

1

Comparação entre os m intervalos baseados em T^2 e pelo método de Bonferroni

$$T^2: l'\bar{X} \pm \sqrt{\frac{p(n-1)}{n-p}} F_{p, n-p}(\alpha) \sqrt{\frac{l'S^*l}{n}}$$

$$\text{Bonferroni: } l'\bar{X} \pm t_{n-1}\left(\frac{\alpha}{2m}\right) \sqrt{\frac{l'S^*l}{n}}$$

$$\Delta = \frac{\text{Amplitude do I Bonferroni}}{\text{Amplitude do I } T^2} = \frac{t_{n-1}\left(\frac{\alpha}{2m}\right)}{\sqrt{\frac{p(n-1)}{n-p} F_{p, n-p}(\alpha)}}$$

$$\Delta \text{ para } \gamma = 1 - \alpha = 0,95, \alpha_i = \frac{0,05}{m}$$

$m = p$				
n	2	4	10	
15	0,88	0,69	0,29	
25	0,90	0,75	0,48	
50	0,91	0,78	0,58	
100	0,91	0,80	0,62	

Os intervalos de Bonferroni são bem mais precisos para $m=p$, principalmente para p grande.

Verifica-se que os intervalos com base em T^2 são mais precisos para m muito maior que p .

Teste de Igualdade de dois vetores de médias

2

$$X_1 \sim N_p(\mu_1, \Sigma) \quad \text{População 1}$$

$$X_2 \sim N_p(\mu_2, \Sigma) \quad \text{População 2}$$

Testar

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

$X_{11}, X_{12}, \dots, X_{1n_1}$ amostra aleatória de tamanho n_1 da população 1, com dist. $N_p(\mu_1, \Sigma)$.

$X_{21}, X_{22}, \dots, X_{2n_2}$ amostra aleatória de tamanho n_2 da população 2, com dist. $N_p(\mu_2, \Sigma)$.

$X_{11}, X_{12}, \dots, X_{1n_1}$ independentes de $X_{21}, X_{22}, \dots, X_{2n_2}$.

$$\bar{X}_1 = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \vdots \\ \bar{X}_{1p} \end{bmatrix} \quad \bar{X}_2 = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \vdots \\ \bar{X}_{2p} \end{bmatrix}$$

vetores de médias amostrais

$\hat{\Sigma}$ matriz de covariância comum é estimada por

$$S_c = \frac{1}{n_1 + n_2 - 2} \left[(n_1 - 1)A_1 + (n_2 - 1)A_2 \right]$$

A_1 = matriz de covariância amostral, amostra da população 1 (denominador $n_1 - 1$)

$$a_{jk} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1ij} - \bar{X}_{1j})(X_{1ik} - \bar{X}_{1k})$$

A_2 = matriz de covariância amostral, amostra da população 2 (denominador $n_2 - 1$)

$$a_{jk} = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2ij} - \bar{X}_{2j})(X_{2ik} - \bar{X}_{2k})$$

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

X_{lij} - valor da variável X_j ($j = 1, 2, \dots, p$) para o i -ésimo elemento amostral ($i = 1, 2, \dots, n_l$) da l -ésima população $l = 1, 2$.

Rejeitamos $H_0: \mu_1 = \mu_2$ ao nível de significância α se

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S_c^{-1} (\bar{X}_1 - \bar{X}_2) \geq$$

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

Uma região de confiança com coeficiente δ para $\mu_1 - \mu_2$ é dada pelo conjunto dos valores de $\mu_1 - \mu_2$ tais que

$$[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)]' S_c^{-1} [\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)] \leq$$

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

Lembrando, uma vez observadas as amostras, esta região é um elipsoide de centro $\bar{X}_1 - \bar{X}_2$.

Exemplo

5

Quatro provas de um teste de inteligência (Informações, Similaridades, Aritmética, Figuras) foram aplicadas a duas amostras de pessoas idosas com e sem um fator senil

Prova	médias amostrais	
	Sem fator senil $n_1 = 37$	Com fator senil $n_2 = 12$
Informações	12,57	8,75
Similaridades	9,57	5,33
Aritmética	11,49	8,50
Figuras	7,97	4,75

$$S_c = \frac{1}{n_1 + n_2 - 2} \left[(n_1 - 1) A_1 + (n_2 - 1) A_2 \right]$$

A_l - matriz de covariância amostral, denominador $n_l - 1$, associada à l -ésima população

$$S_c = \begin{bmatrix} 11,2553 & 9,4042 & 7,1489 & 3,3830 \\ & 13,5318 & 7,3830 & 2,5532 \\ & & 11,5744 & 2,6170 \\ & & & 5,8085 \end{bmatrix}$$

$$S_c^{-1} = \begin{bmatrix} 0,259064 & -0,135783 & -0,058797 & -0,064719 \\ & 0,186449 & -0,038305 & 0,014382 \\ & & 0,150964 & -0,016920 \\ & & & 0,211171 \end{bmatrix} \quad 6$$

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' S_c^{-1} (\bar{x}_1 - \bar{x}_2) = 22,05$$

$$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}(\alpha) = 11,15$$

$$p = 4 \quad n_1 + n_2 - p - 1 = 4$$

$$\alpha = 0,05 \quad F_{4,44}(0,05) \simeq 2,61$$

$$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} = \frac{47.4}{44} = 4,27$$

$$T_{obs}^2 = 22,05 > 11,15$$

Rejeita-se $H_0: \mu_{sf} = \mu_{cf}$.

Os dados sugerem a inexistência de igualdade de médias em pelo menos uma das quatro provas.

Rejeitada H_0 , construir IC para as diferenças de médias, para detectar quais variáveis produziram a rejeição de H_0 (e portanto a desigualdade de médias).

Intervalos de confiança para $l'(\mu_1 - \mu_2)$ pelo método de T^2 de Hotelling são dados por

$$l'(\bar{X}_1 - \bar{X}_2) \pm \sqrt{l' S_c l} \sqrt{\frac{n_1 + n_2}{n_1 n_2} \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F(\alpha)}_{p, n_1 + n_2 - p - 1}$$

com coeficiente global $1 - \alpha$.

No exemplo, construir IC para as componentes de $\mu_1 - \mu_2$

$$\mu_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \end{bmatrix} \quad \mu_2 = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \end{bmatrix} \quad l'(\mu_1 - \mu_2) \text{ para } l' = (1 \ 0 \ 0 \ 0)$$

$$l'(\mu_1 - \mu_2) = \mu_{11} - \mu_{21}$$

$$l'(\bar{X}_1 - \bar{X}_2) = \bar{X}_{11} - \bar{X}_{21}$$

Fixando coef. de conf. global $1 - \alpha = 0,95$

IC para $\mu_{11} - \mu_{21}$:

$$\bar{X}_{11} - \bar{X}_{21} \pm \sqrt{\frac{49}{37,12} \cdot \frac{47,4}{49-4-1} F_{4,44}(0,05)} \underbrace{\sqrt{l' S_c l}}_{11,2553} \quad 3,82 \pm 3,72$$

12,57 - 8,75

$$\mu_{11} - \mu_{21} : [0,10; 7,54]$$

$$\mu_{12} - \mu_{22} : [0,16; 8,32]$$

Coef. de conf. global

$$\mu_{13} - \mu_{23} : [-0,79; 6,77]$$

0,95

$$\mu_{14} - \mu_{24} : [0,54; 5,90]$$

Para a diferença de médias em Aritméticas, o IC contém o valor 0. Os três outros IC indicam diferenças entre as médias populacionais

Pelo método de Bonferroni os IC ficam

$$\bar{X}_{1j} - \bar{X}_{2j} \pm t_{n_1+n_2-2} \left(\frac{\alpha}{2p} \right) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) s_{jj}}$$

s_{jj} - j -ésimo elemento da diagonal principal de S_c , $j = 1, 2, \dots, p$

- A suposição de igualdade de matrizes de covariância é mais forte que no caso univariado porque envolve igualdade de p variâncias e todas as possíveis covariâncias.
- Se $\Sigma_1 \neq \Sigma_2$, não é possível obter uma medida de distância do tipo T^2 cuja distribuição não dependa das matrizes Σ_1 e Σ_2 desconhecidas.
- Existem testes para $H_0: \Sigma_1 = \Sigma_2$ (Tópico 5 do curso) mas exigem distribuição normal multivariada. (Ver pág. 291 Johnson e Wichern).
- O problema de testar igualdade de Vetores de médias sem a suposição de igualdade de variâncias é denominado "The Behrens - Fisher Problem". (Ver pág 142 Mardia, Kent e Bibby).

Obs: Não fazer análises separadas para cada uma das p variáveis pois este procedimento não leva em consideração as correlações entre as p variáveis.

Exercício - Componentes Principais

10

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad \text{Var}(X) = \Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

a) Verifique que as raízes características e os vetores característicos normalizados são

$$\lambda_1 = 5,83 \quad e_1' = [0,383 \quad -0,924 \quad 0]$$

$$\lambda_2 = 2,00 \quad e_2' = [0 \quad 0 \quad 1]$$

$$\lambda_3 = 0,17 \quad e_3' = [0,924 \quad 0,383 \quad 0]$$

b) Considere as transformações

$$Y_1 = e_1' X = 0,383 X_1 - 0,924 X_2$$

$$Y_2 = e_2' X = X_3 \quad Y_3 = e_3' X = 0,924 X_1 + 0,383 X_2$$

Prove que $\text{Var}(Y_1) = \lambda_1$ $\text{Var}(Y_2) = \lambda_2$ e $\text{Var}(Y_3) = \lambda_3$
e que $\text{Var}(Y_1) + \text{Var}(Y_2) + \text{Var}(Y_3) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)$

c) Verifique que $\text{Cov}(Y_j, Y_k) = 0 \quad j \neq k$

As v. Transformadas são denominadas Componentes Principais associadas a Σ .