Inferencia II

MULTIVARIATE ANALYSIS
MESIO (15-16)

PROF. SERGI CIVIT
PROF. MIQUEL SALICRÚ

Función de verosimilitud en dos poblaciones

Sea x_1, …, x_n, y_1, ,…, y_m muestras aleatorias representativas de las variable $X\approx N_p(\mu_1, \Sigma_1)$ e $Y\approx N_p(\mu_2, \Sigma_2)$. En este contexto, la función de densidad conjunta es el producto de marginales:

$$L(x_1,\dots,x_n,y_1,\dots,y_m)=\prod_{i=1}^{n}\left\{\frac{1}{|\Sigma_1|^{1/2}(2\pi)^{p/2}}e^{-\frac{1}{2}(x_i-\mu_1)'\Sigma_1^{-1}(x_i-\mu_1)}\right\}$$

$$\cdot\prod_{i=1}^{m}\left\{\frac{1}{|\Sigma_2|^{1/2}(2\pi)^{p/2}}e^{-\frac{1}{2}(y_i-\mu_2)'\Sigma_2^{-1}(y_i-\mu_2)}\right\}$$

$$=(2\pi)^{-np/2}|\Sigma_1|^{-n/2}\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}(x_i-\mu_1)'\Sigma_1^{-1}(x_i-\mu_1)\right\}$$

$$\cdot(2\pi)^{-mp/2}|\Sigma_2|^{-m/2}\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{m}(y_i-\mu_2)'\Sigma_2^{-1}(y_i-\mu_2)\right\}$$

Test de interés

Test de hipótesis sobre los parámetros.

$$\left.\begin{array}{l} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta \end{array}\right\} \rightarrow \left.\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array}\right\} \quad \left.\begin{array}{l} H_0 : \Sigma_1 = \Sigma_2 \\ H_1 : \Sigma_1 \neq \Sigma_2 \end{array}\right\}$$

Test de la razón de verosimilitud

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(x_1, \ldots, x_n, \theta)}{\max_{\theta \in \Theta} L(x_1, \ldots, x_n, \theta} < c$$

Asintoticamente:

$$-2 \ln \Lambda \approx \chi^2_{\nu - \nu_0}$$

Contraste de medias

$$\left.\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array}\right\} \qquad (\Sigma_1 = \Sigma_2)$$

$$H_0 : \hat{\mu}_1 = \hat{\mu}_2 = \frac{n\overline{X} + m\overline{Y}}{n+m}, \quad \hat{\Sigma}_1 = \hat{\Sigma}_2 = \Sigma_{\text{pooled } (\mu_1 = \mu_2)}$$

$$\left(\sigma^2_{ij} = \sum_{k=1}^{n} (x_{ik} - \frac{n\overline{X}_i + m\overline{Y}_i}{n+m})(x_{jk} - \frac{n\overline{X}_i + m\overline{Y}_i}{n+m}) + \sum_{k=1}^{m} (y_{ik} - \frac{n\overline{X}_i + m\overline{Y}_i}{n+m})(y_{jk} - \frac{n\overline{X}_i + m\overline{Y}_i}{n+m}) \right)$$

$$H_1 : \hat{\mu}_1 = \overline{X}, \ \hat{\mu}_2 = \overline{Y}, \ \hat{\Sigma}_1 = \hat{\Sigma}_2 = \Sigma_{\text{pooled } (\mu_1 \neq \mu_2)} = \frac{nS_1 + mS_2}{n+m} \quad (S_{\text{corregida}} = \frac{nS_1 + mS_2}{n+m-2})$$

$$T^2 = g(\Lambda) = \frac{n \cdot m}{n+m} (\overline{X} - \overline{Y})' S^{-1}_{\text{corregida}} (\overline{X} - \overline{Y}) \approx T^2_{p, n+m-2} \approx \frac{(n+m-2)p}{n+m-p-1} F_{p, n+m-p-1}$$

Matrices de varianzas-covarianzas

(PRACTICA III)

$$\left.\begin{array}{l} H_0 : \Sigma_1 = \Sigma_2 \\ H_1 : \Sigma_1 \neq \Sigma_2 \end{array}\right\}$$

H_0: $\hat{\mu}_1 = \overline{X}, \ \hat{\mu}_2 = \overline{Y}, \ \hat{\Sigma}_1 = \hat{\Sigma}_2 = \dfrac{nS_1 + mS_2}{n+m}$

H_1: $\hat{\mu}_1 = \overline{X}, \ \hat{\mu}_2 = \overline{Y}, \ \hat{\Sigma}_1 = S_1, \ \hat{\Sigma}_2 = S_2$

$$-2\ln\Lambda = -2\ln \frac{|S_1|^{n/2}|S_2|^{m/2}}{\left|\dfrac{nS_1 + mS_2}{n+m}\right|^{(n+m)/2}} \approx \chi^2_{p(p+1)/2}$$

Contraste múltiple de medias 1/3

matriz	orden	media	covarianza	distribución
$\mathbf{X}_1$	$n_1 \times p$	$\overline{\mathbf{x}}_1$	$\mathbf{S}_1$	$N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$
$\mathbf{X}_2$	$n_2 \times p$	$\overline{\mathbf{x}}_2$	$\mathbf{S}_2$	$N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{X}_g$	$n_g \times p$	$\overline{\mathbf{x}}_g$	$\mathbf{S}_g$	$N_p(\boldsymbol{\mu}_g, \boldsymbol{\Sigma})$

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{g} n_i \overline{\mathbf{x}}_i, \quad \widehat{\mathbf{S}} = \frac{1}{n-g}\sum_{i=1}^{g} n_i \mathbf{S}_i,$$

Contraste múltiple de medias 2/3

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g.$$

$$\begin{array}{lll} \mathbf{B} & = \sum_{i=1}^{g} n_i(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})' & \text{(dispersión entre grupos)} \\ \mathbf{W} & = \sum_{i=1}^{g} \sum_{\alpha=1}^{n_i} (\mathbf{x}_{i\alpha} - \overline{\mathbf{x}}_i)(\mathbf{x}_{i\alpha} - \overline{\mathbf{x}}_i)' & \text{(dispersión dentro grupos)} \\ \mathbf{T} & = \sum_{i=1}^{g} \sum_{\alpha=1}^{n_i} (\mathbf{x}_{i\alpha} - \overline{\mathbf{x}})(\mathbf{x}_{i\alpha} - \overline{\mathbf{x}})' & \text{(dispersión total)} \end{array}$$

$$\mathbf{T} = \mathbf{B} + \mathbf{W}.$$

Contraste múltiple de medias 1/3

Si H_0 es cierta,

$$\mathbf{B} \sim W_p(\boldsymbol{\Sigma}, g-1), \ \mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n-g), \ \mathbf{T} \sim W_p(\boldsymbol{\Sigma}, n-1),$$

$\mathbf{B}, \mathbf{W}$ son estocásticamente independientes.

el estadístico

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} \sim \Lambda(p, n-g, g-1).$$

y se rechaza la H_0 cuando Λ es pequeño (Λ - Wilks)

Referencias

Cuadras C (2014). Nuevos metodos de análisis multivariante. CMC Editions.

Anderson, T.W. (2003). An introduction to multivariate statistical analysis. Wiley.

Seber, G.A.F. (1984). Multivariate observations. John Wiley & Sons.