Revista de Matemática: Teoría y Aplicaciones 1994 **1**(1) : 31–40
cimpa – ucr – ccss   issn: 1409-2433

ALGUNAS CONSIDERACIONES SOBRE EL MODELO LINEAL MULTIVARIADO

Jorge Poltronieri Vargas*   Asdrúbal Duarte Esquivel**

Resumen

Se trata el modelo lineal general y las pruebas de hipótesis clásicas. En particular se considera el análisis de varianza a dos factores, dándose explicitamente las formas de las matrices asociadas, así como la estadística del cociente de verosimilitud para las hipótesis nulas. También se obtienen en forma explícita las distribuciones de los estimadores, bajo la hipótesis de normalidad.

Abstract

We consider the general linear model with the classical hypothesis. In particular we consider the variance analysis in two factors and we give explicitly the forms of the asocited matrixes as well as the statistics of the likelihood ratio test for the null hypothesis. We obtain in a explicit form the distributions of the estimator under hypothesis of normality.

1. Introducción

Los resultados teóricos desarrollados en [4] y en [5], se aplican aquí al modelo lineal general multidimensional. Se tratan los resultados clásicos de pruebas de hipótesis, y en particular el análisis de varianza a dos factores.

Consideremos $X_1, \ldots, X_N$ vectores aleatorios independientes tales que :

$$X_\alpha \sim N(\mu_\alpha, \Sigma) \qquad \alpha = 1, \ldots, N.$$

Denotaremos $X = (X_1, \ldots, X_N) \sim N(\Gamma, I \otimes \Sigma)$, donde $\Gamma = (\mu_1, \ldots, \mu_N)$. Observemos que si $X_\alpha \sim N(\mu, \Sigma)$, se tiene que $\Gamma = (\mu, \ldots, \mu)$. Además:

$$I - \frac{1}{N} J \ , \text{ con } \quad J = \mathbf{1}\,\mathbf{1}' \ , \quad \mathbf{1}' = (1, \ldots, 1),$$

*Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. E-Mail: jpvargas@racsa.co.cr

**Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. E-Mail: aduarte@carari.ucr.ac.cr

es una matriz simétrica idempotente; por lo tanto:

$$X(I - \frac{1}{N}J) \sim N(0, (I - \frac{1}{N}J) \otimes \Sigma),$$

pues $(I - \frac{1}{N}J)\Gamma = 0$, y dado que $rang(I - \frac{1}{N}J) = N - 1$,

$$X(I - \frac{1}{N}J)X' = \sum_{\alpha=1}^{N}(X_\alpha - \overline{X})(X_\alpha - \overline{X})' \sim W(\Sigma, N-1),$$

donde $W(\Sigma, N-1)$ designa la distribución Wishart de matriz de varianzas Σ y $N-1$ grados de libertad.

Por otro lado, si $X_\alpha \sim N(\Lambda m_\alpha, \Sigma), \quad \alpha = 1, \dots, N$, son independientes, con m_α un vector r-dimensional y $H = MM' = \sum_{\alpha=1}^{N} m_\alpha m'_\alpha$ no singular, se tiene que $I - M'H^{-1}M$ es idempotente. Se sabe que :

$$X \sim N(\Lambda M, I \otimes \Sigma),$$

$$X(I - M'H^{-1}M) \sim N(0, (I - M'H^{-1}M) \otimes \Sigma),$$

puesto que $\Lambda M(I - M'H^{-1}M) = 0$ y la forma cuadrática:

$$X(I - M'H^{-1}M)X' \sim W(\Sigma, N - r),$$

donde $r = rangH$.

2. El modelo lineal

Consideremos $X_1, \dots, X_N$ variables aleatorias independientes, tales que:

$$X_\alpha \sim N(\beta Z_\alpha, \Sigma), \qquad \alpha = 1, \dots, N,$$

con Z_α vector q-dimensional, β matriz $p \times q$ y Σ matriz de covarianza.

Las matrices β y Σ son desconocidas. Así el modelo se escribe:

$$X \sim N(\beta Z, I \otimes \Sigma),$$

donde $Z = (Z_1, \dots, Z_N)$. Los estimadores de máxima verosimilitud son:

$$\hat{\beta} = XZ'A^{-1}$$

$$N\hat{\Sigma} = \sum_{\alpha=1}^{N}(x_\alpha - \hat{\beta}Z_\alpha)(x_\alpha - \hat{\beta}Z_\alpha)' = \sum_{\alpha=1}^{N} X_\alpha X'_\alpha - \hat{\beta}A\hat{\beta}',$$

donde $A = ZZ'$, i.e. $\hat{\beta} \sim N(\beta, A^{-1} \otimes \Sigma), \quad N\hat{\Sigma} \sim W(\Sigma, N - q)$.

Denotaremos un estimador con subíndice Ω cuando se trate del estimador de máxima verosimilitud de la muestra, y con subíndice ω cuando se trate del estimador de máxima verosimilitud bajo la hipótesis H_0.

Consideremos $Q = XZ'A^{-1}Z \sim N(\beta Z, Z'A^{-1}Z \otimes \Sigma)$, pues $Z'A^{-1}Z$ es idempotente. Suponemos que $\beta = (\beta_1, \beta_2)$ tales que β_1 tiene q_1 columnas y β_2 tiene q_2 columnas. Si deseamos considerar la hipótesis $H_0 : \beta_1 = \beta_1^*$, donde β_1^* es una matriz dada, se tiene que:

$$\hat{\beta}_{1\,\Omega} \sim N(\beta_1, A_{11\cdot 2} \otimes \Sigma),$$

con

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}.$$

Así: $\hat{\beta}_\Omega \sim N(\beta, A^{-1} \otimes \Sigma)$, $\quad (\hat{\beta}_{1\,\Omega} - \beta_1^*)A_{11\cdot 2}(\hat{\beta}_{1\,\Omega} - \beta_1^*) \sim W(\Sigma, q_1)$, con $q_1 = \text{rang}(A_{11\cdot 2})$.

Sea $Y = X - \beta_1^* Z_1 \sim N(\beta_2 Z_2, I \otimes \Sigma)$, entonces:

$$\hat{\beta}_{2\,\omega} = Y(Z_2' A_{22}^{-1}) = (XZ_2' - \beta_1^* A_{12})A_{22}^{-1},$$

donde:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad A_{11} = Z_1 Z_1', \quad A_{22} = Z_2 Z_2', \quad A_{12} = Z_1 Z_2'.$$

Bajo la hipótesis H_0 se tiene:

$$\hat{\beta}_{2\,\omega} \sim N(\beta_2, A_{22}^{-1} \otimes \Sigma),$$

$$N\hat{\Sigma}_\omega = \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' - \hat{\beta}_{2\,\omega} A_{22} \hat{\beta}_{2\,\omega}' = Y(I - Z_2 A_{22}^{-1} Z_2')Y' \sim W(\Sigma, N - q_2),$$

pues $I - Z_2 A_{22}^{-1} Z_2'$ es idempotente de rango $N - q_2$ $(q = q_1 + q_2)$.

Para probar la hipótesis $H_0 : \beta_1 = \beta_1^*$ se utiliza la estadística:

$$U = \frac{|N\hat{\Sigma}_\Omega|}{|N\hat{\Sigma}_\omega|},$$

y se compara con $U_{p,q_1,N-q}(\alpha)$, donde α es el nivel. La cantidad $|N\hat{\Sigma}_\Omega|$ denota el determinante de $N\hat{\Sigma}_\Omega$.

3. Prueba de igualdad de medias

Consideremos $Y_\alpha^{(i)} \sim N(\mu^{(i)}, \Sigma)$ $\quad \alpha = 1, \ldots, N_i, \quad i = 1, \ldots, q$, observaciones de q poblaciones de igual matriz de covarianza. Sea H_0 la hipótesis de igualdad de medias para las q poblaciones, i.e. $H_0 : \mu^{(1)} = \cdots = \mu^{(q)}$. Se define $X_{N_1+\cdots+N_{i-1}+k} = Y_k^{(i+1)}$ y tenemos:

$$\Gamma = (\mu^{(1)}, \ldots, \mu^{(1)}, \ldots, \mu^{(q)}, \ldots, \mu^{(q)}),$$

$$X = (Y_1^{(1)}, \ldots, Y_{N_1}^{(1)}, \ldots, Y_1^{(q)}, \ldots, Y_{N_q}^{(q)}),$$

por lo que:

$$X \sim N(\Gamma, I \otimes \Sigma), \qquad N = \sum_{i=1}^{q} N_i.$$

Se utiliza el modelo lineal, introduciendo las variables Z_α, $\alpha = 1, \ldots, N$, de la forma:

$$Z = (Z_1, \ldots, Z_N) = \begin{bmatrix} 1 & \ldots & 1 & \ldots & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \ldots & 1 & \ldots & 1 \\ 1 & \ldots & 1 & \ldots & 1 & \ldots & 1 \end{bmatrix},$$

y se define $\beta = (\beta_1, \beta_2)$ por:

$$\beta_1 = (\mu^{(1)} - \mu^{(q)}, \ldots, \mu^{(q-1)} - \mu^{(q)}),$$

$$\beta_2 = \mu^{(q)}.$$

Así se obtiene que: $X_\alpha \sim N(\beta Z_\alpha, \Sigma)$ $\alpha = 1, \ldots, N$. La hipótesis H_0 es: $\beta_1 = 0$. La matriz $A = Z\,Z'$ es:

$$\begin{bmatrix} N_1 & \ldots & 0 & N_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & N_{q-1} & N_{q-1} \\ N_1 & \cdots & N_{q-1} & N \end{bmatrix},$$

y

$$C = X\,Z' = (C_1, C_2) = (\sum_\alpha Y_\alpha^{(1)}, \ldots, \sum_\alpha Y_\alpha^{(q-1)}, \sum_{\alpha\, i} Y_\alpha^{(i)}),$$

$$A_{22} = N, \qquad C_2 = \sum_{\alpha\, i} Y_\alpha^{(i)},$$

$$\hat{\beta}_{2\,\omega} = C_2 A_{22}^{-1} = \frac{1}{N} \sum_{\alpha\, i} Y_\alpha^{(i)} = \overline{Y} \sim N(\beta_2, \frac{1}{N}\Sigma),$$

$$N\hat{\Sigma}_\omega = \sum_{\alpha\, i} Y_\alpha^{(i)} Y_\alpha^{(i)\prime} - \beta_{2\,\omega} A_{22} \beta_{2\,\omega}' = \sum_{\alpha\, i} (Y_\alpha^{(i)} - \overline{Y})(Y_\alpha^{(i)} - \overline{Y})'$$

$$= Y(I - \frac{1}{N} J)Y' \sim W(\Sigma, N-1),$$

pues bajo la hipótesis H_0 se tiene $\Gamma(I - \frac{1}{N}J) = 0$.

El estimador $\hat{\beta}_\Omega = C\,A^{-1} = (\overline{Y}^{(1)} - \overline{Y}^{(q)}, \ldots, \overline{Y}^{(q-1)} - \overline{Y}^{(q)}, \overline{Y}^{(q)})$, donde $\overline{Y}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} Y_\alpha^{(i)}$. Además:

$$CA^{-1}C' = \sum_{i=1}^{q} N_i \overline{Y}^{(i)} \overline{Y}^{(i)\prime},$$

$$N\hat{\Sigma}_\Omega = \sum_{\alpha\, i} (Y_\alpha^{(i)} - \overline{Y}^{(i)})(Y_\alpha^{(i)} - \overline{Y}^{(i)})'.$$

Sea H la matriz definida por:

$$H = \begin{bmatrix} J_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_q \end{bmatrix}$$

con $J_i = \frac{1}{N_i}\mathbf{1}_{N_i}\mathbf{1}'_{N_i}$. Así definida H es idempotente, de rango q y además :

$$YH = (\overline{Y}^{(1)}, \ldots, \overline{Y}^{(1)}, \ldots, \overline{Y}^{(q)}, \ldots, \overline{Y}^{(q)}) \sim N(\Gamma, H \otimes \Sigma),$$

pues $\Gamma H = \Gamma$. Por otro lado :

$$N\hat{\Sigma}_\Omega = \sum_{\alpha\, i}(Y_\alpha^{(i)} - \overline{Y}^{(i)})(Y_\alpha^{(i)} - \overline{Y}^{(i)})' \sim W(\Sigma, N-q).$$

Si H_0 es la verdadera hipótesis, la estadística: $U = |N\hat{\Sigma}_\Omega|/|N\hat{\Sigma}_\omega|$ se distribuye como una $U_{p,\, q-1,\, N-q}$.

Observemos que :

- $N\hat{\Sigma}_\omega - N\hat{\Sigma}_\Omega = Y(H - \frac{1}{N}J)Y' \sim W(\Sigma, q-1)$.
- $N\hat{\Sigma}_\omega/(N-q)$ es un estimador sin sesgo de Σ, si H_0 es verdadera.

4. Análisis de varianza

El caso analizado anteriormente puede ser considerado como un análisis de varianza a un factor. Vamos a desarrollar aquí el caso de dos factores con una observación por celda.

Sea Y_{ij} $i = 1, \ldots, r$; $j = 1, \ldots, c$, una variable aleatoria p-dimensional tal que:

$$Y_{ij} = \mu + \lambda_i + \nu_j + E_{ij},$$

donde los $E_{ij} \sim N(0, \Sigma)$ son independientes; μ, λ_i, ν_j son vectores tales que:

$$\sum_{i=1}^{r} \lambda_i = 0 \ , \ \sum_{j=1}^{c} \nu_j = 0.$$

El modelo se escribe :

$$Y = \beta Z + E,$$

donde:

$$Y = (Y_{11}, \ldots, Y_{1c}, \ldots, Y_{r1}, \ldots, Y_{rc}),$$
$$\beta = (\mu, \lambda_1, \ldots, \lambda_r, \nu_1, \ldots, \nu_c),$$
$$E = (E_{11}, \ldots, E_{1c}, \ldots, E_{r1}, \ldots, E_{rc}),$$

y

$$Z = \begin{bmatrix} 1 & \ldots & 1 & \ldots & 1 & \ldots & 1 \\ 1 & \ldots & 1 & \ldots & 0 & \ldots & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \ldots & 1 & \ldots & 1 \\ 1 & \ldots & 0 & \ldots & 1 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 1 & \ldots & 0 & \ldots & 1 \end{bmatrix}.$$

Sea $\overline{Y}_{\cdot\cdot} = \dfrac{1}{rc}\sum_{i\,j} Y_{ij}, \quad \overline{Y}_{i\cdot} = \dfrac{1}{c}\sum_{j} Y_{ij}, \quad \overline{Y}_{\cdot j} = \dfrac{1}{r}\sum_{i} Y_{ij}$, entonces:

$$\sum_{i\,j}(Y_{ij} - \overline{Y}_{\cdot\cdot})(Y_{ij} - \overline{Y}_{\cdot\cdot})' = \sum_{i\,j}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})'$$

$$+c\sum_{i}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})' + r\sum_{j}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})'.$$

Vamos a determinar las matrices asociadas a cada una de las formas cuadráticas.

Sean las matrices:

$$H_1 = \frac{1}{c}\begin{bmatrix} J & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J \end{bmatrix},$$

con $J = \mathbf{1}_c\mathbf{1}_c'$, i.e. H_1 tiene r matrices J en la diagonal;

$$L = \frac{1}{rc}\mathbf{1}_{rc}\mathbf{1}_{rc}',$$

$$H_2 = \frac{1}{r}\begin{bmatrix} P_1 & \dots & P_c \\ \vdots & \ddots & \vdots \\ P_1 & \dots & P_c \end{bmatrix} \quad \text{con} \quad P_i = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix},$$

donde el vector $(1, \dots, 1)$ está situado en la posición i. La matriz P_i es $r \times r$ y $P_iP_j = P_i$. Las matrices H_1, L, H_2 son idempotentes de rangos respectivos r, 1, c. Así tenemos:

$$YL = (\overline{Y}_{\cdot\cdot}, \dots, \overline{Y}_{\cdot\cdot}, \dots, \overline{Y}_{\cdot\cdot}, \dots, \overline{Y}_{\cdot\cdot})$$

$$YH_1 = (\overline{Y}_{1\cdot}, \dots, \overline{Y}_{1\cdot}, \dots, \overline{Y}_{r\cdot}, \dots, \overline{Y}_{r\cdot})$$

$$YH_2 = (\overline{Y}_{\cdot 1}, \dots, \overline{Y}_{\cdot c}, \dots, \overline{Y}_{\cdot 1}, \dots, \overline{Y}_{\cdot c}).$$

Sabemos que $H_1 - L$ es idempotente de rango $r - 1$, y

$$Y(H_1 - L) \sim N(\Gamma(H_1 - L), (H_1 - L) \otimes \Sigma),$$

donde $\Gamma(H_1 - L) = (\lambda_1, \dots, \lambda_1, \dots, \lambda_r, \dots, \lambda_r)$.

La forma cuadrática:

$$B_1 = Y(H_1 - L)Y' = c\sum_{i=1}^{r}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})' \sim W(\Sigma, r-1, \tau),$$

donde el parámetro de decentraje $\tau = \Gamma(H_1 - L)\Gamma' = c\sum_{i=1}^{r} \lambda_i\lambda_i'$.

La matriz $H_2 - L$ es idempotente de rango $c - 1$, y

$$B_2 = Y(H_2 - L)Y' = r\sum_{j=1}^{c}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})' \sim W(\Sigma, c-1, \tau),$$

donde el parámetro de decentraje $\tau = \Gamma(H_2 - L)\Gamma' = r\sum_{j=1}^{c}\nu_j\nu_j'$.

La matriz $I - H_1 - H_2 + L$ es idempotente de rango $(r-1)(c-1)$:

$$Y(I - H_1 - H_2 + L) \sim N(0, (I - H_1 - H_2 + L) \otimes \Sigma),$$

pues $\Gamma(I - H_1 - H_2 + L) = 0$, por lo que:

$$A = Y(I - H_1 - H_2 + L)Y' =$$

$$\sum_{ij}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})' \sim W(\Sigma, (r-1)(c-1)).$$

Notemos que las formas cuadráticas A, B_1, B_2 son independientes, pues el producto de las matrices asociadas es cero ($H_1H_2 = H_1L = H_2L = LL = L$), y que B_1 y B_2 son distribuciones Wishart decentradas.

Si se considera la hipótesis $H_0 : \lambda_i = 0, \;\; i = 1, \ldots, r$, es decir, el efecto del primer factor es nulo, se tiene que $\Gamma(H_1 - L) = 0$. Bajo la hipótesis H_0, la estadística:

$$B_1 = Y(H_1 - L)Y' \sim W(\Sigma, r-1).$$

De esta manera la estadística:

$$U = \frac{|A|}{|A + B_1|} \sim U_{p,\, r-1,\, (r-1)(c-1)}.$$

Si se considera la hipótesis $H_0 : \nu_j = 0, \;\; j = 1, \ldots, c$, es decir, el efecto del segundo factor es nulo, se tiene que $\Gamma(H_2 - L) = 0$, y la estadística:

$$B_2 = Y(H_2 - L)Y' \sim W(\Sigma, c-1).$$

Así la estadística:

$$U = \frac{|A|}{|A + B_2|} \sim U_{p\,,\, c-1\,,\, (r-1)(c-1)}.$$

En los dos casos se rechaza la hipótesis H_0 si :

$$U \leq U_{p\,,\, m\,,\, (r-1)(c-1)}(\alpha)$$

para $m = r - 1$, o $m = c - 1$.

Este resultado se generaliza fácilmente en el caso de n observaciones por celda, $n > 1$.

5. Aplicaciones

En está sección vamos a considerar algunos ejemplos., lo cuales nos ayudarán a comprender, la utilidad de la teoría desarrollada en este trabajo.

1. Primeramente consideraremos un estudio realizado por Bernard (1935) (ver [1]), en el cual realiza $4(= p)$ medidas sobre cráneos egipcios, correspondiendo a $4(= p)$ poblaciones:

Predinástica ($i = 1$), sexta a undésima dinastía ($i = 2$), duodécima y decimotercera dinastía ($i = 3$), y la dinastía Ptoloméica ($i = 4$) .

El número de observaciones por poblaciones $N_1 = 91$, $N_2 = 162$, $N_3 = 70$, $N_4 = 75$. La hipótesis H_0 que consideramos es que las cuatro poblaciones son iguales, es decir, las medias de las poblaciones son idénticas $\mu^{(1)} = \mu^{(2)} = \mu^{(3)} = \mu^{(4)}$.

Los datos sumarizados son:

$$(\bar{Y}^{(1)}, \bar{Y}^{(2)}, \bar{Y}^{(3)}, \bar{Y}^{(4)}) = \begin{pmatrix} 133{,}582 & 134{,}265 & 134{,}371 & 135{,}306 \\ 98{,}307 & 96{,}462 & 95{,}857 & 95{,}040 \\ 50{,}835 & 51{,}148 & 50{,}100 & 52{,}093 \\ 133{,}000 & 134{,}882 & 133{,}642 & 131{,}466 \end{pmatrix}$$

$$N\Sigma_\Omega = \begin{pmatrix} 9661{,}997 & 445{,}573 & 11130{,}623 & 214{,}584 \\ 445{,}573 & 9073{,}115 & 1239{,}211 & 2255{,}812 \\ 1130{,}623 & 1239{,}211 & 3938{,}320 & 1271{,}054 \\ 2148{,}584 & 2255{,}812 & 1271{,}054 & 8741{,}508 \end{pmatrix}.$$

De los datos tenemos:

$$N\hat{\Sigma}_\omega = \begin{pmatrix} 9785{,}178 & 214{,}197 & 1217{,}929 & 2019{,}820 \\ 214{,}197 & 9559{,}460 & 1131{,}716 & 2381{,}126 \\ 2117{,}929 & 1131{,}716 & 4088{,}731 & 1133{,}473 \\ 2019{,}820 & 2381{,}126 & 1133{,}473 & 9382{,}242 \end{pmatrix}.$$

La estadística está dada por:

$$U = \frac{|N\widehat{\Sigma}_\Omega|}{|N\hat{\Sigma}_\omega|} = 0{,}8214344,$$

$N = 398$, $n = 394$, $p = 4$, $q = 4$. Como n es grande, aproximamos $-m\log U_{4,3,394}$ con χ^2_{12}, cuando la hipótesis H_0 es cierta. Así $-m\log U = 77{,}30$ y como $\chi^2_{12}(0{,}01) = 26{,}2$, se rechaza la hipótesis H_0, es decir, hay diferencias significativas entre las poblaciones.

Para el análisis de varianza utilizamos un ejemplo discutido por Anderson(1958). En este ejemplo, se considera como primera componente del vector de observaciones el rendimiento de un campo de cebada; la segunda componente son las mismas medidas hechas al año siguiente ($p = 2$). Los datos aparecen en la tabla siguiente. Los índices columna indican las variedades de cebada y las filas indican las localidades.

Se considera el modelo a dos factores, donde el primer factor es la variedad con $r = 5$ niveles, y el segundo factor es la localidad con $c = 6$ niveles. La hipótesis H_0 que se considera, es que el efecto debido a la variedad es nulo, es decir, dentro del modelo:

$$Y_{ij} = \mu + \lambda_i + \nu_j + E_{ij}$$

los $\lambda_i = 0, \quad i = 1, \ldots, r = 5$. En otras palabras, las observaciones se explican por medio de un valor general μ y el efecto debido a la localidad $\mu_j, \quad j = 1, \ldots, c = 6$.

Así

$$\sum_{i,j} Y_{ij} Y'_{ij} = \begin{pmatrix} 380944 & 315381 \\ 315381 & 277625 \end{pmatrix},$$

	Variedades				
Localidad	M	S	V	T	P
UF	81	105	120	110	98
	81	82	80	87	84
W	147	142	151	192	146
	100	116	112	148	108
M	82	77	78	131	90
	103	105	117	140	130
C	120	121	124	141	125
	99	62	96	126	76
GR	99	89	69	89	104
	66	50	97	62	80
D	87	77	79	102	96
	68	67	67	92	94

Cuadro 1: Variedades de cebada por cada localidad

$$\sum_j (6\bar{Y}_{i\cdot})(6\bar{Y}_{\cdot j})' = \begin{pmatrix} 2157924 & 1844346 \\ 1844346 & 1579583 \end{pmatrix},$$

$$\sum_i (5\bar{Y}_{i\cdot})(5\bar{Y}_{i\cdot})' = \begin{pmatrix} 1874386 & 1560145 \\ 1560145 & 1353727 \end{pmatrix},$$

$$(30\bar{Y}_{\cdot\cdot})(30\bar{Y}_{\cdot\cdot}) = \begin{pmatrix} 10705984 & 9145240 \\ 9145240 & 7812025 \end{pmatrix}.$$

La suma de los cuadrados de los errores:

$$A = \begin{pmatrix} 3279 & 802 \\ 802 & 4017 \end{pmatrix},$$

$$5\sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})' = \begin{pmatrix} 18011 & 7188 \\ 7188 & 10345 \end{pmatrix},$$

$$B = \begin{pmatrix} 2788 & 2550 \\ 2550 & 2863 \end{pmatrix}.$$

La estadística para la hipótesis H_0 es:

$$U = \frac{|A|}{|A+B|} = 0{,}4107.$$

Este resultado se compara con $U_{2,4,20}(0{,}05)$, o bien se compara con el valor:

$$\frac{1-\sqrt{0{,}4017}}{\sqrt{0{,}4017}}\cdot\frac{19}{4}=2{,}66$$

con un $F_{8,38}(0{,}05)=2{,}18$ lo que indica que hay diferencias entre variedades , con un error del 5 %.

Referencias

[1] T.W. Anderson (1958) *An introduction to multivariate statistical analysis.* J. Wiley, N.Y.

[2] J.R. Barra (1971) *Notions fondamentales de statistique mathématique.* Dunod, Paris.

[3] H. Muirhead (1982) *Aspects of multivariate statistical theory.* J. Wiley, N.Y.

[4] J. Poltronieri (1988) *Estudio de formas cuadráticas en el caso multivariado.* In: IV Simposio de Métodos Matemáticos Aplicados a las Ciencias Ed. U.C.R.

[5] J. Poltronieri (1988) *Formas cuadráticas y formas lineales en estadística multivariada.* In: IV Simposio de Métodos Matemáticos Aplicados a las Ciencias Ed. U.C.R.

[6] K. Takeuchi, H. Yanai, B.N. Mukherjee (1984) *The foundations of multivariate analysis.* Wiley Eastern Limited.

[7] M. Tenenhaus, F. Young (1987) *An analysis and synthesis of multiple correspondance analysis, optimal scaling, dual scaling, homogenity analysis and other methods for quantifying categorical multivariate data*, Psychometrica 50(1), pp. 91-119.