

MAT-269: Test de hipótesis sobre Σ

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Test de independencia

Considere la partición:

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Nuestro interés es probar la hipótesis $H_0 : \Sigma_{12} = 0$.

Considere $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ y $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{Q}$ los MLE de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ particionadas como:

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},$$

donde

$$Q_{11} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1)^\top Q_{22} = \sum_{i=1}^n (x_{i2} - \bar{x}_2)(x_{i2} - \bar{x}_2)^\top$$

$$Q_{12} = Q_{21}^\top = \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)^\top$$



Test de independencia

Bajo H_0 se tienen dos muestras independientes $\mathbf{x}_{11}, \dots, \mathbf{x}_{n1}$ desde $N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ y $\mathbf{x}_{12}, \dots, \mathbf{x}_{n2}$ desde $N_{p_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. Además,

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})L_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

De modo que el LRT para H_0 es

$$\begin{aligned}\Lambda &= \frac{\max_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}} L_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})L_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})}{\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \\&= \frac{L_1(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_{11})L_2(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_{22})}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})} \\&= \frac{(2\pi)^{-np_1/2} |\hat{\boldsymbol{\Sigma}}_{11}|^{-n/2} e^{-np_1/2} (2\pi)^{-np_2/2} |\hat{\boldsymbol{\Sigma}}_{22}|^{-n/2} e^{-np_2/2}}{(2\pi)^{-np/2} |\hat{\boldsymbol{\Sigma}}|^{-n/2} e^{-np/2}} \\&= \left(\frac{|\hat{\boldsymbol{\Sigma}}_{11}| |\hat{\boldsymbol{\Sigma}}_{22}|}{|\hat{\boldsymbol{\Sigma}}|} \right)^{-n/2} = \left(\frac{|\mathbf{Q}|}{|\mathbf{Q}_{11}| |\mathbf{Q}_{22}|} \right)^{n/2},\end{aligned}$$

donde $\hat{\boldsymbol{\Sigma}}_{ii} = \mathbf{Q}_{ii}/n$, para $i = 1, 2$.



Ahora, como $\mathbf{Q} > \mathbf{0}$ con prob. 1 y \mathbf{Q}_{11} es no singular, sigue que

$$|\mathbf{Q}| = |\mathbf{Q}_{11}| |\mathbf{Q}_{22} - \mathbf{Q}_{21} \mathbf{Q}_{11}^{-1} \mathbf{Q}_{12}|,$$

haciendo $\mathbf{E} = \mathbf{Q}_{22} - \mathbf{Q}_{21} \mathbf{Q}_{11}^{-1} \mathbf{Q}_{12}$ ($= \mathbf{Q}_{22 \cdot 1}$) y $\mathbf{H} = \mathbf{Q}_{21} \mathbf{Q}_{11}^{-1} \mathbf{Q}_{12}$. Tenemos

$$LR = \Lambda^{2/n} = \frac{|\mathbf{Q}_{22 \cdot 1}|}{|\mathbf{Q}_{22}|} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}.$$

Dado que $\mathbf{Q} = (n-1)\mathbf{S} \sim W_p(n-1, \mathbf{\Sigma})$. Luego, cuando $H_0 : \mathbf{\Sigma}_{12} = \mathbf{0}$ es verdadero $\mathbf{\Sigma}_{22 \cdot 1} = \mathbf{\Sigma}_{22}$ ($= \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}$) y las matrices \mathbf{H} y \mathbf{E} son independientes

$$\mathbf{H} \sim W_{p_2}(p_1, \mathbf{\Sigma}_{22}), \quad \mathbf{E} \sim W_{p_2}(n-1-p_1, \mathbf{\Sigma}_{22}),$$

de ahí que

$$LR \sim \Lambda_{p_2}(p_1, n-p_1-1),$$

y H_0 es rechazado si LR es muy pequeño.



Matriz de dispersión diagonal

Un caso especial de la hipótesis anterior es $H_0 : \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$. De este modo, tenemos:

$$LR = \frac{|\hat{\Sigma}|}{\hat{\sigma}_{11} \cdots \hat{\sigma}_{pp}} = \frac{|Q|}{q_{11} \cdots q_{pp}},$$

donde $q_{rr} = \sum_{i=1}^n (x_{ir} - \bar{x}_r)^2 = n\hat{\sigma}_{rr}$, para $r = 1, \dots, p$. De este modo, $LR \xrightarrow{D} \chi^2(p(p-1)/2)$.

Note que

$$r_{jk} = \frac{q_{jk}}{\sqrt{q_{jj}q_{kk}}} = \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}\hat{\sigma}_{kk}}},$$

es la correlación muestral entre las variables j y k . Además, la matriz de correlación $R = (r_{jk})$ está dada por

$$R = \text{diag}(q_{11}^{-1/2}, \dots, q_{pp}^{-1/2}) Q \text{diag}(q_{11}^{-1/2}, \dots, q_{pp}^{-1/2}).$$

De ahí que

$$LR = \frac{|Q|}{q_{11} \cdots q_{pp}} = |R|,$$

es decir, el estadístico está basado en el determinante de la matriz de correlación.



Test de esfericidad

En práctica, se desea probar hipótesis del tipo $H_0 : \Sigma = \sigma^2 \mathbf{I}$, donde σ^2 no es especificado. Note que bajo H_0 , tenemos

$$\ell(\boldsymbol{\theta}) = -\frac{np}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \text{tr}\{\mathbf{Q} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top\}.$$

Además, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$. Por otro lado,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{np}{2\sigma^2} + \frac{1}{2\sigma^4} \text{tr}\{\mathbf{Q} + n(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})^\top\},$$

de ahí que

$$\hat{\sigma}^2 = \frac{1}{np} \text{tr} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top = \frac{1}{p} \text{tr} \hat{\Sigma}.$$

Por tanto,

$$LR = \Lambda^{2/n} = \frac{|\hat{\Sigma}|}{(\text{tr} \hat{\Sigma}/p)^p} = \frac{|\mathbf{Q}|}{(\text{tr} \mathbf{Q}/p)^p}$$



En efecto,

$$\begin{aligned}\max_{\mu, \Sigma} L(\mu, \Sigma) &= (2\pi)^{-np/2} |\hat{\sigma}^2 \mathbf{I}| \exp \left(-\frac{n}{2\hat{\sigma}^2} \text{tr } \hat{\Sigma} \right) \\ &= (2\pi)^{-np/2} (\text{tr } \hat{\Sigma}/p)^{-np/2} \exp \left(-\frac{np}{2 \text{tr } \hat{\Sigma}} \text{tr } \hat{\Sigma} \right) \\ &= (2\pi)^{-np/2} (\text{tr } \hat{\Sigma}/p)^{-np/2} e^{-np/2}.\end{aligned}$$

Note que $H_0 : \Sigma = \sigma^2 \mathbf{I}$ es equivalente a probar que todos los valores propios λ_j de Σ son iguales, esto es

$$1 = \frac{\text{media geométrica de } \lambda_j}{\text{media aritmética de } \lambda_j} = \frac{\prod_j \lambda_j^{1/p}}{\sum_j \lambda_j/p} = \frac{|\Sigma|^{1/p}}{\text{tr } \Sigma/p},$$

y sustituímos Σ por su MLE $\hat{\Sigma}$, luego verificamos si el estadístico $LR^{1/p}$ es cercano a la unidad.



Test de equicorrelación

Una hipótesis que surge, por ejemplo, en análisis de varianza es:

$$H_0 : \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^\top].$$

Si $\mathbf{S} = (s_{jk})$ es el estimador insesgado de Σ , entonces los MLE de σ^2 y ρ son (ejercicio)

$$\hat{\sigma}^2 = \frac{1}{p} \sum_{j=1}^p s_{jj}, \quad \hat{\sigma}^2 \hat{\rho} = \frac{1}{p(p-1)} \sum_{j \neq k} s_{jk}.$$

Es estadístico de razón de verosimilitudes es

$$LR = \Lambda^{2/n} = \frac{|\mathbf{S}|}{\hat{\sigma}^{2p}(1 - \hat{\rho})^{p-1}\{1 + (p-1)\hat{\rho}\}},$$

que es asintóticamente chi-cuadrado con $p(p+1)/2 - 2$ grados de libertad.



Comparando poblaciones normales

Suponga $\mathbf{x}_1, \dots, \mathbf{x}_{n_1}$ muestra aleatoria desde $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ y una muestra independiente $\mathbf{y}_1, \dots, \mathbf{y}_{n_2}$ desde $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ y considere la hipótesis $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 (= \boldsymbol{\Sigma})$. La función de verosimilitud puede ser escrita como:

$$L_{12}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = L_1(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)L_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2),$$

donde

$$L_i(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = (2\pi)^{-n_i p/2} |\boldsymbol{\Sigma}_i|^{-n_i/2} \exp \left\{ -\frac{1}{2} \text{tr } \boldsymbol{\Sigma}_i^{-1} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_i)(\mathbf{x}_i - \boldsymbol{\mu}_i)^\top \right\},$$

maximizar L_{12} es equivalente a la maximización simultánea de cada L_i ($i = 1, 2$), de modo que L_{12} es maximizada en

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}}, \quad \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{y}}, \quad \hat{\boldsymbol{\Sigma}}_1 = \mathbf{Q}_1/n_1, \quad \hat{\boldsymbol{\Sigma}}_2 = \mathbf{Q}_2/n_2.$$

De este modo,

$$\begin{aligned} L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_2) &= L_1(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)L_2(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2) \\ &= (2\pi)^{-n_1 p/2} |\hat{\boldsymbol{\Sigma}}_1|^{-n_1/2} e^{-n_1 p/2} (2\pi)^{-n_2 p/2} |\hat{\boldsymbol{\Sigma}}_2|^{-n_2/2} e^{-n_2 p/2} \\ &= (2\pi)^{-n p/2} |\hat{\boldsymbol{\Sigma}}_1|^{-n_1/2} |\hat{\boldsymbol{\Sigma}}_2|^{-n_2/2} e^{-n p/2} \end{aligned}$$



Comparando poblaciones normales

Haciendo $\Sigma_1 = \Sigma_2 = \Sigma$, en este caso deseamos maximizar

$$\begin{aligned}\log L_{12}(\mu_1, \mu_2, \Sigma) &= \log L_1(\mu_1, \Sigma) + \log L_2(\mu_2, \Sigma) \\&= c - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \Sigma^{-1} \left\{ \sum_{i=1}^{n_1} (\mathbf{x}_i - \mu_1)(\mathbf{x}_i - \mu_1)^\top \right. \\&\quad \left. + \sum_{i=1}^{n_2} (\mathbf{y}_i - \mu_2)(\mathbf{y}_i - \mu_2)^\top \right\} \\&= c - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \Sigma^{-1} \mathbf{Q} + \frac{1}{2} \operatorname{tr} \Sigma^{-1} \{ n_1 (\bar{\mathbf{x}} - \mu_1)(\bar{\mathbf{x}} - \mu_1)^\top \\&\quad + n_2 (\bar{\mathbf{y}} - \mu_2)(\bar{\mathbf{y}} - \mu_2)^\top \},\end{aligned}$$

donde $c = -\frac{np}{2} \log 2\pi$ y $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2$. Como

$$\operatorname{tr} \left(n_1 \Sigma^{-1} (\bar{\mathbf{x}} - \mu_1)(\bar{\mathbf{x}} - \mu_1)^\top \right) = n_1 (\bar{\mathbf{x}} - \mu_1)^\top \Sigma^{-1} (\bar{\mathbf{x}} - \mu_1) \geq 0,$$

y análogamente para μ_2 . Se tiene que $\log L_{12}(\mu_1, \mu_2, \Sigma)$ es maximizada para cualquier $\Sigma > \mathbf{0}$ cuando $\mu_1 = \bar{\mathbf{x}}$ y $\mu_2 = \bar{\mathbf{y}}$.



Comparando poblaciones normales

De este modo,

$$\log L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \boldsymbol{\Sigma}) \geq \log L_{12}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}),$$

y se desea maximizar

$$\log L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \boldsymbol{\Sigma}) = c - \frac{n}{2} \{ \log |\boldsymbol{\Sigma}| + \text{tr } \boldsymbol{\Sigma}^{-1} \boldsymbol{Q} / n \},$$

y es fácil notar que $\hat{\boldsymbol{\Sigma}} = \boldsymbol{Q} / n$. En efecto,

$$\log L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}) \geq \log L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \boldsymbol{\Sigma}) \geq \log L_{12}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}),$$

de modo que $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\mu}}_2$ y $\hat{\boldsymbol{\Sigma}}$ son los MLE bajo $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, además

$$\log L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}) = c - \frac{n}{2} \{ \log |\hat{\boldsymbol{\Sigma}}| + \text{tr } \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}} \},$$

o bien

$$L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-np/2} |\hat{\boldsymbol{\Sigma}}|^{-n/2} e^{-np/2}.$$



El test de razón de verosimilitudes es

$$\begin{aligned}\Lambda &= \frac{\max_{H_0} L_{12}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})}{\max_{H_1} L_{12}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)} = \frac{L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}})}{L_{12}(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_2)} \\ &= \frac{|\hat{\boldsymbol{\Sigma}}|^{-n/2}}{|\hat{\boldsymbol{\Sigma}}_1|^{-n_1/2} |\hat{\boldsymbol{\Sigma}}_2|^{-n_2/2}} = c_{12} \frac{|\mathbf{Q}_1|^{n_1/2} |\mathbf{Q}_2|^{n_2/2}}{|\mathbf{Q}|^{n/2}},\end{aligned}$$

con $c_{12} = n^{np/2} / (n_1^{n_1 p/2} n_2^{n_2 p/2})$. De este modo

$$LR = -2 \log \Lambda \xrightarrow{D} \chi^2(\nu),$$

con $\nu = p(p+1)/2$ (bajo H_0).



Comparando poblaciones normales

Suponga $\mathbf{x}_1, \dots, \mathbf{x}_{n_1}$ muestra aleatoria desde $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ independiente de la muestra $\mathbf{y}_1, \dots, \mathbf{y}_{n_2}$ desde $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. Se desea probar $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ ¹ asumiendo que $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 (= \boldsymbol{\Sigma})$. Sabemos que

$$\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}/n_1), \quad \mathbf{Q}_1 = (n_1 - 1)\mathbf{S}_1 \sim W_p(n_1 - 1, \boldsymbol{\Sigma}),$$

y $\bar{\mathbf{x}}$ es independiente de \mathbf{Q}_1 . Análogamente $\bar{\mathbf{y}} \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}/n_2)$ y es independiente de $\mathbf{Q}_2 = (n_2 - 1)\mathbf{S}_2 \sim W_p(n_2 - 1, \boldsymbol{\Sigma})$. Dado que $\mathbf{x}_1, \dots, \mathbf{x}_{n_1}$ y $\mathbf{y}_1, \dots, \mathbf{y}_{n_2}$ son independientes, tenemos que

$$\bar{\mathbf{x}} - \bar{\mathbf{y}} \sim N_p\left(\boldsymbol{\delta}, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right),$$

$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 \sim W_p(n_1 + n_2 - 2, \boldsymbol{\Sigma}),$$

y $\bar{\mathbf{x}} - \bar{\mathbf{y}}$ es independiente de \mathbf{Q} .

¹o equivalentemente $H_0 : \boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$.



Comparando poblaciones normales

Defina $S_P = Q/(n_1 + n_2 - 2)$, luego

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \boldsymbol{\delta})^\top \mathbf{S}_P^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}} - \boldsymbol{\delta}) \sim T^2(p, n_1 + n_2 - 2)$$
$$\stackrel{d}{=} \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F(p, n_1 + n_2 - p - 1).$$

Para probar $H_0 : \boldsymbol{\delta} = \mathbf{0}$ usamos el estadístico

$$T_0^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}} - \bar{\mathbf{y}})^\top \mathbf{S}_P^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}),$$

y se rechaza la hipótesis H_0 a un nivel de significancia α si

$$T_0^2 \geq \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F_{1-\alpha}(p, n_1 + n_2 - p - 1).$$

