El tiempo promedio que gasta el bus urbano en la ciudad de Cali es de 70 minutos. Se toma una muestra aleatoria de 12 recorridos; con esos datos se obtuvo una desviación estándar corregida de 8 minutos. ¿Cuál es la probabilidad de que en esa muestra se tenga un tiempo promedio entre 64.94 y 76.84 minutos?

En este ejemplo se tiene:

$$\mu = 70,\ \hat{S} = 8,\ \ n = 12$$

Se define la variable:

$\overline{X}$: tiempo promedio en minutos que gasta el bus urbano en la muestra.

Así, entonces, se ha de usar una distribución de la media muestral correspondiente a la expresión 2.3, porque se desconoce la desviación estándar poblacional,

$$t = \frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}} \text{ con } n\ \text{-}1 = 12 \text{ -}1 = 11 \text{ grados de libertad}$$

En efecto,

$$P(64.94 \leq \overline{X} \leq 96.84) = P\left(\frac{64.94 - 70}{\frac{8}{\sqrt{12}}} \leq \frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}} \leq \frac{76.84 - 70}{\frac{8}{\sqrt{12}}} \right)$$

$$P(64.94 \leq \overline{X} \leq 96.84) = P(-2.19 \leq t \leq 2.96)$$

$$P(64.94 \leq \overline{X} \leq 96.84) = P(t \leq 2.96) - P(t \leq -2.19) = 0.9935 - 0.0255 = 0.968 \cong 96.8\ \%$$

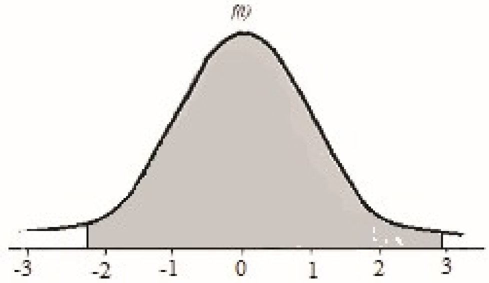

Figura 2.22. Área bajo la curva t-student con n = 11 grados de libertad

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que en esa muestra se obtenga un tiempo promedio de recorrido para el bus urbano entre 64.94 y 76.84 minutos es del 96.8 %. Ver Figura 2.22. En este caso, los valores de probabilidad fueron obtenidos por medio del software libre R.

2.10 El 4 % de los artículos que produce una máquina son defectuosos; se toma una muestra aleatoria de 400 artículos. ¿Cuál es la probabilidad de que más del 5 % de los artículos de la muestra sean defectuosos?

Los datos asociados al problema planteado son:

n=400
p=4 % = 0.04
q=1 – p = 1 – 0.04 = 0.96

Se define la variable:

$\hat{p}$: porcentaje de artículos defectuosos producidos por la máquina en la muestra.

En este problema se ha de utilizar la distribución de la proporción muestral, que sigue una distribución normal estándar cuando el tamaño de la muestra es grande; esta se indicó en la expresión 2.5.

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

En efecto,

$$P(\hat{p} \geq 0.05) = P\left(\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \geq \frac{0.05-p}{\sqrt{\frac{pq}{n}}}\right) = P\left(Z \geq \frac{0.05-0.04}{\sqrt{\frac{0.04*0.96}{400}}}\right)$$

$$P(\hat{p} \geq 0.05) = P(Z \geq 1.02) = 1 - p(Z \leq 1.02) = 1 - 0.8461 = 0.1539 \cong 15.93\ \%$$

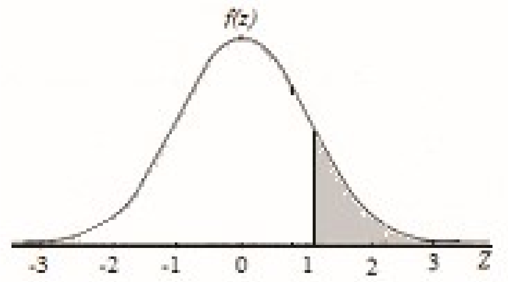

Figura 5.23. Área bajo la curva normal estándar $P(Z \geq 1.02)$

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que el porcentaje de artículos defectuosos encontrados en la muestra supere el 5 % es del 15.39 %.Ver Figura 5.23.

2.11 En la ciudad A para niños de grado quinto de educación básica primaria se tiene un peso promedio $\mu_2 = 35\ kg$, con $\sigma_2^2 = 5$; mientras que para la ciudad B para niños que cursan este grado se tiene un peso promedio $\mu_1 = 45\ kg$, con $\sigma_1^2 = 8$; se toma en la ciudad A una muestra aleatoria de n_2=50, y otra en la ciudad B de n_1=60. ¿Cuál es la probabilidad de que la media muestral del peso de los niños de la ciudad B difiera de la de los niños de la ciudad A en más de 11 kg?

Los datos relacionados con la anterior situación son:

Ciudad A	Ciudad B
$\mu_2 = 35\ kg$	$\mu_1 = 45\ kg$
$\sigma_2^2 = 5$	$\sigma_1^2 = 8$
$n_2 = 50$	$n_1 = 60$

Se define la variable.

$\overline{X}_1 - \overline{X}_2$: diferencia de medias muestrales asociadas con el peso promedio de los niños en las ciudades B y A, respectivamente, en sus correspondientes muestras.

En este caso se ha de utilizar la distribución de la diferencia de medias muestrales establecida en la expresión 2.7.

$$Z=\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Luego, entonces,

$$P(\overline{X}_1-\overline{X}_2\geq 11)=P\left(\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}}}\geq\frac{11-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}}}\right)$$

$$P(\overline{X}_1-\overline{X}_2\geq 11)=P\left(Z\geq\frac{11-(45-35)}{\sqrt{\frac{8}{60}+\frac{5}{50}}}\right)=P\left(Z\geq\frac{1}{\sqrt{0.233}}\right)$$

$$P(\overline{X}_1-\overline{X}_2\geq 11)=P(Z\geq 2.07)=1-P(Z\leq 2.07)=1-0.9808=0.0192\cong 1.92\ \%$$

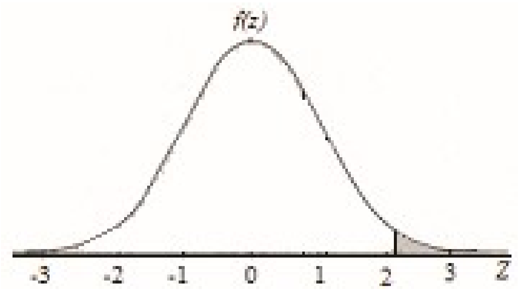

Figura 5.24. Área bajo la curva normal estándar $P(Z\geq 2.07)$

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que la media muestral del peso de los niños de la ciudad B difiera en más de 11 kg de la media muestral del peso de los niños de la ciudad A es de 1.92 %.Ver Figura 5.24.

2.12 Un candidato a la presidencia de la república tiene el 60 % de la intención de voto en los potenciales votantes en el departamento de Nariño, y el 58 % en los del Valle del Cauca; se toma una muestra aleatoria de 400 votantes en

Nariño y de 500 en el Valle del Cauca. ¿Cuál es la probabilidad que la diferencia entre las proporciones muestrales de los potenciales votantes en Nariño y el Valle no superen el 3 %?

Los datos asociados con el anterior problema son:

Población 1: Nariño	Población 2: Valle
$p_1 = 60\ \% = 0.6$	$p_2 = 58\ \% = 0.58$
$n_1 = 400$	$n_2 = 500$

Se define la variable,

$\hat{p}_1 - \hat{p}_2$: diferencia de proporciones muestrales asociadas con la intención de voto de los potenciales votantes en las muestras de los departamentos de Nariño y Valle del Cauca, respectivamente.

En este problema se ha de usar la distribución de la diferencia de proporciones muestrales indicada en la expresión 2.6.

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Por lo tanto,

$$P(\hat{p}_1 - \hat{p}_2 \leq 0.03) = P\left(\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq \frac{0.03 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \right)$$

$$P(\hat{p}_1 - \hat{p}_2 \leq 0.03) = P\left(Z \leq \frac{0.03 - (0.6 - 0.58)}{\sqrt{\frac{0.6 * 0.4}{400} + \frac{0.58 * 0.42}{500}}} \right)$$

$$P(\hat{p}_1 - \hat{p}_2 \leq 0.03) = P(Z \leq 0.3032) = 0.6179 \cong 61.79\%$$

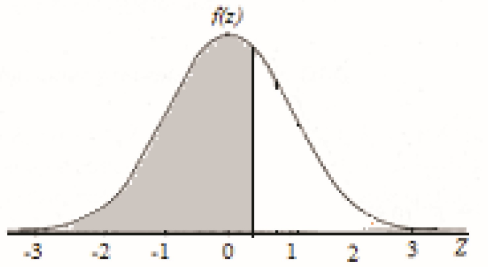

Figura 5.25. Área bajo la curva normal estándar $P(Z \leq 0.3032)$

Fuente: los autores con la ayuda del software libre R.

La probabilidad de que la diferencia entre las proporciones muestrales de los potenciales votantes por el candidato a la presidencia de la república en Nariño y Valle del Cauca no supere el 3 % es del 61.89 %, aproximadamente. Ver Figura 5.25.

Capítulo 3

3.1 Para la variable aleatoria X con distribución exponencial, determinar el estimador máximo verosímil. La función de densidad de probabilidad está dada por:

$$f(x,\lambda)=\begin{cases}\lambda e^{-\lambda x} & \text{si } x \geq 0 \quad \text{con } \lambda > 0 \\ 0 & \text{si } x < 0\end{cases}$$

Donde $\lambda = \theta$ es el parámetro de la variable aleatoria X con modelo exponencial.

Se toma una muestra aleatoria $X_1, X_2, ..., X_n$; la función de verosimilitud es:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

Aquí los x_i son observaciones correspondientes a las variables aleatorias $X_1, X_2, ..., X_n$.

Se trata de encontrar un valor del parámetro de tal manera que se maximice la función de verosimilitud,

$$L(\lambda) = \left(\lambda e^{-\lambda x_1}\right).\left(\lambda e^{-\lambda x_2}\right)....\left(\lambda e^{-\lambda x_n}\right)$$

$$L(\lambda) = \lambda^n e^{-\lambda(x_1 + x_2 + ... + x_n)}$$

Ahora, aplicando el logaritmo natural resulta:

$$Ln\big(L(\lambda)\big) = nLn(\lambda) - \lambda(x_1 + x_2 + \ldots + x_n)$$

A continuación, se realiza el cálculo de la derivada parcial con respecto al parámetro λ.

$$\frac{\partial Ln\big(L(\lambda)\big)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

Igualando a cero, resulta:

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

Por lo tanto,

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Esta última expresión corresponde al estimador de máxima verosimilitud del parámetro λ de una variable aleatoria con distribución exponencial.

3.2 En la empresa azucarera AA, la cantidad de azúcar depositada por la máquina M en cada uno de los paquetes se distribuye normalmente, con desviación estándar de 2 g; de un lote de 500 paquetes se toma una muestra aleatoria de 25 paquetes (bolsas) empacados por tal máquina, y se encuentra un contenido promedio de 2500 g. Construir un intervalo de confianza del 98 % para estimar la verdadera media de empacado en las bolsas en el lote que se haga a través de la máquina M.

En este caso, se tiene:

$$N = 500$$
$$n = 25$$
$$\bar{X} = 2500$$
$$\sigma = 2$$

Además,

$$1-\alpha = 0.98 \rightarrow \alpha = 0.02$$
$$\frac{\alpha}{2} = 0.01$$

En la Figura 5.26 se observa el valor de Z obtenido al leer una tabla normal estándar para un 99 % de probabilidad.

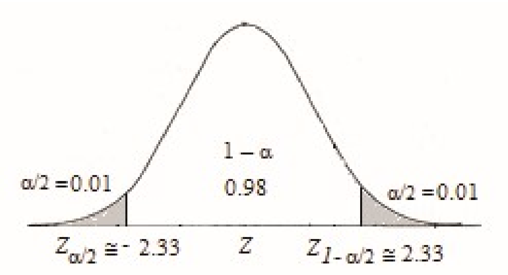

Figura 5.26 Valor de Z en una curva normal estándar
Fuente: los autores con la ayuda del software libre R.

$$Z_{1-\frac{\alpha}{2}} = Z_{0.99} \cong 2.33$$

Luego el intervalo de confianza es:

$$\mu \in \left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right)$$

Reemplazando los valores, resulta:

$$\mu \in \left(2500 - 2.33(\frac{2}{\sqrt{25}}) \sqrt{\frac{500-25}{500-1}}, 2500 + 2.33(\frac{2}{25}) \sqrt{\frac{500-25}{500-1}} \right)$$

Realizando las operaciones de tipo aritmético se obtiene:

$$\mu \in \left(2500 - 2.33(0.4)(0.97565), 2500 + 2.33(0.4)(0.97565) \right)$$

Luego

$$\mu \in \left(2500 - 0.9093, 2500 + 0.9093 \right)$$

Finalmente,

$$\mu \in \left(2499.09, 2500.91 \right)$$

En conclusión, con un nivel de confianza del 98 % se infiere que el promedio poblacional de empacado de las bolsas de azúcar por la máquina M está ente 2499.09 g y 2500.91 g, aproximadamente. Lo anterior indica que en 98 de cada 100 muestras tomadas se encuentra el parámetro media poblacional.

3.3 De un lote de 200 unidades del producto LM se ha seleccionado una muestra aleatoria de 10 unidades, obteniéndose un peso neto promedio de 1000 g, con una desviación estándar corregida de 5 g; si se asume que los datos provienen de una distribución normal, determinar un intervalo de confianza del 95 % para estimar el verdadero peso promedio de las unidades del producto en la población.

$$\hat{S} = 5 \qquad 1-\alpha = 95\,\% = 0.95$$
$$n = 10 \qquad \alpha = 0.05$$
$$\bar{X} = 1000 \qquad \frac{\alpha}{2} = 0.025$$

En la Figura 5.27 se observa el valor de t obtenido al leer una tabla de la distribución *t-student* con $n - 1 = 10 - 1 = 9$ grados de libertad para un 0.975=97.5 % de probabilidad.

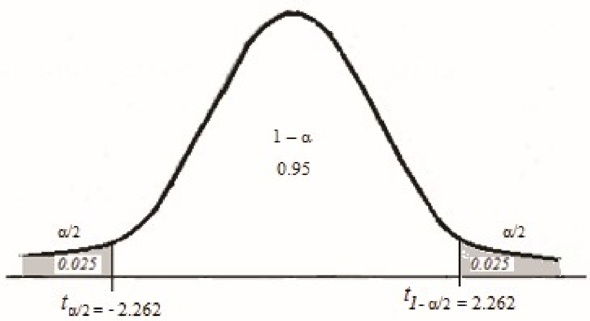

Figura 5.27 Valor de t en una curva t-student con n=9 grados

Fuente: los autores con la ayuda del software libre R.

$$t_{1-\frac{\alpha}{2}} = t_{0.975,9} = 2.262$$

Así, el intervalo de confianza es:

$$\mu \in \left(\bar{X} - t_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{X} + t_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right)$$

Al reemplazar los valores se tiene:

$$\mu \in \left(1000 - 2.262(\frac{5}{\sqrt{10}})\sqrt{\frac{200-10}{200-1}}, 1000 + 2.262(\frac{5}{\sqrt{10}})\sqrt{\frac{200-10}{200-1}}\right)$$

Haciendo las operaciones de tipo aritmético se obtiene:

$$\mu \in (1000 - 2.262(1.5811)(0.9771), 1000 + 2.262(1.5811)(0.9771))$$

Entonces,

$$\mu \in (1000 - 3.4945, 1000 + 3.4945)$$

Por consiguiente,

$$\mu \in (996.50, 1003.49)$$

En conclusión, con un nivel de confianza del 95 % se infiere que el peso promedio poblacional de las unidades del producto LM está entre 996.50 g y 1003.49 g, aproximadamente. Lo anterior indica que en 95 muestras, de cada 100 tomadas, se encuentra el parámetro μ .

3.4 En el departamento de Boyacá, Colombia, se tomó una muestra aleatoria de 500 ciudadanos y se les preguntó si pertenecen o no a la población económicamente activa de este departamento; 350 de los encuestados respondieron que sí pertenecen a esta población. Construir un intervalo de confianza del 99 % para estimar la verdadera proporción de ciudadanos que pertenecen a la población económicamente activa de este departamento.

En primera instancia, se define así *X:* número de ciudadanos en la muestra de 500 que sí pertenecen a la población económicamente activa, luego:

$$n = 500$$

$$\hat{p} = \frac{x}{n} = \frac{350}{500} = 0.7$$

$$\hat{q} = 1 - 0.7 = 0.3$$

Adicionalmente,

$$1 - \alpha = 0.99 \rightarrow \alpha = 0.01$$

$$\frac{\alpha}{2} = 0.005$$

En la Figura 5.28 se observa el valor de Z obtenido al leer una tabla normal estándar para un 0.995 = 99.5 % de probabilidad.

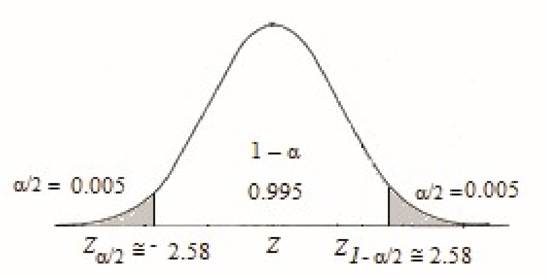

Figura 5.28 Valor de Z en una curva normal estándar

Fuente: los autores con la ayuda del software libre R.

$$Z_{1-\frac{\alpha}{2}} = Z_{0.995} \cong 2.58$$

Luego el intervalo de confianza es:

$$p \in \left(\hat{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \ ,\ \hat{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

Reemplazando los valores resulta:

$$p \in \left(0.7 - 2.58\sqrt{\frac{0.7*0.3}{500}}, 0.7 + 2.58\sqrt{\frac{0.7*0.3}{500}} \right)$$

Realizando las operaciones de tipo aritmético se obtiene:

$$p \in \left(0.7 - 2.58(0.02049), 0.7 + 2.58(0.02049)\right)$$

Luego

$$p \in \left(0.7 - 0.05286, 0.7 + 0.05286\right)$$

En consecuencia,

$$p \in \left(0.6471, 0.7529\right)$$

En conclusión, con un nivel de confianza del 99 % se establece que la verdadera proporción de ciudadanos del departamento de Boyacá se encuentra, aproximadamente, entre el 64.71 % y el 75.29 %. Lo anterior indica que en 99 de cada 100 muestras tomadas se encuentra el parámetro p .

3.5 En un centro de distribución de computadores se ofrecen computadores de dos marcas diferentes en un periodo de tiempo específico; se selecciona aleatoriamente un mes y se encuentra que se venden 350 computadores, de un total de 500, de la marca A, y 333, de un total de 450, de la marca B. Determinar un intervalo de confianza del 98 % para estimar la diferencia entre las verdaderas proporciones de las marcas A y B de computadores que se venden en todo el mercado en ese mes.

En este caso se tiene:

Marca A	Marca B
$n_1 = 500, \quad x_1 = 350$	$n_2 = 450, x_2 = 333$
$\hat{p}_1 = \frac{x_1}{n_1} = \frac{350}{500} = 0.7$	$\hat{p}_2 = \frac{x_2}{n_2} = \frac{333}{450} = 0.74$
$\hat{q}_1 = 1 - 0.7 = 0.3$	$\hat{q}_2 = 1 - 0.74 = 0.26$

Además,

$$1 - \alpha = 0.98 \rightarrow \alpha = 0.02$$

$$\frac{\alpha}{2} = 0.01$$

En la Figura 5.29 se observa el valor de Z obtenido al leer una tabla normal estándar para un 0.99 = 99 % de probabilidad.

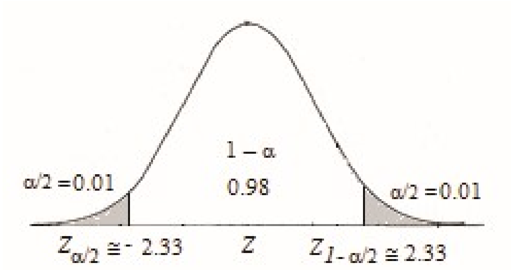

Figura 5.29 Valor de Z en una curva normal estándar

Fuente: los autores con la ayuda del software libre R.

$$Z_{1-\frac{\alpha}{2}} = Z_{0.99} \cong 2.33$$

Luego, en concordancia con la expresión 3.8, el intervalo de confianza es:

$$p_1 - p_2 \in \left((\hat{p}_1 - \hat{p}_2) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \;,\; (\hat{p}_1 - \hat{p}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

Reemplazando los valores, resulta:

$$\left((0.7-0.74)-2.33\sqrt{\frac{(0.7)(0.3)}{500}+\frac{(0.74)(0.26)}{450}}\ ,\ (0.7-0.74)+2.33\sqrt{\frac{(0.7)(0.3)}{500}+\frac{(0.74)(0.26)}{450}}\right)$$

Realizando las operaciones de tipo aritmético se obtiene:

$$p_1-p_2\in\left((0.7-0.74)-2.33\sqrt{0.00042+0.000427}\ ,\ (0.7-0.74)+2.33\sqrt{0.00042+0.000427}\ \right)$$

Luego

$$p_1-p_2\in\left((0.7-0.74)-2.33(0.0291)\ ,\ (0.7-0.74)+2.33(0.0291)\right)$$

En consecuencia,

$$p_1-p_2\in\left(-\,0.04-0.0678\ ,\ \text{-}0.04+0.0678\right)$$

Finalmente,

$$p_1-p_2\in\left(-\,0.1078,\ 0.0278\right)$$

En conclusión, con un nivel de confianza del 98 % se infiere que la verdadera diferencia de proporciones poblacionales se encuentra entre el -10.78 % y el 2.78 %, aproximadamente. Lo anterior indica que en 98 de cada 100 muestras tomadas se encuentra el parámetro p_1-p_2. La verdadera diferencia de proporciones de las marcas A y B de computadores que se venden en todo el mercado en ese mes está entre los porcentajes indicados.

3.6 Se analiza el contenido de oro presente en una aleación. En una muestra de 40 circuitos integrados especiales se encontró un contenido medio de 5.8 u.i de oro, con una desviación estándar de 0.6 u.i; asimismo, se inspecciona el contenido de oro en otra muestra aleatoria de 50 circuitos integrados corrientes, detectándose un contenido promedio de 5 u.i, con una desviación estándar de 0.8 u.i; se supone que las muestra provienen de poblaciones normales. Construir un intervalo de confianza del 95 % para estimar la diferencia de contenidos medios de oro de la primera clase de circuito con respecto a la segunda.

En este caso se tiene:

Clase 1	Clase 2
$n_1=40$	$n_2=50$
$\bar{X}_1=5.8$	$\bar{X}_2=5$
$S_1=0.6$	$S_2=0.8$

El valor de la desviación estándar corregida para la clase 1 "circuitos integrados especiales" se obtiene de la siguiente manera:

$$\hat{S}_1^2 = \frac{n_1}{n_1 - 1} S_1^2 = \frac{40}{39}(0.6)^2 = 0.3692 \quad \rightarrow \quad \hat{S}_1 = 0.6076$$

De modo similar,

$$\hat{S}_2^2 = \frac{n_2}{n_2 - 1} S_2^2 = \frac{50}{49}(0.8)^2 = 0.6530 \quad \rightarrow \quad \hat{S}_2 = 0.8080$$

Además,

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

En la Figura 5.30 se observa el valor aproximado de Z, obtenido al leer una tabla normal estándar para una probabilidad acumulada de 0.975=97.5 %. En este caso, para obtener el intervalo de confianza se recurre a la expresión 3.12.

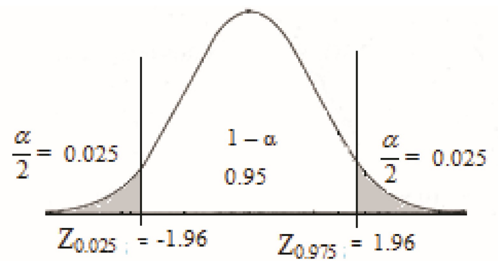

Figura 5.30 Intervalo de confianza sobre la curva normal estándar

Fuente: los autores con la ayuda del *software* libre R.

$$Z_{0.975} = 1.96$$

Luego, usando:

$$\mu_1 - \mu_2 \in \left((\bar{X}_1 - \bar{X}_2) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}} \ ,\ (\bar{X}_1 - \bar{X}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}} \right)$$

Y reemplazando los valores, resulta:

$$\mu_1 - \mu_2 \in \left((5.8 - 5) - 1.96\sqrt{\frac{0.3692}{40} + \frac{0.6530}{50}} \ ,\ (5.8 - 5) + 1.96\sqrt{\frac{0.3692}{40} + \frac{0.6530}{50}} \right)$$

Efectuando las operaciones de tipo aritmético se obtiene:

$$\mu_1 - \mu_2 \in \left((5.8-5)-1.96(0.149298),(5.8-5)+1.96(0.149298)\right)$$

Luego

$$\mu_1 - \mu_2 \in (0.8-0.292624, 0.8+0.292624)$$

Finalmente,

$$\mu_1 - \mu_2 \in (0.5073\ ,\ 1.0926)$$

En conclusión, con un nivel de confianza del 95 % se infiere que la diferencia de medias poblacionales referidas al contenido de oro presente en cada una de las aleaciones de la primera clase de circuitos integrados con respecto a la segunda clase está entre 0.5073 y 1.0926 u.i. Lo anterior indica que en 95 de cada 100 muestras seleccionadas se encuentra el parámetro $\mu_1 - \mu_2$. Se deduce que el contenido promedio de oro en la segunda clase de circuitos es mayor que el contenido promedio de oro en la primera clase de circuitos.

3.7 El siguiente ejemplo ha sido adaptado de Canavos (1988). Un determinado procedimiento produce cierta clase de cojinetes de bola, cuyo diámetro interior es de 5 cm; se selecciona una muestra aleatoria de 12 de esos cojinetes; al medir sus diámetros internos se obtiene una desviación estándar corregida 0.03 cm. Bajo el supuesto de normalidad para los datos, construir un intervalo de confianza del 99 % para estimar la varianza poblacional; asimismo, determinar un intervalo de confianza para estimar la desviación estándar poblacional.

Los requerimientos son los siguientes:

$$\hat{S} = 0.03$$

$$n = 12$$

$$1-\alpha = 0.99 \rightarrow \alpha = 0.01$$

$$\frac{\alpha}{2} = 0.005$$

$$1-\frac{\alpha}{2} = 0.995$$

En la Figura 5.31 se observan los valores de los cuantiles obtenidos al leer una tabla *chi-cuadrado* $n-1 = 12 - 1 = 11$ grados de libertad para un 0.005 =0.5 % y un 0.995 = 99.5 % de probabilidad acumulada, respectivamente.

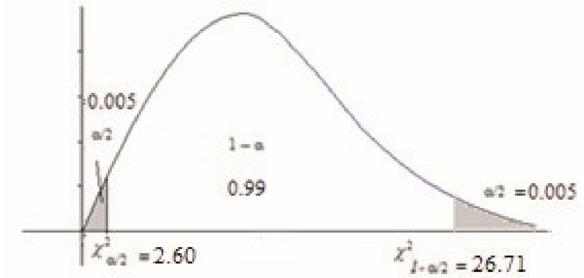

Figura 5.31 Intervalo de confianza sobre la curva chi-cuadrado

Fuente: los autores con la ayuda del *software* libre R.

$$\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.995,11} = 26.757$$

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{0.005,11} = 2.603$$

Luego, usando la expresión 3.13:

$$\sigma^2 \in \left(\frac{(n-1)\hat{S}^2}{\chi^2_{1-\frac{\alpha}{2}}} \; , \; \frac{(n-1)\hat{S}^2}{\chi^2_{\frac{\alpha}{2}}} \right)$$

Y reemplazando los valores, resulta:

$$\sigma^2 \in \left(\frac{(12-1)(0.03)^2}{26.757} \; , \; \frac{(12-1)(0.03)^2}{2.603} \right)$$

Efectuando las operaciones de tipo aritmético se obtiene:

$$\sigma^2 \in \left(\frac{0.0099}{26.757} \; , \; \frac{0.0099}{2.603} \right)$$

Por consiguiente,

$$\sigma^2 \in (0.00037 \; , \; 0.0038)$$

En conclusión, con un nivel de confianza del 99 % se infiere que la varianza poblacional está entre 0.00037 y 0.0038. Lo anterior indica que en 99 de cada 100 muestras seleccionadas se encuentra el parámetro σ^2.

Ahora, para estimar la desviación estándar poblacional se usa la expresión 3.14:

$$\sigma \in \left(\sqrt{\frac{(n-1)\hat{S}^2}{\chi^2_{1-\frac{\alpha}{2}}}} \; , \; \sqrt{\frac{(n-1)\hat{S}^2}{\chi^2_{\frac{\alpha}{2}}}} \right)$$

Al sustituir los valores pertinentes se obtiene:

$$\sigma \in \left(\sqrt{\frac{(12-1)(0.03)^2}{26.757}} \; , \; \sqrt{\frac{(12-1)(0.03)^2}{2.603}} \right)$$

Así, entonces,

$$\sigma \in \left(\sqrt{0.00037} \; , \sqrt{0.0038} \right)$$

Por lo tanto,

$$\sigma \in \left(0.01923 \; , 0.0616 \right)$$

En conclusión, con un nivel de confianza del 99 % se infiere que la desviación estándar poblacional está entre 0.01923 y 0.0616 cm. Lo anterior indica que en 99 de cada 100 muestras seleccionadas se encuentra el parámetro σ.

3.8 Se tiene la creencia de que los egresados de la titulación de Administración de Empresas obtienen un salario promedio mayor que el de los egresados de la titulación de Economía; además, se quiere saber si la variación de sus correspondientes salarios difiere. Para comprobarlo se ha tomado una muestra aleatoria de 10 administradores, obteniéndose una media muestral de 2 600 000 pesos por mes, con una varianza corregida de 1 200 000, y una muestra aleatoria de 13 economistas, obteniéndose un promedio de 2 400 000 pesos por mes, con una varianza de 1 300 000; se supone que los dos conjuntos de datos provienen de muestras independientes seleccionadas de poblaciones normales. Construir un intervalo de confianza del 98 % para estimar el cociente de varianzas poblacionales.

En este caso se tiene:

Administradores	Economistas
$n_1 = 10$	$n_2 = 13$
$\bar{X}_1 = 2\ 600\ 000$	$\bar{X}_2 = 2\ 400\ 000$
$\hat{S}_1^2 = 1\ 200\ 000$	$\hat{S}_2^2 = 1\ 300\ 000$

Además,

$$1-\alpha = 0.98 \rightarrow \alpha = 0.02$$
$$\frac{\alpha}{2} = 0.01$$

En la Figura 5.32 se presentan los valores de los cuantiles obtenidos al leer una tabla *F* de Fisher con n_1-1=10-1=9 grados de libertad para el numerador, y n_2-1=13-1=12 grados de libertad para el denominador; estos corresponden a un 0.01=1 % y 0.99=99 % de probabilidad.

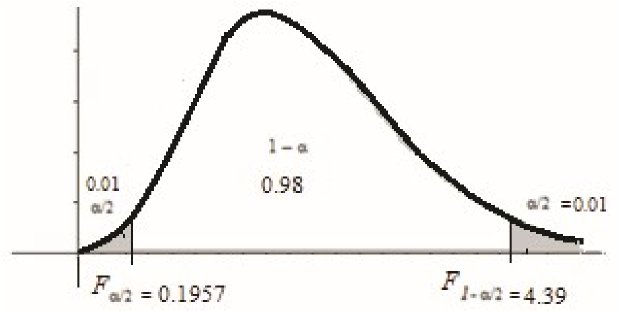

Figura 5.32 Intervalo de confianza sobre la curva de Fisher

Fuente: los autores con la ayuda del *software* libre R.

$$F_{1-\frac{\alpha}{2}} = F_{0.99,9,12} = 4.39$$

$$F_{\frac{\alpha}{2}} = F_{0.01,9,12} = \frac{1}{F_{0.99,12,9}} = \frac{1}{5.11} = 0.1957$$

Luego, usando la expresión 3.15:

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{\hat{S}_1^2}{\hat{S}_2^2 F_{1-\frac{\alpha}{2}}} \, , \, \frac{\hat{S}_1^2}{\hat{S}_2^2 F_{\frac{\alpha}{2}}} \right)$$

Al reemplazar los respectivos valores resulta:

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{1200000}{1300000(4.39)} \, , \, \frac{1200000}{1300000(0.1957)} \right)$$

Efectuando las operaciones de tipo aritmético se obtiene,

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{1\ 200\ 000}{5\ 707\ 000} \ , \ \frac{1\ 200\ 000}{254\ 410} \right)$$

Por consiguiente,

$$\frac{\sigma_1^2}{\sigma_2^2} \in (0.21 \ , \ 4.7167)$$

En conclusión, con un nivel de confianza del 98 % se infiere que el cociente de las varianzas poblacionales está entre 0.21 y 4.7167. Lo anterior indica que en 98 de cada 100 muestras seleccionadas se encuentra el parámetro σ_1^2 / σ_2^2. Como el intervalo de confianza contiene el valor 1, entonces cabe la posibilidad de que las varianzas poblacionales resulten iguales; asimismo, se evidencia que el cociente de varianzas también es inferior a 1; esto indica que la varianza del salario de los administradores es inferior a la de los economistas, y como además el cociente de varianzas es mayor que 1, se tiene que la varianza del salario de los administradores resulta mayor que la de los economistas.

Capítulo 4

4.1 Un fabricante de llantas afirma que sus cauchos duran hasta que, en promedio, rueden 21 000 km, pero un distribuidor potencial de estos considera que duran menos; por eso quiere verificar la afirmación del fabricante, y de un lote de 200 llantas selecciona 11 de manera aleatoria y las pone a rodar, encontrando un promedio de 20 500 kilómetros. Si el fabricante informa que las llantas que ha producido tienen una desviación estándar poblacional de 1000 kilómetros, probar la afirmación del fabricante usando un nivel de significación del 2.5 %.

1. Planteamiento de hipótesis

$$H_o: \mu = 21\ 000$$

$$H_1: \mu < 21\ 000$$

2. Nivel de significación $\alpha = 0.025$

3. La dirección de la prueba y la región crítica se presentan en la Figura 5.33.

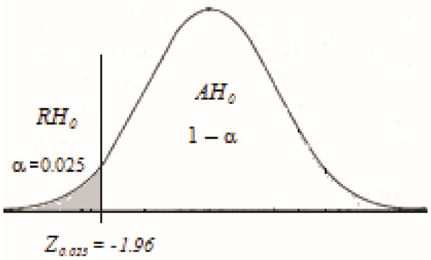

Figura 5.33 Prueba unilateral izquierda para la media poblacional
Fuente: los autores con la ayuda del software libre R.

4. Para la estadística de prueba se usa la siguiente información:

$$\bar{X} = 20\ 500 \qquad \sigma = 1000 \qquad n = 11$$

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{20\ 500 - 21\ 000}{\frac{1000}{\sqrt{11}}\sqrt{\frac{200-11}{200-1}}} = \frac{-500}{315.51(0.9745)} = \frac{-500}{307.4644} = -1.6262$$

5. La estadística de prueba tiene un valor de -1.6262, que es mayor que el valor -1.96 en el modelo teórico correspondiente a una distribución normal estándar; por lo tanto, se ubica en la región AH_0 de aceptación de la hipótesis nula.

6. Decisión: se acepta la hipótesis Ho: $\mu = 21\ 000$, es decir, no hay evidencias suficientes para rechazarla.

7. Por lo tanto, con un nivel de significancia del 2.5 % (confiabilidad del 97.5 %) se concluye que, en promedio, las llantas duran hasta rodar 21 000 kilómetros; en consecuencia, el fabricante tiene la razón.

4.2 Un representante estudiantil afirma que al 10 % de los estudiantes de la universidad B les sirven el almuerzo casi frío; los administradores del restaurante de esta universidad reconocen que algunas veces eso ha sucedido, sin embargo, consideran que el porcentaje es muy inferior al 10 %. En el intento de desvirtuar la afirmación del representante se toma una muestra aleatoria de 150 estudiantes de esta universidad, 12 de los cuales responden que alguna vez han recibido el almuerzo casi frío; probar la afirmación hecha por el representante, usando un nivel de significancia del 3 %.

1. Planteamiento de hipótesis

Ho: $p = 0.1$

H_1: $p < 0.1$

2. Nivel de significancia $\alpha = 0.03$

3. Dirección de la prueba: se trata de una prueba unilateral izquierda; la región crítica se observa en la Figura 5.34.

4. Estadística de prueba

Se tienen los siguientes datos provenientes de la muestra:

$$n = 150 \quad x = 12 \quad \hat{p} = \frac{x}{n} = \frac{12}{150} = 0.08$$

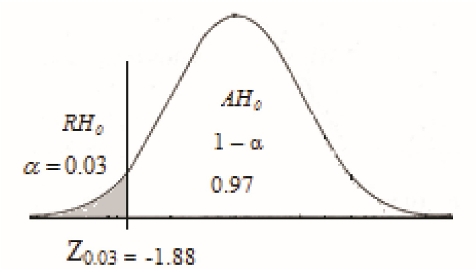

Figura 5.34 Prueba unilateral izquierda para la proporción poblacional

Fuente: los autores con la ayuda del software libre R.

La estadística de prueba por utilizar es:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.08 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{150}}} = \frac{-0.02}{\sqrt{0.0006}} = \frac{-0.02}{0.024494} \cong -0.8165$$

En este caso, se ha de tenido en cuenta que $q_0 = 1 - p_0 = 1 - 0.1 = 0.9$.

5. El valor de la estadística de prueba, -0.8165, cae en la región AH_0, de aceptación de la hipótesis nula H_0, dado que es mayor que el valor de Z=-1.88 en la distribución teórica normal estándar.

6. Decisión: aceptar H_0: $p = 0.1$

7. Finalmente, con un nivel de significancia del 3 % se concluye que la afirmación hecha por el representante estudiantil de que al 10 % de los estudiantes de

la universidad B les sirven el almuerzo casi frío es cierta; no hay evidencias suficientes para rechazar tal afirmación.

4.3 Un estudio sobre el gusto por la práctica deportiva en hombres y mujeres reveló que a 171 de ellos, en una muestra aleatoria de 380, y a 168 de ellas, en una muestra aleatoria de 350, les gusta esta práctica; estos resultados se constituyen en suficiente evidencia para afirmar que el porcentaje de gusto por la práctica deportiva de los hombres difiere del de las mujeres; usar un nivel de significancia del 3 % para realizar la prueba de hipótesis.

1. Planteamiento de hipótesis

Ho: $p_1 = p_2$

H_1: $p_1 \neq p_2$

2. Nivel de significancia $\alpha = 0.03$; $\frac{\alpha}{2} = 0.015$

3. Dirección de la prueba: se trata de una prueba bilateral; la región crítica se indica en la Figura 5.35, donde se presenta una curva normal estándar.

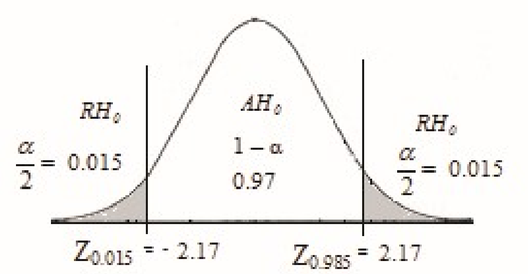

Figura 5.35 Prueba bilateral para la diferencia de proporciones poblacionales

Fuente: los autores con la ayuda del software libre R.

4. Estadística de prueba

En este ejemplo se tiene que:

Hombres	Mujeres
$n_1 = 380, \; x_1 = 171$	$n_2 = 350, \; x_2 = 168$
$\hat{p}_1 = \frac{x_1}{n_1} = \frac{171}{380} = 0.45$	$\hat{p}_2 = \frac{x_2}{n_2} = \frac{168}{350} = 0.48$
$\hat{q}_1 = 1 - 0.45 = 0.55$	$\hat{q}_2 = 1 - 0.48 = 0.52$

La estadística de prueba por utilizar es:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

En primera instancia, se realiza el cálculo siguiente:

$$p = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{380(0.45) + 350(0.48)}{380 + 350} = \frac{171 + 168}{730} = \frac{339}{730} \cong 0.4644$$

Luego, se determina el valor $q = 1 - p = 1 - 0.4644 = 0.5356$

Reemplazando los valores en la estadística de prueba, resulta:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.45 - 0.48}{\sqrt{(0.4644)(0.5356)\left(\frac{1}{380} + \frac{1}{350}\right)}} = \frac{-0.03}{\sqrt{0.001364}} = \frac{-0.03}{0.036932} = -0.8123$$

5. El valor de la estadística de prueba, -0.8123, cae en la región AH_0, de aceptación de la hipótesis nula H_0, puesto que se encuentra entre Z = -2.17 y Z = 2.17 en la distribución teórica normal estándar.

6. Decisión: aceptar H_0: $p_1 = p_2$

7. En consecuencia, con un nivel de significancia del 3 % se concluye que el porcentaje de gusto por la práctica deportiva de los hombres es igual al de las mujeres; no hay evidencias suficientes para aceptar que haya diferencias significativas en tales porcentajes.

4.4 El Director general de Cámaras de Comercio en Colombia desea determinar, al 5 % de significancia, si las utilidades en millones por mes de las pequeñas empresas en Cali y Bucaramanga son iguales; para esto toma una muestra aleatoria de 20 empresas pequeñas en Cali, y otra de 17 empresas pequeñas en Bucaramanga, y encuentra que la utilidad promedio en la primera es de 8 millones de pesos por mes, con una desviación estándar de 500 mil pesos, y en la segunda, de 7.5 millones de pesos por mes, con una desviación estándar de 400 mil pesos.

Esta prueba de hipótesis se ha de abordar de dos formas: *i)* considerando igualdad de varianzas poblacionales y *ii)* considerando varianzas poblacionales diferentes. A continuación, se desarrolla la prueba de hipótesis para el primer caso; el segundo se deja para que el lector se ejercite resolviéndolo.

1. Planteamiento de hipótesis

Ho: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

2. Nivel de significancia $\alpha = 0.05$; $\frac{\alpha}{2} = 0.025$

3. Se trata de una prueba bilateral; la región crítica se observa en la Figura 5.36. En esa figura se han de ubicar los valores del modelo teórico correspondiente a una distribución *t-student* con $n_1+n_2-2=20+17-2=$ 35 grados de libertad, considerando igualdad de varianzas:

$$t_{\frac{\alpha}{2}} = t_{0.025,35} = -2.030 \qquad t_{1-\frac{\alpha}{2}} = t_{0.975,35} = 2.030$$

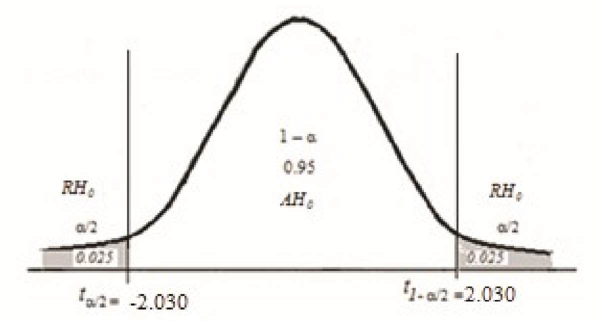

Figura 5.36 Prueba bilateral para la diferencia de medias poblacionales

Fuente: los autores con la ayuda del software libre R.

4. Estadística de prueba

En este ejemplo los datos muestrales son:

Cali	Bucaramanga
$n_1 = 20$	$n_2 = 17$
$\bar{X}_1 = 8$	$\bar{X}_2 = 7.5$
$S_1 = 0.5$	$S_2 = 0.4$

Por tratarse de muestras inferiores a 30 y por ser desconocidas las desviaciones estándar, la estadística de prueba por usar en la prueba de hipótesis es:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

No obstante, para calcular el valor de la estadística se requiere determinar las varianzas corregidas o cuasivarianzas y calcular el valor de S_p; tarea que se desarrolla a continuación:

$$\hat{S}_1^2 = \frac{n_1}{n_1 - 1} S_1^2 = \frac{20}{19}(0.5)^2 = 0.2631$$

$$\hat{S}_2^2 = \frac{n_2}{n_2 - 1} S_2^2 = \frac{17}{16}(0.4)^2 = 0.17$$

$$S_p = \sqrt{\frac{(20-1)(0.2631) + (17-1)(0.17)}{20+17-2}} = \sqrt{\frac{7.7189}{35}} = 0.4696$$

Reemplazando los valores en la estadística de prueba, resulta:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{8 - 7.5}{0.4696\sqrt{\frac{1}{20} + \frac{1}{17}}} = \frac{0.5}{0.4696(0.3298)} = \frac{0.5}{0.154874} = 3.2284$$

5. El valor de la estadística de prueba, 3.2282, cae en la región RH_0, de rechazo de la hipótesis nula H_0, puesto que es mayor que 2.030 en la distribución teórica *t-student,* con 35 grados de libertad.

6. Decisión: rechazar H_0: $\mu_1 = \mu_2$

7. Luego, con un nivel de significancia del 5 %, se concluye que las utilidades promedio en millones de pesos por mes de las pequeñas empresas en Cali y Bucaramanga difieren.

4.5 El diámetro en una muestra aleatoria de 10 varillas de hierro de media pulgada es una variable aleatoria con desviación estándar corregida de 0.2 milésimas de pulgada; el proceso de fabricado de las varillas de hierro se encuentra bajo absoluto control si la varianza es de 0.09, y fuera de control si es mayor; un empleado que lleva el control de calidad sobre el diámetro de las varillas producidas considera que el proceso actualmente está fuera de control; probar esta hipótesis con un nivel de significancia del 5 %.

1. Planteamiento de hipótesis

$Ho: \sigma^2 = 0.09$

$H_1: \sigma^2 > 0.09$

2. Nivel de significancia $\alpha = 0.05$

3. Dirección de la prueba: se trata de una prueba unilateral derecha; la región crítica se observa en la Figura 5.37, donde se presenta una curva asimétrica correspondiente a una distribución *chi-cuadrado* para una probabilidad acumulada de 0.95 con 9 grados de libertad.

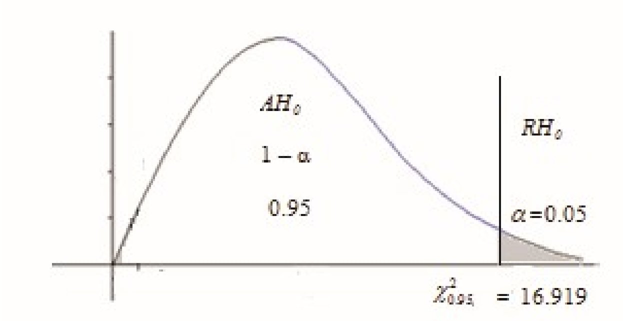

Figura 5.37 Prueba unilateral derecha para la varianza poblacional

Fuente: los autores con la ayuda del software libre R.

4. Estadística de prueba

En este ejemplo los datos muestrales son:

$$n = 10$$

$$\hat{S} = 0.2$$

Reemplazando los valores en la estadística de prueba, resulta:

$$\chi^2 = \frac{(n-1)\hat{S}^2}{\sigma_0^2} = \frac{(10-1)(0.2)^2}{0.09} = \frac{9(0.04)}{0.09} = \frac{0.36}{0.09} = 4$$

5. El valor de la estadística de prueba, 4, cae en la región AH_0, de aceptación de la hipótesis nula H_0, puesto que es menor que el valor teórico, 16.919, en la distribución *chi- cuadrado* con *n*-1=10-1=9 grados de libertad.

6. Decisión: aceptar H_0: $\sigma^2 = 0.09$; en consecuencia, se acepta que la varianza es igual a 0.09.

7. Por consiguiente, con un nivel de significancia del 5 % se concluye que el proceso de fabricación de las varillas en referencia al diámetro está bajo absoluto control; es decir, la consideración hecha por el empleado no tiene fundamento.

4.6 Para comparar la variabilidad en la duración de dos clases de calzado, A y B, se tomaron dos muestras aleatorias; la primera, de 6 pares, proporcionó una desviación estándar de 3 meses, y la segunda, de 5 pares, generó una desviación estándar corregida de 2.5 meses; bajo el supuesto de que las muestras provienen de poblaciones normales independientes, probar la hipótesis de que la varianza de la clase de calzado A es superior a la varianza de la clase de calzado B, usando un nivel de significancia del 5 %.

1. Planteamiento de hipótesis

Ho: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 > \sigma_2^2$

2. Nivel de significancia $\alpha = 0.05$

3. Dirección de la prueba: se trata de una prueba unilateral derecha; la región crítica se presenta en la Figura 5.38, donde se presenta una curva asimétrica correspondiente a una distribución *F* de *Fisher,* con 5 grados de libertad para el numerador y 4 grados para el denominador.

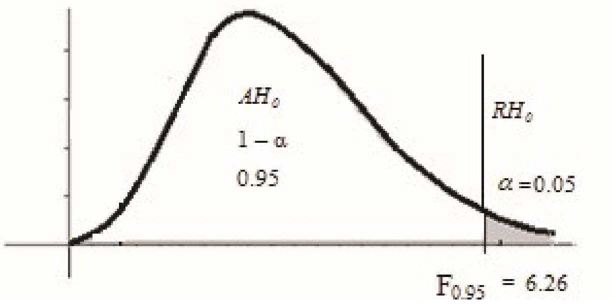

Figura 5.38 Prueba unilateral derecha para la igualdad de varianzas poblacionales

Fuente: los autores con la ayuda del software libre R.

4. Estadística de prueba

En este caso los datos de la muestra son:

Calzado A	Calzado B
$n_1 = 6$	$n_2 = 5$
$S_1 = 3$	$\hat{S}_2 = 2.5$

En primera instancia, conviene calcular la desviación estándar corregida correspondiente a los datos de la primera muestra:

$$\hat{S}_1^2 = \frac{n_1}{n_1 - 1} S_1^2 = \frac{6}{5}(3)^2 = 10.8$$

Reemplazando estos resultados en la estadística de prueba, se tiene:

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{10.8}{(2.5)^2} = \frac{10.8}{6.25} = 1.728$$

5. El valor de la estadística de prueba, 1.728, cae en la región AH_0, de aceptación de la hipótesis nula H_0, puesto que es menor que el valor teórico, 6.26, en la distribución teórica *F* de *Fisher* con n_1-1=6-1=5 grados de libertad para el numerador y n_2 -1 = 5 -1 = 4 grados de libertad para el denominador.

6. Decisión: aceptar H_0: $\sigma_1^2 = \sigma_2^2$; implica que se acepta que las varianzas poblacionales son iguales.

7. Así, entonces, con un nivel de significancia del 5 % se concluye que la varianza de la duración de la clase de calzado A es igual a la varianza (poblacional) de la duración de la clase de calzado B; no hay evidencias suficientes para afirmar que las varianzas poblacionales sean distintas.

4.7 Se tiene una población finita de 600 personas; se seleccionó una muestra piloto del 2 % de la población total y se obtuvo una media muestral de 300 dólares para el salario promedio mensual; también se obtuvo de la muestra piloto una desviación estándar corregida de 50 dólares. Si se admite un error del 3 % de la media muestral, calcular el tamaño muestral al nivel de confianza del 95 %.

En este caso, los datos son:

$$1-\alpha=0.95 \qquad N=600$$
$$\alpha=0.05 \rightarrow \frac{\alpha}{2}=0.025 \qquad \hat{S}=50$$
$$e=9$$

Conviene en este caso utilizar la expresión 4.10; sin embargo, hace falta el valor de Z, que se obtiene de una distribución normal estándar para una probabilidad acumulada del 0.975, obteniéndose un valor de 1.96, como se observa en la Figura 5.39.

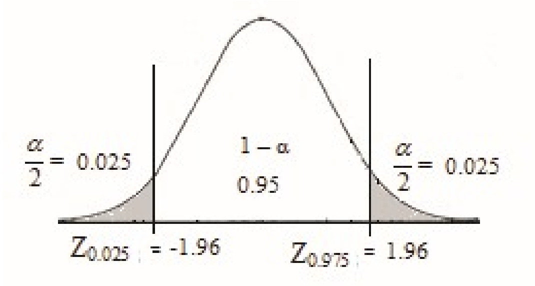

Figura 5.39 Valor de Z para un nivel de confianza del 95 %

Fuente: los autores con la ayuda del software libre R.

Reemplazando los valores, resulta:

$$n=\frac{NZ^2\hat{S}^2}{(N-1)e^2+Z^2\hat{S}^2}=\frac{600(1.96)^2(50)^2}{599(9)^2+(1.96)^2(50)^2}=\frac{600(3.8416)(2500)}{599(81)+(3.8416)(2500)}$$

$$n=\frac{5762400}{48519+9604}=\frac{5762400}{58123}\cong 99.14$$

En el contexto anterior, el tamaño de la muestra para estimar el salario promedio mensual, usando un nivel de confianza del 95 % es de 99 individuos. Esta cantidad de individuos se han de seleccionar apropiadamente usando métodos aleatorios de muestreo.

4.8 Se requiere indagar el mercado para el producto H; se sabe que el 60 % de los individuos de la ciudad A utilizan diariamente este producto; calcular el tamaño de la muestra si se desea que el error máximo absoluto sea del 4 % con un nivel de confianza del 96 %.

En este caso, se tiene que:

$$1-\alpha = 0.96$$

$$\alpha = 0.04 \quad \rightarrow \quad \frac{\alpha}{2} = 0.02$$

$$e = 0.04$$

$$p = 0.6$$

En estas circunstancias se recomienda utilizar la expresión 4.1; el valor de Z=2.055 se obtiene de una distribución normal estándar para una probabilidad acumulada del 0.98, como se indica en la Figura 5.40.

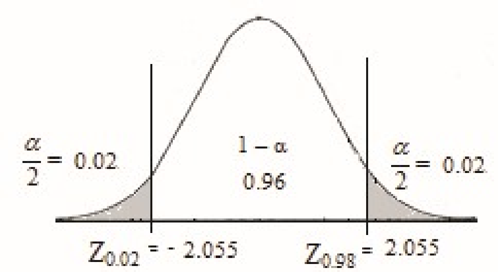

Figura 5.40 Valor de Z para un nivel de confianza del 96 %

Fuente: los autores con la ayuda del software libre R.

Reemplazando los valores, resulta:

$$n = \frac{Z^2 pq}{e^2} = \frac{(2.055)^2(0.6)(0.4)}{(0.04)^2} = \frac{(4.223)(0.24)}{0.0016} = \frac{1.01352}{0.0016} \cong 633.45$$

En estas circunstancias, el tamaño de la muestra para indagar el mercado del producto H en la ciudad A es de 633 individuos, admitiendo un error del 4 % y un nivel de confianza del 96 %. Este número de individuos se han de seleccionar de forma pertinente, utilizando un método de muestreo apropiado.

GLOSARIO DE SÍMBOLOS

N: tamaño de la población o número de individuos que la conforman

n: tamaño de la muestra o número de individuos que la constituyen

E: conjunto

X, Y, Z: variables numéricas o aleatorias

Ω : espacio muestral o conjunto de todos los resultados posibles de un experimento aleatorio

$\emptyset$: conjunto vacío

$\leq$, $\geq$: menor o igual, mayor o igual, respectivamente

$\neq$: diferente

$\cong$: aproximadamente igual

$\Im$: sigma álgebra o familia de subconjuntos del espacio muestral

P: medida de probabilidad

$(\Omega, \Im, P)$: indica un espacio de probabilidad.

R: conjunto de los números reales

β: sigma álgebra de *Borel* sobre el conjunto de los números reales

(R, β): espacio medible sobre los números reales

$X^{-1}(E)$: denota la imagen inversa del conjunto *E* a través de la variable aleatoria *X*

R_X : simboliza el rango de la variable aleatoria *X*

$P_X(B)$: probabilidad inducida por la variable aleatoria *X*

$f_X(x)$: función de probabilidad o de densidad de probabilidad

$\sum_{x_i \in R_X} f(x_i) = 1$: suma de los valores de la función *f* sobre el rango de la variable aleatoria *X*

$F_X(x) = P(X \leq x)$: función de distribución de probabilidad de la variable aleatoria *X*

$X : N(0,1)$: la variable aleatoria *X* tiene distribución normal estándar

μ : media o promedio poblacional

σ^2 : varianza poblacional

σ : desviación estándar poblacional

CV: coeficiente de variación poblacional

p : proporción poblacional

$\overline{X}$: media o promedio muestral

S^2 : varianza muestral

S: desviación estándar muestral

$\hat{S}^2$: varianza corregida o cuasivarianza

$\hat{S}$: desviación estándar corregida o cuasidesviación estándar

cv : coeficiente de variación muestral

$\hat{p}$: proporción o porcentaje muestral

$\binom{N}{n}$: combinatorio de los N elementos tomados de n en n.

$n!$: factorial de n

$x \in R$: x pertenece al conjunto de los números reales

$E(X)$: valor esperado o media de la variable aleatoria X

$Var(X)$: varianza de la variable aleatoria X

$F(z) = \Phi(z)$: distribución normal estándar

$P(Z \leq 1.64)$: probabilidad de Z menor o igual al valor 1.64 en la normal estándar

$\chi^2_{0.99,15} = 30.578$: valor de la *chi-cuadrado* para una probabilidad del 99 % y 15 grados

$t_{0.95,24} = 1.71$: valor de la *t-student* para una probabilidad del 95 % y 24 grados de libertad

$F_{0.99,7,4} = 14.98$: valor de la F de Fisher para una probabilidad del 99 % con 7 y 4 grados

θ : parámetro en el espacio de parámetros $\Theta \subseteq R^n$ como subconjunto de R a la n

T, $\hat{\theta}$: estimador del parámetro θ

$1 - \alpha$: nivel de confianza

$\hat{p}_1 - \hat{p}_2$: diferencia de proporciones o de porcentajes muestrales

$p_1 - p_2$: diferencia de proporciones o de porcentajes poblacionales.

$\overline{X}_1 - \overline{X}_2$: diferencia de promedios o de medias muestrales

$\mu_1 - \mu_2$: diferencia de promedios o de medias poblacionales

$\prod_{i=1}^{n} f(x_i)$: productoria desde i=1 hasta *n* de los valores de la función f en x_i

$Z = \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$: estadística de prueba para el test de la media con σ conocida

$t = \dfrac{\bar{X} - \mu_0}{\frac{\hat{S}}{\sqrt{n}}}$: estadística de prueba para el test de la media con σ desconocida

$Z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$: estadística de prueba para el test de la proporción

$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$: estadística de prueba para el test de la diferencia de proporciones

$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$: estadística de prueba para el test de la diferencia de medias

$\chi^2 = \dfrac{(n-1)\hat{S}^2}{\sigma_0^2}$: estadística de prueba para el test de la varianza

$F = \dfrac{\hat{S}_1^2}{\hat{S}_2^2}$: estadística de prueba para el cociente de varianzas

$e = \hat{p} - p$: error máximo absoluto para la proporción

$n = \dfrac{Z^2 pq}{e^2}$: tamaño de la muestra para la proporción en población infinita

$n = \dfrac{NZ^2 pq}{(N-1)e^2 + Z^2 pq}$: tamaño de la muestra para la proporción en población finita

$e = \bar{X} - \mu$: error máximo absoluto para la media

$n = \dfrac{Z^2 \sigma^2}{e^2}$: tamaño de la muestra para la media en población infinita con σ conocida

$n = \dfrac{Z^2 \hat{S}^2}{e^2}$: tamaño de la muestra para la media en población infinita, con σ desconocida

$n = \dfrac{N Z^2 \sigma^2}{(N-1)e^2 + Z^2 \sigma^2}$: tamaño de la muestra de la media en población finita, σ conocida

$n = \dfrac{N Z^2 \hat{S}^2}{(N-1)e^2 + Z^2 \hat{S}^2}$: tamaño de la muestra, en población finita, σ desconocida