Técnicas de Inferencia Estadística II

Tema 4. Contrastes para la mediana y otros cuantiles

M. Concepción Ausín
Universidad Carlos III de Madrid

Grado en Estadística y Empresa
Curso 2014/15

Contenidos

Introducción

- En muchas situaciones que se observan frecuentemente en la práctica, utilizando las técnicas de bondad de ajuste del tema 3, se concluye que una muestra dada no sigue una distribución normal.
- En este caso, los contrastes paramétricos para poblaciones normales analizados en el tema 2 no son válidos.
- En estas situaciones todavía se pueden plantear y contrastar hipótesis acerca de algunas características importantes de la distribución poblacional. En particular, una de las medidas de mayor interés en la práctica es la mediana.

Introducción

- Para simplificar el problema, consideraremos en este tema exclusivamente variables continuas.

- Recordamos que la mediana poblacional de una variable aleatoria continua, X, es el único valor, $Q_{0,5}$, que divide la distribución en dos partes con la mitad de probabilidad:

$$\Pr(X \geq Q_{0,5}) = \Pr(X \leq Q_{0,5}) = \frac{1}{2}$$

- Por tanto, para variables continuas la mediana es el único valor en el que la función de distribución es igual a 0.5:

$$F(Q_{0,5}) = \frac{1}{2}$$

- La mediana poblacional se estima con la mediana muestral (Me) que divide a la muestra ordenada en dos mitades iguales.

Contraste de los signos para la mediana

Suponemos una muestra aleatoria simple $(X_1, \ldots, X_n)$ de una población desconocida con función de distribución continua.

Queremos resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : Q_{0,5} = m_0 & \qquad H_0 : Q_{0,5} = m_0 & \qquad H_0 : Q_{0,5} = m_0 \\ H_1 : Q_{0,5} \neq m_0 & \qquad H_1 : Q_{0,5} < m_0 & \qquad H_1 : Q_{0,5} > m_0 \end{array}$$

Si la hipótesis nula es cierta, el número de observaciones en la muestra mayores que m_0 debe se aproximadamente el mismo que el número observaciones menores que m_0.

Así, denotamos con $(+)$ si la observación está por encima de m_0 y con (-) en caso contrario.

Contraste de los signos para la mediana

Considerando las variables:

$$Y_i = \left\{ \begin{array}{ll} 1, & \text{si } (+) \\ 0, & \text{si } (\text{-}) \end{array} \right\}$$

para $i = 1, \ldots, n$, podemos considerar la variable "Número de observaciones menores que m_0" que se obtiene mediante $\sum_{i=1}^{n} Y_i$.

Observamos que si la hipótesis nula es cierta,

$$\Pr(Y_i = 1) = \Pr(+) = P(X \geq m_0) = 0{,}5$$

Así, podemos definir el siguiente estadístico de contraste:

$$\sum_{i=1}^{n} Y_i \sim_{H_0} Bin\,(n, 0{,}5)$$

Contraste de los signos para la mediana

Cálculo del p-valor

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} > m_0$

$$\text{p-valor} = \Pr\left(Bin(n, 0{,}5) \geq \sum_{i=1}^{n} y_i\right)$$

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} < m_0$

$$\text{p-valor} = \Pr\left(Bin(n, 0{,}5) \leq \sum_{i=1}^{n} y_i\right)$$

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} \neq m_0$

$$\text{p-valor} = \text{mín}\left\{2\Pr\left(Bin(n, 0{,}5) \geq \sum_{i=1}^{n} y_i\right), 2\Pr\left(Bin(n, 0{,}5) \leq \sum_{i=1}^{n} y_i\right)\right\}$$

Contraste de los signos para la mediana

Ejemplo 4.1.

Un pequeño comercio desea saber si la mediana del gasto por cliente en la tienda es mayor de los 48 euros que espera de beneficio. El gasto de los últimos 10 clientes observados ha sido: 80, 75, 65, 84, 40, 60, 49, 50, 38 y 39 euros. Contrastar esta afirmación al nivel $\alpha = 0.05$.

Contraste de los signos para una proporción

Utilizando el test de los signos se pueden resolver también contrastes sobre una proporción.

Consideramos que se tiene una muestra $(Y_1, Y_2, \ldots, Y_n)$ de una población Bernouilli, $Bin(1, p)$:

$$Y = \begin{cases} 1, & \text{con probabilidad } p \\ 0, & \text{con probabilidad } 1 - p \end{cases}$$

Se pueden resolver contrastes para la proporción:

$$\begin{array}{lll} H_0 : p = p_0 & H_0 : p = p_0 & H_0 : p = p_0 \\ H_1 : p \neq p_0 & H_1 : p > p_0 & H_1 : p < p_0 \end{array}$$

utilizando el estadístico de contraste:

$$\sum_{i=1}^{n} Y_i \sim_{H_0} Bin\,(n, p_0)$$

Contraste de los signos para una proporción

Ejemplo 4.2.

Las calificaciones ordenadas obtenidas por doce estudiantes de un curso han sido: 2.7 4.2 5.3 6.1 6.7 7.2 8.5 8.7 8.9 9.5 9.6 9.7 9.8 10. Contrastar mediante el test de los signos la hipótesis de que la proporción de estudiantes que aprueban es superior al 75 %

Contraste de los signos para una proporción

Ejemplo 4.3.

Una agencia de viajes asegura que en una cierta localidad turística más del 75 % de los días de verano se registra una temperatura máxima inferior a 30 grados. Contrastar mediante el test de los signos dicha hipótesis al 90 %, utilizando la siguiente muestra de temperaturas máximas registradas en 10 días de verano: 30, 32, 36, 29, 27, 28, 31, 27, 31, 29.

Contraste de los signos para una proporción

Cuando el tamaño muestral, n, es grande, se puede aproximar la distribución binomial por una normal:

$$\sum_{i=1}^{n} Y_i \sim_{H_0} Bin\,(n, p_0) \to N(np_0, np_0(1 - p_0))$$

Luego, se tiene que:

$$\hat{p} = \frac{\sum\limits_{i=1}^{n} Y_i}{n} \to_{H_0} N\left(p_0, \frac{p_0(1 - p_0)}{n}\right)$$

y podemos usar el siguiente estadístico de contraste:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to_{H_0} N(0, 1)$$

donde,

$$\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}.$$

Ejemplo 4.4.

Se trabaja con la hipótesis de que uno de cada diez varones manifiesta algún tipo de daltonismo.

- *Elegidos 400 varones, se detectan 50 daltónicos. Con un nivel de significación del 10 %, ¿se puede rechazar la hipótesis de partida? ¿se obtendrá la misma conclusión si el nivel de significación es del 2 %?*

Contraste de los rangos signados de Wilcoxon

Suponemos una muestra aleatoria simple $(X_1, \ldots, X_n)$ de una población desconocida con función de distribución continua y simétrica.

Queremos resolver contrastes sobre la mediana poblacional:

$$\begin{array}{lll} H_0 : Q_{0,5} = m_0 & \qquad H_0 : Q_{0,5} = m_0 & \qquad H_0 : Q_{0,5} = m_0 \\ H_1 : Q_{0,5} \neq m_0 & \qquad H_1 : Q_{0,5} < m_0 & \qquad H_1 : Q_{0,5} > m_0 \end{array}$$

Para resolver estos contrastes se calculan las diferencias:

$$|X_1 - m_0|, |X_2 - m_0|, \ldots, |X_n - m_0|$$

y las ordenamos en orden creciente, asignando a cada X_i su rango.

Por ejemplo, si $n = 4$ y se tiene:

$$|X_2 - m_0| < |X_3 - m_0| < |X_1 - m_0| < |X_4 - m_0|,$$

entonces, los rangos de X_i, para $i = 1, \ldots, 4$, son:

$$rg(X_1) = 3; \quad rg(X_2) = 1; \quad rg(X_3) = 2; \quad rg(X_4) = 4.$$

Contraste de los rangos signados de Wilcoxon

El estadístico de Wilcoxon, T^+ es:

$$T^+ = \text{Suma de los rangos de las } X_i \text{ mayores que } m_0$$

$$T^- = \text{Suma de los rangos de las } X_i \text{ menores que } m_0$$

Claramente, bajo H_0, es de esperar que: T^+ y T^- sean iguales. Además,

$$T^+ + T^- = \sum_{i=1}^{n} i = \frac{n\,(n+1)}{2},$$

por lo que basta tener en consideración T^+, cuya distribución es conocida.

Contraste de los rangos signados de Wilcoxon

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} > m_0$

$$\text{p-valor} = \Pr\left(T^+ > t^+\right)$$

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} < m_0$

$$\text{p-valor} = \Pr\left(T^+ < t^+\right)$$

- $H_0 : Q_{0,5} = m_0 \quad vs \quad H_1 : Q_{0,5} \neq m_0$

$$\text{p-valor} = \text{mín}\left\{2\Pr(T^+ < t^+), \quad 2\Pr(T^+ > t^+)\right\}$$

donde t^+ es la suma de los rangos de las observaciones mayores que m_0.