UFOP/ICEB/DEEST
EST022 - Estatística Multivariada II
Prof. Tiago Martins Pereira

Máximo da função de verossimilhança para o modelo restrito: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k = \mathbf{\Sigma}$

Demonstração: Considere amostras aleatórias $K \geqslant 2$ populações normais p-variadas $N_p(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$ de tamanho n_k, $k = 1, ..., K$. A amostra da k-ésima população é dada por $\boldsymbol{X}_{k1}, \boldsymbol{X}_{k2}, \cdots, \boldsymbol{X}_{kn_k}$, em que $\boldsymbol{X}_{kj} \in \mathbb{R}^p$ corresponde à j-ésima unidade da k-ésima população. Considere também que $n = \sum_{k=1}^{K} n_k$.
O nosso interesse é testar:

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k = \mathbf{\Sigma} \quad \text{vs} \quad H_a : \text{pelo menos uma difere}$$

Temos que, a função de verossimilhança para o modelo irrestrito é dada por:

$$\boldsymbol{L}_\Omega(\boldsymbol{X}; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k) = (2\pi)^{-\frac{np}{2}} \prod_{k=1}^{K} |\mathbf{\Sigma}_k|^{-\frac{n_k}{2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k)^t \mathbf{\Sigma}_k^{-1} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k) \right\}$$

de forma que a função suporte será dada por:

$$g(\boldsymbol{X}; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k) = -\frac{np}{2} \ln(2\pi) - \sum_{k=1}^{K} \frac{n_k}{2} \ln |\mathbf{\Sigma}_k| - \frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{n_k} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k)^t \mathbf{\Sigma}_k^{-1} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k)$$

Utilizando os mesmos passos algébricos mostrados anteriormente, podemos mostrar que a expressão

$$\sum_{k=1}^{K} \sum_{j=1}^{n_k} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k)^t \mathbf{\Sigma}_k^{-1} (\boldsymbol{X}_{kj} - \boldsymbol{\mu}_k)$$

pode ser reescrita como

$$\sum_{k=1}^{K} \operatorname{tr}[(n_k - 1) \mathbf{\Sigma}_k^{-1} \boldsymbol{S}_k + n_k \mathbf{\Sigma}_k^{-1} (\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)^t],$$

em que a matriz

$$\boldsymbol{S}_k = \frac{1}{n_k - 1}\sum_{j=1}^{n_k}(\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)(\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)^t$$

representa o estimador não-viesado da covariância para a k-ésima população. Temos então que a função suporte pode ser reescrita como:

$$g(\boldsymbol{X};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) = -\frac{np}{2}\ln(2\pi) - \sum_{k=1}^{K}\frac{n_k}{2}\ln|\boldsymbol{\Sigma}_k| - \frac{1}{2}\sum_{k=1}^{K}\mathrm{tr}[(n_k-1)\boldsymbol{\Sigma}_k^{-1}\boldsymbol{S}_k] - \frac{1}{2}\sum_{k=1}^{k}\mathrm{tr}[n_k\boldsymbol{\Sigma}_k^{-1}(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)^t]$$

Derivando a função g em relação a $\boldsymbol{\mu}_k$, temos:

$$\frac{\partial g}{\partial \boldsymbol{\mu}_k} = n_k\boldsymbol{\Sigma}_k^{-1}(\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k), \quad \text{para} \;\; k = 1, ..., K$$

Igualando a zero, temos o estimador de máxima verossimilhança para $\boldsymbol{\mu}_k$:

$$\begin{aligned}\frac{\partial g}{\partial \boldsymbol{\mu}_k} = \mathbf{0} \quad &\Rightarrow \quad n_k\boldsymbol{\Sigma}_k^{-1}(\bar{\boldsymbol{X}}_k - \hat{\boldsymbol{\mu}}_k) = 0 \\ &\Rightarrow \quad \bar{\boldsymbol{X}}_k - \hat{\boldsymbol{\mu}}_k = 0 \\ &\Rightarrow \quad \hat{\boldsymbol{\mu}}_k = \bar{\boldsymbol{X}}_k\end{aligned}$$

■

Derivando em relação a $\boldsymbol{\Sigma}_k$, obtemos o estimador de máxima verossimilhança para a matriz de covariâncias:

$$\frac{\partial g}{\partial \boldsymbol{\Sigma}_k} = -\frac{n_k}{2}\boldsymbol{\Sigma}_k^{-1} + \frac{(n_k-1)}{2}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{S}_k\boldsymbol{\Sigma}_k^{-1} + \frac{n_k}{2}\boldsymbol{\Sigma}_k^{-1}(\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)^t\boldsymbol{\Sigma}_k^{-1}$$

Igualando a zero e substituindo o estimador de $\boldsymbol{\mu}_k$:

$$\begin{aligned}-\frac{n_k}{2}\hat{\boldsymbol{\Sigma}}_k^{-1} + \frac{(n_k-1)}{2}\hat{\boldsymbol{\Sigma}}_k^{-1}\boldsymbol{S}_k\hat{\boldsymbol{\Sigma}}_k^{-1} + \frac{n_k}{2}\hat{\boldsymbol{\Sigma}}_k^{-1}(\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)(\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)^t\hat{\boldsymbol{\Sigma}}_k^{-1} &= \mathbf{0} \\ -\frac{n_k}{2}\hat{\boldsymbol{\Sigma}}_k^{-1} + \frac{(n_k-1)}{2}\hat{\boldsymbol{\Sigma}}_k^{-1}\boldsymbol{S}_k\hat{\boldsymbol{\Sigma}}_k^{-1} &= \mathbf{0}\end{aligned}$$

Pré e pós multiplicando ambos os lados da equação acima por $\hat{\boldsymbol{\Sigma}}_k$, e resolvendo para $\hat{\boldsymbol{\Sigma}}_k$, temos:

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_k\hat{\boldsymbol{\Sigma}}_k^{-1}\hat{\boldsymbol{\Sigma}}_k &= \frac{(n_k-1)}{n_k}\hat{\boldsymbol{\Sigma}}_k^{-1}\hat{\boldsymbol{\Sigma}}_k\boldsymbol{S}_k\hat{\boldsymbol{\Sigma}}_k^{-1}\hat{\boldsymbol{\Sigma}}_k \\ \hat{\boldsymbol{\Sigma}}_k &= \frac{n_k-1}{n_k}\boldsymbol{S}_k = \frac{1}{n_k}\sum_{j=1}^{n_k}(\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)(\boldsymbol{X}_{kj} - \bar{\boldsymbol{X}}_k)^t\end{aligned}$$

$\blacksquare$

Substituindo os estimadores de máxima verossimilhança do vetor de médias e matriz de covariâncias na função de verossimilhança, obtemos seu máximo para o espaço irrestrito:

$$\begin{aligned}
\boldsymbol{L}_{\Omega}(\boldsymbol{X};\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}}_k) &= (2\pi)^{-\frac{np}{2}}\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{-\frac{n_k}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\mathrm{tr}[(n_k-1)\hat{\boldsymbol{\Sigma}}_k^{-1}\boldsymbol{S}_k+n_k\hat{\boldsymbol{\Sigma}}_k^{-1}(\bar{\boldsymbol{X}}_k-\bar{\boldsymbol{X}}_k)(\bar{\boldsymbol{X}}_k-\bar{\boldsymbol{X}}_k)^t]\right\}\\
&= (2\pi)^{-\frac{np}{2}}\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{-\frac{n_k}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\mathrm{tr}\left[\ (n_k-1)\frac{n_k}{(n_k-1)}\boldsymbol{S}_k^{-1}\boldsymbol{S}_k\right]\right\}\\
&= (2\pi)^{-\frac{np}{2}}\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{-\frac{n_k}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{K}n_kp\right\}\\
&= (2\pi)^{-\frac{np}{2}}\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{-\frac{n_k}{2}}\exp\left\{-\frac{np}{2}\right\}
\end{aligned}$$

$\blacksquare$

Sob H_0, a função de verossimilhança para o modelo restrito é dada por:

$$\boldsymbol{L}_{\Omega_0}(\boldsymbol{X};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{n_k}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k)\right\}$$

e a função suporte, por:

$$g(\boldsymbol{X};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}) = -\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{n_k}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k)$$

Da mesma forma, tem-se:

$$\sum_{k=1}^{K}\sum_{j=1}^{n_k}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{kj}-\boldsymbol{\mu}_k) = \sum_{k=1}^{K}\mathrm{tr}[(n_k-1)\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_k+n_k\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)^t]$$

Logo, a função suporte pode ser reescrita como:

$$g(\boldsymbol{X};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}) = -\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_{k=1}^{K}\mathrm{tr}[(n_k-1)\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_k]-\frac{1}{2}\sum_{k=1}^{K}\mathrm{tr}[n_k\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)^t]$$

Derivando a função suporte em relação ao vetor de médias temos:

$$\frac{\partial g}{\partial\boldsymbol{\mu}_k} = n_k\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}_k-\boldsymbol{\mu}_k)\quad \text{para}\ \ k=1,2,...,K$$

Igualando a zero, temos o estimador de máxima verossimilhança da média da k-ésima população, dado por:

$$\begin{aligned}
\frac{\partial g}{\partial \boldsymbol{\mu}_k} = \mathbf{0} \quad & \Rightarrow \quad n_k \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}_k - \hat{\boldsymbol{\mu}}_k) = \mathbf{0} \\
& \Rightarrow \quad \bar{\boldsymbol{X}}_k - \hat{\boldsymbol{\mu}}_k = \mathbf{0} \\
& \Rightarrow \quad \hat{\boldsymbol{\mu}}_k = \bar{\boldsymbol{X}}_k
\end{aligned}$$

Derivando a função suporte em relação a $\boldsymbol{\Sigma}$, temos:

$$\frac{\partial g}{\partial \boldsymbol{\Sigma}} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \sum_{k=1}^{K} \frac{(n_k - 1)}{2} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_k \boldsymbol{\Sigma}^{-1} + \sum_{k=1}^{K} \frac{n_k}{2} \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)(\bar{\boldsymbol{X}}_k - \boldsymbol{\mu}_k)^t \boldsymbol{\Sigma}^{-1}$$

Igualando a zero e substituindo o estimador de máxima verossimilhança da média:

$$\begin{aligned}
-\frac{n}{2}\hat{\boldsymbol{\Sigma}}^{-1} + \sum_{k=1}^{K} \frac{(n_k - 1)}{2} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{S}_k \hat{\boldsymbol{\Sigma}}^{-1} + \sum_{k=1}^{K} \frac{n_k}{2} \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)(\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)^t \hat{\boldsymbol{\Sigma}}^{-1} &= \mathbf{0} \\
-\frac{n}{2}\hat{\boldsymbol{\Sigma}}^{-1} + \sum_{k=1}^{K} \frac{(n_k - 1)}{2} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{S}_k \hat{\boldsymbol{\Sigma}}^{-1} &= \mathbf{0}
\end{aligned}$$

Pré e pós multiplicando ambos os lados da equação acima por $\hat{\boldsymbol{\Sigma}}_k$, e resolvendo para $\hat{\boldsymbol{\Sigma}}_k$, temos:

$$\begin{aligned}
\sum_{k=1}^{K} (n_k - 1) \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{S}_k \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}} &= n \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}} \\
\hat{\boldsymbol{\Sigma}} &= \frac{\sum_{k=1}^{K} (n_k - 1) S_k}{n} = \frac{\sum_{k=1}^{K} n_k \hat{\boldsymbol{\Sigma}}_k}{n}
\end{aligned}$$

Assim, o máximo da função de verossimilhança para o modelo restrito será dado então, por:

$$\begin{aligned}
\boldsymbol{L}_{\Omega_0}(\boldsymbol{X}; \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}) &= (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{K} \mathrm{tr}[(n_k - 1)\hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{S}_k] - \frac{1}{2} \sum_{k=1}^{K} \mathrm{tr}[n_k \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)(\bar{\boldsymbol{X}}_k - \bar{\boldsymbol{X}}_k)^t] \right\} \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{K} \mathrm{tr}\left[(n_k - 1) \frac{n}{\sum_{k=1}^{K} (n_k - 1)} \boldsymbol{S}_k^{-1} \boldsymbol{S}_k \right] \right\} \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} n \frac{\sum_{k=1}^{K} (n_k - 1)}{\sum_{k=1}^{K} (n_k - 1)} \mathrm{tr}(\boldsymbol{I}) \right\} \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{ -\frac{np}{2} \right\}
\end{aligned}$$

A estatística do teste de razão de verossimilhança é dada por:

$$
\begin{aligned}
\Lambda &= \frac{\boldsymbol{L}_{\Omega_0}(\boldsymbol{X};\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}})}{\boldsymbol{L}_{\Omega}(\boldsymbol{X};\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\Sigma}}_k)} = \frac{(2\pi)^{-\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}}\exp\left\{-\frac{np}{2}\right\}}{(2\pi)^{-\frac{np}{2}}\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{-\frac{n_k}{2}}\exp\left\{-\frac{np}{2}\right\}} \\
&= \frac{\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{\frac{n_k}{2}}}{|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}
\end{aligned}
$$

De forma que,

$$
-2\ln(\Lambda) = -2\ln\left(\frac{\prod_{k=1}^{K}|\hat{\boldsymbol{\Sigma}}_k|^{\frac{n_k}{2}}}{|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}\right) = -\left[\sum_{k=1}^{K} n_k\ln|\hat{\boldsymbol{\Sigma}}_k| - n\ln|\hat{\boldsymbol{\Sigma}}|\right]
$$

possui assintoticamente distribuição qui-quadrado, sob H_0, com $f = \frac{(k-1)p(p+1)}{2}$ graus de liberdade.

Dessa forma, podemos formular um teste de hipóteses

$$
H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma} \quad \text{vs} \quad H_a : \text{pelo menos uma difere}
$$

Ao nível de significância α, rejeitamos H_0 em favor de H_a se observarmos

$$
\chi_2 = -\left[\sum_{k=1}^{K} n_k\ln|\hat{\boldsymbol{\Sigma}}_k| - n\ln|\hat{\boldsymbol{\Sigma}}|\right] > \chi^2_f(\alpha)
$$

de acordo com o Resultado 02. $\chi^2_f(\alpha)$ denota o $100(1-\alpha)$-ésimo percentil superior de uma distribuição χ^2_f, sendo $f = \frac{(k-1)p(p+1)}{2}$ gl.