Lista 1 - Estatística Não-Paramétrica - 1/2015 - 30/03/2015

Augusto Cesar Ribeiro Nunes - 13/0103004

23 de março de 2015

Este trabalho está disponível em um repositório do github:

Questão 15

Use the data in Table 5.8 to create a kernel density estimate for each group of subjects. Do the estimated densities appear to represent the same underlying distribution?

Carregamento e descrição dos dados sobre os sujeitos dos Tipos A e B

```
typeA <- c(3.6,2.6,4.7,8.0,3.1,8.8,4.6,5.8,4.0,4.6)
typeB <- c(16.2,17.4,8.5,15.6,5.4,9.8,14.9,16.6,15.9,5.3,10.5)
summary(typeA)
```

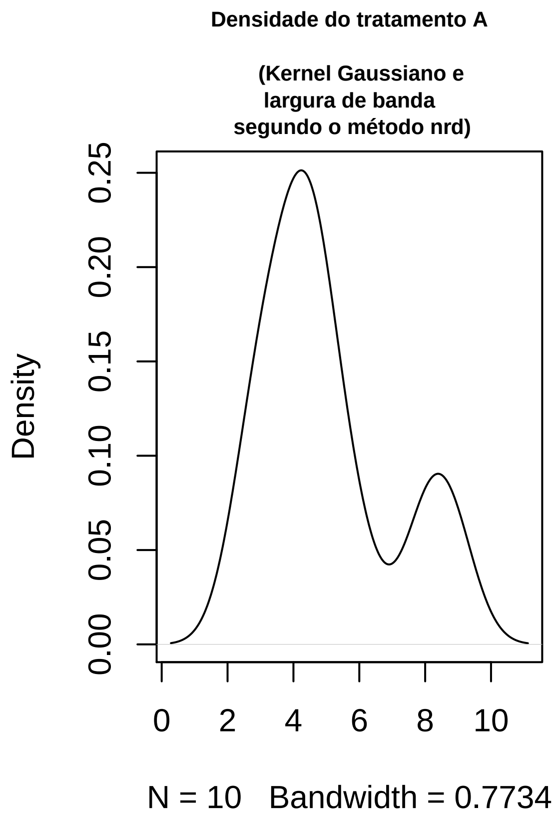
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.600   3.700   4.600   4.980   5.525   8.800
```

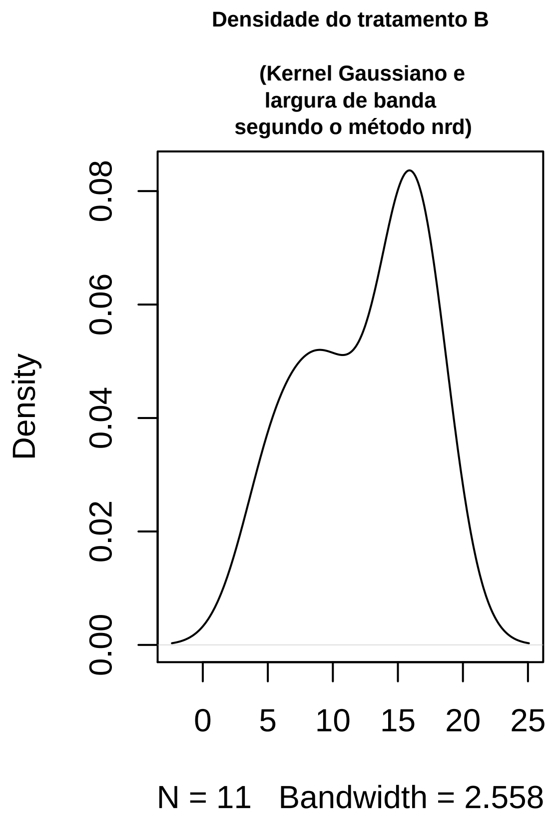
```
summary(typeB)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.30    9.15   14.90   12.37   16.05   17.40
```

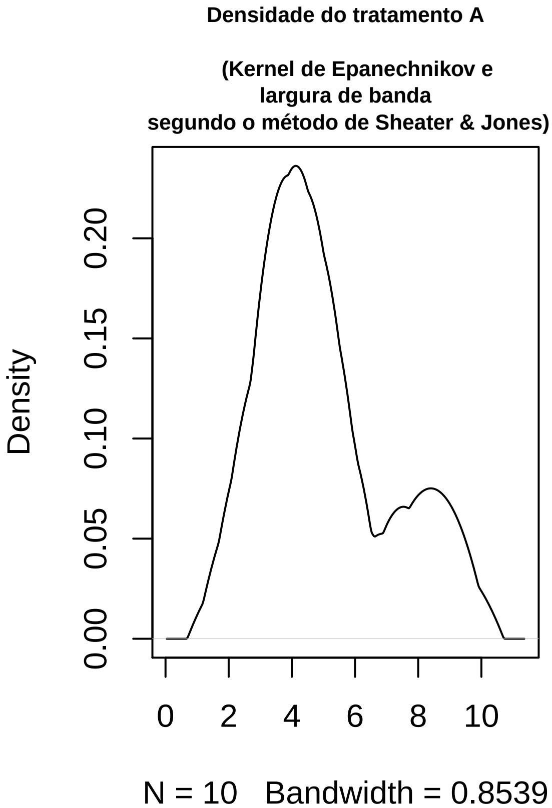
Gráfico das duas densidades

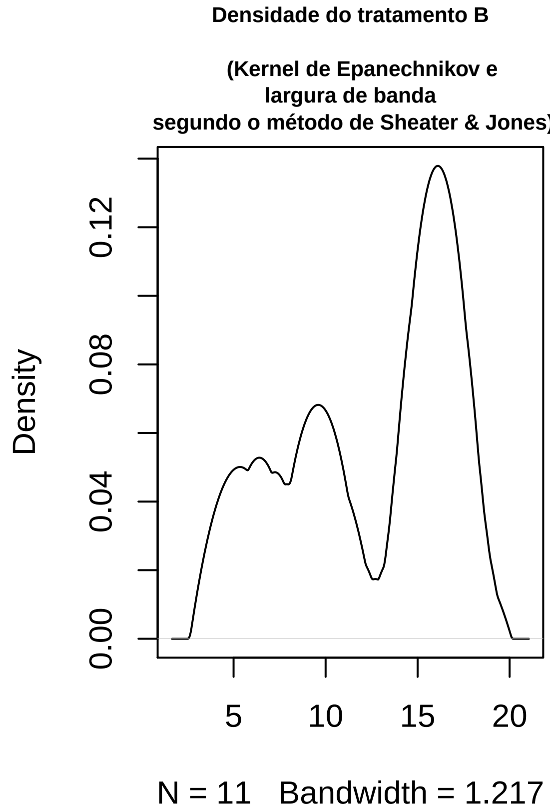
```
par(mfrow=c(1,2), cex.main=0.7)
plot(density(typeA), main="Densidade do tratamento A\n
     (Kernel Gaussiano e \nlargura de banda\n segundo o método nrd)")
plot(density(typeB), main="Densidade do tratamento B\n
     (Kernel Gaussiano e \nlargura de banda\n segundo o método nrd)")
```






```
plot(density(typeA, kernel="epanechnikov", bw="SJ"), main="Densidade do tratamento A\n
     (Kernel de Epanechnikov e \nlargura de banda\n segundo o método de Sheater & Jones)")
plot(density(typeB, kernel="epanechnikov", bw="SJ"), main="Densidade do tratamento B\n
     (Kernel de Epanechnikov e \nlargura de banda\n segundo o método de Sheater & Jones)")
```





Claramente, as densidades estimadas não representam a mesma população, já que:

- As amplitudes das densidades e de x são claramente diferentes;
- As modas são diferentes, na verdade a distribuição do tratamento A é bimodal (mistura de distribuições?), e a do tratamento B possivelmente não.
- As larguras de banda são sensivelmente diferentes entre si.

Questão 17

Show that the asymptotic MISE is minimized when

$$h = \left(\frac{\int k^2}{n\sigma_k^4 \int (f'')^2} \right)^{\frac{1}{5}}$$

De fato, devemos mostrar que $\frac{\partial MISE(f,\hat{f})}{\partial h} = 0$ e que $\frac{\partial^2 MISE(f,\hat{f})}{\partial h^2} > 0.$ Temos

$$\frac{\partial MISE(f,\hat{f})}{\partial h} \simeq \frac{-1 \int k^2 h^{-2}}{n} + \sigma_k^4 h^3 \int (f'')^2$$

O sinal $\simeq$ justifica-se pois usamos uma aproximação por séries de Taylor de ordem 1 para a expressão

$$MISE = \int \left(\frac{1}{nh} f(x) \int k^2(y) dy - \left[\frac{(f'')^2 h^4}{4} \sigma_k^4 \right] dx \right)$$

Resolvendo a expressão acima para h, chegamos a

$$h^5 \simeq \left(\frac{\int k^2}{n\sigma_k^4 \int (f'')^2}\right)$$

Para mostrar a segunda parte da questão, note que todas as quantidades envolvidas são ≥ 0, isto é:

- $\int K > 0 \Rightarrow \int K^2 > 0$, por definição da função Kernel;
- n, σ_k^4 e h^3 também são positivos, por definição.

Então $\frac{\partial^2 MISE(f,\hat{f})}{\partial h^2} > 0$. Logo,

$$h^5 \simeq \left(\frac{\int k^2}{n\sigma_k^4 \int (f'')^2}\right)$$

é o mínimo da Média do Erro Quadrático Integrado.$\square$

Questão 18

If f is the density for a normal random variable with mean 0 and variance σ^2, show

$$\int (f'')^2 = \frac{3}{8\sqrt[2]{\pi}\sigma^5}$$

De fato, temos

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

E então

$$f'' = -1/2\,\frac{\sqrt{2}}{\sigma^2\sqrt{\sigma^2}}\mathrm{e}^{-1/2\,\frac{x^2}{\sigma^2}} + 1/2\,\frac{x^2\sqrt{2}}{\sigma^4\sqrt{\sigma^2}}\mathrm{e}^{-1/2\,\frac{x^2}{\sigma^2}}$$

E, finalmente,

$$\int (f'')^2 = 1/2\,\frac{1}{\sigma^{10}}\left(-1/2\,\sigma^2 x^3 \mathrm{e}^{-\frac{x^2}{\sigma^2}} - 1/2\,\sigma^2\left(-1/2\,\sigma^2 x \mathrm{e}^{-\frac{x^2}{\sigma^2}} + 1/4\,\sigma^3\sqrt{\pi}\mathrm{erf}\,(x/sigma)\right) + 1/2\,\sigma^5\sqrt{\pi}\mathrm{erf}\,(x/\sigma)\right)$$

Que é diferente do resultado esperado.

Questão 19

_Using the normal kernel K with $\sigma_k = 1$ and assuming the underlying data come from a normal population with mean 0 and variance σ^2, show the optimal bandwidth is

$$h = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{\frac{-1}{5}}$$

NÃO FIZ