

# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

---

#### Exercício 1

Considere os dados do Projeto Qualidade de Vida (Arquivo Dados- CEA15P01-QV-Sessões) para as variáveis do questionário *FACIT-F* avaliadas nas sessões de quimioterapia 1, 5 e 8.

Considerando a primeira sessão de quimioterapia, as seguintes pontuações foram obtidas pelas pacientes ( $n=30$ ) nos itens do questionário *FACIT-F* relacionados ao Bem Estar Físico (variável *BEFi*):

28 23 24 28 24 22 20 28 19 27 27 27 25 22 28 28 23 26 26 23 28 27 27 28 26 28 25 23 24 18

Os valores desta variável variam de 0 a 28 pontos (quanto maior melhor é a sensação de bem-estar).

(a) Calcule (à mão) os quartis  $Q_1$ ,  $Q_2$  e  $Q_3$  para a variável *BEFi*.

SOLUÇÃO: o primeiro passo é ordenar os dados.

Posição	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Valor	18	19	20	22	22	23	23	23	23	24	24	24	25	25	26

Posição	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Valor	26	26	27	27	27	27	27	28	28	28	28	28	28	28	28

O tamanho da amostra é  $n=30$

A posição de  $Q_1$  é dada por  $0,25(n+1)=0,25 \times 31=7,75$  logo  $Q_1=(23+23)/2=23$

A posição de  $Q_2$  é dada por  $0,50(n+1)=0,50 \times 31=15,5$  logo  $Q_2=26$

A posição de  $Q_3$  é dada por  $0,75(n+1)=0,75 \times 31=23,25$  logo  $Q_3=28$

(b) Construa (à mão) o *boxplot* da variável *BEFi*. Comente.

SOLUÇÃO:

Distância interquartil =  $Q_3 - Q_1=5$

$LI= Q_1-1,5(Q_3 - Q_1)=23-1,5 \times 5=23-7,5=15,5$

$LS= Q_3+1,5(Q_3 - Q_1)=28+7,5=35,5$

Mínimo Típico=18 (maior do que  $LI$ )

Máximo Típico=28 (menor do que  $LS$ )

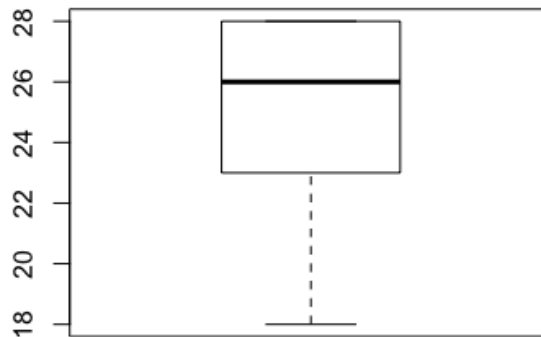
Assim, o *boxplot* para dados de *BEFi* é como a seguir:

# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

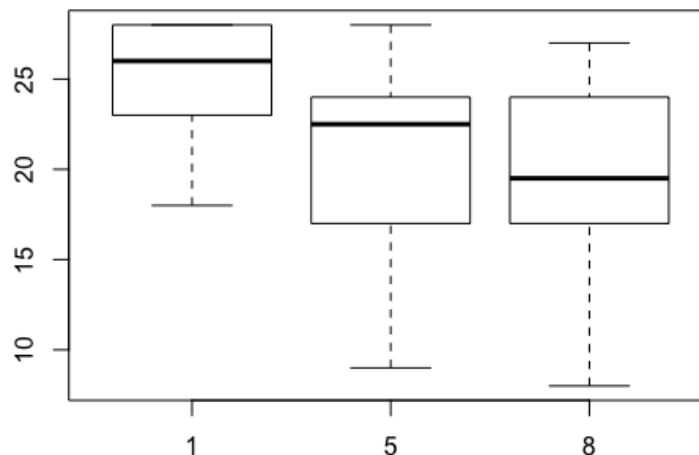
---



Apesar de esperarmos 75% das observações inferiores ou iguais a  $Q_3$ , neste conjunto o terceiro quartil coincide com o máximo típico, ou seja, 100% dos valores são inferiores ao valor 28. Isso ocorreu porque foram observados oito valores iguais a 28. Além disso, observa-se uma distribuição bastante assimétrica, mas sem valores atípicos.

(c) Usando o *Rcmdr*, construa o *boxplot* da variável *BEFi* por sessão de quimioterapia. Comente.

- Carregar o arquivo CEAI2015-QV-Sessoes.xls no *Rcmdr*: Dados-> Importar arquivos de dados -> do arquivo Excel
- Transformar variável “sessão” em fator: Dados-> Modificação de variáveis no conjunto de dados -> Converter variável numérica para fator
- Construir *boxplot* da variável “BEFi” por “sessão”: Gráficos-> Boxplot-> Gráfico por grupos



# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

---

Do gráfico dos *boxplots* observa-se que a mediana do Bem Estar Físico (variável *BEFi*) diminui ao longo do tempo, isto é, à medida que as sessões aumentam. Também se nota um aumento na dispersão dos valores observados.

#### Exercício 2

Considere o arquivo BATATABT que contém dados de um experimento realizado com o objetivo de verificar como a temperatura afeta as batatas durante o armazenamento. Para isso foram injetadas bactérias em 54 batatas em 3 níveis de quantidade: baixa (1), média (2) e alta (3). As batatas foram armazenadas por cinco dias em dois níveis de temperatura: 10 graus C (1) e 16 graus C (2). A variável observada foi o diâmetro (em *mm*) da parte apodrecida da batata.

(a) Classifique as variáveis do estudo

- Bactéria (Nível de quantidade de bactéria): qualitativa ordinal.
- Temperatura (Nível de temperatura de armazenamento): qualitativa ordinal.
- Diâmetro (Diâmetro (em *mm*) da parte apodrecida da batata): quantitativa contínua.

(b) Utilizando medidas descritivas e gráficos, verifique se a quantidade de bactéria injetada parece influenciar no diâmetro da parte apodrecida da batata. Comente.

- Carregar o arquivo BATATABT.xls no *Rcmdr*: Dados-> Importar arquivos de dados -> do arquivo Excel
- Transformar variável “Bactéria” em fator: Dados-> Modificação de variáveis no conjunto de dados -> Converter variável numérica para fator -> Defina nomes
- Medidas descritivas da variável “Diâmetro” por “Bactéria”: Estatísticas-> Resumos-> Resumos numéricos-> Resuma por grupos
- Construir *boxplot* da variável “Diâmetro” por “Bactéria”: Gráficos-> Boxplot-> Gráfico por grupos

Resumos numéricos do diâmetro da parte apodrecida da batata, por quantidade de bactéria injetada:

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	Diâmetro:n
baixa	5.277778	4.198117	9.00	0.7954326	0	0.00	6.5	9.00	10	18
média	9.166667	6.617712	10.75	0.7219323	0	4.00	7.5	14.75	23	18
alta	13.777778	7.712141	12.50	0.5597522	2	7.25	14.0	19.75	26	18

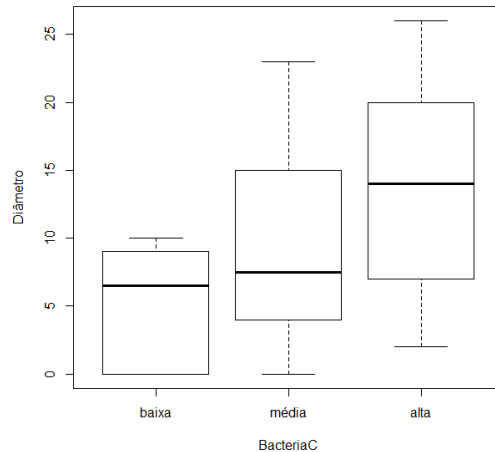
A média do diâmetro (em *mm*) e a variabilidade (*sd* e *DI*) da parte apodrecida da batata é maior quando o nível da quantidade de bactérias injetadas aumenta. O menor valor de *cv* ocorre para a categoria *alta* da variável quantidade injetada de bactéria, ou seja, em relação à média, a variabilidade é menor para essa categoria.

# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

*Boxplot*



Podemos observar que a distribuição do diâmetro da parte apodrecida parece não ser simétrica para os níveis 1-baixa e 2-média da quantidade injetada de bactéria.

Por fim, podemos concluir que quanto maior a quantidade de bactérias injetada, maior será o diâmetro (em *mm*) da parte apodrecida da batata.

(c) Repita o item (a) para a temperatura de armazenamento. Comente.

- Transformar variável “Temperatura” em fator: Dados-> Modificação de variáveis no conjunto de dados -> Converter variável numérica para fator -> Use números
- Medidas descritivas da variável “Diâmetro” por “Temperatura”: Estatísticas-> Resumos-> Resumos numéricos-> Resuma por grupos
- Construir *boxplot* da variável “Diâmetro” por Temperatura”: Gráficos-> Boxplot-> Gráfico por grupos

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	Diâmetro:n
1	5.444444	4.317525	7.0	0.7930149	0	2	5	9.0	15	27
2	13.370370	7.270320	10.5	0.5437635	0	8	14	18.5	26	27

A média do diâmetro (em *mm*) da parte apodrecida da batata assim como *sd* e *DI* são maiores quanto maior é a temperatura de armazenamento. Em termos de variabilidade observou-se um *cv* menor para o nível 2 de temperatura.

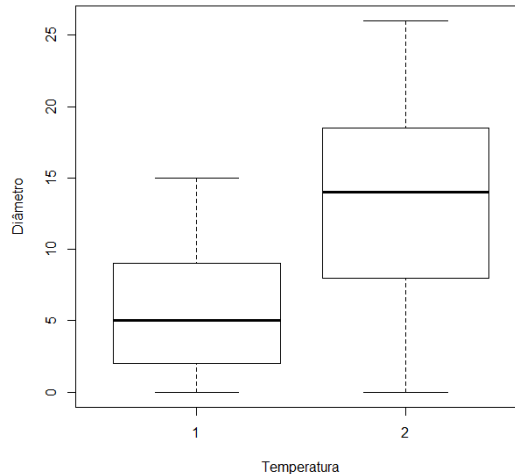
# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

---

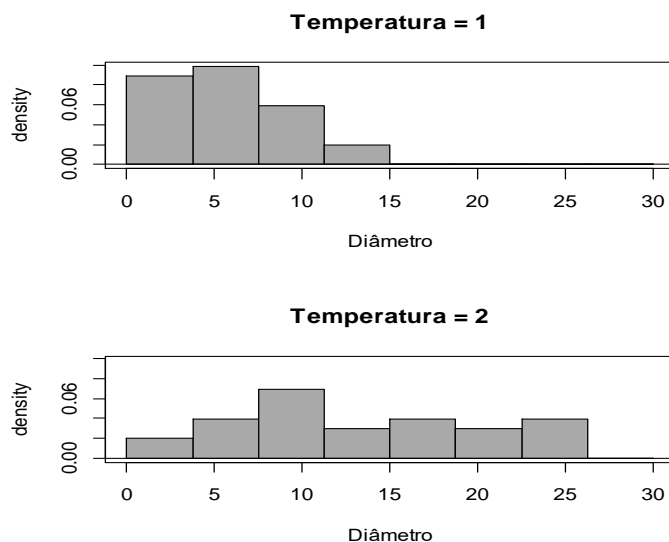
*Boxplot*



Podemos observar que a distribuição do diâmetro da parte apodrecida aparenta ser simétrica para os dois níveis de temperatura. Por fim, podemos concluir que quanto mais alta a temperatura de armazenamento, maior será o diâmetro (em *mm*) da parte apodrecida da batata.

(d) Obtenha o histograma da variável diâmetro da parte apodrecida da batata, por temperatura.

Construir histograma da variável “Diâmetro” : Gráficos-> Histograma-> Gráfico por grupos



# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

#### Exercício 3

Num estudo foram realizadas as análises químicas dos principais óxidos (% em peso) em rochas de duas regiões. Os dados a seguir apresentam os valores de  $SiO_2$  (dióxido de silício) encontrados nas amostras.

Faixas	Frequência	
	Região 1	Região 2
0  — 20	5	2
20  — 40	2	3
40  — 60	13	7
60  — 80	5	7
80  — 90	3	15
90  — 100	1	4

(a) Qual é a variável em estudo? Classifique-a.

Quantidade (% em peso) de  $SiO_2$  (dióxido de silício). Quantitativa contínua.

(b) Construa dois histogramas, um para cada região e compare.

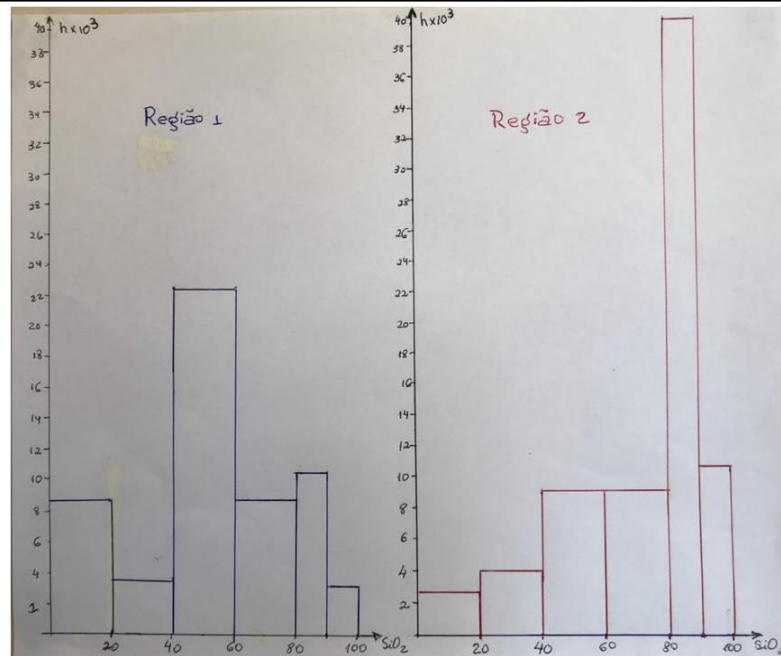
Os histogramas devem ser construídos baseados na densidade ( $h$ ), pois as amplitudes dos intervalos de classe são diferentes.

Faixas	Região 1				Região 2			
	$f$	$fr$	base	$h=fr/base$	$f$	$fr$	base	$h=fr/base$
0  — 20	5	0,172	20	0,0086	2	0,053	20	0,0027
20  — 40	2	0,069	20	0,0035	3	0,079	20	0,00395
40  — 60	13	0,448	20	0,0224	7	0,184	20	0,0092
60  — 80	5	0,172	20	0,0086	7	0,184	20	0,0092
80  — 90	3	0,103	10	0,0103	15	0,395	10	0,0395
90  — 100	1	0,034	10	0,0034	4	0,105	10	0,0105
	29				38			

# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO



Na Região 1 a maior densidade de valores de  $SiO_2$  (dióxido de silício) se encontra na faixa de 40 — 60, enquanto que na Região 2 está na faixa 80 — 90.

**Obs.:** No *Rcmdr* não há opção para construção de histograma com intervalos de classe diferentes. Mas, pode-se instalar o pacote *agricolae* e rodar no *Rcmdr* o seguinte programa em que `frequency=3`, indica que se considera a densidade.

```
library(agricolae)
regiao1<- c(0,20,40,60,80,90,100)
regiao2<- c(0,20,40,60,80,90,100)
fr1<-c(5,2,13,5,3,1)
fr2<-c(2,3,7,7,15,4)
par(mfrow=c(1,2))
graph.freq(regiao1,counts=fr1,frequency=3, ylim = c(0, 0.04))
graph.freq(regiao2,counts=fr2,frequency=3, ylim = c(0, 0.04))
```

# MAE116 - Noções de Estatística

## Grupo B/D - 2º semestre de 2020

### Lista de exercícios 2 - Estatística Descritiva II - CLASSE - GABARITO

---

