

Teste de Hipóteses para a matriz de correlação

$R = \{\rho_{ij}\} \quad i, j = 1, 2, \dots, p$ matriz de correlação populacional

$$\underline{X} \sim N_p(\mu, \Sigma)$$

$H_0: R = I$ (Neste caso é equivalente a testar que as componentes de X são independentes)

$H_a: R \neq I$

$\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ amostra aleatória do vetor aleatório X .

Teste da Razão de Verossimilhanças

μ, Σ parâmetros $p + p + \frac{p(p-1)}{2} = \frac{p(p+3)}{2}$ parâmetros

$\downarrow \quad \downarrow \quad \downarrow$
médias variâncias covariâncias

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\underline{x}_i - \mu)' \Sigma^{-1} (\underline{x}_i - \mu) \right]$$

Ω - espaço paramétrico $\frac{p(p+3)}{2}$ dimensional

$$\sup L(\mu, \Sigma) = L(\hat{\mu}, \hat{\Sigma})$$

$\hat{\mu}, \hat{\Sigma}$ estimadores de max. veross. de μ e Σ

$$= \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})' \hat{\Sigma}^{-1} (x_i - \bar{x}) \right]$$

$$\sum_{i=1}^n (x_i - \bar{x})' \hat{\Sigma}^{-1} (x_i - \bar{x}) = \text{tr} \left(\sum_{i=1}^n (x_i - \bar{x})' \hat{\Sigma}^{-1} (x_i - \bar{x}) \right) =$$

$$= \sum_{i=1}^n \text{tr} \left[\underbrace{(x_i - \bar{x})'}_A \underbrace{\hat{\Sigma}^{-1} (x_i - \bar{x})}_B \right] = \sum_{i=1}^n \text{tr} \left[\underbrace{(x_i - \bar{x})}_{p \times 1} \underbrace{(x_i - \bar{x})'}_{1 \times p} \underbrace{\hat{\Sigma}^{-1}}_{p \times p} \right]$$

$$= \text{tr} \left\{ \underbrace{\sum_{i=1}^n \left[(x_i - \bar{x}) (x_i - \bar{x})' \hat{\Sigma}^{-1} \right]}_{n \hat{\Sigma}} \right\} = \text{tr} (n \mathbb{I}_p) = np$$

$$= \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} np \right]$$

O subespaço de Ω correspondendo a H_0 é

$$\Omega_0 = \left\{ -\infty < \mu_j < \infty, 0 < \sigma_j^2 < \infty, \sigma_{jk} = 0 \text{ } j \neq k, j, k = 1, 2, \dots, p \right\}$$

$2p$ parâmetros

Sob H_0 , a função de verossimilhança é o produto de n densidades normal p -variadas com matriz de covariâncias diagonal D

$$D = \begin{bmatrix} \sigma_1^2 & 0 & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{bmatrix} = \Sigma$$

$$L_0(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} \left(\prod_{j=1}^p \sigma_j^2 \right)^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' D^{-1} (x_i - \mu) \right]$$

$$\sum_{i=1}^n (x_i - \mu)' D^{-1} (x_i - \mu) = \sum_{i=1}^n \begin{bmatrix} x_{i1} - \mu_1 \\ x_{i2} - \mu_2 \\ \vdots \\ x_{ip} - \mu_p \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ & & \ddots \\ 0 & & & 1/\sigma_p^2 \end{bmatrix} \begin{bmatrix} x_{i1} - \mu_1 \\ x_{i2} - \mu_2 \\ \vdots \\ x_{ip} - \mu_p \end{bmatrix}$$

$$= \sum_{i=1}^n \left[\frac{(x_{i1} - \mu_1)^2}{\sigma_1^2} + \frac{(x_{i2} - \mu_2)^2}{\sigma_2^2} + \dots + \frac{(x_{ip} - \mu_p)^2}{\sigma_p^2} \right]$$

$$\sup L_0(\mu, \Sigma) \text{ é obtido para } \hat{\mu} = \bar{X} \text{ e } \hat{\sigma}_j^2 = \sum_{i=1}^n \frac{(x_{ij} - \bar{x}_j)^2}{n}$$

$$j = 1, 2, \dots, p$$

$$\sup L_0(\mu, \Sigma) = L_0(\bar{x}, \hat{\Sigma}_0)$$

$$\hat{\Sigma}_0 = \begin{bmatrix} \hat{\sigma}_1^2 & & 0 \\ & \hat{\sigma}_2^2 & \\ 0 & & \ddots & \hat{\sigma}_p^2 \end{bmatrix}$$

↑ unidim

$$= \frac{1}{(2\pi)^{np/2} \left(\prod_{j=1}^p \hat{\sigma}_j^2 \right)^{n/2}} \left\{ -\frac{1}{2} \sum_{i=1}^n \left[\frac{(x_{i1} - \bar{x}_1)^2}{\hat{\sigma}_1^2} + \frac{(x_{i2} - \bar{x}_2)^2}{\hat{\sigma}_2^2} + \dots + \frac{(x_{ip} - \bar{x}_p)^2}{\hat{\sigma}_p^2} \right] \right\}$$

$$= \frac{1}{(2\pi)^{np/2} \left(\prod_{j=1}^p \hat{\sigma}_j^2 \right)^{n/2}} \exp \left\{ -\frac{1}{2} \left[\frac{n \hat{\sigma}_1^2}{\hat{\sigma}_1^2} + \frac{n \hat{\sigma}_2^2}{\hat{\sigma}_2^2} + \dots + \frac{n \hat{\sigma}_p^2}{\hat{\sigma}_p^2} \right] \right\}$$

$$= \frac{1}{(2\pi)^{np/2} \left(\prod_{j=1}^p \hat{\sigma}_j^2 \right)^{n/2}} \exp \left(-\frac{np}{2} \right).$$

Estatística da Razão de Verossimilhança

$$\Lambda = \frac{L_0(\bar{x}, \hat{\Sigma}_0)}{L(\hat{\mu}, \hat{\Sigma})} = \frac{|\hat{\Sigma}|^{n/2}}{\left(\prod_{j=1}^p \hat{\sigma}_j^2 \right)^{n/2}}$$

Observar que

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_1 & & 0 \\ & \hat{\sigma}_2 & \\ 0 & & \ddots \\ & & & \hat{\sigma}_p \end{bmatrix} \hat{R} \begin{bmatrix} \hat{\sigma}_1 & & 0 \\ & \hat{\sigma}_2 & \\ 0 & & \ddots \\ & & & \hat{\sigma}_p \end{bmatrix} = \Delta \hat{R} \Delta$$

$$\hat{R} = \{r_{jk}\} \quad j, k = 1, 2, \dots, p \quad \text{matriz de correlações}$$

$$|\hat{\Sigma}| = |\Delta \hat{R} \Delta| = |\Delta| |\hat{R}| |\Delta| = \prod_{j=1}^p \hat{\sigma}_j^2 |\hat{R}|$$

$$L = \left[\frac{|\hat{\Sigma}|}{\prod_{j=1}^p \hat{\sigma}_j^2} \right]^{n/2} = \left[\frac{\prod_{j=1}^p \hat{\sigma}_j^2 |\hat{R}|}{\prod_{j=1}^p \hat{\sigma}_j^2} \right]^{n/2} = |\hat{R}|^{n/2}$$

Sob H_0 , $-2 \ln L$ tem assintoticamente distribuição χ^2 com $g.l. = \dim(\Omega) - \dim(\Omega_0) = \frac{p(p+3)}{2} - 2p = \frac{p(p-1)}{2}$.

Rejeitamos H_0 ao nível α se $-2 \ln |\hat{R}|^{n/2} > \chi^2_{\frac{p(p-1)}{2}, \alpha}$

$\chi^2_{\frac{p(p-1)}{2}, \alpha}$ quantil de ordem $1-\alpha$ da dist. χ^2 com $\frac{p(p-1)}{2}$ g.l.

Bartlett sugere o uso da estatística

$$T = - \left(n - 1 - \frac{2p+5}{6} \right) \ln |R|$$

que sob H_0 tem distribuição assintótica $\chi^2_{\frac{p(p-1)}{2}}$

Rejeitamos H_0 se $T > \chi^2_{\frac{p(p-1)}{2}, \alpha}$.

Exemplo

Em um estudo sobre alterações de arteriosclerose, um pesquisador encontrou a seguinte matriz de correlação para alterações de arteriosclerose na aorta abdominal, aorta torácica e artéria coronariana para uma amostra de 54 pessoas.

$$\hat{R} = \begin{bmatrix} 1 & 0,69 & 0,43 \\ & 1 & 0,30 \\ & & 1 \end{bmatrix}$$

Teste a hipótese de independência entre os graus de alterações nas três artérias na população.

$$H_0: R = I$$

$$H_a: R \neq I$$

Supondo que as três variáveis têm distribuição N_3 , $R = I \Rightarrow X_1, X_2, X_3$ são independentes

$$|\hat{R}| = 0,427 \quad 0 \leq |\hat{R}| \leq 1$$

$$\Lambda = |\hat{R}|^{n/2} = 0,427^{27}$$

$$-2 \ln \Lambda = -54 \ln 0,427 = 45,95$$

$$\frac{p(p-1)}{2} = 3 \text{ gl} \quad -2 \ln \Lambda > \chi^2_{3,0,001}$$

Rejeita-se H_0 .

Estatística de Bartlett

$$\begin{aligned} T &= - \left(n-1 - \frac{2p+5}{6} \right) \ln 0,427 = -51,17(-0,85) = \\ &= 43,49 \end{aligned}$$

Rejeitamos H_0 .

4 - Inferência para o vetor de Médias μ $X \sim N(\mu, \Sigma)$

Resultados do Tópico 3

$\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ amostra aleatória da dist. $N_p(\mu, \Sigma)$.

a) $\bar{X} \sim N_p(\mu, \frac{1}{n} \Sigma)$

$$S = \frac{1}{n} \sum_{i=1}^n (\tilde{x}_i - \bar{X})(\tilde{x}_i - \bar{X})'$$

b) $(n-1)S^* = nS \sim W(\Sigma, n-1)$ $S^* = \frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \bar{X})(\tilde{x}_i - \bar{X})'$

c) $\bar{X}$ e S são independentes

Resultado - Distribuição T^2 de Hotelling

Seja Y uma variável aleatória com distribuição $F_{p, n-p}$. Nessas condições, a variável aleatória

$$X = \frac{(n-1)p}{n-p} Y \quad \left(\overset{T^2}{Y} = \frac{n-p}{(n-1)p} \overset{T^2}{X} \right)$$

tem distribuição T^2 de Hotelling com parâmetros $n-1$ e p . (Ver pag. 74 - Mardia, Kent e Bibby)

$$X \sim T^2_{n-1, p}$$

Resultado

A distribuição T^2 de Hotelling com parâmetros m e p pode ser obtida como

$$T^2 = m (X - \mu)' W^{-1} (X - \mu)$$

onde $X \sim N_p(\mu, \Sigma)$ e $W \sim \text{Wishart}(\Sigma, m)$
com X e W independentes.

Resultado

Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ são vetores aleatórios independentes identicamente distribuídos com distribuição $N_p(\mu, \Sigma)$ e se $n-1 \geq p$ então

$$T^2 = n (\bar{X} - \mu)' S^{*-1} (\bar{X} - \mu) \sim T_{n-1, p}^2$$

Prova: $\bar{X} \sim N(\mu, \frac{1}{n} \Sigma)$

$$(\bar{X} - \mu) \sim N_p(0, \frac{1}{n} \Sigma)$$

$$\sqrt{n} (\bar{X} - \mu) \sim N_p(0, \Sigma)$$

$$(n-1) S^* \sim W(\Sigma, n-1)$$

$\bar{X}$ e S^* independentes

$$\begin{aligned} \therefore & \underbrace{(n-1)}_m \underbrace{\sqrt{n} (\bar{X} - \mu)' [(n-1) S^*]^{-1} \sqrt{n}}_{N_p(0, \Sigma)} \underbrace{(\bar{X} - \mu)}_{W(\Sigma, n-1)} \\ & \quad \quad \quad \downarrow \\ & = n (\bar{X} - \mu)' S^{*-1} (\bar{X} - \mu) \end{aligned}$$

$$\sim T_{n-1, p}^2$$

10

$X \sim N(\mu, \Sigma)$ μ e Σ desconhecidos

$$H_0: \mu = \mu_0 \quad H_a: \mu \neq \mu_0$$
$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \mu_0 = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{p0} \end{bmatrix}$$

vetor de constantes especificadas

Sob $H_0, \mu = \mu_0$ e $X_1, X_2, \dots, X_n$ a.a da dist $X \sim N(\mu, \Sigma)$

$$T^2 = n(\bar{X} - \mu_0)' S^{*-1} (\bar{X} - \mu_0) \sim T_{n-1, p}^2$$

Grandes valores de T^2 levam à rejeição de H_0 .

Rejeita-se H_0 ao nível α se $T^2 > \frac{(n-1)p F_{p, n-p}(\alpha)}{n-p}$

onde $F_{p, n-p}(\alpha)$ é o quantil de ordem $1-\alpha$ da dist. F com p graus de liberdade no numerador e $n-p$ no denominador

justificativa

$P(T^2 > t_c) = \alpha$ t_c quantil de ordem $1-\alpha$ da dist. T^2 H_0

$$P\left(\frac{n-p}{(n-1)p} T^2 > \frac{n-p}{(n-1)p} t_c\right) = \alpha = P\left(T^2 > \underbrace{\frac{(n-1)p}{n-p} \frac{n-p}{(n-1)p} t_c}_{F_c}\right)$$

Exemplo

X_1 - nota de Aptidão Verbal

X_2 - nota de Raciocínio Abstrato

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mu, \Sigma) \quad \mu \text{ e } \Sigma \text{ desconhecidos}$$

Para uma amostra de $n=101$ alunos observou-se

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 55,24 \\ 34,97 \end{bmatrix}$$

e a matriz de covariância amostral é

$$S^* = \frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_i - \bar{x})(\tilde{x}_i - \bar{x})' = \begin{bmatrix} 210,54 & 126,99 \\ 126,99 & 119,68 \end{bmatrix}$$

Testar ao nível $\alpha = 0,01$ a hipótese de que o vetor de médias é $\mu_0 = \begin{bmatrix} 60 \\ 50 \end{bmatrix}$.

$$\begin{array}{l|l} H_0: \mu = \begin{bmatrix} 60 \\ 50 \end{bmatrix} & T^2 = n(\bar{x} - \mu_0)' S^{*-1} (\bar{x} - \mu_0) \\ H_a: \mu \neq \begin{bmatrix} 60 \\ 50 \end{bmatrix} & = 101 \begin{bmatrix} 55,24 - 60 & 34,97 - 50 \end{bmatrix} \times \\ & = \begin{bmatrix} 0,01319 & -0,01400 \\ -0,01400 & 0,02321 \end{bmatrix} \begin{bmatrix} 55,24 - 60 \\ 34,97 - 50 \end{bmatrix} = 357,43 \end{array}$$

$$p=2 \quad n-p=99 \quad F_{2,60}(0,01) = 4,98 \Rightarrow F_{2,99}(0,01) < 4,98$$

Rejeita-se H_0 para $T^2 > \frac{100 \cdot 2}{99} \cdot 4,98 \simeq 10.$

Rejeitamos $H_0.$