2
Distribuciones usuales de muestreo

En este capítulo se presentan las distribuciones de probabilidad asociadas con los estimadores o estadísticas más frecuentes, entre ellas, la distribución normal, la *chi cuadrado*, la *t-student* y la de *Fisher;* estas distribuciones posibilitan el estudio de variables aleatorias como la media, la varianza y la proporción muestral, la diferencia de medias y de proporciones muestrales y el cociente de varianzas muestrales. Asimismo, se proporcionan ejemplos de aplicación alusivos.

2.1. Distribuciones de probabilidad de uso frecuente en inferencia estadística

Algunos de los aspectos teóricos considerados en esta sección son asumidos o adaptados de los conceptos expuestos al respecto por diversos autores: Alvarado & Obagi (2008), Bickel & Doksum (1977), Blanco (2004), Burbano y Valdivieso (2015), Canavos (1988), Freund y Miller (2000), Gutiérrez *et al.* (2008), Lindgren (1993), Mayorga (2003), Nieves, Sánchez & Clicero (2010) y Shao (1999), por citar algunos.

2.1.1 Distribución normal

Se dice que una variable aleatoria X sigue una distribución normal de parámetros μ y σ, donde μ es un número real y σ es un número real positivo, si su función de densidad es (Bickel & Doksum, 1977; Burbano *et al.*, 2015):

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x\in R$$

La representación gráfica de esta función de densidad se indica en la Figura 2.1; la curva de esta densidad suele denominarse campana de Gauss, y el área bajo la curva tiene un valor de 1.

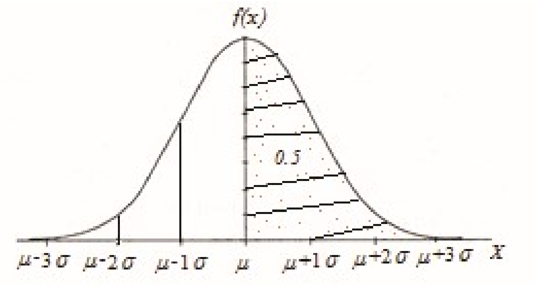

Figura 2.1 Función de densidad de la variable aleatoria X normal

Fuente: los autores apoyados en el software libre R.

Además, la función *f*, por tratarse de una función de densidad, cumple con las dos siguientes condiciones:

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$

En cuanto al valor esperado y la varianza de la variable aleatoria *X* se cumple:

i) $E(X) = \mu$

ii) $Var(X) = \sigma^2$

En el caso particular de que $\mu = 0$ y $\sigma = 1$, se tiene la densidad de la distribución normal estándar; esta queda expresada de la siguiente manera:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right], \quad x \in R$$

La mencionada densidad se deduce realizando la siguiente estandarización de la variable involucrada:

$$z = \frac{x-\mu}{\sigma}$$

$$dz = \frac{1}{\sigma}dx$$

La función de densidad de la distribución normal estándar satisface la siguiente igualdad:

$$\int_{-\infty}^{\infty} f(z)dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz = 1$$

La densidad de la distribución normal estándar para la variable aleatoria Z es:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Sin pérdida de generalidad, esta función se escribe de la siguiente forma:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right], \quad x \in R$$

Donde $f(x)$ es una función de densidad de probabilidad para la variable aleatoria X. La función de distribución de probabilidad para la variable aleatoria Z con distribución normal estándar es la siguiente:

$$F(z) = P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}t^2\right] dt$$

En la Figura 2.2 se observa la función de densidad de la distribución normal estándar; se trata de una curva simétrica con respecto al eje vertical, dado que esta función de densidad satisface que $f(-x) = f(x)$, hecho que también indica que *f* es una función par. El área bajo la curva *f* es igual a 1, en la parte izquierda se ubica un área igual a 0.5=50 % y en la parte derecha se tiene un área de 0.5 como resultado de la simetría. Los valores en el eje horizontal para la variable estandarizada están aproximadamente entre -3.6 y 3.6.

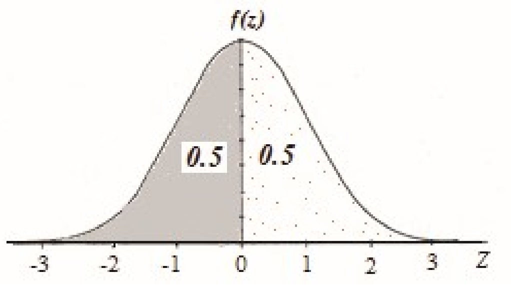

Figura 2.2. Función de densidad de la variable aleatoria Z normal estándar

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.1. En la Figura 2.3 se indica el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad $P(Z \leq 2.14)$; esta ha de leerse en una tabla para la distribución normal estándar.

$$F(2.14) = P(Z \leq 2.14) = \Phi(2.14) = 0.9838$$

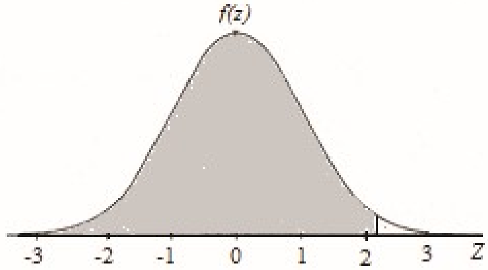

Figura 2.3 Área bajo la curva normal estándar $P(Z \leq 2.14)$

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

Ejemplo 2.2. En la Figura 2.4 se ha sombreado el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad,

$$F(-1.22) = P(Z \leq -1.22) = \Phi(-1.22) = 0.1112$$

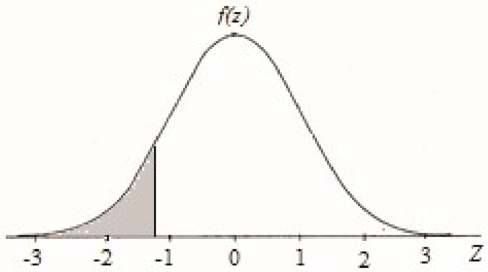

Figura 2.4. Área bajo la curva normal estándar $P(Z \leq -1.22)$

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.3. En la Figura 2.5 se muestra el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad

$$P(Z \leq 0.86) = 0.8051 = 80.51\ \%$$

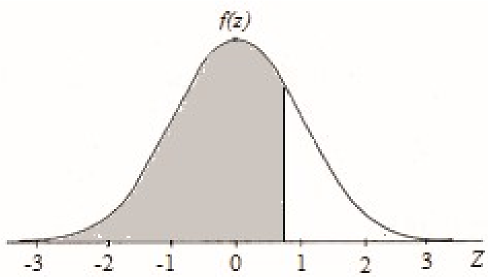

Figura 2.5 Área bajo la curva normal estándar $P(Z \leq 0.86)$

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

Ejemplo 2.4. En la Figura 2.6 se indica el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad

$$P(Z \geq 1.42) = 1 - P(Z \leq 1.42) = 1 - 0.9222 = 0.0778 = 7.78\ \%$$

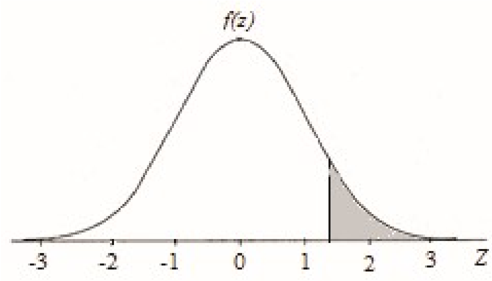

Figura 2.6. Área bajo la curva normal estándar $P(Z \geq 1.42)$

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

Ejemplo 2.5. En la Figura 2.7 se presenta el valor del área bajo la curva normal estándar equivalente al cálculo de la probabilidad

$$P(-1.38 \leq Z \leq 2.47) = P(Z \leq 2.47 - P(Z \leq -1.38)$$

$$P(-1.38 \leq Z \leq 2.47) = 0.9932 - 0.0838 = 0.9094$$

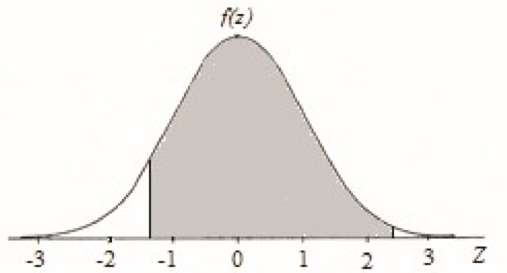

Figura 2.7. Área bajo la curva normal estándar $P(-1.38 \leq Z \leq 2.47)$

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.6. En la fábrica A se ha establecido que la duración (en horas) de los bombillos que se producen se distribuye normalmente con media de 800 horas y desviación estándar de 50 horas. Si de forma aleatoria se toma un bombillo de la producción:

a) ¿Cuál es la probabilidad de que dure máximo 842 horas?
b) ¿Cuál es la probabilidad de que dure por lo menos 718 horas?
c) ¿Cuál es la probabilidad de que dure entre 782 y 917 horas?

$\mu = 800$ horas, $\sigma = 50$ horas

X : duración en horas de un bombillo cualquiera producido por la fábrica A.

a) $P(X \leq 842) = P\left(\dfrac{X-\mu}{\sigma} \leq \dfrac{842-\mu}{\sigma}\right) = P\left(Z \leq \dfrac{842-800}{50}\right)$

$$P(X \leq 840) = P\left(Z \leq \frac{42}{50}\right) = P(Z \leq 0.84) = 0.7995 = 79.95\ \%$$

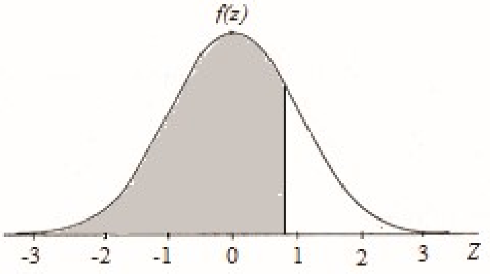

Figura 2.8. Área bajo la curva normal estándar $P(Z \leq 0.84)$

Fuente: los autores apoyados en el *software* libre R.

La probabilidad de que un bombillo escogido aleatoriamente de la producción presente una duración máxima de 842 horas es del 79.95 %. Una representación de este caso se indica en la Figura 2.8.

b) $P(X \geq 718) = P\left(\dfrac{X-\mu}{\sigma} \geq \dfrac{718-\mu}{\sigma}\right) = P\left(Z \geq \dfrac{718-800}{50}\right)$

$$P(X \geq 718) = P(Z \geq -1.64) = 1 - P(Z \leq -1.64) = 1 - 0.0505 = 0.9495 = 94.95\ \%$$

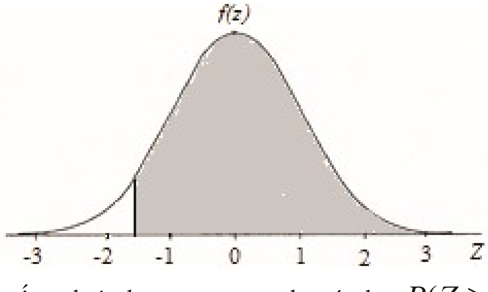

Figura 2.9. Área bajo la curva normal estándar $P(Z \geq -1.64)$

Fuente: los autores apoyados en el *software* libre R.

La probabilidad de que un bombillo escogido de forma aleatoria de esa producción dure por lo menos 718 horas es del 94.95 %. Una representación de la anterior situación se tiene en la Figura 2.9.

c) $P(782 \leq X \leq 917) = P\left(\dfrac{782-\mu}{\sigma} \leq \dfrac{X-\mu}{\sigma} \leq \dfrac{917-\mu}{\sigma}\right)$

$$P(782 \leq X \leq 903) = P\left(\frac{782-800}{50} \leq Z \leq \frac{917-800}{50}\right) = P(-0.36 \leq Z \leq 2.34)$$

$$P(782 \leq X \leq 917) = 0.9904 - 0.3594 = 0.631 = 63.1\ \%$$

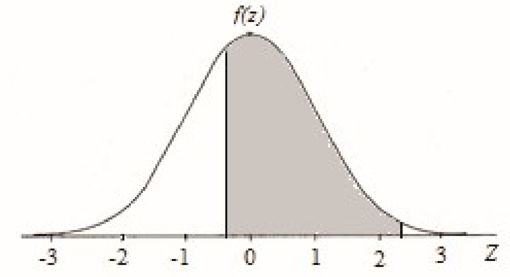

Figura 2.10. Área bajo la curva normal estándar $P(-0.36 \leq Z \leq 2.34)$

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

La probabilidad de que un bombillo escogido aleatoriamente de la producción dure entre 782 y 917 horas es de 63.1 % (ver Figura 2.10).

2.1.2 Distribución chi-cuadrado

Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes, cada una con distribución normal estándar, $\mu = 0$ y $\sigma = 1$, entonces la siguiente variable aleatoria tiene distribución *chi-cuadrado* con *n* grados de libertad:

$$\chi^2 = X_1^2 + X_2^2 +, ..., + X_n^2 = \sum_{i=1}^{n} X_i^2$$

Es decir, una variable aleatoria *chi-cuadrado* es la suma de *n* variables aleatorias en las condiciones mencionadas, cada una de ellas elevada al cuadrado.

Una variable aleatoria *X* tiene distribución *chi-cuadrado* con *n* grados de libertad si su función de densidad es:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\dfrac{x}{2}\right) & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Donde $\Gamma(.)$ es la función gamma dada por:

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \exp(-t)dt \text{ con } \alpha > 0$$

La representación gráfica de la función de densidad para una variable aleatoria *X* que tiene distribución chi-cuadrado con *n* grados de libertad se indica en la Figura 2.11; se trata de una curva asimétrica hacia la derecha; en esta, el área bajo la curva tiene valor de 1.

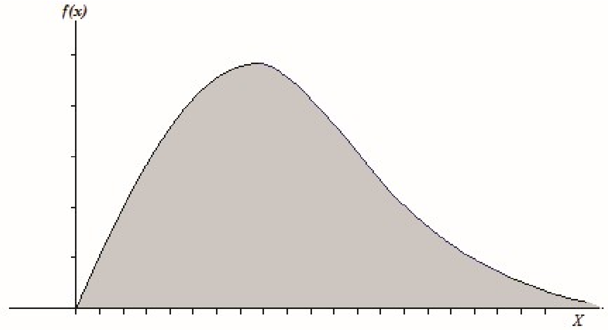

Figura 2.11. Área bajo la curva chi-cuadrado con *n* grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Por supuesto, la función *f*, por tratarse de una función de densidad de probabilidad, cumple con las dos siguientes condiciones:

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2}\right) dx = 1$

De la función gamma se deducen los siguientes resultados:

$$\Gamma(n) = \int_0^{\infty} t^{n-1} \exp(-t)dt = (n-1)! \text{ siempre que } n \text{ sea entero positivo}$$

$$\Gamma(2) = \int_0^{\infty} t^{2-1} \exp(-t)dt = (2-1)! = 1! = 1$$

$$\Gamma(3) = \int_0^{\infty} t^{3-1} \exp(-t)dt = (3-1)! = 2! = 2$$

Asimismo, es posible demostrar que se cumple la siguiente igualdad:

$$\Gamma(\tfrac{1}{2}) = \int_0^{\infty} t^{\frac{1}{2}-1} \exp(-t)dt = \sqrt{\pi}$$

Si *X* es una variable aleatoria con distribución *chi-cuadrado* con *n* grados de libertad, entonces se deduce que:

i) $E(X) = n$

ii) $Var(X) = 2n$

La función de distribución de probabilidad para la variable aleatoria *X* con distribución *chi-cuadrado* con *n* grados de libertad es la siguiente:

$$P(X \leq x) = \begin{cases} \displaystyle\int_0^x \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} \exp\left(-\frac{t}{2}\right) dt & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Ejemplo 2.7. En la Figura 2.12 se indica el valor del área bajo la curva de la densidad *chi-cuadrado* equivalente al cálculo de la probabilidad $P(X \leq 18.307)$ con *n*=10 grados de libertad; suele denotarse con $\chi^2_{0.95,10}$ la distancia sobre el eje horizontal cuya área bajo la curva es de 0.95; esta se lee en una tabla para la distribución *chi-cuadrado*.

$$P(X \leq 18.307) = 0.95$$

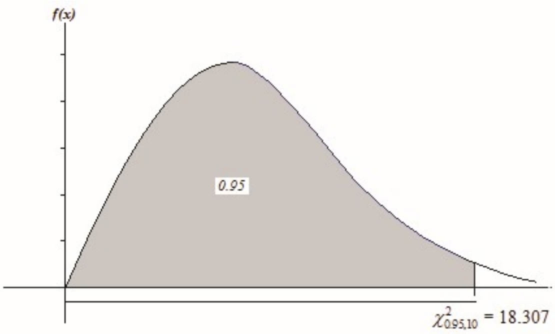

Figura 2.12 Área bajo la curva chi-cuadrado con n = 10 grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.8. En la Figura 2.13 se indica el valor del área bajo la curva de la densidad chi-cuadrado, equivalente al cálculo de la probabilidad $P(X \leq 5.226)$ con n = 12 grados de libertad; $\chi^2_{0.05,12}$ denota una distancia sobre el eje horizontal; esta se lee en una tabla para la distribución *chi-cuadrado*.

$$P(X \leq 5.226) = 0.05$$

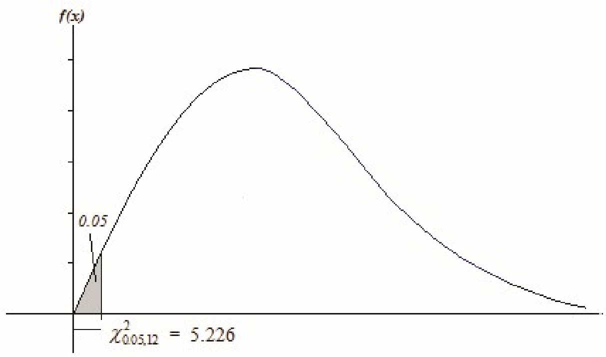

Figura 2.13 Área bajo la curva chi-cuadrado con n = 12 grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.9. A continuación se presentan algunos valores de la distancia *chi-cuadrado* para valores específicos de probabilidad o del área bajo la curva de la densidad *chi-cuadrado;* estos son susceptibles de ser verificados en la tabla de esta distribución (ver Blanco, 2004):

$\chi^2_{0.99,15} = 30.578$

$\chi^2_{0.975,15} = 27.488$

$\chi^2_{0.90,8} = 13.36$

$\chi^2_{0.01,20} = 8.26$

$\chi^2_{0.025,14} = 5.629$

Ahora, si Z es una variable aleatoria con distribución normal estándar, $\mu = 0$ y $\sigma = 1$, entonces la variable aleatoria Z^2 tiene distribución *chi-cuadrado* con n = 1 grados de libertad. Lo anterior quiere decir que si X es una variable aleatoria con distribución normal con media μ y desviación estándar σ,

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \text{ implica que } Z^2 = \left(\frac{X - \mu}{\sigma}\right)^2$$

tiene distribución *chi-cuadrado* con n = 1 grados de libertad. Además, es posible mostrar que la variable aleatoria

$$\frac{(n-1)\hat{S}^2}{\sigma^2}$$

tiene distribución *chi-cuadrado* con n - *1* grados de libertad.

Si $X_1, X_2, ..., X_n$ es una muestra aleatoria para estudiar la variable aleatoria X, con distribución normal, entonces, las variables aleatorias $\overline{X}$ y $\hat{S}^2$ son independientes.

2.1.3 Distribución t-student

Si Z es una variable aleatoria con distribución normal estándar, $\mu = 0$ y $\sigma = 1$, y si Y es una variable aleatoria con distribución *chi-cuadrado* con n grados de libertad, pero Z y Y son independientes, entonces, la variable siguiente se denomina variable aleatoria *t-student* con n grados de libertad:

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

Es decir, una variable aleatoria t-student es el cociente entre una variable normal estándar y la raíz cuadrada de una variable chi-cuadrado dividida por sus grados de libertad.

Una variable aleatoria X tiene distribución *t-student* con n grados de libertad si su función de densidad de probabilidad es:

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} \quad \text{con} \quad \infty < x < \infty$$

Donde $\Gamma(.)$ es la función gamma definida en el apartado anterior.

La representación gráfica de la función de densidad para una variable aleatoria X que tiene distribución *t-student* con n grados de libertad se indica en la Figura 2.14; se trata de una curva simétrica similar a la curva normal estándar, pero más aplanada y con sus colas un poco más altas que aquella; el área bajo la curva también tiene un valor de 1.

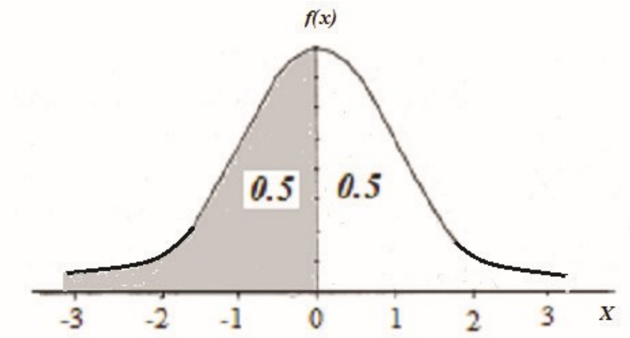

Figura 2.14. Área bajo la curva t-student con *n* grados de libertad
Fuente: los autores apoyados en el *software* libre R.

Por supuesto, la función *f*, por tratarse de una función de densidad de probabilidad, cumple con las dos siguientes condiciones:

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dx = 1$

Si X es una variable aleatoria con distribución *t-student* con n grados de libertad, entonces, se deduce que:

i) $E(X) = 0$

ii) $Var(X) = \dfrac{n}{n-2}$ con $n \geq 3$

La función de distribución de probabilidad para la variable aleatoria X con distribución *t-student* con n grados de libertad es la siguiente:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dt$$

Ejemplo 2.10. En la Figura 2.15 se indica el valor del área bajo la curva de la densidad t-student equivalente al cálculo de la probabilidad $P(t \leq 1.812)$ con n=10 grados de libertad; el valor denotado con $t_{0.95,10}$ corresponde a un valor sobre el eje horizontal; la probabilidad se ha de leer en una tabla para la distribución *t-student*.

$$P(t \leq 1.812) = 0.95$$

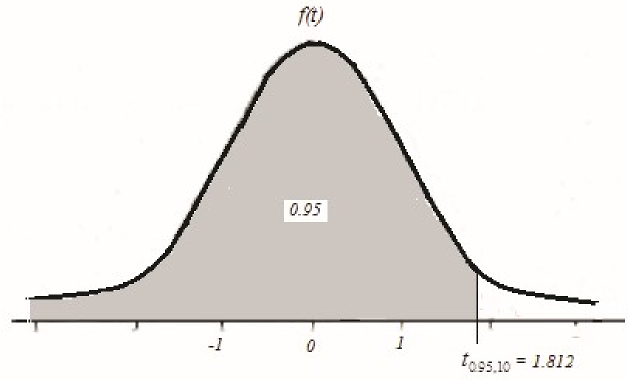

Figura 2.15. Área bajo la curva t-student con n = 10 grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.11. En la Figura 2.16 se indica el valor del área bajo la curva de la densidad t-student, equivalente al cálculo de la probabilidad $P(t \leq -2.131)$ con n = 15 grados de libertad; el valor denotado con $t_{0.025,15}$ corresponde a un valor sobre el eje horizontal; la probabilidad se lee en una tabla para la distribución *t-student*.

$$P(t \leq -2.131) = 0.025$$

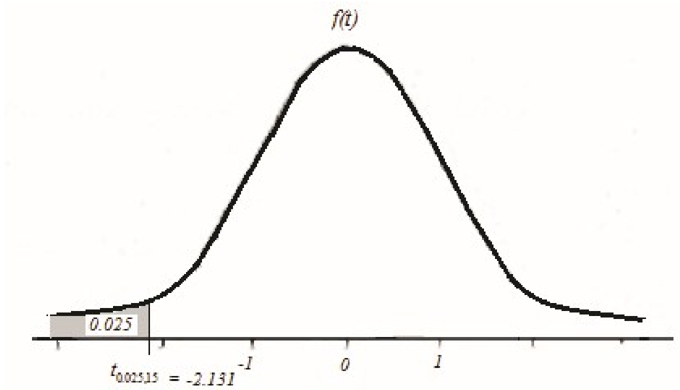

Figura 2.16. Área bajo la curva t-student con *n = 15* grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.12. En la Figura 2.17 se indica el valor del área bajo la curva de la densidad t-student, equivalente al cálculo de la probabilidad $P(-1.697 \leq t \leq 2.042)$ con n = 30 grados de libertad; en este caso se aplica el siguiente criterio para leer la tabla de la distribución *t-student*.

$$P(-1.697 \leq t \leq 2.042) = P(t \leq 2.042) - P(t \leq -1.697) = 0.975 - 0.05 = 0.925$$

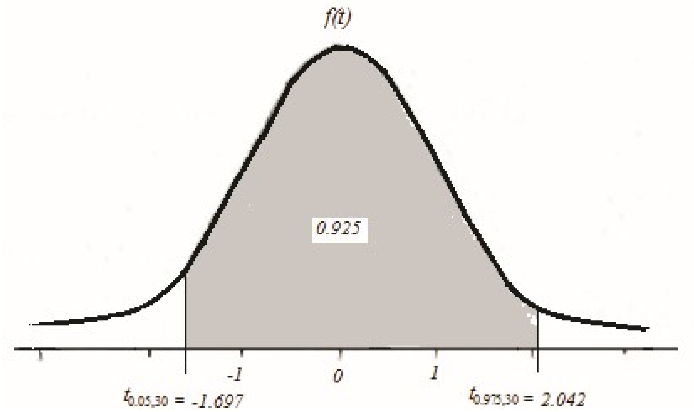

Figura 2.17. Área bajo la curva t-student con *n = 30* grados de libertad

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

Ejemplo 2.13. Enseguida, se presentan algunos valores sobre el eje horizontal, correspondientes a valores específicos de probabilidad o del área bajo la curva de la densidad *t-student;* estos son susceptibles de verificar en la tabla de esta distribución.

$$t_{0.99,12} = 2.681$$

$$t_{0.01,14} = -2.624$$

$t_{0.975,80} = 1.990$

$t_{0.95,24} = 1.71$

2.1.4 Distribución F de Fisher

Si X es una variable aleatoria con distribución *chi-cuadrado* con *m* grados de libertad, y si Y es otra variable aleatoria con distribución *chi-cuadrado* con *n* grados de libertad, pero *X* y *Y* son variables aleatorias independientes, entonces, la variable siguiente se denomina variable aleatoria de Fischer, con *m* y *n* grados de libertad, respectivamente:

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

Es decir, una variable aleatoria de *Fisher* es el cociente entre dos variables chi-cuadrado, cada una dividida por sus grados de libertad.

Una variable aleatoria *X* tiene distribución de *Fisher* con *m* y *n* grados de libertad si su función de densidad de probabilidad es:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\left(\frac{m}{2}-1\right)} (n+mx)^{-\left(\frac{m+n}{2}\right)} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Donde $\Gamma(.)$ es la función gamma definida en los apartados anteriores.

La representación gráfica de la función de densidad para una variable aleatoria *X* que tiene distribución de *Fisher* con *m* y *n* grados de libertad se indica en la Figura 2.18; se trata de una curva asimétrica hacia la derecha; en esta, el área bajo la curva tiene un valor de 1.

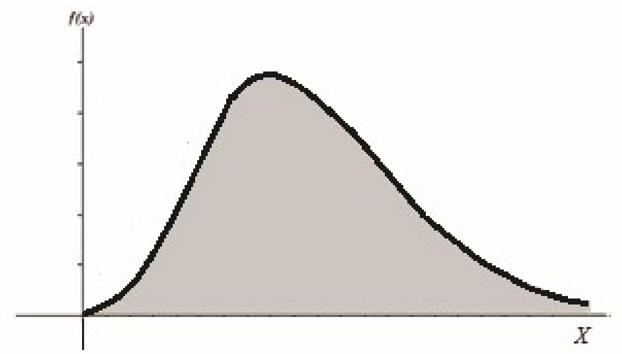

Figura 2.18. Área bajo la curva de Fisher con *m* **y** *n* grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Por supuesto, la función *f*, por tratarse de una función de densidad de probabilidad, cumple con las siguientes dos condiciones:

$i)\quad f(x) \geq 0$

$$ii)\quad \int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\left(\frac{m}{2}-1\right)} (n+mx)^{-\left(\frac{m+n}{2}\right)} dx = 1$$

Si X es una variable aleatoria con distribución de *Fisher* con m y n grados de libertad, entonces, se deduce que:

$$i)\ E(X) = \frac{n}{n-2} \quad \text{con } n > 2$$

$$ii)\ Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{con } n > 4$$

La función de distribución de probabilidad para la variable aleatoria X con distribución de *Fisher* con m y n grados de libertad es la siguiente:

$$P(X \leq x) = \begin{cases} \displaystyle\int_{0}^{x} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} t^{\left(\frac{m}{2}-1\right)} (n+mt)^{-\left(\frac{m+n}{2}\right)} dt & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Ejemplo 2.14. En la Figura 2.19 se indica el valor del área bajo la curva de la densidad de *Fisher,* equivalente al cálculo de la probabilidad $P(X \leq 3.07)$ con

m = 8 y n = 10 grados de libertad; suele denotarse con $F_{0.95,8,10}$ la distancia sobre el eje horizontal, cuya área bajo la curva es de 0.95; esta ha de leerse en una tabla para la distribución de *Fisher*.

$$P(X \leq 3.07) = 0.95$$

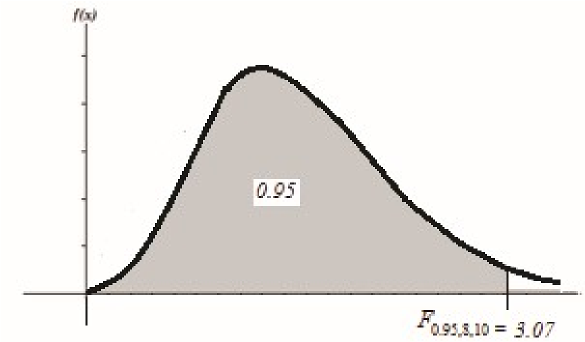

Figura 2.19 Área bajo la curva de Fisher con m = *8* **y** n = 10 grados de libertad

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

Ejemplo 2.15. En la Figura 2.20 se observa el valor del área bajo la curva de la densidad de *Fisher,* equivalente al cálculo de la probabilidad $P(X \leq 0.2159)$ con m = 11 y n = 9 grados de libertad; se trata de calcular $F_{0.01,11,9}$, correspondiente a la distancia sobre el eje horizontal cuya área bajo la curva es de 0.01; esta no se puede leer directamente en una tabla de la distribución de *Fisher;* en este caso, se ha de recurrir a la siguiente fórmula de interpolación:

$$F_{0.01,11,9} = \frac{1}{F_{0.99,9,11}} = \frac{1}{4.63} \cong 0.2159$$

$$P(X \leq 0.2159) = 0.01$$

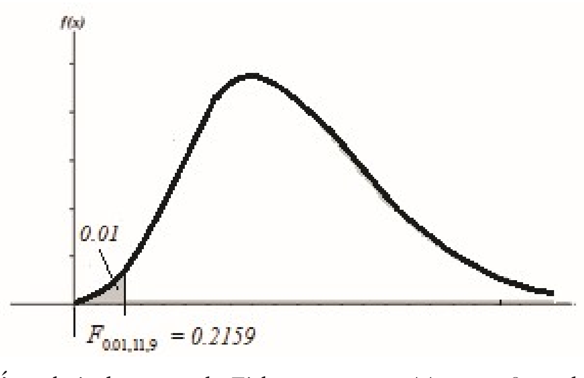

Figura 2.20 Área bajo la curva de Fisher con m = *11* **y** n = 9 grados de libertad

Fuente: los autores apoyados en el *software* libre R.

Ejemplo 2.16. Enseguida se presentan algunos valores de la distancia sobre el eje horizontal para valores específicos de probabilidad o del área bajo la curva de la densidad Fisher; estos se determinan con la tabla de esta distribución (ver Blanco, 2004).

$$F_{0.99,7,4} = 14.98$$

$$F_{0.95,6,8} = 3.58$$

$$F_{0.01,5,9} = \frac{1}{F_{0.99,9,5}} = \frac{1}{10.16} \cong 0.0984$$

$$F_{0.05,6,12} = \frac{1}{F_{0.95,12,6}} = \frac{1}{4.00} = 0.25$$

2.2. Distribuciones muestrales

En esta sección se presentan las distribuciones correspondientes a la media, proporción y varianza muestral, además de las distribuciones de la diferencia de proporciones, medias y cociente de varianzas muestrales. Algunos de los aspectos teóricos considerados en esta sección son asumidos o adaptados de los conceptos expuestos al respecto por diversos autores, entre ellos: Bickel & Doksum (1977), Canavos (1988), Lindgren (1993), Shao (1999), Freund y Miller (2000), Mayorga (2003), Blanco (2004), Gutiérrez *et al.* (2008) y Burbano y Valdivieso (2015).

2.2.1 Distribución de la media muestral

Si $X_1, X_2, ..., X_n$ para $n > 1$ es una muestra aleatoria con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$, entonces $\bar{X}$ es una variable aleatoria; esta tiene distribución de probabilidad y para determinarla se ha de tener presente que se está trabajando con una muestra de variables aleatorias independientes e idénticamente distribuidas, por consiguiente:

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{E(X_1) + E(X_2) + ... + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{Var(X_1) + Var(X_2) + ... + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Luego

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

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Usando el teorema central del límite, se tiene que la siguiente variable aleatoria Z tiene distribución normal estándar:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Así, entonces,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \tag{2.1}$$

La expresión 2.1 se utiliza para una población infinita con desviación estándar conocida. Ahora, si el muestreo se realiza sin reemplazo en una población finita de tamaño N, entonces, en concordancia con Freund y Miller (2000), se deduce que:

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

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

Aplicando el teorema central del límite a la variable promedio muestral, se obtiene que la variable aleatoria Z siguiente tiene distribución normal estándar:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Luego

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)} \sim N\,(0,1) \tag{2.2}$$

La expresión 2.2 se utiliza para una población finita de tamaño N con desviación estándar conocida.

Ahora, si la desviación estándar σ es desconocida, pero constante, entonces la variable aleatoria Z tiene distribución normal estándar:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Por otro lado, la siguiente variable aleatoria tiene distribución chi-cuadrado con *n-1* grados de libertad cuando se muestrea de una población normal:

$$\frac{(n-1)\hat{S}^2}{\sigma^2}$$

Luego el cociente entre la variable normal estándar anterior (expresión 2.1) y la raíz cuadrada de la variable aleatoria con distribución chi-cuadrado con *n-1* grados de libertad dividida por sus *n-1* grados de libertad tiene distribución t-student con *n-1* grados de libertad, así:

$$t = \frac{Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{(n-1)\hat{S}^2}{\sigma^2} / (n-1)}} = \frac{\bar{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}}$$

En consecuencia,

$$t = \frac{\bar{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}} \qquad (2.3)$$

tiene distribución t-student con n -1 grados de libertad; además, la expresión 2.3 se utiliza con frecuencia para *n* menor que 30.

Para una muestra aleatoria proveniente de una población finita de tamaño *n* con desviación estándar desconocida, entonces, se deduce que

$$t = \frac{\bar{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \qquad (2.4)$$

corresponde a una variable aleatoria con distribución *t-student* con *n-1* grados de libertad; esta se utiliza con frecuencia para *n* menor que 30 con poblaciones finitas y con desviación estándar poblacional desconocida.

Ejemplo 2.17. De los datos de la variable X (utilidades en millones de pesos por día de las cinco empresas más eficientes en la ciudad de Paipa, Boyacá-Colombia, en el año 2014), considerados en el *ejemplo 1.11,* a saber: 12, 11, 10, 9, 8, obtener todas las muestras de tamaño n=2 seleccionas por medio de un muestreo aleatorio simple sin reemplazo, determinar el promedio para cada muestra, construir la distribución de probabilidad para la variable aleatoria "media muestral", obtener su valor esperado y su varianza y verificar si se cumplen las igualdades para la esperanza matemática y la varianza presentadas por Freund y Miller (2000).

Como se tiene que $N = 5$ y $n = 2$, entonces, el número total de muestras distintas que son posibles de 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:

$$M_1=\{12,11\},M_2=\{12,10\},M_3=\{12,9\},M_4=\{12,8\},M_5=\{11,10\},$$

$$M_6=\{11,9\},M_7=\{11,8\},M_8=\{10,9\},M_9=\{10,8\},M_{10}=\{9,8\}$$

Sus correspondientes medias muestrales son:

$$\overline{X}_1=\frac{12+11}{2}=11.5,\overline{X}_2=\frac{12+10}{2}=11,\overline{X}_3=\frac{12+9}{2}=10.5,\overline{X}_4=\frac{12+8}{2}=10,\overline{X}_5=\frac{11+10}{2}=10.5$$

$$\overline{X}_6=\frac{11+9}{2}=10,\overline{X}_7=\frac{11+8}{2}=9.5,\overline{X}_8=\frac{10+9}{2}=9.5,\overline{X}_9=\frac{10+8}{2}=9,\overline{X}_{10}\frac{9+8}{2}=8.5$$

Los anteriores valores indican que la media muestral es una variable aleatoria que cambia de valor a medida que se cambia de muestra (con tamaño de muestra constante).

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

$$f\left(\overline{X}=11.5\right)=\frac{1}{10},\ f\left(\overline{X}=11\right)=\frac{1}{10},\ f\left(\overline{X}=10.5\right)=\frac{2}{10},\ f\left(\overline{X}=10\right)=\frac{2}{10},$$

$$f\left(\overline{X}=9.5\right)=\frac{2}{10},\ f\left(\overline{X}=9\right)=\frac{1}{10},\ f\left(\overline{X}=8.5\right)=\frac{1}{10}$$

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

$$E(\bar{X}) = 11.5\left(\frac{1}{10}\right) + 11\left(\frac{1}{10}\right) + 10.5\left(\frac{2}{10}\right) + 10\left(\frac{2}{10}\right) + 9.5\left(\frac{2}{10}\right) + 9\left(\frac{1}{10}\right) + 8.5\left(\frac{1}{10}\right)$$

$$E(\bar{X}) = \frac{11.5 + 11 + 21 + 20 + 19 + 9 + 8.5}{10} = \frac{100}{10} = 10$$

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

$$E(\bar{X}^2) = (11.5)^2\left(\frac{1}{10}\right) + 11^2\left(\frac{1}{10}\right) + (10.5)^2\left(\frac{2}{10}\right) + 10^2\left(\frac{2}{10}\right) + (9.5)^2\left(\frac{2}{10}\right) + 9^2\left(\frac{1}{10}\right) + (8.5)^2\left(\frac{1}{10}\right)$$

$$E(\bar{X}^2) = \frac{(11.5)^2 + 11^2 + 2(10.5)^2 + 2(10)^2 + 2(9.5)^2 + 9^2 + (8.5)^2}{10}$$

$$E(\bar{X}^2) = \frac{132.25 + 121 + 220.5 + 200 + 180.5 + 81 + 72.25}{10} = \frac{1007.5}{10} = 100.75$$

Con los resultados anteriores, se calcula así la varianza:

$$Var(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}))^2 = 100.75 - (10)^2 = 100.75 - 100 = 0.75$$

En el ejemplo 1.11 ya se obtuvieron el valor esperado y la varianza de la variable utilidades; estos fueron:

$$\mu = 10$$

$$\sigma^2 = 2$$

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

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

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2}{2}\left(\frac{5-2}{5-1}\right) = \frac{3}{4} = 0.75$$

2.2.2 Distribución de la proporción muestral

Si X es una variable aleatoria con distribución binomial de parámetro p, donde este parámetro indica la probabilidad de éxito en una muestra aleatoria de tamaño n (de variables aleatorias Bernoulli), entonces, para la variable aleatoria

X: número de éxitos en la muestra de la característica A de interés,

$$E(X) = np$$

$$Var(X) = np(1-p)$$

Considerando la variable

$$\hat{p} = \frac{X}{n}$$

su valor esperado y varianza son los siguientes:

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{Var(X)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

Luego

$$E(\hat{p}) = p$$

$$Var(\hat{p}) = \frac{pq}{n}$$

Donde

$$q = 1 - p$$

Para un tamaño de muestra n "grande", se aplica el teorema central del límite, obteniéndose:

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - np}{\sqrt{npq}}$$

Al realizar operaciones de tipo aritmético, la expresión anterior se transforma en:

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

Así, entonces,

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1) \tag{2.5}$$

La expresión 2.5 corresponde a una variable aleatoria con distribución normal estándar para *n* grande; esta involucra a la variable aleatoria proporción muestral.

2.2.3 Distribución de la diferencia de proporciones muestrales

Si se tienen dos poblaciones binomiales con parámetros p_1 y p_2, respectivamente, y se ha seleccionado una muestra aleatoria de tamaño n_1 en la primera población, y una muestra aleatoria de tamaño n_2 en la segunda, de forma que estas sean independientes, entonces, para las variables aleatorias:

X_1: número de éxitos en la primera muestra de la característica A de interés y

X_2: número de éxitos en la segunda muestra de la característica A de interés

se definen las proporciones muestrales correspondientes de la siguiente manera:

$$\hat{p}_1 = \frac{X_1}{n_1} \quad \text{y} \quad \hat{p}_2 = \frac{X_2}{n_2}$$

Sus correspondientes valores esperados y sus varianzas son:

$$E(\hat{p}_1) = p_1$$

$$Var(\hat{p}_1) = \frac{p_1 q_1}{n_1}$$

Donde $q_1 = 1 - p_1$

$$E(\hat{p}_2) = p_2$$

$$Var(\hat{p}_2) = \frac{p_2 q_2}{n_2}$$

Donde $q_2 = 1 - p_2$

Se define la variable aleatoria denominada diferencia de proporciones muestrales; esta se denota con $\hat{p}_1 - \hat{p}_2$, y su valor esperado es:

$$E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$$

Si estas variables aleatorias $\hat{p}_1$ y $\hat{p}_2$ son independientes, entonces, su varianza equivale a:

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

Para tamaños n_1 y n_2 "grandes", se aplica el teorema central del límite, obteniéndose,

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

Así, entonces, la siguiente variable aleatoria tiene distribución normal estándar:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1) \tag{2.6}$$

La expresión 2.6 corresponde a una variable aleatoria con distribución normal estándar para n_1 y n_2 grandes; esta involucra a la variable aleatoria diferencia de proporciones muestrales.

2.2.4 Distribución de la diferencia de medias muestrales

Dadas dos poblaciones normales, se desea estudiar una determinada característica de tipo cuantitativo; para esto, se selecciona una muestra aleatoria de tamaño n_1 en la primera población y una muestra aleatoria de tamaño n_2 en la segunda, de forma que estas resulten independientes, y se obtienen sus respectivos promedios, para definir una nueva variable aleatoria llamada diferencia de medias muestrales. A continuación, se determina la distribución de probabilidad de esa nueva variable.

Si $X_{11}, X_{12}, ..., X_{1n_1}$ para $n_1 > 1$ es la muestra aleatoria extraída de la primera población con $E(X_{1i}) = \mu_1$ y $Var(X_{1i}) = \sigma_1^2$, entonces, $\bar{X}_1$ es una variable aleatoria; asimismo, si $X_{21}, X_{22}, ..., X_{2n_2}$, para $n_2 > 1$, es la muestra aleatoria extraída de la segunda población con $E(X_{2j}) = \mu_2$ y $Var(X_{2j}) = \sigma_2^2$, entonces, $\bar{X}_2$ es otra variable aleatoria; estas variables se expresan así:

$$\bar{X}_1 = \sum_{i=1}^{n_1} \frac{X_{1i}}{n_1} \qquad \bar{X}_2 = \sum_{j=1}^{n_2} \frac{X_{2j}}{n_2}$$

Luego la diferencia $\bar{X}_1 - \bar{X}_2$ es una nueva variable aleatoria. El valor esperado y la varianza de estas variables son, respectivamente:

$$E(\bar{X}_1) = \mu_1 \qquad Var(\bar{X}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{X}_2) = \mu_2 \qquad Var(\bar{X}_2) = \frac{\sigma_2^2}{n_2}$$

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

Por la independencia de las variables $\bar{X}_1$ y $\bar{X}_2$ resulta que:

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si las varianzas son conocidas, entonces, usando el teorema central del límite se tiene que la siguiente variable aleatoria Z tiene distribución normal estándar:

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

Así, entonces,

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \tag{2.7}$$

La expresión 2.7 se utiliza para trabajar con dos poblaciones infinitas que presenten desviaciones estándar conocidas.

Ahora, si para cada población la desviación estándar es desconocida y las muestras tienen tamaño inferior a 30 (muestras pequeñas), se maneja el supuesto de que las desviaciones son iguales, en este caso es posible probar que la siguiente variable aleatoria tiene distribución *t-student* con n_1+n_2-2 grados de libertad,

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{2.8}$$

En la expresión 2.8, el valor S_p se obtiene a través de la siguiente expresión:

$$S_p = \sqrt{\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}}$$

Pero si en cada población la desviación estándar es desconocida y las muestras tienen tamaño inferior a 30 (muestras pequeñas), también es posible suponer que las desviaciones estándar son distintas; en este caso se deduce que la siguiente variable aleatoria tiene distribución *t-student* con *g* grados de libertad:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \tag{2.9}$$

Para la expresión 2.9, los grados de libertad *g* se obtienen por medio de la siguiente expresión (Gutiérrez *et al.*, 2008):

$$g = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{S}_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{\hat{S}_2^2}{n_2}\right)^2}{n_2 + 1}} - 2$$

En cambio, si en cada población la desviación estándar es desconocida y las muestras tienen tamaño mayor o igual a 30 (muestras grandes), sea para varianzas iguales o desiguales, se deduce que la siguiente variable aleatoria (expresión 2.10) tiene distribución normal estándar:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \sim N(0,1) \tag{2.10}$$

2.2.5 Distribución del cociente de varianzas muestrales

Dadas dos poblaciones normales, si $X_{11}, X_{12}, ..., X_{1n_1}$ para $n_1 > 1$ es una muestra aleatoria extraída de la primera población con $E(X_{1i}) = \mu_1$ y $Var(X_{1i}) = \sigma_1^2$, entonces, $\hat{S}_1^2$ es una variable aleatoria; asimismo, si $X_{21}, X_{22}, ..., X_{2n_2}$ para $n_2 > 1$ es la muestra aleatoria extraída de la segunda población con $E(X_{2j}) = \mu_2$

y $Var(X_{2j}) = \sigma_2^2$, entonces, $\hat{S}_2^2$ es otra variable aleatoria; estas variables se expresan de la siguiente manera:

$$\hat{S}_1^2 = \sum_{i=1}^{n_1} \frac{(X_{1i} - \overline{X}_1)^2}{n_1 - 1} \qquad \hat{S}_2^2 = \sum_{j=1}^{n_2} \frac{(X_{2j} - \overline{X}_2)^2}{n_2 - 1}$$

Además, las siguientes variables aleatorias tienen distribución chi-cuadrado con n_1-1 y n_2-1 grados de libertad, respectivamente:

$$\frac{(n_1 - 1)\hat{S}_1^2}{\sigma_1^2} \quad \text{y} \quad \frac{(n_2 - 1)\hat{S}_2^2}{\sigma_2^2}$$

Por definición, al cociente entre las dos variables aleatorias anteriores, cada una dividida entre sus grados de libertad, le corresponde una distribución de Fisher, escrita como sigue:

$$F = \frac{\dfrac{\dfrac{(n_1 - 1)\hat{S}_1^2}{\sigma_1^2}}{n_1 - 1}}{\dfrac{\dfrac{(n_2 - 1)\hat{S}_2^2}{\sigma_2^2}}{n_2 - 1}} = \frac{\dfrac{\hat{S}_1^2}{\sigma_1^2}}{\dfrac{\hat{S}_2^2}{\sigma_2^2}} = \frac{\hat{S}_1^2 \sigma_2^2}{\hat{S}_2^2 \sigma_1^2}$$

Así, entonces, la variable de la expresión 2.11 tiene distribución de Fischer con n_1-1 grados de libertad para el numerador y n_2-1 grados de libertad para el denominador.

$$F = \frac{\hat{S}_1^2 \sigma_2^2}{\hat{S}_2^2 \sigma_1^2} \tag{2.11}$$

Para el caso particular en que las varianzas sean iguales, resulta que

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

con n_1-1 grados de libertad para el numerador y n_2-1 grados de libertad para el denominador.

Actividades para el estudio independiente 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) $P(0.29 \leq Z \leq 2.48)$

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?

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 por lo menos 486 g?
b) ¿Cuál es la probabilidad de que su peso sea máximo 480 g?
c) ¿Cuál es la probabilidad de que su peso esté entre 476 y 550 g?
d) ¿Cuál es la probabilidad de que su peso sea menos de 400 g?
e) ¿Cuál es la probabilidad de que su peso sea más de 440 g?

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

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

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

2.5 Obtener las probabilidades y graficar:

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

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

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

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

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

2.7 Para la variable X (gastos en transporte por día [miles de pesos] de los cinco mejores estudiantes de la universidad B en la ciudad de Tunja, 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).

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.

2.9 El tiempo promedio que gasta el bus urbano en la ciudad de Cali es de 70 minutos. Si se toma una muestra aleatoria de 12 recorridos y con esos datos se obtiene 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?

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?

2.11 En la ciudad A, para niños de grado quinto de educación básica primaria se tiene un peso promedio de 35 kg con varianza de 5 , mientras que en la ciudad B, para niños que cursan el mismo grado, se tiene un peso promedio de 45 kg con varianza de 8. En la ciudad A se toma una muestra aleatoria de 50, y en la ciudad B otra de 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?

2.12 Un candidato a la presidencia de la república tiene el 60 % de la intención de voto en el departamento de Nariño, y el 58 % en el Valle del Cauca, en

Colombia. Si se toma una muestra aleatoria de 400 votantes en Nariño y de 500 en el Valle del Cauca, ¿cuál es la probabilidad de que la diferencia entre las proporciones muestrales de los potenciales votantes en Nariño y el Valle no superen el 3 %?

Ejercicios para el capítulo 2

2.1. Leer en una tabla normal estándar los siguientes valores de probabilidad y elaborar la representación gráfica correspondiente:

a) $P(Z \le 2.16)$

b) $P(Z \ge 1.62)$

c) $P(Z < -0.025)$

d) $P(Z > 2.33)$

e) $P(Z < -1.47)$

f) $P(Z = -0.94)$

g) $P(Z < 5.4)$.

h) $P(Z > 3.8)$.

i) $P(Z < -4.2)$.

2.2. El salario de los trabajadores de la empresa J&T se distribuye normalmente con media de 800 (en miles de pesos) y desviación estándar de 15; si se escoge aleatoriamente a un trabajador de esta empresa, ¿cuál es la probabilidad de que su salario sea superior a 813, inferior a 819 o esté entre 812 y 832?

2.3. El peso de un producto de la marca H es una variable con distribución normal, con media de 1000 g y desviación estándar de 5 g; si se escoge aleatoriamente una unidad del producto de la marca H, ¿cuál es la probabilidad de que su peso sea superior a 1009 g?

2.4. En un proceso productivo se sabe que el 96 % de los artículos de la marca MM resultan de buena calidad; si se selecciona una muestra aleatoria de 100 unidades de estos, ¿cuál es la probabilidad de que más de 94 unidades de esta marca resulten de buena calidad?

2.5. Una fábrica produce unidades de medicamento del tipo A para vacunos; la probabilidad de que cualquier unidad producida por la fábrica no resulte efectiva es del 3 %; si se escoge una muestra aleatoria de 90 unidades, ¿cuál es la probabilidad de que, por lo menos, 5 no resulten efectivas?

2.6. El peso en un grupo de aves sigue una distribución normal con media 2000 g y desviación estándar de 100 g; si se escoge aleatoriamente un ave de ese grupo, ¿cuál es la probabilidad de que su peso sea mayor que 1857 g?

2.7. Si X es una variable aleatoria con distribución normal estándar, mostrar que su función de densidad de probabilidad cumple la siguiente igualdad:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = 1$$

2.8. Leer en una tabla los siguientes valores de probabilidad correspondientes a una distribución chi-cuadrado y elaborar la representación gráfica correspondiente:

a) $\chi^2_{0.95,16} = ?$

b) $\chi^2_{0.99,20} = ?$

c) $\chi^2_{0.975,6} = ?$

d) $\chi^2_{0.01,9} = ?$

e) $\chi^2_{0.025,8} = ?$

2.9. Si Z es una variable aleatoria con distribución normal estándar, $\mu = 0$ y $\sigma = 1$, demostrar que Z^2 es una variable aleatoria con distribución *chi-cuadrado* con $n=1$ grados de libertad.

2.10. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria para estudiar la variable aleatoria X, con distribución normal con media μ y desviación estándar σ, deducir que la variable aleatoria siguiente:

$$\frac{(n-1)\hat{S}^2}{\sigma^2}$$

tiene distribución chi-cuadrado con $n - 1$ grados de libertad.

2.11. Si X es una variable aleatoria con distribución chi-cuadrado con n grados de libertad, usar la función generadora de momentos para deducir que:

i) $E(X) = n$

ii) $Var(X) = 2n$

2.12. Si X es una variable aleatoria con distribución chi-cuadrado con n grados de libertad, mostrar que su función de densidad de probabilidad cumple la siguiente igualdad:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

2.13. Leer en una tabla los siguientes valores de probabilidad correspondientes a una distribución t-student y elaborar la representación gráfica correspondiente:

a) $t_{0.975,8} = ?$

b) $t_{0.01,14} = ?$

c) $t_{0.95,40} = ?$

d) $t_{0.025,14} = ?$

2.14. Leer en una tabla los siguientes valores de probabilidad correspondientes a una distribución de Fisher y elaborar la representación gráfica correspondiente:

a) $F_{0.99,6,13} = ?$

b) $F_{0.95,10,4} = ?$

c) $F_{0.01,12,12} = ?$

d) $F_{0.05,9,13} = ?$

2.15. Si se realiza un muestreo sin reemplazo de una población finita de tamaño *N*, entonces, deducir que

$$E(\overline{X}) = \mu$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

2.16. La duración promedio de 1000 teclados es de 40 meses, con una desviación estándar de 10 meses; determinar: *i)* la probabilidad de que en una muestra aleatoria de 64 teclados se tenga una duración promedio comprendida entre 38 y 42 meses, *ii)* hallar la probabilidad de que en una muestra aleatoria de 36 teclados la duración promedio sea inferior a 39 meses.

2.17. El 98 % de los artículos que produce una máquina son de buena calidad. Si se toma una muestra aleatoria de 800 artículos: *i)* ¿Cuál es la probabilidad de que más del 98 % de los artículos de la muestra sean de buena calidad? y *ii)* ¿Cuál es la probabilidad de que menos del 3 % de los artículos de la muestra sean defectuosos?

2.18. En la ciudad A los adultos mayores de 60 años tienen un peso promedio de 65 kg, con una varianza poblacional de 10, mientras que para los de la ciudad B el peso promedio es de 66 kg, con una varianza poblacional de 12; si se toma una muestra aleatoria de 40 individuos en la ciudad A y otra de 45 individuos en la ciudad B: *i)* ¿Cuál es la probabilidad de que la media muestral del peso de adultos de la ciudad A difiera de la de los adultos de la ciudad B en más de 13 kg? y *ii)* ¿Cuál es la probabilidad de que la diferencia absoluta de las medias muestrales no superen los 9 kg?