Inferência a Respeito dos Parâmetros da Distribuição Normal Multivariada

Vicente G. Cancho
garibay@icmc.usp.br

Inferência Para o Vetor de Parâmetros

Definição

Dizemos que um vetor aleatório $\boldsymbol{X}$ de ordem $(p \times 1)$ tem distribuição normal multivariada (ou p-variada) com vetor de médias $\boldsymbol{\mu}$ e matriz de variâncias e covariâncias Σ de ordem $p \times p$, se sua função de densidade conjunta é dada por

$$f(\boldsymbol{x}) = ((2\pi)^p|\Sigma|)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \ \boldsymbol{x} \in R^p. \quad (1)$$

onde $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ e $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$.

Notação: $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Propriedades da distribuição normal multivariada

Se $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\boldsymbol{X}_1$ e $\boldsymbol{X}_2$ é a partição de $\boldsymbol{X}$ de ordem $q \times 1$ e $(p-q) \times 1$ respectivamente e suponha que $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são particionados de acordo. Isto é

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

Então

(i) $X_j \sim N(\mu_j, \sigma_{jj})$, $j = 1, \ldots, p$.

(ii) $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \sim N_q(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}^\top)$

(iii) A distribuição condicional $\boldsymbol{X}_1$ dado $\boldsymbol{X}_2 = \boldsymbol{x}_2$ é

$$N_p\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

(iv) $(\boldsymbol{X} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_{(p)}$

EMV de $\boldsymbol{\mu}$ e Σ

Suponha que $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ é i-ésima amostra de uma população normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de variância-covariancia $\boldsymbol{\Sigma}$, ou seja $\boldsymbol{X}_i \overset{i.i.d}{\sim} N_p(, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.
A função de verossimilhança de $\boldsymbol{\mu}$ e Σ dada a amostra observada é dada por

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} ((2\pi)^p |\Sigma|)^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right\}$$

A função log-verossimilhança

$$\ell((\boldsymbol{\mu}, \boldsymbol{\Sigma})) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \sum_{i=1}^{n} \frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})$$

Estimador de máxima verossimilhança do vetor de médias

O EMV do vetor médias é o valor de $\boldsymbol{\mu}$ que minimiza

$$\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) = \sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i - 2n\bar{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + n\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

onde $\bar{x} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i$ é o vetor de medias amostrais.
Tomando a derivada em relação a $\boldsymbol{\mu}$, obtém-se

$$\frac{\partial \ell((\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -2n\boldsymbol{\Sigma}^{-1}\bar{x} + 2n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \iff \hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}.$$

vetor de medias amostrais $\bar{x}$ é o EMV do vetor de médias populacional $\boldsymbol{\mu}$

Estimador de máxima verossimilhança da matriz de covariâncias

O EMV da matriz de covariância é o valor de $\mathbf{\Sigma}$ que minimiza

$$-n\log(|\Sigma^{-1}|) + \sum_{i=1}^{n} tr\left(\mathbf{\Sigma}^{-1}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^{\top}\right)$$

$\widehat{\boldsymbol{\mu}} = \bar{x}$ o vetor de medias amostrais.
Tomando a derivada em relação ao Σ^{-1} obtém-se

$$\frac{\partial \ell((\boldsymbol{\mu}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}^{-1}} = -n\mathbf{\Sigma} + \sum_{i=1}^{n}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^{\top}$$

a matriz de coveriâncias $\widehat{\Sigma} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^{\top}$ é o EMV de Σ.

Suponha que $\boldsymbol{X}_1, \ldots, \boldsymbol{X} > n$ é uma a.a. de uma população $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Sejam

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{X})(\boldsymbol{X}_i - \bar{X})^\top$$

o vetor de médias e a matriz de covariâncias amostrais.

Teorema (3)

$\bar{X}$ e $\bar{S}$ são independente com

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$$

$$(n-1)\boldsymbol{S} = \sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{X})(\boldsymbol{X}_i - \bar{X})^\top \sim W_p(n-1, \Sigma)$$

Corolário (2)

A estatística T^2 de Hotelling para uma amostra multivariada é

$$T^2 = n(\boldsymbol{X} - \boldsymbol{\mu})^\top S^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$$

e tem-se

$$\frac{n-p}{p}\frac{n}{n-1}(\boldsymbol{X} - \boldsymbol{\mu})^\top S^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim F_{p,n-p}$$

Dado um modelo estatístico, com função distribuição $\boldsymbol{P_\theta}$, com $\boldsymbol{\theta} \in \Theta \subset R^p$, tem-se interesse em testar a seguintes hipóteses:

$$H_0 : \boldsymbol{\theta} \in \Theta_0, \quad vs \quad H_1 : \boldsymbol{\theta} \in \Theta_1$$

tal que $\Theta_0 \cup \Theta_1 = \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$. A estatística da razão de verossimilhança, ou às vezes chamada de estatística de Wilks, é então

$$W = -2 \log \left(\frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})} \right)$$

A hipótese nula é rejeitada se o valor observado de W for grande. Em alguns casos, a distribuição exata de W em H0 pode ser avaliada. Em outros casos, o teorema de Wilks afirma que para n grande,

$$W \xrightarrow{d} \chi^2_\nu,$$

onde ν é o número de parâmetros livres em H_1, mas não em H_0.

Testes de hipóteses

Suponha que tem-se interesse em testar hipóteses sobre os parâmentros da distribuição normal multivariada, baseado uma amostra aleatória, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, de uma população normal com vetor de médias $\boldsymbol{\mu}$ e matriz de variâncias $\boldsymbol{\Sigma}$.

1. $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ quando Σ é conhecido

Neste caso, sabemos a distribuição exata da estatística da razão de verossimilhança

$$W = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \Sigma^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) \sim \chi^2_{(p)}$$

sob H_0.

Note que

$$-2 \log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = n \log(2\pi\Sigma) + ntr(\Sigma^{-1}\boldsymbol{S}) + n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \Sigma^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu})$$

Vimos que o EMV irrestrito de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ é $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}$ e

$$-2\log(\sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = n\log(2\pi\Sigma) + ntr(\Sigma^{-1}\boldsymbol{S})$$

O EMV de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ sob H_0 é $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_0$ e

$$-2\log(\sup_{\boldsymbol{\theta}\in\Theta_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = n\log(2\pi\Sigma)+ntr(\Sigma^{-1}\boldsymbol{S})+n(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)^\top\Sigma^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)$$

Dai tem-se

$$W = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top\Sigma^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

que sob H_0 tem distribuição qui-quadrado com p graus de liberdade.

2. $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ quando Σ é desconhecido

- Sabemos que, a EMV de $\boldsymbol{\mu}$ e Σ irrestrito são $\hat{\boldsymbol{\mu}}$ e $\widehat{\Sigma} = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{x})(\boldsymbol{x}_i - \bar{x})^\top = S_0 = \frac{n-1}{n}\boldsymbol{S}$.
- O EMV de $\boldsymbol{\mu}$ e Σ sob H_0 são $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0$ e

$$\hat{\Sigma}_0 = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^\top = S_0 + \boldsymbol{\delta}\boldsymbol{\delta}_0^\top$$

. onde $\boldsymbol{\delta} = \sqrt{n}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)$. A estatística da razão de verossimilhança é então

$$W = n\log|S_0 + \boldsymbol{\delta}\boldsymbol{\delta}_0^\top| - n\log(S_0)$$

- Note que W é uma função crescente monótona de

$$\delta^\top \boldsymbol{S}^{-1}\delta = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

que é a estatística T^2 de Hotelling.

3. $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$ vs $H_1 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$, $\boldsymbol{\mu}$ desconhecido

Temos a estatística da razão de verossimilhança

$$W = -n \log\left(|\mathbf{\Sigma}_0^{-1}\boldsymbol{S}_0|\right) - np + tr(\mathbf{\Sigma}_0^{-1}\boldsymbol{S}_0)$$

Esse é o caso em que a distribuição exata de W é difícil de determinar. Para n grande, a distribuição de W é tem distribuição qui-quadrado com $p(p+1)/2$ graus de liberdade

Exemplo

Com a finalidade de verificar se o rendimento médio (RM), nas disciplinas de matemática, ciências, inglês e espanhol, dos alunos do último ano do ensino médio são as mesmas aos alunos de anos anteriores, foram escolhidos aleatoriamente 10 alunos cujos rendimentos resultaram

Aluno	Mat	Ciência	Inglês	Espanhol
1	81	89	73	74
2	73	79	73	74
3	61	86	81	81
4	55	70	76	73
5	61	71	61	66
6	52	70	56	58
7	56	74	56	56
8	65	87	73	69
9	54	76	69	72
10	48	71	62	63

A nota média em cada uma das disciplinas em anos anteriores foi 60 pontos. Isso mudou? Use $\alpha = 0,05$.

Teste de Hipóteses

- X_1 : RM em Matemática
- X_2 : RM em Ciências
- X_3 : RM em Inglês,
- X_4 : RM em Espanhol.

Se $E(X_j) = \mu_j$ representa a RM média de todos alunos na disciplina j. Então as hipóteses de interesse é

$$H_0 := \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} 60 \\ 60 \\ 60 \\ 60 \end{pmatrix} = \boldsymbol{\mu}_0 \ vs \ H_1 := \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} \neq \begin{pmatrix} 60 \\ 60 \\ 60 \\ 60 \end{pmatrix} = \boldsymbol{\mu}_0$$

Seja $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ uma amostra aleatória de uma população com vetor de médias $\boldsymbol{\mu}$ e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$. Sejam $\overline{\boldsymbol{X}}$ e $\boldsymbol{S}$ o vetor de médias amostrais e a matriz de variâncias e covariâncias amostrais.

Hipóteses de interesse: $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

Supondo que a amostra vem de uma população $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, tem-se que sob H_0

$$T^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top S^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

Assim, rejeitamos H_0 a um nível de significância α se

$$T^2_{obs} = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top S^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p}\ F^{-1}_{p,n-p}(1-\alpha)$$

em que $F^{-1}_{p,n-p}(1-\alpha)$ é o quantil $1-\alpha$ da distribuição $F_{p,n-p}$.

Dos dados do Exemplo tem-se

$$\overline{\boldsymbol{x}} = \begin{pmatrix} 60.6 \\ 77.3 \\ 68.0 \\ 68.6 \end{pmatrix} \quad S = \begin{pmatrix} 102.04 & 56.69 & 41.22 & 39.49 \\ 56.69 & 56.46 & 42.00 & 35.36 \\ 41.22 & 42.00 & 75.78 & 65.11 \\ 39.49 & 35.36 & 65.11 & 61.38 \end{pmatrix}$$

A estatística observada

$$T^2_{obs} = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top S^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) = 151,1352$$

Para um nível de significância de 5% tem-se

$$\frac{(n-1)p}{n-p} \; F^{-1}_{p,n-p}(1-\alpha) = \frac{9 \times 4}{10-4} \; F^{-1}_{4,6}(0,95) = 27,18 < 151,1352$$

Rejeita-se H_0, ao nível de significância de 5%.

Testes de hipóteses para vetor de médias:Exemplo

Equivalentemente

$$\frac{n-p}{(n-1)p}T^2 = \frac{6}{9\times 4}151,1352 = 25,1892 > F^{-1}_{4,6}(0,95) = 4,534$$

```
> library(ICSNP)
> HotellingsT2(x, mu = c(60,60,60,60), test = "f")
Hotelling's one sample T2-test

data:  x
T.2 = 25.189, df1 = 4, df2 = 6, p-value = 0.0006802
alternative hypothesis: true location is not equal to c(60,60,60,60)
```

Testes de hipóteses para vetor de médias:Exemplo

Equivalentemente

$$\frac{n-p}{(n-1)p}T^2 = \frac{6}{9\times 4}151,1352 = 25,1892 > F^{-1}_{4,6}(0,95) = 4,534$$

```
> library(ICSNP)
> HotellingsT2(x, mu = c(60,60,60,60), test = "f")
Hotelling's one sample T2-test

data:  x
T.2 = 25.189, df1 = 4, df2 = 6, p-value = 0.0006802
alternative hypothesis: true location is not equal to c(60,60,60,60)
```

Para grandes amostras

```
> HotellingsT2(x, mu = c(60,60,60,60), test = "chi")

Hotelling's one sample T2-test
data:  x
T.2 = 151.14, df = 4, p-value < 2.2e-16
alternative hypothesis: true location is not equal to c(60,60,60,60)
```

Definição

A região de confiança de 100$(1-\alpha)$*% determinado pela amostra aleatória* $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ *é denotado por* $\mathcal{R}_{\boldsymbol{X}}$ *é*

$$P\left(\mathcal{R}_{\boldsymbol{X}} \text{ contém o verdadeiro valor } \boldsymbol{\theta}\right) = 1 - \alpha.$$

onde P *é a distribuição de probabilidade de* $\boldsymbol{X}$ *e* $\boldsymbol{\theta} \in \Theta \subset R^2$.

Das propriedades do EMV dos parâmetros da distribuição normal multivariada tem-se

$$\frac{n(n-p)}{(n-1)p}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^\top S^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \sim F_{p,n-p}$$

é uma quantidade pivotal.

Região de Confiança

Daí

$$P\left(n(\overline{\boldsymbol{X}}-\boldsymbol{\mu})^\top S^{-1}(\overline{\boldsymbol{X}}-\boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F^{-1}_{p,n-p,}(1-\alpha)\right) = 1-\alpha,$$

onde $F^{-1}_{p,n-p,}(1-\alpha)$ é o quantil $1-\alpha$ da distribuição F com p e n-p graus de liberdade.
Temos que

$$\mathcal{A} = \left\{\boldsymbol{\mu} \in R^p : n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})^\top S^{-1}(\overline{\boldsymbol{x}}-\boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F^{-1}_{p,n-p,}(1-\alpha)\right\}$$

é a **região com** $100(1-\alpha)\%$ **de confiança** para $\boldsymbol{\mu}$.
Se $\boldsymbol{\mu}_0 \in \mathcal{A}$, não rejeitamos $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$; caso contrário, rejeitamos H_0.

Região de Confiança

- Para $p > 4$ não se pode fazer gráficos da região de confiança para $\boldsymbol{\mu}$.
- Não entanto, pode-se calcular os eixos da elipsoide de confiança e seus tamanhos relativos, os quais são determinados pelos autovalores λ_i e os autevetores $\boldsymbol{e}_i$ de $\boldsymbol{S}$.
- Os tamanhos dos semi-eixos de

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^\top S^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq c^2 = \frac{(n-1)p}{n-p} F^{-1}_{p,n-p,}(1-\alpha)$$

são determinados por

$$\frac{\sqrt{\lambda_i} c}{\sqrt{n}} = \sqrt{\lambda_i} \sqrt{\frac{(n-1)p}{(n-p)n} F^{-1}_{p,n-p,}(1-\alpha)}$$

- Começando do centro, determinado por $\bar{X}u$ os eixos de elipsóide são:

$$\sqrt{\lambda_i} \sqrt{\frac{(n-1)p}{(n-p)n} F^{-1}_{p,n-p,}(1-\alpha)} \boldsymbol{e}_i$$

Região de Confiança:Exemplo

Obter uma região de 95% de confiança para $\boldsymbol{\mu}$ e verifique se o ponto $\boldsymbol{\mu}_0 = (13, 4)^\top$ pertence a mesma. Considere a seguinte informação/

$$\bar{\boldsymbol{x}} = \begin{pmatrix} 10 \\ 2 \end{pmatrix}, \; S = \begin{pmatrix} 1 & -0,5 \\ -0,5 & 1 \end{pmatrix}, \; S^{-1} = \frac{1}{3}\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \; e \; n = 2.$$

Os autovalores e autovetores de S, são:

$$\lambda_1 = 1.5 \boldsymbol{e}_1 = (-0.7071068, -0.7071068)^\top$$
$$\lambda_1 = 0.5 \boldsymbol{e}_1 = (0.7071068 - 0.7071068)^\top$$

Região de Confiança:Exemplo

A elipse de 95% confiança para $\boldsymbol{\mu}$ consiste de todos os valores (μ_1, μ_2) que satisfazem

$$\begin{pmatrix} 10 - \mu_1 & 3 - \mu_2 \end{pmatrix} \frac{1}{3} \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 10 - \mu_1 \\ 3 - \mu_2 \end{pmatrix} \leq \frac{2(2)}{1} \times 199,5.$$

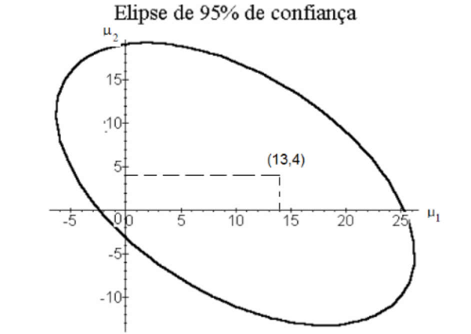
ou

$$4(10 - \mu_1) + 4(10 - \mu_1)(3 - \mu_2) + 4(3 - \mu_2)^2 \leq 708$$

Para verificar se o ponto $\boldsymbol{\mu}_0^\top = (13, 4)$ pertence a elipse, calcula-se

$$4(10 - 13)^2 + 4(10 - 13)(3 - 4) + 4(3 - 4)^2 = 52 \leq 708$$

Região de Confiança:Exemplo



Intervalo de confiança simultânea

- Suponha $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ é uma amostra aleatória de $N_p(\boldsymbol{\mu}, \Sigma)$ e seja
$$Z_j = a_1 X_{j1} + \cdots + a_p X_{jp} = \boldsymbol{a}\boldsymbol{X}_j, \; j = 1, \ldots, n$$
ou

- a média e variância amostral dos valores observados $z_1, \ldots, z_n$ são respectivamente

$$\bar{z} = a_1 \bar{\boldsymbol{X}}_1 + \cdots + a_p \bar{\boldsymbol{X}}_p = \boldsymbol{a}^\top \bar{\boldsymbol{X}} \; e, \; s_Z^2 = \boldsymbol{a}^\top S \boldsymbol{a}$$

- Pode-se mostrar que

$$T = \frac{\bar{Z} - \mu_Z}{s_z/\sqrt{n}} = \frac{\sqrt{n}(\boldsymbol{a}^\top \bar{\boldsymbol{X}} - \boldsymbol{a}^\top \boldsymbol{\mu})}{\sqrt{\boldsymbol{a}^\top S \boldsymbol{a}}} \sim t(n-1)$$

Intervalo de confiança simultânea

Um intervalo de 100(1-α)% para $\boldsymbol{a}^\top \boldsymbol{\mu}$ baseada na distribuição t-Student é dado por

$$\left(\boldsymbol{a}^\top \bar{\boldsymbol{X}} - \frac{\sqrt{\boldsymbol{a}^\top S \boldsymbol{a}}}{\sqrt{n}} F^{-1}_{T_{(n-1)}}(1-\alpha/2)\ ,\ \boldsymbol{a}^\top \bar{\boldsymbol{X}} + \frac{\sqrt{\boldsymbol{a}^\top S \boldsymbol{a}}}{\sqrt{n}} F^{-1}_{T_{(n-1)}}(1-\alpha/2) \right) \tag{2}$$

onde $F^{-1}_{T_{(n-1)}}(1-\alpha/2)$ é o quantile $1-\alpha/2$ da distribuição t-Student.
O intervalo pode ser interpretado como intervalos sobre componentes do vetor de médias. Por exemplo, fazendo-se 1 na i-ésima posição e zero as outras posições de $\boldsymbol{a}^\top$, então (2) da o intervalo de confiança para μ_i, $i = 1, \ldots, p$

Intervalo de confiança simultânea

Não entanto o coeficiente de confiança para todos os intervalos tomados simultaneamente não é $1-\alpha$. Para corrigir esta imperfeição demonstra-se (Johnson e Wichern, 1988) que para garantir o coeficiente nominal de confiança simultâneo de $1-\alpha$. para a cobertura dos valores paramétricos é necessário recorrer à distribuição T^2 de Hotelling. Este resultado está apresentado a seguir

$$\left(\boldsymbol{a}^\top \bar{\boldsymbol{X}} - \sqrt{\frac{(n-1)p}{(n-p)n}(\boldsymbol{a}^\top S \boldsymbol{a}) q_f} \ , \ \boldsymbol{a}^\top \bar{\boldsymbol{X}} + \sqrt{\frac{(n-1)p}{(n-p)n}(\boldsymbol{a}^\top S \boldsymbol{a}) q_f} \right)$$

onde $q_f = F^{-1}_{F_{(p,n-p)}}(1-\alpha/2)$ é quantil $1-\alpha/2$ da distribuição F com p e n-p graus de liberdade.
Esses intervalos são chamados de intervalos de T^2-Hotelling

Intervalo de confiança simultânea

Note que sucessivas escolhas de $\boldsymbol{a}_1^\top = (1, 0, \ldots, 0)$, $\boldsymbol{a}_2^\top = (1, 0, \ldots, 0), \ldots, \boldsymbol{a}_p^\top = (0, 0, \ldots, 1)$

$$\left(\bar{X}_1 - \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{11}}{n}}, \bar{X}_1 + \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{11}}{n}}\right)$$
$$\left(\bar{X}_2 - \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{22}}{n}}, \bar{X}_2 + \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{22}}{n}}\right)$$
$$\vdots$$
$$\left(\bar{X}_p - \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{pp}^2}{n}}, \bar{X}_p + \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{pp}}{n}}\right)$$

vai conter todos simultaneamente com probabilidade $1 - \alpha$

Uma escolha cuidadosa de $\boldsymbol{a}^\top$ tal como $\boldsymbol{a}^\top = (1, -1, 0, \ldots, 0)$ para os intervalos-T^2 nos permite testar hipóteses especiais

$$\bar{X}_1 - \bar{X}_2 \mp \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{11} + s_{22} + 2s_{12}}{n}}$$

Exemplo

Considere os dados a continução e contrua um intervalo de 90% de confiança para vetor de médias de X_1 e X_2.

x_1 :	1,43	1,62	2,46	2,48	2,97	4,03	4,47	5,76	6,61	6,68	6,79	7,46	7,88	8,92	9,42
x_2 :	-0,69	-5,00	-1,13	-5,20	-6,39	2,87	-7,88	-3,97	2,32	-3,24	-3,56	1,61	-1,87	-6,60	-7,74

Intervalo de confiança simultânea: Exemplo

Dos dados tem-se $n = 15$ e $p = 2$

$$\bar{\boldsymbol{x}} = \begin{pmatrix} 5,26 \\ -3,09 \end{pmatrix}, \; S = \begin{pmatrix} 7,12 & -0,72 \\ -0,72 & 12,43 \end{pmatrix} \; e \; S^{-1} = \begin{pmatrix} 0,1412 & 0,0082 \\ 0,0082 & 0,0809 \end{pmatrix}$$

e

$$q_f = \frac{(n-1)p}{n-p} F^{-1}_{F_{(p,n-p)}}(0,95) = 5,95$$

Se $\boldsymbol{a}^\top = (1,0)$ tem-se: $\bar{X}_1 \mp \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{11}}{n}}$

$$5,26 \mp \sqrt{5,95} \sqrt{\frac{7,12}{15}}$$

Dai tem-se um intervalo de 90% de confiança para μ_1

$$(3,579; 9,941)$$

Similarmente, um intervalo de 90% de confiança para μ_2

$$(-5,31; -0,87)$$

O intervalo de 100(1-α)% de confiança para $\mu_1 - \mu_2$ é dada por

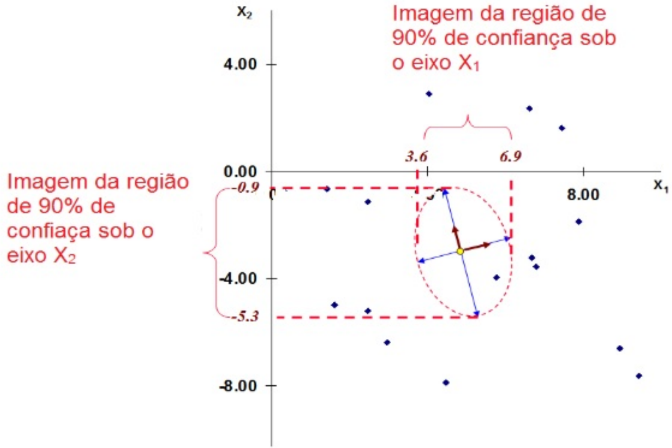
$$\bar{X}_1 - \bar{X}_2 \mp \sqrt{\frac{(n-1)p}{(n-p)} q_f} \sqrt{\frac{s_{11} + s_{22} - 2s_{12}}{n}}$$

Para $1 - \alpha = 0,95$ tem-se

$$5,26 - (-3,09) \mp \sqrt{5,95} \sqrt{\frac{7,12 + 12,42 - 2(9,72)}{15}} = 8,35 \mp 2,885$$

O intervalo de 90% de confiança para $\mu_1 - \mu_2$ é dada por

$$(5,465; 11,235)$$



Note-se que as projeções da elipse para os eixos fazem os intervalos de confiança simultâneos

O intervalo univarido de 100(1-α)% de confiança para μ_1 é dado por

$$\left(\bar{x}_1 - F^{-1}_{T_{(n-1)}}\sqrt{\frac{s_{11}}{n}}, \bar{x}_1{}' + F^{-1}_{T_{(n-1)}}\sqrt{\frac{s_{11}}{n}}\right)$$

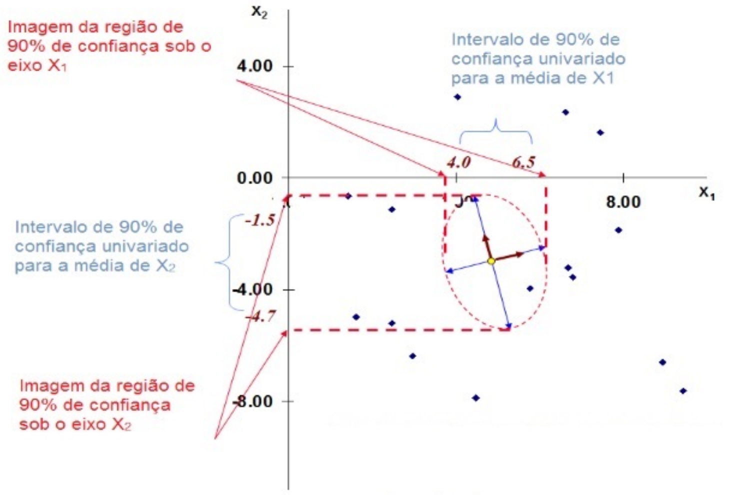
O intervalo univarido de 90% de confiança para μ_1 é dado por

$$\left(5,26 - 1,76\sqrt{\frac{7,12}{15}}, 5,26 + 1,76\sqrt{\frac{7,12}{15}}\right) = 5,26 \mp 1,213 = (4,047, 6,473)$$

Similarmente, O intervalo univarido de 90% de confiança para μ_3 é dado por

$$\left(-3,09 - 1,76\sqrt{\frac{12,43}{15}}, -3,09 + 1,76\sqrt{\frac{12,43}{15}}\right) = -3,09 \mp 1,603$$
$$= (-4,69, -1,497)$$

Intervalo de confiança simultânea: Exemplo



Note que nos IC univariados os comprimentos são mais curtos aos que consideram a correlação de X_1 e X_2

Intervalo de confiança simultânea

Método de Bonferroni

- Muitas vezes um pequeno numero de intervalos de confiança é requerido.
- Nesse caso pode-se ter uma melhor opção do que as comparações simultâneas.
- Intervalos de confiança mais curto (mais curtos) do que intervalo simultâneo de T^2.
- Essa alternativa de intervalo é conhecido por método de Bonferroni

$$\left(\bar{X}_j - F^{-1}_{T_{(n-1)}}(1-\frac{\alpha}{2p})\sqrt{\frac{s_{jj}}{n}}, \bar{X}_j + F^{-1}_{T_{(n-1)}}(1-\frac{\alpha}{2p})\sqrt{\frac{s_{jj}}{n}}\right), j = 1, \ldots, p$$

Intervalo de confiança simultânea: Exemplo

Para os dados do exemplo anterior obtenha os IC de Bonferroni

$$\left(\bar{X}_1 - F^{-1}_{T_{(n-1)}}(1 - \frac{\alpha}{2p})\sqrt{\frac{s_{11}}{n}}, \bar{X}_1 + F^{-1}_{T_{(n-1)}}(1 - \frac{\alpha}{2p})\sqrt{\frac{s_{jj}}{n}}\right),$$

$$(5,26 - 2,145\sqrt{\frac{7,12}{15}}, 5,26 + 2,145\sqrt{\frac{7,12}{15}}) = (5,26 - 1,478; 5,26 + 1,478)$$

Um Intervalo de 90% de confiança para μ_1 é; (3,782 ; 6,782).
Similarmente, um IC para μ_2 b

$$(-3,09 - 2,145\sqrt{\frac{12,43}{15}}; -3,09 + 2,145\sqrt{\frac{12,43}{15}}) = -3,09 \mp 1,953;$$

Um Intervalo de 90% de confiança para μ_2 é: (-5,043 , -1,137).

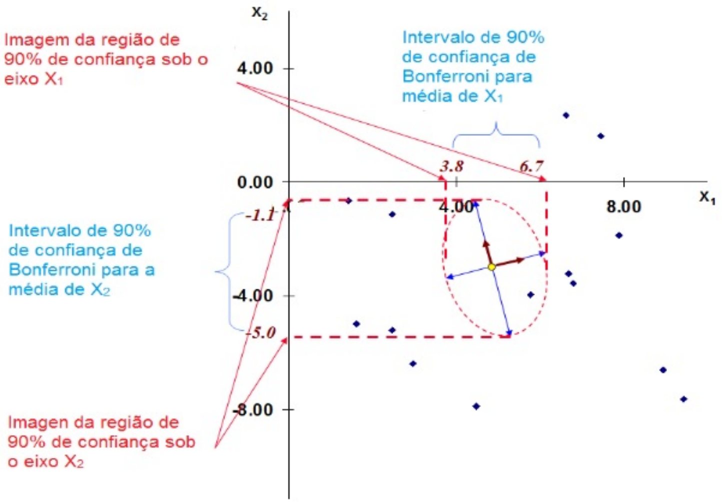


Imagem da região de 90% de confiança sob o eixo X1
Intervalo de 90% de confiança de Bonferroni para média de X1
Intervalo de 90% de confiança de Bonferroni para a média de X2
Imagen da região de 90% de confiança sob o eixo X2
X2
X1
4.00
0.00
-1.1
-4.00
-5.0
-8.00
3.8
6.7
4.00
8.00

Inferência para amostras grandes ao respeito do vetor de uma média populacional

Seja $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ uma amostra aleatória de uma população com vetor de médias $\boldsymbol{\mu}$ e matriz variância covariância $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}$ é definida positiva).

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \approx \chi^2_{(p)}$$

- $\approx$ significa aproximadamente
- $P\left(n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi^2_{(p)}(\alpha)\right) \approx 1 - \alpha$
- Quando $n \longrightarrow \infty$

$$\frac{(n-1)p}{n-p} F_{p,n-p} \longrightarrow \chi^2_{(p)}.$$

Inferência para amostras grandes ao respeito do vetor de uma média

Para $n - p$ grande

1. Teste de hipótese

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \ vs \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

   Rejeita H_0 se $T^2 > \chi^2(\alpha)$ onde $\chi^2(\alpha)$ é o quantil α-superior da distribuição qui-quadrado com p graus de liberdade.

2. Intervalos T^2 simultâneos:

$$\boldsymbol{a}^\top \bar{\boldsymbol{X}} \mp \sqrt{\chi^2(\alpha)} \sqrt{\frac{\boldsymbol{a}^\top \boldsymbol{S}^{-1} \boldsymbol{a}}{n}}$$

3. Região de confiança $\boldsymbol{\mu}$

$$(\bar{\boldsymbol{X}} - \mu)^\top \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \mu) \leq \frac{\chi^2_{(p)}}{n}$$

Suponha temos coletado uma amostra aleatória de 107 observações e $p = 5$ características e calculamos as seguintes estatísticas:

Variável	Média	Desvio Padrão
X_1	58,6	4,44
X_2	19,3	2,39
X_3	101,5	8,87
X_4	67,0	3,99
X_5	42,5	4,69

Construir os 5 intervalos simultâneos de 90% de confiança para as médias indivíduais μ_i, $i = 1, \ldots, 5$
Temos amostras grandes ($n = 107 >> p = 5$), usaremos a abordagem de amostras grandes para construir os intervalos simultâneos de 90%.

Inferência para amostras grandes ao respeito do vetor de uma média:Exemplo

$$\mu_1 = \bar{x}_1 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{11}}{n}} = 58{,}6 \pm \sqrt{9{,}236}\,\frac{4{,}44}{\sqrt{107}} = 58{,}6 \pm 6{,}034 \Rightarrow \mu_1 : (52{,}565; 64{,}635)$$

$$\mu_2 = \bar{x}_2 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{22}}{n}} = 19{,}3 \pm \sqrt{9{,}236}\,\frac{2{,}39}{\sqrt{107}} = 19{,}3 \pm 3{,}248 \Rightarrow \mu_2 : (16{,}052; 22{,}548)$$

$$\mu_3 = \bar{x}_3 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{33}}{n}} = 101{,}3 \pm \sqrt{9{,}236}\,\frac{8{,}87}{\sqrt{107}} = 101{,}3 \pm 12{,}056 \Rightarrow \mu_3 : (89{,}444; 113{,}556)$$

$$\mu_4 = \bar{x}_4 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{44}}{n}} = 67{,}0 \pm \sqrt{9{,}236}\,\frac{3{,}99}{\sqrt{107}} = 67{,}0 \pm 5{,}423 \Rightarrow \mu_4 : (61{,}577; 72{,}423)$$

$$\mu_5 = \bar{x}_5 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{55}}{n}} = 42{,}5 \pm \sqrt{9{,}236}\,\frac{4{,}69}{\sqrt{107}} = 42{,}5 \pm 6{,}374 \Rightarrow \mu_4 : (36{,}126; 48{,}874)$$

Inferência para amostras grandes ao respeito do vetor de uma média:Exemplo

Se usamos esses intervalos para testar a hipóteses nula

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix} = \begin{pmatrix} 55 \\ 25 \\ 111 \\ 59 \\ 34 \end{pmatrix}$$

Ao nível de confiança de 90%, poderia rejeitar a hipóteses nula? (Por que?)