FEBRERO 2018. EJERCICIO 1

Suponga que en un muestreo aleatorio simple de tamaño 50, sobre una población $N_3(\mu, V)$, se han obtenido los siguientes resúmenes estadísticos:

$$\bar{x} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \qquad S = \begin{pmatrix} 5 & & \\ -2 & 4 & \\ -3 & 1 & 2 \end{pmatrix}$$

(a) Contraste la hipótesis de que la media poblacional tiene componentes nulas.

(b) ¿Qué decidiría sobre la hipótesis de esfericidad?

(a) Para realizar este contraste planteamos las siguientes hipótesis:

$$H_0: \mu = (0 \ 0 \ 0)', \quad V = \text{cualquiera}$$
$$H_1: \mu \neq (0 \ 0 \ 0)', \quad V = \text{cualquiera}$$

Teniendo en cuenta que nuestra población es normal, obtendremos el estadístico de contraste mediante el método de razón de verosimilitudes, obteniendo el máximo de la función soporte para una normal, bajo H_0 y bajo H_1. Utilizaremos para ello los estimadores MV de μ y V, que son $\bar{x}$ y S, respectivamente.

El estimador que se obtiene aplicando el método de razón de verosimilitudes viene dado por:

$$\lambda = 2\left(L(H_1) - L(H_0)\right)$$

que para este contraste es:

$$\lambda = n \log \frac{|S_0|}{|S|}$$

donde S_0 es la varianza generalizada bajo H_0. Se puede demostrar que este estadístico se puede escribir en función del estadístico T^2 de Hotelling, dado que:

$$\frac{|S_0|}{|S|} = 1 + \frac{T^2}{n-1}$$

luego será:

$$\lambda = n\left(\log 1 + \frac{T^2}{n-1}\right)$$

Como tenemos una muestra grande ($n > 30$) y se sabe que para n grande es

$$\log\left(1 + \frac{a}{n}\right) \approx \frac{a}{n}$$

entonces se puede escribir:

$$\lambda \approx n \cdot \frac{T^2}{n-1} \approx T^2 \quad , \text{ pues } \frac{n}{n-1} \approx 1$$

siendo T^2 el estadístico de Hotelling que viene dado por:

$$T^2 = (n-1)(\bar{x} - \mu_0)' \cdot S^{-1} \cdot (\bar{x} - \mu_0) \quad , \quad T^2 \sim T^2(p, n-1)$$

En nuestro caso tenemos:

$$n = 50 \qquad \bar{x} = (2 \ \ 1 \ \ 3)' \qquad \mu_0 = (0 \ \ 0 \ \ 0)' \quad \text{y}$$

$$S = \begin{pmatrix} 5 & -2 & -3 \\ -2 & 4 & 1 \\ -3 & 1 & 2 \end{pmatrix} \qquad \text{luego:}$$

$$T^2 = 49 \cdot (2 \ \ 1 \ \ 3) \cdot \begin{pmatrix} 5 & -2 & -3 \\ -2 & 4 & 1 \\ -3 & 1 & 2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = 4949$$

El valor obtenido para T^2 es lo suficientemente grande como para rechazar la hipótesis nula, en consecuencia, admitimos que la media para esta población no es $\bar{x} = (0 \ \ 0 \ \ 0)'$.

NOTA.- Podríamos haber utilizado la relación existente entre la distribución T^2 de Hotelling y la F de Fisher:

$$F = \frac{n-p}{p(n-1)} \cdot T^2(p, n-1)$$

que en este caso hubiera conducido al valor 1582'3 ($n = 50$, $p = 3$) y observaríamos que el contraste se rechazaría si $F > F_{p, n-p, \alpha}$, es decir si:

$$1582'3 > F_{3, 47, \alpha}$$

y se observa que $F = 1582'3$ es mayor que $F_{3,47}$ para los niveles de significación habituales. Sin embargo, entendemos que el valor obtenido para T^2 es lo suficientemente grande como para rechazar la hipótesis nula planteada.

(b) El contraste de esfericidad plantea las mismas hipótesis que el contraste de independencia, es decir, que la matriz de covarianzas es diagonal, con el añadido de que supone que todas las variables tienen la misma varianza y están incorreladas. Esto equivale a decir que la matriz V_0 es de la forma $\sigma^2 I$. Las hipótesis son, por tanto:

$$H_0: \ V = \sigma^2 I, \quad \mu = \text{cualquiera}$$
$$H_1: \ V \text{ y } \mu = \text{cualquiera}$$

De nuevo utilizamos el método de razón de verosimilitudes, calculando el máximo del soporte bajo Ho y bajo H_1. Bajo Ho se tiene que $V = \sigma^2 I$ y μ será sustituida por $\bar{x}$, que es su estimador MV. Bajo H_1, μ y V se sustituyen por $\bar{x}$ y S, sus estimadores MV. La aplicación del método nos lleva a:

$$\lambda = 2\left(L(H_1) - L(H_0) \right)$$

que en este caso viene dado por $\lambda = np \log \hat{\sigma}^2 - n \log |S|$, donde $\hat{\sigma}^2$ es el estimador MV de σ^2 que se obtiene derivando en $L(H_0)$ respecto de σ^2 y viene dado por $\hat{\sigma}^2 = trS/p$. Por tanto:

$$\lambda = np \log \frac{trS}{p} - n \log |S|$$

Con los datos de nuestro enunciado es:

$$n = 50 \qquad p = 3 \qquad trS = 11 \qquad |S| = 3$$

luego:

$$\lambda = 50 \cdot 3 \cdot \log \frac{11}{3} - 50 \cdot \log 3 = 139'96$$

Este estadístico se distribuye según χ^2_g donde $g = \frac{p(p+1)}{2} - 1$ son los grados de libertad, que para $p = 3$ es $g = 5$.
La hipótesis nula será rechazada cuando sea $\lambda > \chi^2_{5,\alpha}$. El valor obtenido es claramente significativo pues a los niveles de significación habituales los valores χ^2_5 son mucho menores.
Se concluye el rechazo de la hipótesis de que las variables tengan la misma varianza y estén incorreladas.