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

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

Demonstração: Seja $\boldsymbol{X}_1, \ \boldsymbol{X}_2, \ \cdots, \ \boldsymbol{X}_n$ uma amostra aleatória de uma distribuição $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. O nosso interesse é testar:

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0 \quad \text{vs} \quad H_a : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0$$

Vimos que o máximo da função de verossimilhança para o modelo irrestrito é dado por

$$\boldsymbol{L}_\Omega(\boldsymbol{X}; \hat{\boldsymbol{\mu}}, \hat{\mathbf{\Sigma}}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{S}_n|^{-\frac{n}{2}} \exp\left\{-\frac{np}{2}\right\}$$

Para o modelo restrito, sob H_0, a verossimilhança normal fica,

$$\boldsymbol{L}(\boldsymbol{X}; \boldsymbol{\mu}, \mathbf{\Sigma}_0) \quad = \quad \frac{1}{(2\pi)^{np/2} |\mathbf{\Sigma}_0|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=i}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu})^t \, \mathbf{\Sigma}_0^{-1} \, (\boldsymbol{X}_j - \boldsymbol{\mu})\right\}.$$

de forma que, a função suporte pode ser escrita como:

$$g(\boldsymbol{X}; \boldsymbol{\mu}, \mathbf{\Sigma}_0) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Sigma}_0| - \frac{1}{2}\sum_{j=i}^{n} (\boldsymbol{X}_j - \boldsymbol{\mu})^t \, \mathbf{\Sigma}_0^{-1} \, (\boldsymbol{X}_j - \boldsymbol{\mu})$$

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

$$\frac{\partial g\left(\boldsymbol{X}; \boldsymbol{\mu}, \mathbf{\Sigma}_0\right)}{\partial \boldsymbol{\mu}} = \mathbf{\Sigma}_0^{-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})$$

Igualando a zero, obtemos,

$$
\begin{aligned}
\boldsymbol{\Sigma}_0{}^{-1}\sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0} \quad &\Rightarrow \quad \sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j-\sum_{j=1}^{n}\hat{\boldsymbol{\mu}}=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j-n\hat{\boldsymbol{\mu}}=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j=n\hat{\boldsymbol{\mu}} \quad \therefore \quad \hat{\boldsymbol{\mu}}=\frac{\sum_{j=1}^{n}\boldsymbol{X}_j}{n}=\bar{\boldsymbol{X}}
\end{aligned}
$$

■

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

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

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

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

Assim,

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

■

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

$$
\begin{aligned}
\Lambda &= \frac{\boldsymbol{L}_{\Omega_0}(\boldsymbol{X};\hat{\boldsymbol{\mu}},\boldsymbol{\Sigma}_0)}{\boldsymbol{L}_{\Omega}(\boldsymbol{X};\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})} = \frac{(2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}_0|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right]\right\}}{(2\pi)^{-\frac{np}{2}}|\boldsymbol{S}_n|^{-\frac{n}{2}}\exp\left\{-\frac{np}{2}\right\}} \\
&= \frac{|\boldsymbol{\Sigma}_0|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right]\right\}}{|\boldsymbol{S}_n|^{-\frac{n}{2}}\exp\left\{-\frac{np}{2}\right\}} \\
&= |\boldsymbol{\Sigma}_0^{-1}|^{\frac{n}{2}}|\boldsymbol{S}_n|^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right]\right\}\exp\left\{\frac{np}{2}\right\} \\
&= |\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n|^{\frac{n}{2}}\exp\left\{-\frac{n\mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right]+np}{2}\right\}
\end{aligned}
$$

De forma que,

$$
\begin{aligned}
-2\ln(\Lambda) &= -2\ln\left(\left|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right|^{\frac{n}{2}}\exp\left\{\frac{-n\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n\right)+np}{2}\right\}\right) \\
&= n\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n) - n\ln|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n| - np
\end{aligned}
$$

De acordo com o Resultado 02, temos que, assintoticamente, $-2\ln(\Lambda)\dot{\sim}\chi^2_{r-s}$, sendo r o número de parâmetros estimados no espaço irrestrito e s o número de parâmetros estimados no modelo restrito. No modelo irrestrito para uma população, estimamos p médias e $\dfrac{p(p+1)}{2}$ variâncias e covariâncias, totalizando $r = \dfrac{p(p+3)}{2}$ parâmetros. Neste caso, no modelo restrito, estimamos somente p médias. Assim, temos que $s = p$ parâmetros e $f = r - s = \dfrac{p(p+3)}{2} - p = \dfrac{p(p+1)}{2}$.

Dessa forma, podemos formular um teste de hipóteses

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \quad \text{vs} \quad H_a : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$$

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

$$\chi = n\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n) - n\ln|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{S}_n| - np > \chi^2_f(\alpha)$$

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 $f = \dfrac{p(p+1)}{2}$ graus de liberdade.