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

Máximo da função de verossimilhança para o modelo restrito: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$

Demonstração: Seja $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ \cdots,\ \boldsymbol{X}_n$ uma amostra aleatória de uma distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Vimos que a função de verossimilhança para o modelo normal multivariado é dada por:

$$\boldsymbol{L}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j-\boldsymbol{\mu})\right\}$$

e, que a função suporte pode ser escrita como:

$$\begin{aligned} g(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\right] \\ &- \frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t\right] \end{aligned}$$

Derivando a função suporte em relação à matriz de covariâncias $\boldsymbol{\Sigma}$, temos

$$\begin{aligned} \frac{\partial g(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} &= -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\boldsymbol{\Sigma}^{-1} \\ &+ \frac{n}{2}\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})(\bar{\boldsymbol{X}}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1} \end{aligned}$$

Igualando a zero e substituindo o valor hipotético de $\boldsymbol{\mu}$, obtemos,

$$-\frac{n}{2}\hat{\boldsymbol{\Sigma}}^{-1}+\frac{1}{2}\hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}+\frac{n}{2}\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)^t\hat{\boldsymbol{\Sigma}}^{-1} = \mathbf{0}$$

Pré e pós multiplicando a equação acima por $\hat{\boldsymbol{\Sigma}}$, temos:

$$-\frac{n}{2}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}+\frac{1}{2}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}+\frac{n}{2}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)^t\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}} = \mathbf{0}$$

Simplificando,

$$-n\hat{\boldsymbol{\Sigma}} + \sum_{j=1}^{n} \left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)^t + n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t = \mathbf{0}$$

De onde vem que,

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}} &= \frac{\sum_{j=1}^{n} \left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)^t}{n} + \frac{n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t}{n} \\
&= \frac{\sum_{j=1}^{n} \left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)^t}{n} + (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \\
&= \boldsymbol{S}_n + (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \\
&= \frac{\sum_{j=1}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^t}{n}
\end{aligned}$$

$\blacksquare$

Vimos também, que podemos reescrever a função de verossimilhança da seguinte maneira:

$$\boldsymbol{L}(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})^t \right) \right] \right\}$$

De forma que, o máximo da função de verossimilhança para o modelo restrito, considerando $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ é dado por:

$$\boldsymbol{L}_{\Omega_0}(\boldsymbol{X}; \boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{n}\sum_{j=1}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^t \right)^{-1} \left(\sum_{j=1}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^t \right) \right] \right\}$$

Assim,

$$\begin{aligned}
\boldsymbol{L}_{\Omega_0}(\boldsymbol{X}; \boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}) &= (2\pi)^{-\frac{np}{2}} |\boldsymbol{S}_n + (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t|^{-\frac{n}{2}} \exp\left\{ -\frac{n}{2}\mathrm{tr}(I) \right\} \\
&= (2\pi)^{-\frac{np}{2}} |\boldsymbol{S}_n + (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t|^{-\frac{n}{2}} \exp\left\{ -\frac{np}{2} \right\}
\end{aligned}$$

$\blacksquare$

A razão entre os máximos das funções de verossimilhança para os espaços restrito e irrestrito fica então,

$$
\begin{aligned}
\Lambda &= \frac{\boldsymbol{L}_{\Omega_0}(\boldsymbol{X};\boldsymbol{\mu}_0,\hat{\boldsymbol{\Sigma}})}{\boldsymbol{L}_{\Omega}(\boldsymbol{X};\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})} = \frac{(2\pi)^{-\frac{np}{2}}|\boldsymbol{S}_n+(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)^t|^{-\frac{n}{2}}\exp\left\{-\dfrac{np}{2}\right\}}{(2\pi)^{-\frac{np}{2}}|\boldsymbol{S}_n|^{-\frac{n}{2}}\exp\left\{-\dfrac{np}{2}\right\}} \\
&= \frac{|\boldsymbol{S}_n+(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)^t|^{-\frac{n}{2}}}{|\boldsymbol{S}_n|^{-\frac{n}{2}}} \\
&= \frac{|\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}}}{|\boldsymbol{S}_n|^{-\frac{n}{2}}}
\end{aligned}
$$

De forma que,

$$
-2\ln(\Lambda) = -2\ln\left(\left[\frac{|\hat{\boldsymbol{\Sigma}}|}{|\boldsymbol{S}_n|}\right]^{-\frac{n}{2}}\right) = n\left(\ln\left[\frac{|\hat{\boldsymbol{\Sigma}}|}{|\boldsymbol{S}_n|}\right]\right) = n\left(\ln(|\hat{\boldsymbol{\Sigma}}|)-\ln(|\boldsymbol{S}_n|)\right)
$$

Observe, agora, que:

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu}_0)(\boldsymbol{X}_j-\boldsymbol{\mu}_0)^t \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t+\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^t+2\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^t\right] \\
&= \boldsymbol{S}_n+\boldsymbol{r}\boldsymbol{r}^t
\end{aligned}
$$

sendo $\boldsymbol{r}=\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0$. Assim, temos que:

$$
\begin{aligned}
-2\ln(\Lambda) &= n\left[\ln(|\hat{\boldsymbol{\Sigma}}|)-\ln(|\boldsymbol{S}_n|)\right] \\
&= n\left[\ln(|\boldsymbol{S}_n+\boldsymbol{r}\boldsymbol{r}^t|)-\ln(|\boldsymbol{S}_n|)\right] \\
&= n\left[\ln(|\boldsymbol{S}_n|\cdot|\boldsymbol{I}+\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t|)-\ln(|\boldsymbol{S}_n|)\right] \\
&= n\left[\left(\ln(|\boldsymbol{S}_n|)+\ln(|\boldsymbol{I}+\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t|)-\ln(|\boldsymbol{S}_n|)\right)\right] \\
&= n\left[\ln(|\boldsymbol{I}+\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t|)\right]
\end{aligned}
$$

Temos que, se a matriz $\boldsymbol{A}$ é simétrica e posto$(\boldsymbol{A})=1$, então $|\boldsymbol{I}+\boldsymbol{A}|=1+\mathrm{tr}(\boldsymbol{A})$. Como posto$(\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t)=$ posto$(\boldsymbol{r}\boldsymbol{r}^t))=$ posto$(\boldsymbol{r})=1$,

$$
|\boldsymbol{I}+\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t| = 1+\mathrm{tr}(\boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t) = 1+\mathrm{tr}(\boldsymbol{r}^t\boldsymbol{S}_n^{-1}\boldsymbol{r}) = 1+\boldsymbol{r}^t\boldsymbol{S}_n^{-1}\boldsymbol{r}
$$

de forma que,

$$-2\ln(\Lambda) = n\left[\ln(|\boldsymbol{I} + \boldsymbol{S}_n^{-1}\boldsymbol{r}\boldsymbol{r}^t|)\right] = n\left[\ln(1 + \boldsymbol{r}^t\boldsymbol{S}_n^{-1}\boldsymbol{r})\right]$$

Como $\boldsymbol{S}_n^{-1} = \dfrac{n}{n-1}\boldsymbol{S}^{-1}$, podemos expressar $\boldsymbol{r}^t\boldsymbol{S}_n^{-1}\boldsymbol{r}$ por $\dfrac{n}{n-1}\boldsymbol{r}^t\boldsymbol{S}^{-1}\boldsymbol{r}$, de forma que $-2\ln(\Lambda)$ fica igual a,

$$-2\ln(\Lambda) = n\left[\ln\left(1 + \frac{n(\boldsymbol{r}^t\boldsymbol{S}^{-1}\boldsymbol{r})}{n-1}\right)\right]$$

Rejeitamos H_0 se a estatística acima for um valor demasiadamente grande. Isto é equivalente a dizer que, rejeita-se H_0 se o valor de

$$n\left(\boldsymbol{r}^t\boldsymbol{S}^{-1}\boldsymbol{r}\right) = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) = T^2$$

for demasiadamente grande. A declaração acima leva imediatamente a um teste das hipóteses

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{vs} \quad H_a : \boldsymbol{\mu} = \boldsymbol{\mu}_0.$$

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

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$$

de acordo com a proposição 01. $F_{p,n-p}(\alpha)$ denota o $100(1-\alpha)$-ésimo percentil superior de uma distribuição $F[p;(n-p)]$.

■