

# Distribuição Normal Multivariada–II

Vicente G. Cancho  
garibay@icmc.usp.br

# Distribuição Normal Multivariada

## Definição

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

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

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

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

# Propriedades da distribuição normal multivariada

Se  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  e  $\mathbf{X}_1$  e  $\mathbf{X}_2$  é a partição de  $\mathbf{X}$  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

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

Então

- (i)  $X_j \sim N(\mu_j, \sigma_{jj}), j = 1, \dots, p.$
- (ii)  $\mathbf{Y} = \mathbf{A}\mathbf{X} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$
- (iii) A distribuição condicional  $\mathbf{X}_1$  dado  $\mathbf{X}_2 = \mathbf{x}_2$  é

$$N_p(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

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

# Distribuição de Wishart

- A distribuição de Wishart (Wishart, 1928) é uma família de distribuições para matrizes simétricas positivas definidas.
- Seja  $\mathbf{X}_1, \dots, \mathbf{X}_n$   $n$  vetores independentes  $N_p(\mathbf{0}, \Sigma)$  e forma uma matriz de dados  $p \times n$ ,  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ . Diz-se que a distribuição de uma matriz aleatória  $p \times p$ ,  $\mathbf{M} = \mathbf{X}\mathbf{X}^\top = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top$  tem a distribuição de Wishart

## Definição

A matriz aleatória  $\mathbf{M} = \mathbf{X}\mathbf{X}^\top = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top$  tem distribuição de Wishart com  $n$  graus de liberdade e matriz de covariância  $\Sigma$  e denotado por  $\mathbf{M} \sim W_p(n, \Sigma)$ . Para  $n \geq p$ , a função de densidade de probabilidade de  $\mathbf{M}$  é

$$p(\mathbf{M}|n, \Sigma) = \frac{|\mathbf{M}|^{\frac{(n-p-1)}{2}} \exp\{-\text{tr}(\mathbf{M}\Sigma^{-1})/2\}}{2^{p(n-1)/2} \pi^{p(p-1)/4} |\Sigma|^{(n-1)/2} \prod_{i=1}^p \Gamma((n-i)/2)}$$

# Distribuição de Wishart

- Em particular se  $\mathbf{M} \sim W_p(n, \sigma^2)$ , então  $\mathbf{M}/\sigma^2 \sim \chi^2(p)$ .
- Quando  $\Sigma = \mathbf{I}_p$ .  $W_p(n, \mathbf{I}_p)$  é chamado da distribuição Wishart padrão.

## Proposição (1)

- 1 Para  $\mathbf{M} \sim W_p(n, \Sigma)$  e  $\mathbf{B}_{p \times m}$ ,  $\mathbf{B}^\top \mathbf{M} \mathbf{B} \sim W_m(n, \mathbf{B}^\top \Sigma \mathbf{B})$ .
- 2 Para  $\mathbf{M} \sim W_p(n, \Sigma)$  com  $\Sigma > 0$   $\Sigma^{-1/2} \mathbf{M} \Sigma^{-1/2} \sim W_p(n, \mathbf{I}_p)$ .
- 3 Se  $\mathbf{M}_i$  são independentes  $W_p(n_i, \Sigma)$  ( $i = 1, \dots, k$ ). Então  $\sum_{i=1}^k \mathbf{M}_i \sim W_p(r, \Sigma)$ , onde  $r = \sum_{i=1}^k n_i$ .
- 4 Para  $\mathbf{M}_n \sim W_p(n, \Sigma)$ , então  $E[\mathbf{M}_n] = n\Sigma$ .
- 5 Se  $\mathbf{M}_1 \sim W_p(n_1, \Sigma)$  e  $\mathbf{M}_2$  são independentes e satisfaz  $\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{M} \sim W_p(n, \Sigma)$  então  $\mathbf{M}_2 \sim W_p(n - n_1, \Sigma)$ .
- 6  $\mathbf{M}_n/n \xrightarrow{P} \Sigma$  (LGN)

## Corolário (1)

Se  $\mathbf{M} \sim W_p(n, \Sigma)$  e  $\mathbf{a} \in R^p$  é tal que  $\mathbf{a}^\top \Sigma \mathbf{a} \neq \mathbf{0}$ , então

$$\frac{\mathbf{a}^\top \mathbf{M} \mathbf{a}}{\mathbf{a}^\top \Sigma \mathbf{a}} \sim \chi^2_{(n)}.$$

## Teorema (1)

Se  $\mathbf{M} \sim W_p(n, \Sigma)$  e  $\mathbf{a} \in R^p$  e  $n > p - 1$

$$\frac{\mathbf{a}^\top \Sigma^{-1} \mathbf{a}}{\mathbf{a}^\top \mathbf{M} \mathbf{a}} \sim \chi^2_{(n-p+1)}.$$

# A distribuição $T^2$ de Hottelling

## Definição

Suponha que  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  e é independente de  $\mathbf{M} \sim W_p(n, \boldsymbol{\Sigma})$  então

$$n(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim T_{n,p}^2.$$

- A estatística  $T^2$  de Hotelling desempenha um papel semelhante na análise multivariada ao da estatística t-Student na análise estatística univariada.
- Ou seja, a aplicação da estatística  $T^2$  de Hotelling é de grande importância prática no teste de hipóteses sobre a média de uma distribuição normal multivariada quando a matriz de covariância é desconhecida.

# A distribuição $T^2$ de Hotteling

## Teorema (3)

Se  $n > p - 1$

$$\frac{n - p + 1}{np} T_{n,p}^2 = F_{p,n-p+1}$$

Um caso especial é quando  $p = 1$ , onde o Teorema anterior indica que  $T_{n,1}^2 = F_{1,n}$



# A função de verossimilhança

Suponha que  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$  é  $i$ -ésima amostra de uma população normal multivariada com vetor de médias  $\boldsymbol{\mu}$  e matriz de variância-covariância  $\Sigma$ , ou seja  $\mathbf{X}_i \stackrel{i.i.d}{\sim} N_p(\boldsymbol{\mu}, \Sigma)$ .

A função de verossimilhança de  $\boldsymbol{\mu}$  e  $\Sigma$  dada a amostra observada é dada por

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

A função log-verossimilhança

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

# Estimador de máxima verossimilhança do vetor de médias

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

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

onde  $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i$  é o vetor de medias amostrais.

Tomando a derivada em relação a  $\mu$ , obtém-se

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

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

# Estimador de máxima verossimilhança da matriz de covariâncias

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

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

$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$  o vetor de medias amostrais.

Tomando a derivada em relação ao  $\Sigma^{-1}$  obtém-se

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

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

# Distribuição Amostral multivariada: $\bar{\mathbf{X}}$ e $\mathbf{S}$

Suponha que  $\mathbf{X}_1, \dots, \mathbf{X}_n$  é uma a.a. de uma população  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Sejam

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

o vetor de médias e a matriz de covariâncias amostrais.

## Teorema (3)

$\bar{\mathbf{X}}$  e  $\bar{\mathbf{S}}$  são independentes com

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$$
$$n\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})^\top \sim W_p(n-1, \boldsymbol{\Sigma})$$

## Corolário (2)

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

$$(n-1)(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) = n(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top \mathbf{S}_n^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim T_{n-1,p}^2.$$

Relembre que  $\mathbf{S}_n = \frac{n}{n-1}\mathbf{S}$  é o estimador não viciado de  $\Sigma$ .

# Distribuição Amostral multivariada: $\bar{\mathbf{X}}$ e $\mathbf{S}$

**Prova:** do Teorema (3). Note que

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \mathbf{x}_1 + \cdots + \frac{1}{n} \mathbf{x}_n$$

é combinação linear de vetores aleatórios com distribuição normal multivariada. Daí o vetor de médias amostrais tem distribuição normal multivariada.

Da definição da distribuição Wishart e a independência de  $\bar{\mathbf{X}}$  e  $\bar{\mathbf{S}}$

$$\underbrace{\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top}_{M \sim W_p(n, \Sigma)} = \underbrace{n(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top}_{M_1 \sim W_p(1, \Sigma)} + \underbrace{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})^\top}_{M_2 \sim W_p(n-1, \Sigma)}$$

# Distribuição Amostral multivariada: $\bar{\mathbf{X}}$ e $\mathbf{S}$

Considere um novo conjunto de vetores aleatórios  $\mathbf{Y}_i$  ( $i = 1, \dots, n$ ) de uma combinação linear de  $\mathbf{X}_i$ s por uma matriz ortogonal  $\mathbf{D}$  satisfazendo

$$\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_n), \quad \mathbf{d}_1 = \frac{1}{n} \mathbf{1}_n$$
$$\mathbf{D}\mathbf{D}^\top = \mathbf{D}^\top \mathbf{D} = \mathbf{I}_n.$$

Seja

$$\mathbf{Y}_j = \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \mathbf{d}_{ji} = \sum_{i=1}^n \tilde{\mathbf{X}}_i \mathbf{d}_{ji} = \tilde{\mathbf{X}} \mathbf{d}_j$$

onde  $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n)$  é uma matriz aleatória de ordem  $p \times n$ . Daí,

❶  $\mathbf{Y}_j$  são normalmente distribuídos;

❷ 
$$E(\mathbf{Y}_j \mathbf{X}_k) = \begin{cases} \boldsymbol{\Sigma} & \text{se } j \neq k \\ \mathbf{0} & \text{se } j = k \end{cases}$$

❸ 
$$\sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top = \sum_{i=1}^n \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top$$

❹ 
$$\mathbf{Y}_1 \mathbf{Y}_1^\top = n(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top.$$

# Distribuição Amostral multivariada: $\bar{\mathbf{X}}$ e $\mathbf{S}$

- De (1) e (2) tem-se  $\mathbf{Y}_1$  e  $\mathbf{Y}_j$  ( $j \leq 2$ ) são independentes;
- Note que

$$\begin{aligned}\sum_{i=1}^n \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top &= \sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top = \mathbf{Y}_1 \mathbf{Y}_1^\top + \sum_{i=2}^n \mathbf{Y}_i \mathbf{Y}_i^\top \\ &= n(\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top + \sum_{i=2}^n \mathbf{Y}_i \mathbf{Y}_i^\top\end{aligned}$$

- Dai,  $\sum_{i=2}^n \mathbf{Y}_i \mathbf{Y}_i^\top = \sum (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$  que é independente de  $\mathbf{Y}_1$ .



# Avaliando a suposição de normalidade

- Suponha que  $\mathbf{X}_1, \dots, \mathbf{X}_n$  seja uma amostra de um vetor aleatório de uma população com distribuição normal multivariada.
- A qualidade dos métodos de inferência baseados nesta amostra dependem de quão próxima da distribuição normal está a sua população
- Sendo assim, são necessários procedimentos para detectar casos onde os dados exibem desvios moderados ou extremos da suposição de normalidade multivariada.

# Avaliando a suposição de normalidade

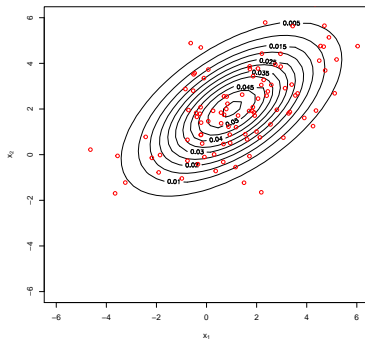
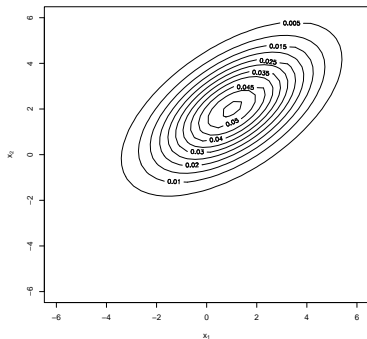
Das propriedades da normal multivariada vistas anteriormente, a avaliação da suposição de normalidade multivariada depende das seguintes questões

- As distribuições marginais dos componentes do vetor aleatório  $\mathbf{X}$  apresentam seguir uma distribuição normal?
- Os gráficos de dispersão de pares de componentes aleatórios (variáveis) apresentam a forma elíptica esperada de uma distribuição normal?
- Existem alguns pontos aberrantes (outliers) que devem ser investigados?

# Avaliando a suposição de normalidade

## Exemplo

Considera uma amostra de tamanho  $n=100$ , de  $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$  com  $\boldsymbol{\mu} = (1, 2)^\top$  e  $\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}$



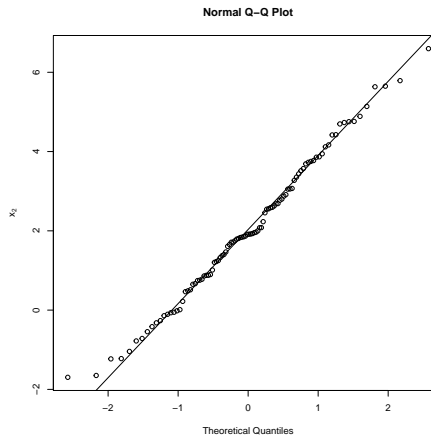
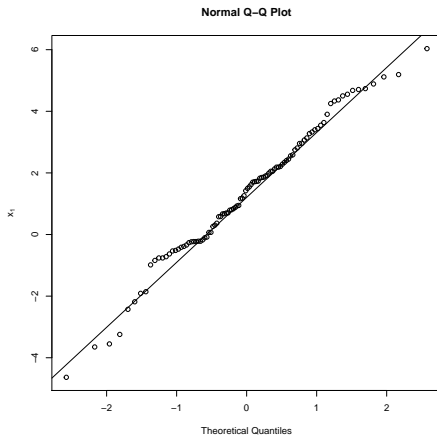
```
library(mvtnorm)
# vetor de médias
mu = matrix(c(1,2), nrow=2)
# matriz de variâncias e covariâncias:
Sigma =matrix(c(4,2,2,3), ncol=2)
## Grafico da f.d.p e contornos
x1=seq(-6, 6, length= 40)
x2 = x1
f = matrix(0, nrow=length(x1), ncol=length(x2))
for (i in 1:length(x1)){
  for (j in 1: length(x2)){
    f[i,j]=dmvnorm(c(x1[i],x2[j]), mean=mu, sigma=Sigma)}}

persp(x1, x2, f, theta = 70, phi = 30, col = "lightblue", ticktype = "detailed"
contour(x1, x2, f, draw=T, nlevels=20, labcex=0.8,
xlab=expression(x[1]),ylab=expression(x[2]))

X=rmvnorm(n=100, mean=mu, sigma=Sigma, method="chol")
points(X,col="red",lwd=2)
```

- Gráficos de quantil-quantil (QQplots) também podem ser utilizados para verificar a suposição de normalidade univariada.
- Em análise multivariada, esses gráficos avaliam as distribuições marginais dos vetores aleatórios a partir da amostra disponível dos seus componentes
- Gráficos de quantil-quantil comparam os quantis da amostra com o quantis esperados da situação onde a amostra segue uma distribuição normal.
- Nesse caso, desvios de normalidade são detectados no gráfico a partir da observação de desvios de uma reta

# QQ plot: Exemplo



# Testes de hipóteses para verificação de normalidade univariada

Testes de hipóteses também podem ser utilizados para avaliar a suposição de normalidade para as distribuições marginais do vetor aleatório. Alguns exemplos:

- Kolmogorov-Smirnov (*ksnormTest* no R): teste não paramétrico que compara a máxima diferença entre a função de distribuição acumulada dos dados e a distribuição normal teórica.
- Shapiro-Wilk (*shapiro.test* no R): teste não paramétrico baseado numa estatística que mede o quão longe os dados estão do que seria esperado no caso de normalidade.

```
> shapiro.test(X[,1])
Shapiro-Wilk normality test
data:  X[, 1]
W = 0.98589, p-value = 0.3675
> shapiro.test(X[,2])
Shapiro-Wilk normality test
data:  X[, 2]
W = 0.99049, p-value = 0.7038
```

# Avaliação da Normalidade bivariada

- Em trabalhos práticos, geralmente é suficiente investigar as distribuições uni e bivariadas subjacentes.
- Se as observações foram geradas de uma distribuição normal multivariada, cada distribuição bivariada será normal e os contornos de densidade constante serão elipses.
- O diagrama de dispersão deve-se ajustar a essa estrutura exibindo uma forma elíptica



- Além disso, vimos que se  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , então

$$P \left[ (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \leq \chi^2_{(2)}(0.50) \right] = 0,50$$

- A grosso modo, devemos esperar que cerca de 50% das observações caiam na região (no elipsóide)

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \leq \chi^2_{(2)}(0.50)$$

- Assim um procedimento útil, embora não exato, é comparar as proporções de pontos dentro de um contorno com a probabilidade teórica correspondente.

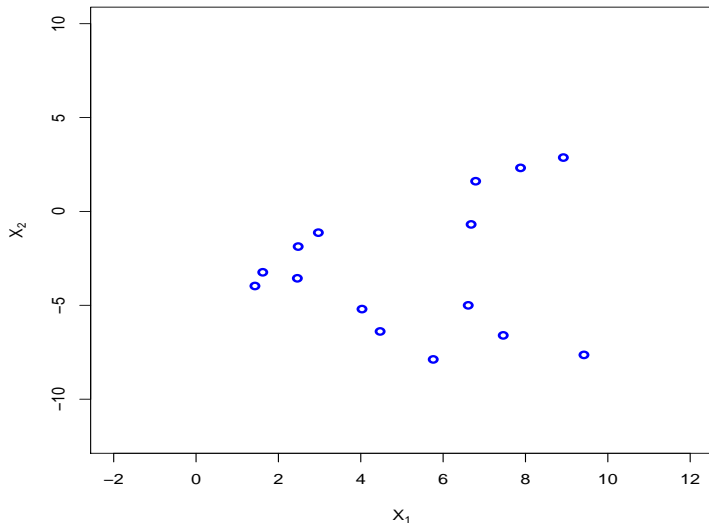
# Avaliação da Normalidade bivariada: Exemplo

Suponha temos uma amostra ordenada de 15 observações de alguma variável aleatória  $X_1$  e  $X_2$ :

$x_{j1}$	$x_{j2}$
5,76	-3,97
6,68	-3,24
6,79	-3,56
7,88	-1,87
2,46	-1,13
2,48	-5,20
2,97	-6,39
4,47	-7,88
1,62	-5,00
1,43	-0,69
7,46	1,61
8,92	-6,60
6,61	2,32
4,03	2,87
9,42	-7,64

Será que esses dados foram retirados de uma população normal bivariada?

# Diagrama de dispersão



- Um método um pouco mais formal para julgar a normalidade bivariada é baseado no quadrado da distância generalizada

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad i = 1, \dots, n.$$

- Esse procedimento não é limitado ao caso  $p = 2$ , pode ser usado. para  $p > 2$ .

# Avaliação de normalidade bivariada

- Quando a população é de fato normal multivariada e ambos  $n$  e  $n - p$  são maiores que 25 ou 30, cada uma das distâncias generalizadas quadradas devem comportar-se segundo uma distribuição de qui-quadrado com  $p$  graus de liberdade.
- Apesar dessas distâncias não serem independentes ou terem distribuição exata de Qui-quadrado, é útil construir um gráfico como se fossem.
- O gráfico resultante é chamado gráfico qui-quadrado.

- Ordene as distâncias quadradas amostrais obtendo

$$d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(n)}$$

- Construa o gráfico de dispersão dos pontos:  $(d_{(i)}, q_i)$ , onde  $P(\chi_p^2 \leq q_i) = \frac{i-0,5}{n}$
- A nuvem de pontos deve se ajustar a uma reta. Um padrão diferente sugere falta de normalidade

# Exemplo

Considerando o exemplo anterior temos:

$$\bar{X} = \begin{bmatrix} 5.27 \\ -3.09 \end{bmatrix}, \quad S = \begin{bmatrix} 7.12 & -0.73 \\ -0.73 & 12.43 \end{bmatrix}$$

$$\text{and } S^{-1} = \begin{bmatrix} 0.141257554 & 0.008249313 \\ 0.008249313 & 0.080967253 \end{bmatrix}$$

Daí temos

$$d_j = (\mathbf{x}_j - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad j = 1, \dots, 15.$$

Por exemplo  $j=1$

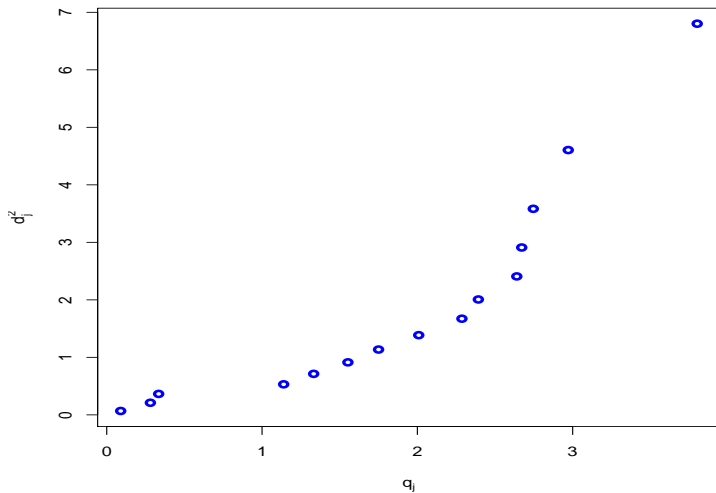
$$d_1 = [5, 76 - 5, 27, -3, 97 - (-3, 09)] \begin{bmatrix} 0.141257554 & 0.008249313 \\ 0.008249313 & 0.080967253 \end{bmatrix} \\ \begin{bmatrix} 5, 76 - 5, 27 \\ -3, 09 - (-3, 09) \end{bmatrix} = 0, 090$$

# Exemplo

$x_{j1}$	$x_{j2}$	$d_{2j}$	$(j - 0.5)/n$	$\chi^2_{(2)}[(j - 0.5)/n]$
5,76	-3,97	0,090	0,033	0,068
6,68	-3,24	0,281	0,100	0,211
6,79	-3,56	0,333	0,167	0,365
7,88	-1,87	1,138	0,233	0,531
2,46	-1,13	1,336	0,300	0,713
2,48	-5,20	1,548	0,367	0,914
2,97	-6,39	1,739	0,433	1,136
4,47	-7,88	2,005	0,500	1,386
1,62	-5,00	2,279	0,567	1,672
1,43	-0,69	2,400	0,633	2,007
7,46	1,61	2,622	0,700	2,408
8,92	-6,60	2,686	0,767	2,911
6,61	2,32	2,737	0,833	3,584
4,03	2,87	2,976	0,900	4,605
9,42	-7,64	3,819	0,967	6,802



# Exemplo



# Transformações de normalização

- Se a suposição de normalidade dos dados não é plausível, que estratégia adotar?
- (i) Usar técnicas estatísticas apropriadas para dados não-normais, após verificar a distribuição plausível para os dados (Poisson, Gamma, etc.)
- (ii) Transformar os dados para uma nova escala, sob a qual a suposição de normalidade é plausível.
- Aqui, somente trataremos da transformação, pois técnicas a serem estudadas a seguir são voltadas para dados normais.
- Transformações não são nada mais do que uma forma de reescrever os dados numa unidade diferente.

# Transformações úteis

Escala original	Escala transformada
Contagens ( $y$ )	$\sqrt{y}$
Proporções ( $\hat{p}$ )	$\text{logit}(\hat{p}) = \frac{1}{2} \log \left( \frac{\hat{p}}{1-\hat{p}} \right)$
Correlação ( $r$ )	Fisher $Z(r) = \frac{1}{2} \log \left( \frac{1+\hat{r}}{1-\hat{r}} \right)$

# Transformações de normalização

- Em muitas situações a escolha para melhorar a aproximação normal não é óbvia. Para tais casos é conveniente deixar que os dados mostrem uma transformação. Uma família útil de transformações para esse propósito é a família de transformações de potências :  $x^\lambda$ .
- As transformações de potência só estão definidas para variáveis positivas. Porém, isso não é tão restritivo quanto parece, porque uma única constante pode ser adicionada a cada observação no conjunto de dados se alguns dos valores observados forem negativos

- Box e Cox (1964) consideraram a seguinte família de transformações de potência modificada:

$$x^\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \ln(x), & \text{se } \lambda = 0 \end{cases}$$

que é contínua em  $\lambda$  para  $x > 0$ .

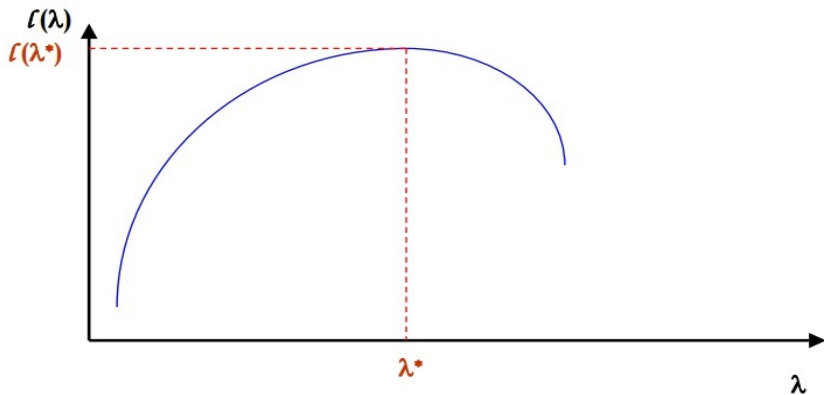
Dada a amostra  $x_1, \dots, x_n$  escolhe-se  $\lambda$  de modo a maximizar:

$$\ell(\lambda) = -\frac{1}{n} \left[ \sum_{j=1}^n (x_j^\lambda - \bar{x}^\lambda)^2 \right] + (\lambda - 1) \sum_{j=1}^n \ln x_j$$

onde  $\bar{x}^\lambda = \sum_{j=1}^n x_j^\lambda / n$

- Para a escolha de  $\lambda$  construir um malha de valores de  $(\lambda, \ell(\lambda))$  onde  $\lambda \in [-2, 2]$  e  $\ell(\lambda)$  é valor máximo em  $\lambda$ ,
- plotar os pares  $(\lambda, \ell(\lambda))$ .

# Transformações de normalização



# Transformações de normalização

- A transformação obtida geralmente melhora a aproximação à normalidade.
- Porém, não há garantias de que mesmo a melhor escolha de  $\lambda$  produzirá um conjunto de dados transformados que seja adequado à suposição de normalidade.
- Os resultados obtidos por uma transformação selecionada de acordo com esse procedimento devem ser cuidadosamente examinados para possíveis violações da suposição de normalidade.
- Essa recomendação de fato vale para qualquer transformação usada.

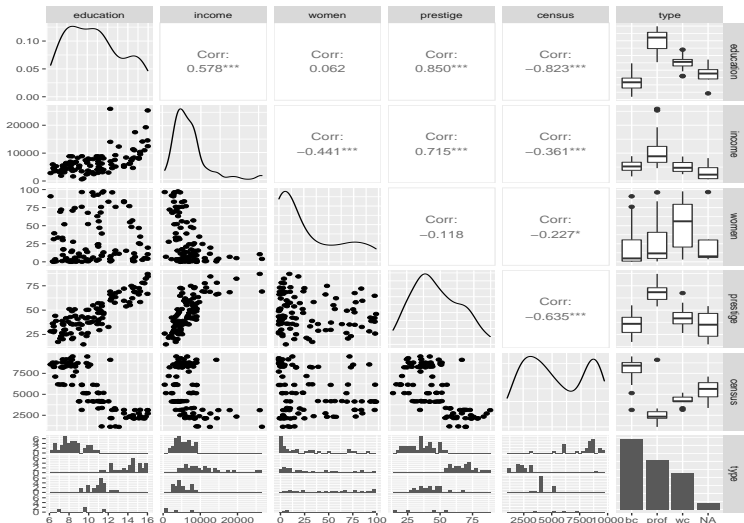


# Transformação de dados: Exemplo

Do pacote *car* do R consideramos o conjunto de dados "Prestige" que possui 102 linhas e 6 colunas.

- education: Average education of occupational incumbents, years, in 1971.
- income: Average income of incumbents, dollars, in 1971.
- women: Percentage of incumbents who are women.
- prestige: Pineo-Porter prestige score for occupation, from a social survey conducted in the mid-1960s.
- census Canadian Census occupational code.
- type: Type of occupation. A factor with levels (note: out of order): bc, Blue Collar; prof, Professional, Managerial, and Technical; wc, White Collar.

# Transformação de dados: Exemplo



# Transformação de dados: Exemplo

A função `PowerTransform` do pacote *car* contém o procedimento inferencial, para  $\lambda$  na transformação de potência de Box-Cox.

```
> dados=data.frame(education, income, prestige, census)
> m=powerTransform(dados)
> summary(m)
```

bcPower Transformations to Multinormality

	Est Power	Rounded Pwr	Wald Lwr Bnd	Wald Up Bnd
education	0.8095	1.00	0.2831	1.3359
income	0.2082	0.33	0.0354	0.3810
prestige	0.9549	1.00	0.6012	1.3085
census	0.6922	0.50	0.3877	0.9966

Likelihood ratio test that transformation parameters are equal to 0  
(all log transformations)

	LRT	df	pval
LR test, lambda = (0 0 0 0)	56.34121	4	1.7006e-11

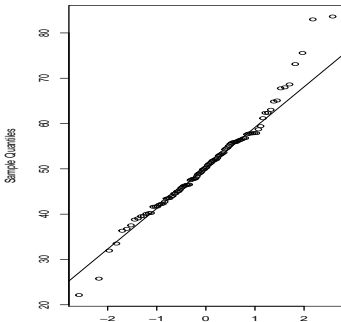
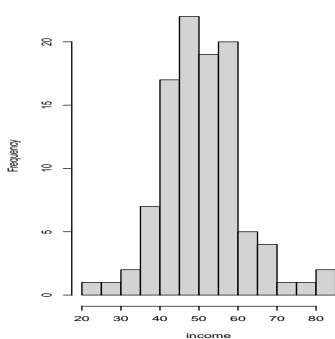
Likelihood ratio test that no transformations are needed

	LRT	df	pval
LR test, lambda = (1 1 1 1)	71.05226	4	1.3656e-14

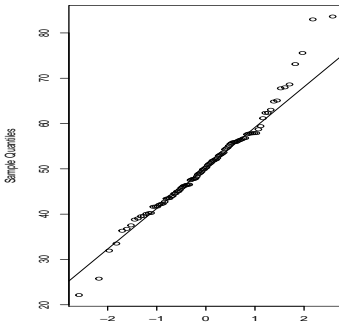
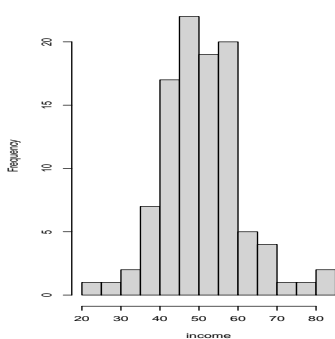
```
> testTransform(m, c(1,0.3,1,.5))
```

	LRT	df	pval
LR test, lambda = (1 0.3 1 0.5)	3.029857	4	0.55284

# Transformação de dados: Exemplo



# Transformação de dados: Exemplo



```
hist(bcPower(income,m$roundlam[2]),main='',xlab='income',nclass=10,)  
qqnorm(bcPower(income,m$roundlam[2]),main='',xlab='')  
qqline(bcPower(income,m$roundlam[2]))
```