DISTRIBUCION JI-CUADRADO E INFERENCIAS SOBRE LA VARIANZA

Sección V.2
(c)
Milton
De la página 211 a la 215

En este capítulo presentamos otra importante familia de variables aleatorias continuas. Estas variables aleatorias, llamadas ji-cuadrado, se utilizan en muchos de los diferentes entornos estadísticos. Aquí se presenta su aplicación para sacar conclusiones relacionadas con la varianza de la población.

7.1 DISTRIBUCION JI-CUADRADO Y ESTIMACION DE INTERVALO DE LA VARIANZA POBLACIONAL

Ya se ha hablado de que la varianza muestral S^2 es un estimador puntual lógico de la varianza de la población σ^2. Es posible demostrar que dividiendo $\Sigma(X - \bar{X})^2$ por $n - 1$ en la definición de S^2, obtenemos un estimador insesgado de σ^2. Las estimaciones de σ^2 obtenidas en repetidas muestras de la misma población fluctúan alrededor de la verdadera varianza de la población. Por el momento, las estimaciones de σ^2 interesan en relación a que son necesarias para evaluar el estadístico T utilizado para hacer inferencias sobre la media de la población.

En algunas ocasiones el interés no se centra en la media sino en la misma varianza. Por lo tanto, no sólo es necesario poder hallar una estimación puntual de σ^2, sino también construir un intervalo de confianza o hacer un test de hipótesis en relación a este parámetro.

Ejemplo 7.1.1. El cobre, mineral requerido en algún grado por la mayoría de las plantas, se considera un micronutriente. Su concentración en una planta se mide en parte por millón, y se determina quemando totalmente la planta y analizando las cenizas. La concentración de cobre varía de una especie a otra. Se diseña un experimento para estimar esta variabilidad.

Ejemplo 7.1.2. El Lhasa Apso es un perro de ascendencia tibetana. La raza fue introducida en Inglaterra de 1921, y criada para exhibición por primera vez en 1933. La alzada deseada en los machos es de 10.5 pulgadas. En todo caso, hay una buena cantidad de variabilidad en la altura. Un criador está intentando reducir esta variabilidad en los animales mediante una crianza selectiva.

Ejemplo 7.1.3. Al manufacturar un fármaco, la potencia varía de uno a otro. No se debe permitir que esta variabilidad sea demasiado grande. Una varianza muy grande puede dar como resultado que algunos lotes contengan fármaco demasiado débil para que sea efectivo, mientras que otros contengan fármaco demasiado fuerte y potencialmente peligroso. Se realizan pruebas periódicas para controlar la varianza del fármaco que está siendo producido.

En cada uno de estos tres ejemplos se requiere que se infiera la varianza de una población. Para construir un intervalo de confianza de σ^2 o para contrastar hipótesis relativas a su valor, es necesario introducir otra familia de variables aleatorias continuas. Las características generales de los miembros de esta familia, llamada *variable-ji-cuadrado* (X^2), son las siguientes:

1. Hay un número infinito de variables aleatorias ji-cuadrado, cada una identificada por un parámetro γ, llamados *grados de libertad*. El parámetro γ es siempre un entero positivo. La notación X^2_γ, designa una variable ji-cuadrado con γ grados de libertad.
2. Cada variable ji-cuadrado es continua.
3. La gráfica de la densidad de cada variable ji-cuadrado es una curva asimétrica que presenta generalmente la forma que aparece en la Figura 7.1.
4. Las variables ji-cuadrado no pueden tomar valores negativos.
5. El parámetro γ es al mismo tiempo un parámetro de forma y un parámetro de situación en el que

$$E[X^2_\gamma] = \gamma \qquad \text{y} \qquad \text{Var } X^2_\gamma = 2\gamma$$

Es decir, el valor medio de una variable aleatoria ji-cuadrado es el mismo que sus grados de libertad y su varianza es el doble de sus grados de libertad.

En la Tabla VIII del Apéndice B se presenta una recopilación parcial de valores de la función de distribución acumulativa con varios grados de libertad para las variables

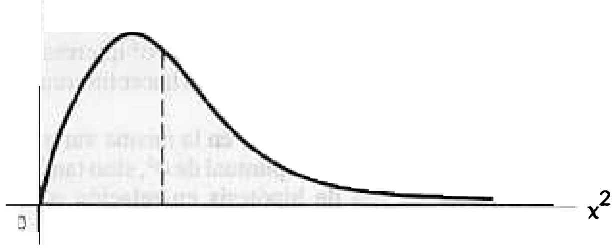

Figura 7.1
Curva ji-cuadrado típica. El punto de equilibrio de la curva es γ, sus grados de libertad.

ji-cuadrado. En esta tabla los grados de libertad aparecen encabezando las filas, las probabilidades encabezando las columnas, y los puntos asociados con esas posibilidades en el cuerpo de la tabla.

Ejemplo 7.1.4. Considérese una variable aleatoria ji-cuadrado con diez grados de libertad. Esta variable aleatoria se indica por X^2_{10}. Su valor medio es 10, el mismo que sus grados de libertad, y su varianza es 20, el doble de sus grados de libertad.

a) $P[X^2_{10} \leq 2.56] = F(2.56) = 0.01$.
b) $P[X^2_{10} \geq 16] = 1 - F(16) = 1 - 0.90 = 0.1$.
c) El punto con el 5% del área a su derecha y el 95% a su izquierda se muestra en la Figura 7.2. Su valor numérico, hallado en la fila 10 y la columna 0.95 de la tabla de ji-cuadrado, es 18.3.

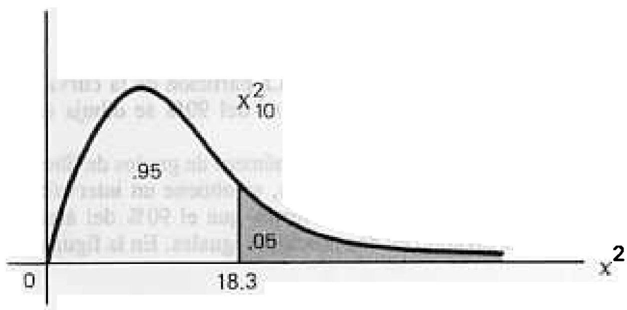

Figura 7.2
$P[X^2_{10} \geq 18.3] = 0.05$.

d) El punto con el 95% del área a su derecha y el 5% a su izquierda se halla en la fila 10 y en la columna 0.05. Su valor numérico es 3.94.
e) $P[3.25 \leq X^2_{10} \leq 20.5] = 0.95$ (Ver Figura 7.3).

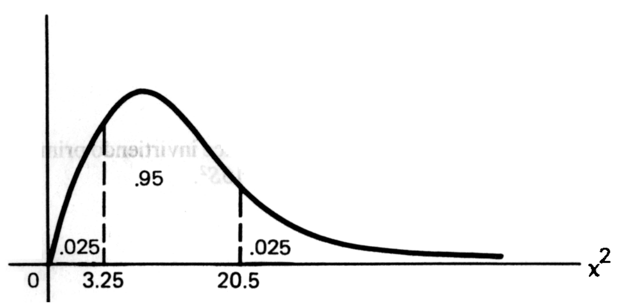

Figura 7.3
$P[3.25 \leq X^2_{10} \leq 20.5] = 0.95$.

Para construir un intervalo de confianza de σ^2 necesitamos una variable aleatoria que contenga este parámetro en su expresión y cuya distribución se conozca. Si es posible hallar dicha variable aleatoria, se puede utilizar un argumento algebraico similar al utilizado para construir un intervalo de confianza de μ para determinar los límites de confianza de σ^2. El Teorema 7.1.1 nos proporciona la variable aleatoria necesaria.

Teorema 7.1.1. Sea $X_1, X_2, X_3, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución que es normal, con media μ y varianza σ^2. La variable aleatoria $(n-1)S^2/\sigma^2$ está distribuida como una variable aleatoria ji-cuadrado con $n-1$ grados de libertad.

El Ejemplo 7.1.5 ilustra cómo se utiliza este teorema para construir un intervalo de confianza de σ^2.

Ejemplo 7.1.5. Se extrajo una muestra aleatoria de 16 plantas para estimar la varianza en la concentración de cobre en las plantas halladas en el New River Valley. Las plantas fueron quemadas, y se analizaron las cenizas. Se obtuvieron las siguientes observaciones con respecto a X, concentraciones de cobre (en partes por millón) (supóngase que X está normalmente distribuida):

5	3	34	18	27	14
8	50	38	43	35	
20	70	25	60	19	

Para estos datos $\Sigma x = 469$, $\Sigma x^2 = 19{,}407$ y $s^2 = 377.30$. La partición de la curva X^2_{15} que se necesita para construir un intervalo de confianza de σ^2 del 90% se dibuja en la Figura 7.4.

Obsérvese que, puesto que la muestra es de tamaño 16, el número de grados de libertad es 15, el tamaño de la muestra menos 1. Como se hizo antes, se obtiene un intervalo de confianza del 90% mediante la partición de la curva, de forma que el 90% del área se encuentre en el centro y el 10% restante dividido en dos porciones iguales. En la figura 7.4 se ve claro que

$$P[7.26 \le X^2_{15} \le 25] = 0.90$$

En este caso, la variable aleatoria ji-cuadrado es

$$\frac{(n-1)S^2}{\sigma^2} = \frac{15S^2}{\sigma^2}$$

Por lo tanto, sabemos que

$$P\left[7.26 \le \frac{15S^2}{\sigma^2} \le 25\right] = .90$$

Nuestro objetivo es despejar σ^2 en medio de la desigualdad. Esto se hace invirtiendo primero cada trozo de la desigualdad y multiplicando, a continuación, por $15S^2$.

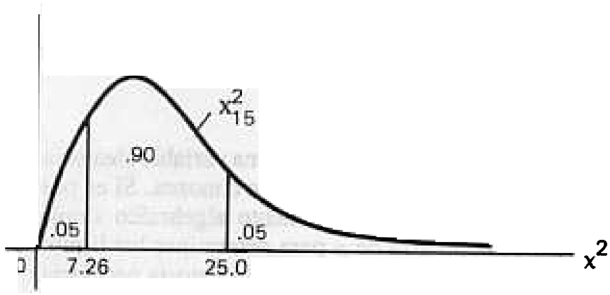

Figura 7.4
Partición de X^2_{15} para obtener un intervalo de confianza de σ^2 del 90%.

Esto da como resultado la siguiente secuencia de pasos:

$$P\left[7.26 \le \frac{15S^2}{\sigma^2} \le 25\right] = .90$$

$$P\left[\frac{1}{25} \le \frac{\sigma^2}{15S^2} \le \frac{1}{7.26}\right] = .90$$

$$P\left[\frac{15S^2}{25} \le \sigma^2 \le \frac{15S^2}{7.26}\right] = .90$$

El límite inferior del intervalo de confianza es

$$\frac{15S^2}{25} = \frac{15(377.30)}{25.0} = 226.38$$

y el límite superior es

$$\frac{15S^2}{7.26} = \frac{15(377.30)}{7.26} = 779.55$$

Tenemos un 90% de seguridad de que la verdadera varianza en la concentración de cobre en el New River Valley está entre 226.38 y 779.54. Para hallar un intervalo de confianza del 90%, de la desviación típica de la población, necesitamos solamente sacar la raíz cuadrada de esos límites numéricos. De este modo podemos decir que tenemos un 90% de confianza en que la verdadera desviación típica de la población esté entre $\sqrt{226.38} = 15.05$ y $\sqrt{779.54} = 27.92$ partes por millón.

La técnica utilizada en el Ejemplo 7.1.5 puede ser generalizada para obtener cualquier nivel de confianza deseado eligiendo adecuadamente los puntos de la tabla de ji-cuadrado. Obsérvese que, dado que la distribución ji-cuadrado no es simétrica, en la tabla deben leerse dos puntos diferentes. El punto que aparece en el límite inferior se designará mediante χ_1^2. Es el más grande de los dos puntos leídos en la tabla. El punto que aparece en el límite superior se indicará mediante χ_2^2. Los límites de confianza generales se dan en el Teorema 7.1.2.

Teorema 7.1.2. Intervalo de confianza de σ^2. Sea $X_1, X_2, X_3, \ldots, X_n$ una muestra aleatoria de tamaño n de una distribución que es normal con media μ y varianza σ^2. Los límites inferior y superior, respectivamente, de un intervalo de confianza de σ^2 son

$$L_1 = \frac{(n-1)S^2}{\chi_1^2} \qquad L_2 = \frac{(n-1)S^2}{\chi_2^2}$$

donde los puntos χ^2 son puntos de la distribución ji-cuadrado con $n - 1$ grados de libertad.

Al igual que en el caso de la distribución T, puede hallarse la aproximación de los puntos ji-cuadrado para muestras grandes utilizando la distribución normal estándar. La forma de hacerlo se muestra en el Ejercicio 5 al final de esta sección. Este ejercicio posibilita la construcción de intervalos de confianza sobre σ^2 para muestras mayores que las que pueden manejarse utilizando la tabla de ji-cuadrado.