

Inferência para a Distribuição Normal Multivariada

Prof. Caio Azevedo

Verossimilhança

- Temos que $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \text{tr}[(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})] = \text{tr}[\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']$.
- Seja $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (ou seja, uma matriz de dados). A verossimilhança de uma amostra aleatória de tamanho n é dada por

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp \left\{ -0,5 \left[\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right] \right\}$$

- Note que (somando-se e subtraindo-se $\bar{\mathbf{x}}$ em cada “()”)
$$\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'$$
,
em que $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$.

Estatística suficiente

- A verossimilhança pode ser reescrita como:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp \left\{ -0,5 \operatorname{tr} \left[\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \right] \right. \\ \left. + 0,5n \times \operatorname{tr} \left[\boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})' \right] \right\}$$

- Assim, pelo critério da fatoração temos que

$(\sum_{i=1}^n \mathbf{X}_i, \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})')$ é uma estatística suficiente para $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$.

Estimadores de máxima verossimilhança

- Por outro lado, temos que a log-verossimilhança pode ser escrita das duas formas seguintes (definindo $\mathbf{V} = \boldsymbol{\Sigma}^{-1}$):

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{const} + \frac{n}{2} \ln |\mathbf{V}| - 0,5 \left[\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \mathbf{V} (\mathbf{x}_i - \boldsymbol{\mu}) \right] \quad (1)$$

$$= \text{const} + \frac{n}{2} \ln |\mathbf{V}| - 0,5 \text{tr} \left[\mathbf{V} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}) (\mathbf{x}_i - \boldsymbol{\mu})' \right] \quad (2)$$

- Derivando (1) com relação à $\boldsymbol{\mu}$ e (2) com relação à $\boldsymbol{\Sigma}$, vem que

Estimadores de máxima verossimilhança

$$S(\boldsymbol{\mu}) = \frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\frac{1}{2} \sum_{i=1}^n [2\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})] = n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}$$

$$S(\mathbf{V}) = \frac{n}{2} \left(\frac{\partial}{\partial \mathbf{V}} \ln |\mathbf{V}| \right) - \frac{1}{2} \frac{\partial}{\partial \mathbf{V}} \text{tr} [(n-1)\mathbf{S}^2 \mathbf{V}] = \frac{n}{2}(\mathbf{V}^{-1})' - \frac{1}{2}(n-1)\mathbf{S}^2$$

em que $\mathbf{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Assim, igualando cada uma das derivadas acima à $\mathbf{0}_{(p \times 1)}$ e $\mathbf{0}_{(p \times p)}$, respectivamente, e resolvendo o sistema resultante em relação aos parâmetros, temos que:

$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$; $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Contudo, em geral, utiliza-se \mathbf{S}^2 para estimar $\boldsymbol{\Sigma}$ (por ser não viciado), enquanto que $\mathcal{E}(\hat{\boldsymbol{\Sigma}}) = \frac{(n-1)}{n} \boldsymbol{\Sigma}$. Pode-se provar ainda que $\bar{\mathbf{X}} \perp \mathbf{S}^2$ (Teorema de Basu).

Distribuições amostrais dos $\hat{\mu}$ e \mathbf{S}^2

- Temos que $\hat{\mu} \sim N_p(\mu, \Sigma/n)$ (exercício). Assim,
 $\hat{\mu}_j \sim N(\mu_j, \sigma_j^2/n), j = 1, 2, \dots, p$, em que $\hat{\mu}_j = \bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ji}$.
- Temos que $(n-1)\mathbf{S}^2 \sim W_p(n-1, \Sigma)$. Assim,
 $\frac{(n-1)S_j^2}{\sigma_j^2} \sim \chi_{(n-1)}^2, j = 1, 2, \dots, p$, em que $S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ji} - \hat{\mu}_j)^2$,
é o j -ésimo elemento da diagonal principal de \mathbf{S}^2 .

Distribuições amostrais dos $\hat{\mu}$ e S^2

- Densidade da distribuição de Wishart ($\mathbf{\Sigma} \sim W_p(n, \mathbf{V})$) (pesquisar sobre ela):

$$f(\mathbf{\Sigma}) = \frac{|\mathbf{\Sigma}|^{\frac{n-p-1}{2}} e^{-\frac{\text{tr}(\mathbf{V}^{-1}\mathbf{\Sigma})}{2}}}{2^{\frac{np}{2}} |\mathbf{V}|^{\frac{n}{2}} \Gamma_p\left(\frac{n}{2}\right)}$$

em que $\Gamma_p(k)$ é a função gama p-variada.

- Seja $\mathbf{X}_i \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{V})$ e defina $\mathbf{\Sigma} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i'$, então $\mathbf{\Sigma} \sim W_p(n, \mathbf{V})$ (obs: esta matriz \mathbf{V} não é a mesma definida anteriormente).

Inferência para cada componente

- $IC(\mu_j, \gamma) = \left[\bar{X}_j - t_{\frac{1+\gamma}{2}} \sqrt{\frac{S_j^2}{n}}, \bar{X}_j + t_{\frac{1+\gamma}{2}} \sqrt{\frac{S_j^2}{n}} \right]$, em que $P(T \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $T \sim t_{(n-1)}$.
- $IC(\sigma_j^2, \gamma) = \left[\frac{(n-1)S_j^2}{q_{\frac{1+\gamma}{2}}}, \frac{(n-1)S_j^2}{q_{\frac{1-\gamma}{2}}} \right]$, em que $P(Q \leq q_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, e $P(Q \leq q_{\frac{1-\gamma}{2}}) = \frac{1-\gamma}{2}$ $Q \sim \chi_{(n-1)}^2$.
- Testes de hipóteses para cada componente podem ser feitos do modo usual. Contudo, para as componentes de μ veremos formas alternativas.

Inferência para combinações lineares do vetor de médias:

$$R_{(1 \times p)} \mu$$

- Temos que $R\hat{\mu} \sim N(R\mu, R\Sigma R'/n)$ (exercício).
- Assim, analogamente ao caso anterior, temos que

$$IC(R\mu, \gamma) = \left[R\bar{X} - t_{\frac{1+\gamma}{2}} \sqrt{\frac{RS^2R'}{n}}, R\bar{X} + t_{\frac{1+\gamma}{2}} \sqrt{\frac{RS^2R'}{n}} \right], \text{ em que}$$

$$P(T \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}, T \sim t_{(n-1)}.$$

Região de confiança para μ

- Pode-se provar que :

- $(\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi_{(p)}^2$ e $Q = n(\bar{\mathbf{X}} - \mu)' \Sigma^{-1} (\bar{\mathbf{X}} - \mu) \sim \chi_{(p)}^2$
(prova:fgm).
- $T = n(\bar{\mathbf{X}} - \mu)' (\mathbf{S}^2)^{-1} (\bar{\mathbf{X}} - \mu) \sim T^2$ de Hotelling e
 $F = \frac{n-p}{(n-1)p} T \sim F_{(p, n-p)}$.
- Seja $\mathbf{X} \sim N_p(\mu, \Sigma)$ e $\mathbf{W} \sim W_p(k, \Sigma)$, $\mathbf{X} \perp \mathbf{W}$. Assim
 $Y = (\mathbf{X} - \mu)' (\mathbf{W}/k)^{-1} (\mathbf{X} - \mu) \sim T^2$ de Hotelling.
- Um região de confiança é um conjunto, digamos $R(\mu, \gamma)$ contido em \mathcal{R}^p , tal que $P(R(\mu, \gamma) \leq q_\gamma) = \gamma$, em que q é um quantil apropriado da distribuição associada à $R(\mu, \gamma)$.

Região de confiança para μ

- Então, se f_γ for um quantil da distribuição $F_{(p, n-p)}$, tal que $P(F \leq f_\gamma) = \gamma$, uma região de confiança γ para μ é o conjunto de todas as p -uplas, digamos μ^* , tais que

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu}^*)' (\mathbf{s}^2)^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}^*) \leq \frac{p(n-1)}{(n-p)} f_\gamma$$

- Para uma dada amostra basta calcular a forma quadrática acima (para todas as p -uplas) e ver quais satisfazem à inequação.
- Mais detalhes (incluindo $p \geq 3$): Johson and Wichern (2002). No R: função “ellipse” do pacote “ellipse”.

Simulação de um vetor aleatório $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} \in \mathcal{R}^p$, $\boldsymbol{\Sigma}$ positiva definida.
- Sabemos que $\mathbf{X} = \boldsymbol{\Psi}\mathbf{Z} + \boldsymbol{\mu}$, $\boldsymbol{\Psi} = \text{Cholesky}(\boldsymbol{\Sigma})$,
 $\mathbf{Z} = (Z_1, \dots, Z_p)$, $Z_i \stackrel{i.i.d.}{\sim} N(0, 1)$, $i = 1, \dots, p$.
- $\mathbf{x} = \boldsymbol{\Psi}\mathbf{F}_Z^{-1}(\mathbf{z}) + \boldsymbol{\mu}$, $\mathbf{z} = (z_1, \dots, z_p)$, em que
 $\mathbf{F}^{-1}(\mathbf{z}) = (F_Z^{-1}(z_1), \dots, F_Z^{-1}(z_p))$ (vetor coluna). No R,
“manualmente”:

```
m.u <- cbind(runif(n)); m.z <- qnorm(m.u)
for (i in 2:p)
  {m.u <- cbind(runif(n))
  m.z <- cbind(m.z, qnorm(m.u))}; m.x <- t(chol(m.sigma)%*%t(m.z)
```



Cont.

- Também no R, mais simples: função “mvrnorm ” do pacote “MASS”.

`mvrnorm`(`n`= “tamanho da amostra”, `mu`= “vetor de médias”,
`Sigma`= “matriz de covariâncias”)

Verificação da normalidade multivariada de uma matriz de dados

- $X_{ji} \sim N(\mu_j, \sigma_j^2), j = 1, 2, \dots, p, i = 1, \dots, n.$
- $n(\mathbf{X}_i - \bar{\mathbf{X}})' (\mathbf{S}^2)^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \approx \chi_p^2,$ para n suficientemente grande.
- Construir histogramas e/ou gráficos de quantis-quantis com envelope para cada variável e para a forma quadrática, teste de Kolmogorov-Smirnov.

Procedimento para se gerar o gráfico de envelopes (quantil-quantil)

- 1) Simule n variáveis aleatórias independentes de interesse ($N(0, 1)$ ou $F_{(p, n-p)}$). Repita este processo m vezes.
- 2) Ao final teremos uma matriz com valores simulados dessas variáveis aleatórias, digamos V_{ij} , $i=1, \dots, n$, (tamanho da amostra) $j=1, \dots, m$ (réplica).

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nm} \end{bmatrix}$$

Cont.

- 3) Dentro de cada amostra, ordena-se, de modo crescente, os valores simulados, obtendo-se $v_{(i)j}^*$ (estatísticas de ordem):

$$\mathbf{V}^* = \begin{bmatrix} v_{(1)1} & v_{(1)2} & \dots & v_{(1)m} \\ v_{(2)1} & v_{(2)2} & \dots & v_{(2)m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{(n)1} & v_{(n)2} & \dots & v_{(n)m} \end{bmatrix}$$

- 4) Pode-se obter os limites $v_{(i)l} = \min_{1 \leq j \leq m} v_{(i)j}$ e $v_{(i)s} = \max_{1 \leq j \leq m} v_{(i)j}$,
 $i = 1, 2, \dots, n$.

Cont.

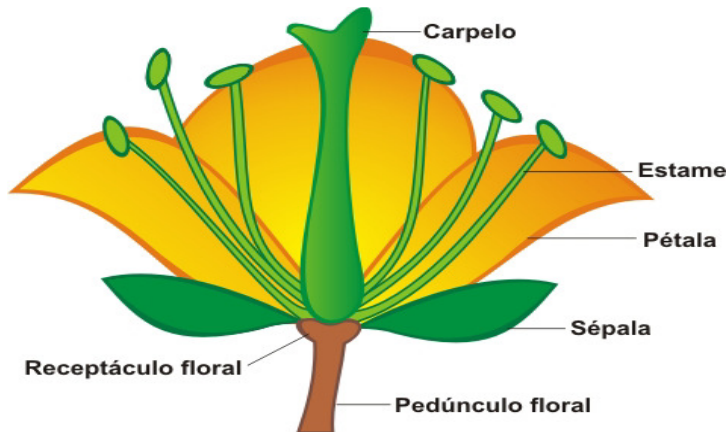
- 5) Porém, na prática considera-se $v_{(i)l} = \frac{v_{(i)(2)} + v_{(i)(3)}}{2}$ e $v_{(i)s} = \frac{v_{(i)(m-2)} + v_{(i)(m-1)}}{2}$ (para se gerar limites de confiança), em que $v_{(i)(r)}$ é a r -ésima estatística de ordem dentro de cada linha, $i = 1, 2, \dots, n$.

- Além disso, consideramos como a linha de referência

$$v_{(i)} = \frac{1}{m} \sum_{j=1}^m v_{(i)j}, i = 1, 2, \dots, n.$$

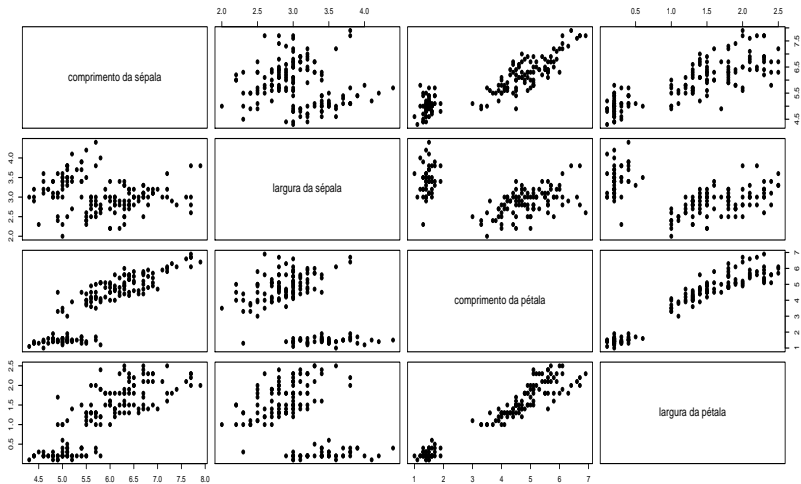
Dados da íris “de Fisher”

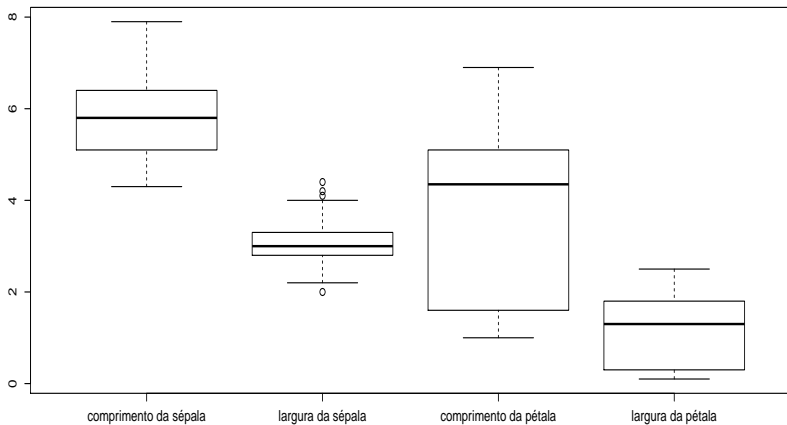
- Os dados consistem de 50 unidades amostrais de três espécies (setosa, virginica, versicolor) de íris (uma espécie de planta), ou seja, temos um total de 150 unidades amostrais.
- De cada uma delas mediu-se quatro variáveis: comprimento e largura da sépala (CS, LS) e comprimento e largura da pétala (CP,LP).
- Objetivo original: quantificar a variação morfológica em relação à essas espécies com bases nas quatro variáveis de interesse.

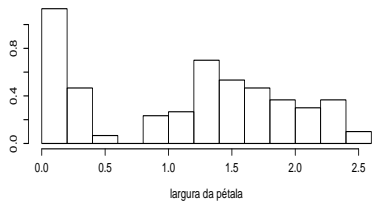
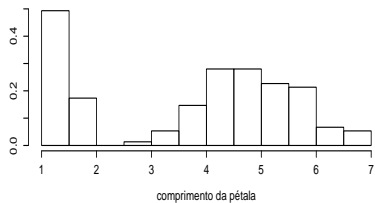
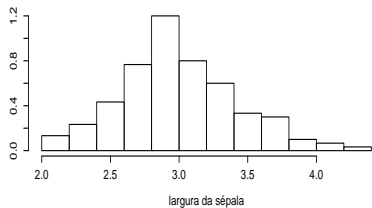
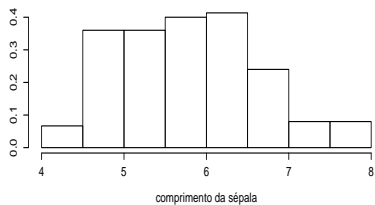


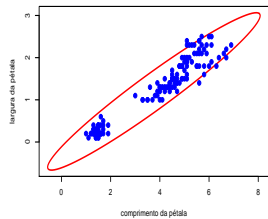
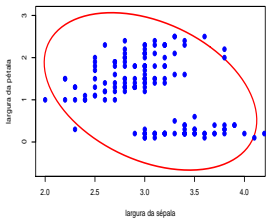
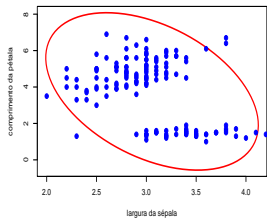
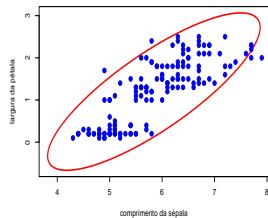
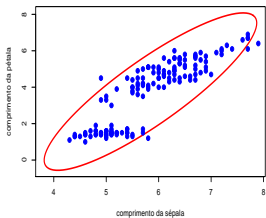
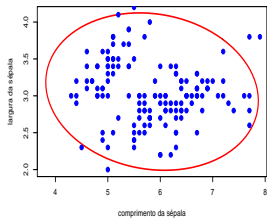
Cont.

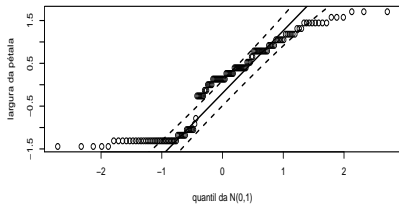
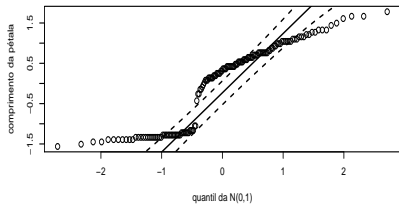
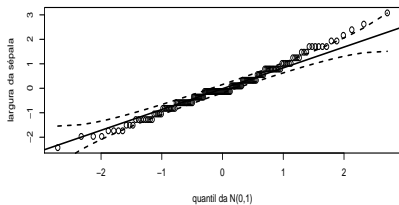
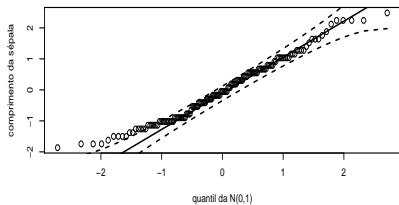
- Seja Y_{ijk} : o valor da k -ésima variável ($k=1,2,3,4$), para o j -ésimo indivíduo ($j=1,\dots,50$) do i -ésimo grupo ($i = 1, 2, 3$).
- Por enquanto, vamos desconsiderar as espécies de Íris, ou seja, vamos considerar $\mathbf{Y}_j = (Y_{j1}, Y_{j2}, Y_{j3}, Y_{j4}) \stackrel{ind.}{\sim} N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, seguindo a ordem apresentada no slide anterior.

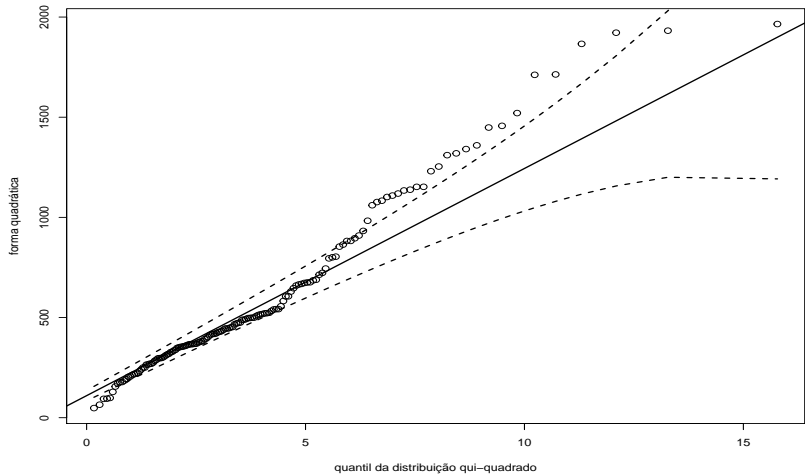












- Testes de Kolmogorov-Smirnov (p -valor entre parênteses):
 - CS: $D= 0,089$ ($p = 0,1891$).
 - LS: $D= 0,106$ ($p = 0,0723$).
 - CP: $D= 0,198$ ($p < 0,0001$).
 - LP: $D= 0,173$ ($p < 0,0003$).