FEBRERO 2016. EJERCICIO 1

En un muestreo aleatorio simple de tamaño 80, sobre una población $N_3(\mu, V)$ se han obtenido los siguientes resúmenes estadísticos:

$$\bar{x} = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} \qquad S = \begin{pmatrix} 10 & 6 & 7 \\ & 8 & 5 \\ & & 9 \end{pmatrix}$$

(a) ¿Qué justificación tiene el uso del estadístico T^2 de Hotelling para el contraste de hipótesis sobre la media poblacional? ¿Avalan los datos la hipótesis de que $\mu = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$?

(b) Contraste la hipótesis de que las componentes son independientes y con varianza común.

(a) El contraste sobre la media poblacional viene dado por las siguientes hipótesis:

$$H_0: \mu = \mu_0, \quad V = \text{cualquiera}$$
$$H_1: \mu \neq \mu_0, \quad V = \text{cualquiera}$$

El estadístico de contraste, obtenido por el método de razón de verosimilitudes viene dado por:

$$\lambda = 2\big(L(H_1) - L(H_0)\big) \qquad (1)$$

donde $L(H_0)$ es el máximo de la función soporte para una población normal bajo H_0, que se obtiene sustituyendo en el soporte μ por μ_0 y V por su estimador MV, que es S. Lo mismo para $L(H_1)$ que es el máximo del soporte bajo H_1 y que se obtiene sustituyendo en el soporte μ y V por sus respectivos estimadores MV, que son $\bar{x}$ y S, respectivamente. Una vez sustituidas en (1) y realizadas las operaciones oportunas se obtiene:

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

donde S_0 es la varianza generalizada bajo H_0.

Se puede demostrar que:

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

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

$$T^2 = (n-1)(\bar{x}-\mu_0)' S^{-1} (\bar{x}-\mu_0)$$

Así, el estadístico λ quedaría:

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

Como nuestra muestra es grande y sabemos que para n grande se verifica que:

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

podemos escribir:

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

De aquí que podamos utilizar en nuestro contraste el estadístico T^2 de Hotelling como valor de λ.

Con los datos del problema tenemos:

$$n = 80 \qquad (\bar{x}-\mu_0)' = (1 \ -2 \ -2) \qquad y \qquad S = \begin{pmatrix} 10 & 6 & 7 \\ 6 & 8 & 5 \\ 7 & 5 & 9 \end{pmatrix}$$

con lo que:

$$T^2 = 79 \cdot (1 \ -2 \ -2) \begin{pmatrix} 10 & 6 & 7 \\ 6 & 8 & 5 \\ 7 & 5 & 9 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix} =$$

$$= \frac{79}{174} \cdot (1 \ -2 \ -2) \cdot \begin{pmatrix} 47 & -19 & -26 \\ -19 & 41 & -8 \\ -26 & -8 & 44 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix} =$$

$$= \frac{79}{174} \cdot (137 \ -85 \ -98) \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix} = \frac{79}{174} \cdot 503 = 228'37$$

La hipótesis nula se rechazará en el caso de que T^2 sea lo suficientemente grande. No obstante podemos utilizar la relación que hay entre la T^2 de Hotelling y la F de Fisher:

$$F = \frac{n-p}{p(n-1)} \cdot T^2(n-1, p) = \frac{80-3}{3 \cdot 79} \cdot 228'37 = 74'20$$

De esta manera, la hipótesis nula será rechazada cuando sea

$$F = 74'20 > F_{p,n-p,\alpha} \rightarrow 74'20 > F_{3,77;\alpha}$$

(b) La hipótesis de que las variables son independientes y con varianza común es la hipótesis de esfericidad. Esta hipótesis equivale a suponer que la matriz de covarianzas es de la forma $\sigma^2 I$. Por tanto, las hipótesis son:

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

De nuevo, aplicando el método de razón de verosimilitudes, calculamos el máximo del soporte bajo Ho y bajo H_1, sustituyendo $V = \sigma^2 I$ y $\mu = \bar{x}$ para obtener $L(H_0)$ y $V = S$, $\mu = \bar{x}$, para obtener $L(H_1)$. Tras las operaciones se llega a:

$$\lambda = 2\left(L(H_1) - L(H_0)\right) =$$
$$= np \log \hat{\sigma}^2 - n \log |S|$$

donde $\hat{\sigma}^2$ es el estimador MV de σ^2 que viene dado por:

$$\hat{\sigma}^2 = \frac{trS}{p}$$

obtenido derivando respecto de σ^2 en la expresión de $L(H_0)$.

Así pues:

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

Con los datos del enunciado será:

$$n = 80 \qquad p = 3 \qquad trS = 27 \qquad |S| = 174$$

$$\lambda = 80 \cdot 3 \cdot \log \frac{27}{3} - 80 \cdot \log 174 = 114'61$$

Este estadístico se distribuye como una χ^2_g, donde g son los grados de libertad que vienen dados por:

$$g = \frac{p(p+1)}{2} - 1 = \frac{3 \cdot 4}{2} - 1 = 5$$

por lo que la hipótesis nula será rechazada si:

$$\lambda = 114'61 > \chi^2_{5;\alpha}$$ y en tal caso rechazaremos que las variables tengan la misma varianza.