

Muestras aleatorias y distribuciones de muestreo

Definición 1.1 Si las variables aleatorias X_1, X_2, \dots, X_n tiene la misma función (densidad) de probabilidad que la de la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces X_1, X_2, \dots, X_n forman un conjunto de n variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una muestra aleatoria de la población.

En el contexto de la definición 7.1, la función (densidad) conjunta de probabilidad de X_1, x_2, \dots, X_n es la función de verosimilitud de la muestra dada por

$$L(x; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

en donde $x = \{x_1, x_2, \dots, x_n\}$ denota los datos muestreados. Cuando las realizaciones x se conocen, $L(x; \theta)$ es una función del parámetro desconocido θ .

Ejemplo 1.1 Se ilustrará el concepto de muestra aleatoria dada en la definición 7.1 mediante lo siguiente: Sea X_1, X_2, \dots, X_n una muestra aleatoria de n variables aleatorias IID de una población cuya distribución de probabilidad es exponencial con densidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty.$$

Cuando se observa X_1 y se registra su realización x_1 ,

$$f(x_1; \theta) = \frac{1}{\theta} e^{-x_1/\theta}, \quad 0 < x_1 < \infty.$$

Ahora se observa X_2 y se registra su realización x_2 . Dado que X_1 y X_2 son estadísticamente independientes y tienen las mismas densidades marginales,

$$f(x_2|x_1) = f(x_2; \theta) = \frac{1}{\theta} e^{-x_2/\theta}, \quad 0 < x_2 < \infty.$$

La función de densidad conjunta de X_1 y X_2 es

$$f(x_1, x_2; \theta) = f(x_1; \theta) f(x_2; \theta) = \frac{1}{\theta^2} e^{-(x_1+x_2)/\theta}, \quad 0 < x_i < \infty, i = 1, 2.$$

Por lo tanto, se desprende que para una muestra aleatoria de tamaño n

$$L(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\theta} e^{-(x_1 + x_2 + \dots + x_n)/\theta}, \quad 0 < x_i < \infty, i = 1, 2, \dots, n.$$

■

1.3 Distribuciones de muestreo de estadísticas

Definición 1.2 Un parámetro es una caracterización numérica de las distribuciones de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés. Por ejemplo, cuando se especifica el valor del parámetro de escala exponencial θ , se describe de manera completa la función de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{(-x/\theta)}.$$

La oración describe de manera completa, sugiere que una vez que se conoce el valor de θ entonces, puede formularse cualquier proposición probabilística de interés.

Dado que los parámetros son prácticamente inherentes a todos los modelos de probabilidad, es imposible calcular las probabilidades deseadas sin un conocimiento del valor de estos. Es por esta razón que la noción de una estadística y su distribución de muestreo es muy importante en inferencia estadística. Antes de dar la definición de una estadística, debe notarse que desde un punto de vista clásico (no bayesiano), un parámetro se considera como una constante fija cuyo valor se desconoce. Desde una perspectiva bayesiana un parámetro siempre es una variable aleatoria con algún tipo de distribución de probabilidad.

Definición 1.3 Una estadística es cualquier función de las variables aleatorias que se observaron en la muestra de manera que esta función no contiene cantidades desconocidas.

De manera general, denótese una estadística por $T = u(X)$. Dado que T es una función de variables aleatorias, es en sí misma una variable aleatoria, y su valor específico $t = u(x)$ puede determinarse cuando se conozcan las realizaciones x de X . Si se emplea una estadística T para estimar un parámetro desconocido θ , entonces T recibe el nombre de **estimador** de θ , y el valor específico de t como un resultado de los datos muestrales recibe el nombre de **estimación** de θ . Esto es, un estimador es una estadística que identifica al mecanismo funcional por medio del cual, una vez que las observaciones en la muestra se realizan, se obtiene una estimación.

Un parámetro es una constante pero una estadística es una variable aleatoria. Además, un valor del parámetro descrito describe de manera completa un modelo de probabilidad; ningún valor de la estadística puede desempeñar tal papel si cada uno de estos depende del valor de las observaciones de las muestras.

Definición 1.4 La distribución de muestreo de una estadística T es la distribución de probabilidad de T que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño n , provenientes de la población de interés.

La distribución de muestreo de una estadística hace posible este tipo de análisis de probabilidad, esencial para valorar el riesgo inherente cuando se formulan ingerencias.

Teorema 1.1 Sea X_1, X_2, \dots, X_n un conjunto de n variables aleatorias independientes cada una con función generadoras de momentos $m_{X_1}(t), m_{X_2}(t), \dots, m_{X_n}(t)$. Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde a_1, a_2, \dots, a_n son constantes, entonces:

$$m_Y(t) = m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t).$$

Demostración.- Mediante el empleo de la definición y la hipótesis de independencia, se tiene

$$\begin{aligned} m_Y(t) &= E \left\{ e^{[t(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} e^{[t(a_2 X_2)]} \cdots e^{[t(a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} \right\} E \left\{ e^{[t(a_2 X_2)]} \right\} \cdots E \left\{ e^{[t(a_n X_n)]} \right\} \\ &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t). \end{aligned}$$

■

De esta forma, la función generadora de momentos de una combinación lineal de n variables aleatorias independientes es el producto de las correspondientes funciones generadoras de momentos con argumentos iguales a las constantes de tiempo t .

Teorema 1.2 Sea X_1, X_2, \dots, X_n un conjunto de variables aleatorias normalmente distribuidas con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$ para $i = 1, 2, \dots, n$. Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde a_1, a_2, \dots, a_n son constantes, entonces Y es una variable aleatoria con distribución normal y media

$$E(Y) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

y con varianza

$$Var(Y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2.$$

Demostración.- Dado que X_i se encuentra normalmente distribuida, su función generadora de momentos es

$$m_{X_i}(t) = e^{\left(\mu_i t + \frac{\sigma_i^2 t^2}{2} \right)}.$$

De acuerdo con el teorema 7.1, la función generadora de momentos de Y es

$$\begin{aligned} m_Y(t) &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t) \\ &= e^{\left(\mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2} \right)} e^{\left(\mu_2 a_2 t + \frac{a_2^2 \sigma_2^2 t^2}{2} \right)} \cdots e^{\left(\mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2} \right)} \\ &= e^{\left[t \sum_{i=1}^n a_i \mu_i + \frac{\left(\sum_{i=1}^n a_i^2 \sigma_i^2 \right) t^2}{2} \right]}. \end{aligned}$$

Por lo tanto, Y se encuentra normalmente distribuida con media $\sum_{i=1}^n a_i \mu_i$ y varianza $\sum_{i=1}^n a_i^2 \sigma_i^2$. ■

Del teorema 7.2 se desprende que si $a_i = 1$ para $i = 1, 2, \dots, n$, entonces la suma de variables aleatorias independientes normalmente distribuidas también posee una distribución normal con media y varianza igual a la suma de las medias y las varianzas de cada una de las variables aleatorias. El resultado anterior se conoce como la propiedad aditiva de la distribución normal. Debe notarse que la hipótesis de normalidad no es necesaria para obtener las fórmulas de la media y la varianza de Y en el teorema 7.2. De hecho, con base en el teorema 6.1, si X_1, X_2, \dots, X_n es un conjunto de n variables aleatorias IID con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$, $i = 1, 2, \dots, n$ entonces para $Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$,

$$E(Y) = \sum_{i=1}^n a_i \mu_i$$

y

$$Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

en donde a_1, a_2, \dots, a_n son constantes.

1.4 La distribución de muestreo de \bar{X}

Sea X_1, X_2, \dots, X_n una muestra aleatoria que consiste en n variables aleatorias IID tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \dots, n$. Entonces la estadística

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

se define como la media de las n variables aleatorias IID o, sencillamente, media muestral. Nótese que una vez que se conocen las realizaciones x_1, x_2, \dots, x_n de X_1, X_2, \dots, X_n , respectivamente, realización \bar{x} de \bar{X} se obtiene promediando los datos muestrales.

Si en, $E(Y) = \sum_{i=1}^n a_i \mu_i$ y $Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2$, $a_i = 1/n$, $i = 1, 2, \dots, n$ entonces el valor esperado y la varianza de \bar{X} son

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

y

$$Var(\bar{X}) = \sum_{i=1}^n \frac{1}{n^2} \sigma^2 = n \left(\frac{\sigma^2}{n^2} \right) = \frac{\sigma^2}{n},$$

respectivamente, en donde μ y σ^2 son la media y la varianza de la distribución de la población a partir de la cual se obtuvo la muestra. Luego de esta última ecuación de $Var(\bar{X})$ la desviación estándar de \bar{X} es

$$d.e.(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

la cual, en algunas ocasiones recibe el nombre de **error estándar de la media**. Si el tamaño de la muestra crece, la precisión de la media muestral para estimar la media poblacional aumenta.

Teorema 1.3 Sea X_1, X_2, \dots, X_n una muestra aleatoria que consiste en n variables aleatorias independientes y normalmente distribuidas con media $E(X_i) = \mu$ y varianzas $Var(X_i) = \sigma^2$, $i = 1, 2, \dots, n$. Entonces la

distribución de la media muestral \bar{X} es normal con media μ y varianza σ^2/n .

Demostración.- Este teorema es un corolario del teorema 7.2. Esto es, sea $a_i = 1/n$; dado que las medias y las varianzas son iguales, respectivamente, la función generadora de momentos de \bar{X} es

$$\begin{aligned} m_{\bar{X}}(t) &= e^{\left(t \sum_{i=1}^n \frac{1}{n} \mu + \frac{t^2 \sum_{i=1}^n \frac{1}{n^2} \sigma^2}{2} \right)} \\ &= e^{\left(t\mu + \frac{t^2 \sigma^2}{2n} \right)}, \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria normalmente distribuida con media μ y varianza σ^2/n . De esta forma, la **función de densidad de probabilidad de \bar{X} cuando se muestrea una población cuya distribución es normal**, está dada por

$$f(\bar{x}, \mu, \sigma/\sqrt{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{\left[-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right]}, \quad -\sigma < \bar{x} < \sigma.$$

■

Ejemplo 1.2 Demostrar que si X_1, X_2, \dots, X_n son n variables aleatorias independientes exponencialmente distribuidas con función de densidad de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0,$$

entre \bar{X} tiene una distribución gama.

Demostración.- Recuerdese que la función generadora de momentos de una variable aleatoria exponencialmente distribuida es $(1 - \theta t)^{-1}$. De esta manera para cada X_i de la muestra,

$$m_{X_i}(t) = (1 - \theta t)^{-1}.$$

Del teorema 7.1 con $a_i = \frac{1}{n}$, $i = 1, 2, \dots, n$ se desprende que la función generadora de momentos de la media muestral \bar{X} es

$$\begin{aligned} m_{\bar{X}}(t) &= m_{X_1}\left(\frac{t}{n}\right) m_{X_2}\left(\frac{t}{n}\right) \dots m_{X_n}\left(\frac{t}{n}\right) \\ &= \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \dots \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \\ &= \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-n}. \end{aligned}$$

Pero la expresión anterior es la función generadora de momentos de una distribución gama con parámetro de forma n y parámetro de escala θ/n . De acuerdo con lo anterior, cuando se muestrea una población cuya distribución de probabilidad es exponencial, la densidad de probabilidad de \bar{X} está dada por

$$f(\bar{x}; n, \theta/n) = \frac{n^n}{\Gamma(n)\theta^n} \bar{x}^{n-1} e^{-n\bar{x}/\theta}, \quad \bar{x} > 0.$$

■

Nótese que si en las expresiones $E(X) = \alpha\theta$ y $\text{Var}(X) = \alpha\theta^2$ se reemplaza α con n y θ con θ/n se obtiene

$$E(\bar{X}) = n \frac{\theta}{n} = \theta$$

y

$$\text{Var}(\bar{X}) = n \frac{\theta^2}{n^2} = \frac{\theta^2}{n},$$

como era de esperarse ya que θ y θ^2 son la media y la varianza de una variable aleatoria con distribución exponencial.

De la sección 5.5, recuérdese que si el parámetro de forma de una distribución gama tiene un valor grande, entonces los valores de probabilidad para una variable aleatoria gama pueden aproximarse, en forma adecuada, por una distribución normal. Dado que Γ^m , muestrear una distribución exponencial con parámetro θ , \bar{X} tiene una distribución gama con media θ , y desviación estándar $\frac{\theta}{\sqrt{n}}$. Entonces, para n grande

$$Z = \frac{\bar{X} - \theta}{\frac{\theta}{\sqrt{n}}}.$$

es, en forma aproximada, $N(0, 1)$.

Para ser que para un valor grande n , la distribución de \bar{X} es aproximadamente normal. De hecho, no importa el tipo de modelo de probabilidad a partir del cual se obtenga la muestra; muestras la media y la varianza existan, la distribución de muestreo de \bar{X} se encontrará aproximada por $N(\mu, \sigma/\sqrt{n})$ para valores grandes de n . Lo anterior constituye uno de los más importantes teoremas en inferencia estadística, y se conoce como **teorema central del límite**.

Teorema 1.4 Sean X_1, X_2, \dots, X_n variables aleatorias IID con una distribución de probabilidad no especificada y que tiene una media μ y varianza σ^2 finita. El promedio muestral $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a ∞ . En otras palabras, la variable aleatoria $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiene como límite una distribución normal estándar.

Demostración.- Se quiere demostrar que la función generadora de momentos de $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiende a la de una distribución normal estándar conforme n tiende al infinito. Sean

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, 2, \dots, n$$

y

$$Y = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Dado que

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \right) = \frac{1}{n} \frac{1}{\frac{\sigma}{\sqrt{n}}} \sum_{i=1}^n (X_i - \mu) = \frac{1}{n} \frac{1}{\frac{\sigma}{\sqrt{n}}} (n\bar{X} - n\mu) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Entonces,

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i.$$

Como resultado se tiene que la función generadora de momentos de Y es igual a la función generadora de momentos de $(1/\sqrt{n}) \sum_{i=1}^n Z_i$. Del teorema 7.1,

$$\begin{aligned} m_Y(t) &= [m_{Z_i}(t/\sqrt{n})]^n \\ &= \left\{ E \left[e^{\frac{tZ_i}{\sqrt{n}}} \right] \right\}^n, \end{aligned}$$

dado que las Z_i son variables aleatorias independientes.

Al expandir (tZ_i/\sqrt{n}) en una serie de Taylor:

$$e^{\frac{tZ_i}{\sqrt{n}}} = 1 + \frac{t}{\sqrt{n}}Z_i + \frac{t^2}{2n}Z_i^2 + \frac{t^3}{3!n^{\frac{3}{2}}}Z_i^3 + \dots$$

Si se toma los valores esperados y se recuerda que $E(Z_i) = 0$ y $Var(Z_i) = 1$, $i = 1, 2, \dots, n$ se tiene

$$E \left[e^{\frac{tZ_i}{\sqrt{n}}} \right] = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}E(Z_i^3) + \dots$$

De acuerdo con lo anterior

$$\begin{aligned} m_Y(t) &= \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}E(Z_i^3) + \dots \right]^n \\ &= \left\{ 1 + \frac{1}{n} \left[\frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \dots \right] \right\}^n \\ &= \left(1 + \frac{u}{n} \right)^n \end{aligned}$$

donde

$$u = \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \dots$$

Ahora

$$\lim_{n \rightarrow \infty} m_Y(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n} \right)^n.$$

Pero por definición

$$\lim_{n \rightarrow \infty} \left(1 + \frac{u}{n} \right) = e^u.$$

Lo anterior da como resultado una situación idéntica a la que se tiene en la demostración del teorema 5.1. Esto es, conforme $n \rightarrow \infty$, todos los términos en u , excepto el primero, tienden hacia cero debido a que todos tienen potencias positivas de n en sus denominadores. Por lo tanto, se puede deducir que

$$\lim_{x \rightarrow \infty} m_Y(t) = e^{\frac{t^2}{2}},$$

o la distribución límite de $Y = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ es la normal estándar para valores grandes de n . ■

La esencia del teorema central del límite recae en el hecho de que para n grande, la distribución $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ es, en forma aproximada, normal con media cero y desviación estándar uno, sin importar cuál sea el modelo de probabilidad a partir del que se obtuvo la muestra. Debe notarse que si el modelo de probabilidad de la población es semejante a una distribución normal (esto es, si es simétrico y existe una concentración relativamente alta alrededor del punto de simetría), la aproximación normal será buena aun para pequeñas muestras. Por otro lado, si el modelo de la población tiene muy poco parecido a una distribución normal (por ejemplo existe una alta asimetría), la aproximación normal sólo será adecuada para valores relativamente grandes de n . En muchos casos, puede concluirse de forma segura, que la aproximación será buena mientras $n > 30$. Por lo tanto, la variable aleatoria

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

se emplea para formular inferencias acerca de μ cuando se conoce el valor de la varianza población σ^2 .

1.5 La distribución de muestro de S^2

Otra estadística importante empleada para formular inferencias con respecto a las varianzas de la población es la varianza denotada por S^2 . Para comenzar, es necesario suponer que μ es conocida y σ^2 no. Así, σ^2 se encuentra definida por

$$S^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n},$$

en donde X_1, X_2, \dots, X_n constituye una muestra aleatoria de una distribución normal con media μ y varianza σ^2 desconocida. Para determinar una distribución de muestreo que permita hacer inferencias sobre σ^2 con base en S^2 definida por (7.13), se enuncia y demuestra el siguiente teorema.

Teorema 1.5 Sean X_1, X_2, \dots, X_n una muestra aleatoria de una distribución normal con media μ y varianza σ^2 . La distribución de la variable aleatoria

$$Y = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$$

es del tipo chi-cuadrada con n grados de libertad.

Demostración.- Dado que $X_i \sim N(\mu, \sigma)$, $i = 1, 2, \dots, n$, $Z_i = \frac{X_i - \mu}{\sigma}$ define n variables aleatorias normales estándar independientes, se tiene:

$$Y = \sum_{i=1}^n Z_i^2.$$

Del teorema 7.1,

$$\begin{aligned} m_Y(t) &= m_{Z_1^2}(t) m_{Z_2^2}(t) \dots m_{Z_n^2}(t) \\ &= (1 - 2t)^{-\frac{1}{2}} (1 - 2t)^{-\frac{1}{2}} \dots (1 - 2t)^{-\frac{1}{2}} \end{aligned}$$

dado que el cuadrado de una variable aleatoria normal estándar tiene una distribución chi-cuadrada con un grado de libertad (véase el ejemplo 5.14). De esta forma se tiene

$$m_Y(t) = (1 - 2t)^{-\frac{n}{2}}$$

que es la función generadora de momentos de una distribución chi-cuadrada con n grados de libertad. De acuerdo con lo anterior, $Y \sim X_n^2$. ■

Desde un punto de vista práctica, la varianza muestra tal como se encuentra definida por

$$S^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$$

tiene poco uso, ya que es muy raro que se conozca el valor de la media poblacional μ . De acuerdo con lo anterior, si se muestra una distribución normal con media μ y varianza σ^2 , la varianza muestral se define por

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1}$$

En el capítulo 8 se verá por qué se emplea el divisor $(n - 1)$. Para determinar la distribución de muestreo de S^2 , y con base en una muestra aleatoria proveniente de una distribución normal, debe tomarse en cuenta el promedio de la muestra \bar{X} . Como resultado se tiene que la distribución de muestreo de $\frac{(n - 1)S^2}{\sigma^2}$ es también una distribución chi-cuadrada con $n - 1$ grados de libertad. Para ello probaremos

primero un teorema útil que involucra suma de dos variables aleatorias independientes chi-cuadrada y entonces se escribirá la anterior expresión en una forma equivalente, con objeto de aprovechar este teorema.

Teorema 1.6 Si X_1 y X_2 son dos variables aleatorias independientes y cada una tiene una distribución chi-cuadrada con v_1 y v_2 grados de libertad respectivamente, entonces:

$$Y = X_1 + X_2$$

también tiene una distribución chi-cuadrada con $v_1 + v_2$ grados de libertad.

Demostración.- Del teorema 7.1, la función generadora de momentos de Y es

$$\begin{aligned} m_Y(t) &= m_{X_1}(t)m_{X_2}(t) \\ &= (1 - 2t)^{-v_1/2}(1 - 2t)^{-v_2/2} \\ &= (1 - 2t)^{-(v_1+v_2)/2} \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria chi-cuadrada con $v_1 + v_2$ grados de libertad. ■

Ahora se deducirá la distribución de muestreo de $\frac{(n-1)S^2}{\sigma^2}$. De

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$$

se tiene que

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

pero

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu - \bar{X} + \mu)^2 \\ &= \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

De esta forma

$$(n-1)S^2 + n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2.$$

Al dividir ambos miembros de la expresión anterior por la varianza poblacional σ^2 se tiene

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

Del teorema 7.5 se desprende que $\sum_{i=1}^n (X_i - \mu)^2 / \sigma^2$ tiene una distribución chi-cuadrada con n grados de libertad. De manera similar, $[(\bar{X} - \mu) / (\sigma / \sqrt{n})]^2$ también posee una distribución chi-cuadrada con un grado de libertad, dado que $(\bar{X} - \mu) / (\sigma / \sqrt{n})$ es $N(0, 1)$. Por lo tanto, si se supone que $(n-1)S^2 / \sigma^2$ y $[(\bar{X} - \mu) / (\sigma / \sqrt{n})]^2$ son variables aleatorias independientes, entonces por el teorema 7.6, cuando se muestrea una población cuya distribución es normal con media y varianza desconocida, la distribución de $(n-1)S^2 / \sigma^2$, es chi-cuadrada con n -grados de libertad. La función de densidad de probabilidad de $Y = (n-1)S^2 / \sigma^2$ se desprende de (5.58) y está dada por:

$$f(y; n-1) = \begin{cases} \frac{1}{\Gamma[(n-1)/2] 2^{(n-1)/2}} y^{[(n-1)/2]-1} e^{-y/2} & y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Nótese, que dado $Y \sim X_{n-1}^2$, $E(Y) = n-1$ y $Var(Y) = 2(n-1)$. Además, ya que $Y = (n-1)S^2 / \sigma^2$, $S^2 = \sigma^2 Y / (n-1)$. Por lo tanto

$$E(S^2) = E\left(\frac{\sigma^2 Y}{n-1}\right) = \frac{\sigma^2}{n-1} E(Y) = \sigma^2,$$

y

$$Var(S^2) = Var\left(\frac{\sigma^2 Y}{n-1}\right) = \frac{\sigma^4}{(n-1)^2} Var(Y) = \frac{2\sigma^4}{n-1}.$$

1.6 La distribución t de Student

Se recordará que cuando se muestrea una distribución normal con desviación estándar conocida σ , la distribución de $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$ es $N(0, 1)$. Desde un punto de vista práctico, la necesidad de conocer σ impide formular inferencias con respecto a μ debido a que generalmente no se conoce el valor de la desviación estándar de la población. Dada la disponibilidad de una muestra aleatoria, el camino lógico que se sigue en este caso es reemplazar σ con una estimación s , que es el valor de la desviación estándar muestral S . Desafortunadamente, cuando lo anterior se lleva a cabo, la distribución $(\bar{X} - \mu) / (S / \sqrt{n})$ no es $N(0, 1)$. Sin embargo es posible determinar la distribución de muestreo exacta cuando se muestra $N(\mu, \sigma)$, con μ y σ^2 desconocidos. Acá se examinará los aspectos teóricos de lo que se conoce como la distribución t de Student.

Supóngase que se realiza un experimento en que se observan dos variables aleatorias X y Z ; X tiene una distribución chi-cuadrada con v grados de libertad y Z una distribución normal con media cero y desviación estándar uno. Sea T otra variable aleatoria que es función de X y Z , de manera tal que

$$T = \frac{Z}{\sqrt{\frac{X}{v}}}.$$

Puesto que los valores de Z se encuentran entre $(-\infty, \infty)$ y los valores de X son positivos. El valor

$$t = \frac{z}{\sqrt{\frac{x}{v}}}$$

recibe el nombre de valor de la variable aleatoria de t de Student. Lo anterior lleva al siguiente teorema.

Teorema 1.7 Sea Z una variable aleatoria normal estándar y X una variable aleatoria chi-cuadrada con v grados de libertad. Si Z y X son independientes, entonces la variable aleatoria

$$t = \frac{z}{\sqrt{\frac{x}{v}}}$$

tiene una distribución t de Student con v grados de libertad y una función de densidad de probabilidad dada por

$$f(t, v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \left[1 + \left(\frac{t^2}{v}\right)\right]^{-\frac{(v+1)}{2}}, \quad -\infty < t < \infty, v > 0.$$

Deducción.- Sea T una variable aleatoria definida por $T = \frac{Z}{\sqrt{X/v}}$. Considere la densidad de probabilidad de T cuando X se mantiene fija en un valor x . Dado que

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

la densidad de probabilidad condicional de

$$T = \frac{Z}{(x/v)^{1/2}}$$

se obtiene al considerar la relación inversa

$$Z = (x/v)^{1/2} T$$

y al sustituir en $f_Z(z)$, en donde el jacobiano de la transformación es

$$\frac{dz}{dt} = (x/v)^{1/2}.$$

De esta forma

$$f(t|x) = \frac{(x/v)^{1/2}}{\sqrt{2\pi}} e^{-xt^2/2v}, \quad -\infty < t < \infty, \quad x > 0.$$

Se sabe que la densidad conjunta de T y X es

$$f(t, x) = f(t|x)f_X(x).$$

Dado que $X \sim X_v^2$,

$$f_X(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{(v-1)/2} e^{-x/2}, \quad x > 0.$$

De esta forma

$$\begin{aligned} f(t, x) &= \frac{1}{\sqrt{2\pi} 2^{v/2} \Gamma(v/2)} x^{(v-1)/2} e^{-\frac{x}{2} - \frac{xt^2}{2v}} \\ &= c_1 x^{(v-1)/2} e^{-c_2 x/2}, \end{aligned}$$

en donde $c_1 = \frac{1}{\sqrt{2\pi v} 2^{v/2} \Gamma(v/2)}$ y $c_2 = [1 + (t^2/v)]$. Integrando $f(t, x)$ con respecto a x , se obtiene la función de densidad de probabilidad de la distribución t de Student. De acuerdo con lo anterior

$$\begin{aligned}
 f_T(t) &= c_1 \int_0^\infty x^{(v-1)/2} e^{-c_2/2} dx \\
 &= c_1 \int_0^\infty (2y/c_2)^{(v-1)/2} e^{-y} (2/c_2) dy, \text{ donde } y = c_2 x/2 \text{ y } dx = (2/c_2) dy \\
 &= c_1 (2/c_2)^{(v+1)/2} \int_0^\infty y^{(v-1)/2} e^{-y} dy \\
 &= c_1 (2/c_2)^{(v+1)/2} \Gamma[(v+1)/2] \\
 &= \frac{1}{\sqrt{2\pi v} 2^{v/2} \Gamma(v/2)} \cdot \frac{2^{(v+1)/2}}{[1 + (t^2/v)]^{(v+1)/2}} \Gamma[(v+1)/2] \\
 &= \frac{\Gamma[(v+1)/2]}{\sqrt{\pi v} \Gamma[v/2]} \left[1 + \left(\frac{t^2}{v} \right) \right]^{-\frac{(v+1)}{2}} \quad -\infty < t < \infty.
 \end{aligned}$$

■

Conforme se tiene un número mayor de grados de libertad, la distribución t de Student tiende hacia la normal estándar.

Puede demostrarse que el valor esperado de T es

$$E(t) = 0, \quad v > 1$$

y la varianza está dada por

$$Var(T) = \frac{v}{v-2}, \quad v > 2.$$

En la tabla F se encuentra los valores cuantiles $t_{1-\alpha, 2}$ tales que:

$$P(T \leq t_{1-\alpha, v}) = \int_{-\infty}^{t_{1-\alpha, v}} f(t; v) dt = 1 - \alpha, \quad 0 \leq \alpha \leq 1.$$

para los distintos valores de v y de las proporciones acumulativas seleccionadas $1 - \alpha$. Por ejemplo, si $v = 15$,

$$\begin{aligned}
 P(T \leq t_{0.90, 15}) &= P(T \leq 1.341) = 0.90, \\
 P(T \leq t_{0.95, 15}) &= P(T \leq 1.753) = 0.95, \\
 P(T \leq t_{0.99, 15}) &= P(T \leq 2.602) = 0.99.
 \end{aligned}$$

Dado que la distribución t es simétrica con respecto al cero, para $\alpha > 0.5$ los valores cuantiles $t_{1-\alpha, v}$ serán negativos pero sus magnitudes serán las mismas que las de los correspondientes valores que se encuentran en el lado derecho. De esta forma, para $v = 15$,

$$\begin{aligned}
 P(T \geq t_{0.10, 15}) &= P(T \leq -1.341) = 0.10, \\
 P(T \geq t_{0.05, 15}) &= P(T \leq -1.753) = 0.05, \\
 P(T \geq t_{0.01, 15}) &= P(T \leq -2.602) = 0.01.
 \end{aligned}$$

Desde un punto de vista práctico, es muy poca la ganancia que se tiene al emplear la distribución t de Student en lugar de la normal estándar cuando $v \geq 30$.

Recuérdese que para formular inferencias con respecto a μ cuando el muestreo se lleva a cabo sobre una distribución normal con media y varianza desconocida, se necesita determinar la distribución $(\bar{X} -$

$\mu)(S/\sqrt{n})$. Cuando se muestrea una distribución $N(\mu, \sigma)$ se sabe, del teorema 7.3, que la distribución de $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ es $N(0, 1)$. Para la misma condición, se sabe que, de

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

y del teorema 7.6, la distribución de $(n-1)S^2/\sigma^2$ es chi-cuadrada con $n-1$ grados de libertad. Dado que puede demostrarse que \bar{X} y S^2 son independientes, del teorema 7.7 se desprende que la distribución de

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \frac{1}{n-1}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma}{\sqrt{S^2}}, \quad \text{o} \quad T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}},$$

es la t de Student con $n-1$ grados de libertad.

1.7 La distribución de la diferencia entre dos medias muestrales

Se necesita obtener la distribución \bar{X} e \bar{Y} cuando el muestreo se lleva a cabo sobre dos poblaciones normales independientes con varianzas iguales. Si se supone que el valor de la varianza σ^2 se conoce del teorema 7.3, se sabe que la distribución de \bar{X} es normal con media μ_X y varianza $\frac{\sigma^2}{n_X}$. La distribución de \bar{Y} también es normal pero con media μ_Y y varianza $\frac{\sigma^2}{n_Y}$. Dado que \bar{X} y \bar{Y} son variables aleatorias independientes normalmente distribuidas, si $a_1 = 1$ y $a_2 = -1$ en el teorema 7.2, la distribución de $\bar{X} - \bar{Y}$ también es normal con media $\mu_X - \mu_Y$ y varianza $\left(\frac{\sigma^2}{n}\right) + \left(\frac{\sigma^2}{n_Y}\right) = \sigma^2 \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$. Por tanto, si se conoce el valor de σ^2 , la distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

es $N(0, 1)$ con varianzas iguales.

En este desarrollo se supone que el valor de σ^2 era conocido. Sin embargo, es poco probable conocer el valor de σ^2 para una situación real. Así pues, debe obtenerse la distribución de $\bar{X} - \bar{Y}$ cuando el muestreo se lleve a cabo sobre dos poblaciones normales independientes con varianzas iguales pero desconocidas. Para cada una de las dos muestras aleatorias, pueden definirse las varianzas muestrales S_X^2 y S_Y^2 dadas por $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$. Dado que $\frac{(n_X-1)S_X^2}{\sigma^2}$ y $\frac{(n_Y-1)S_Y^2}{\sigma^2}$ son dos variables independientes chi-cuadrada, con n_X-1 y n_Y-1 grados de libertad respectivamente, por el teorema 7.6, la distribución de

$$W = \frac{(n_X-1)S_X^2}{\sigma^2} + \frac{(n_Y-1)S_Y^2}{\sigma^2}$$

también es chi-cuadrada con $n_X + n_Y - 2$ grados de libertad. De la expresión $T = \frac{Z}{\sqrt{X/v}}$ se desprende el hecho de que el cociente de Z en

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

y la raíz cuadrada de W dividida entre sus grados de libertad tiene una distribución t de Student con $n_X + n_Y - 2$ grados de libertad. Esto es,

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{\sigma^2} \frac{1}{n_X + n_Y - 2}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2} \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

o

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}},$$

en donde

$$S_p^2 = \frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2}.$$

Que en general recibe el nombre de **estimador combinado de la varianza común** σ^2 .

Si las varianzas σ_X^2 y σ_Y^2 no son iguales, pero conocen sus valores, el problema es sencillo. La distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

aún es $N(0, 1)$.

1.8 La distribución F