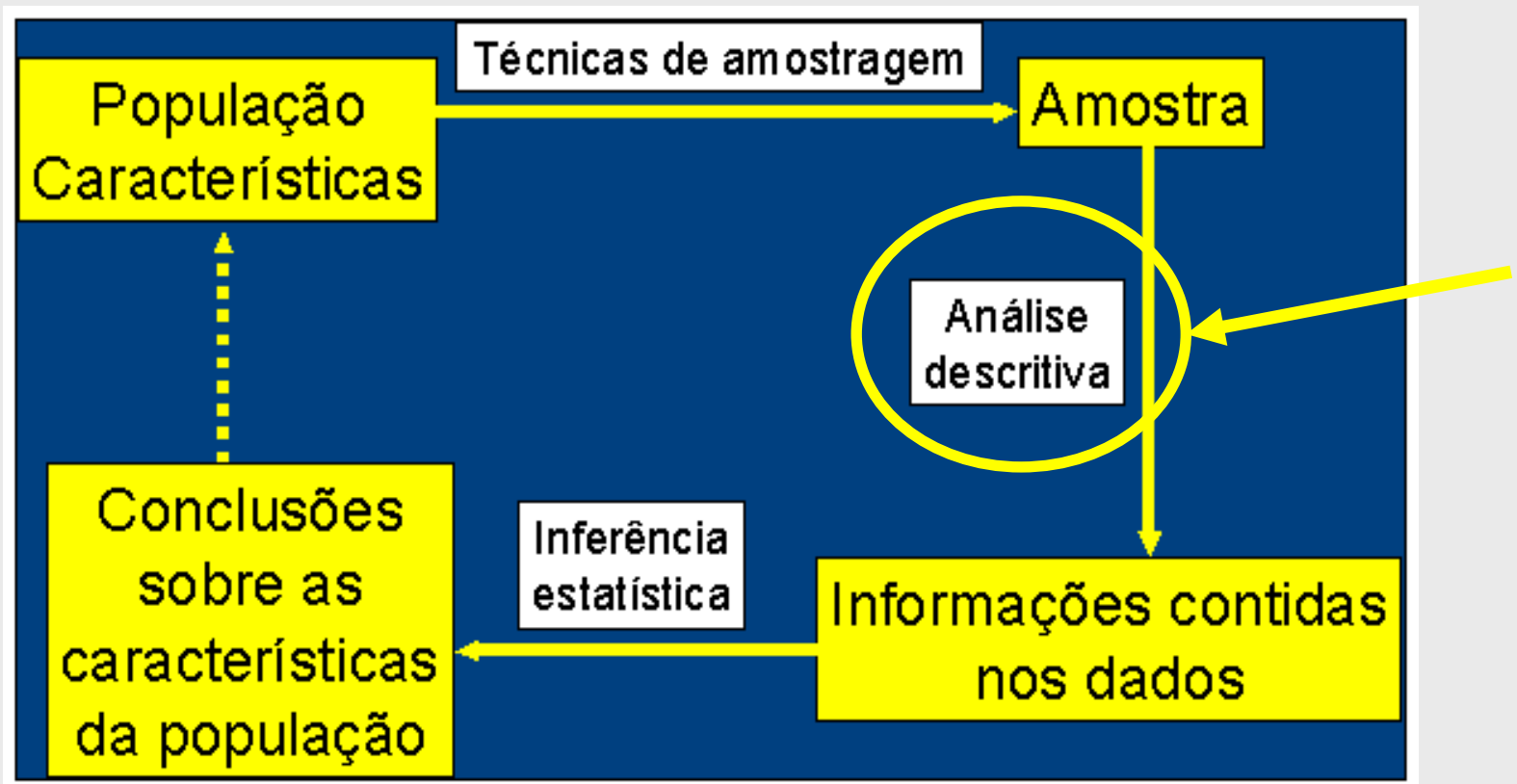


Medidas separatrizes

ACH0021 – Tratamento e Análise de Dados/Informações
Prof. Regis Rossi A. Faria
1º sem. 2020



Etapas da análise estatística



ANÁLISE DESCRITIVA

- **tabelas**
- **gráficos**
- **medidas**
 - **média, mediana, moda**
 - **desvio-padrão, coeficiente de variação**
 - **percentis, quartis, decis**

MEDIDAS SEPARATRIZES

- **medidas que dividem a distribuição em partes iguais**
- **servem para descrever posições numa distribuição de dados**

MEDIDAS SEPARATRIZES

Quartil

Valores da variável que dividem a distribuição em quatro partes iguais.

	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	
25%	25%	25%	25%	

Q1: deixa abaixo 25% das observações

25%	75%
-----	-----

Q2: deixa abaixo 50% das observações

50%	50%
-----	-----

Q3: deixa abaixo 75% das observações

75%	25%
-----	-----

Quartis

	$\frac{1}{4}$		$\frac{1}{2}$		$\frac{3}{4}$	
25%		25%		25%		25%

Q1: deixa abaixo 25% das observações

25%	75%
-----	-----

Posição do Q1 (1º quartil) = $(n+1)/4$

Quartis

	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	
25%	25%	25%	25%	

Q2: deixa abaixo 50% das observações

50%	50%
-----	-----

$$\text{Posição do Q2 (2º quartil)} = (n+1)/2$$

Quartis

	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	
25%	25%	25%	25%	

Q3: deixa abaixo 75% das observações

75%	25%
-----	-----

$$\text{Posição do Q3 (3º quartil)} = \frac{3}{4}(n+1)$$

Cálculo dos quartis

9 10 17 29 31 53 55 122 129 140 336 → n=11

Onde ficam os quartis?

Cálculo dos quartis

9 10 17 29 31 53 55 122 129 140 336 → n=11

Posição do Q1 = $(11+1)/4 = 3^{\text{a}}$ posição → Q1 = 17

Posição do Q2 = $(11+1)/2 = 6^{\text{a}}$ posição → Q2 = 53

Posição do Q3 = $[3 \cdot (11+1)]/4 = 9^{\text{a}}$ posição → Q3 = 129

Cálculo dos quartis

10 19 20 21 25 30 31 33 37 61 77 88 91 → n=13

Onde ficam os quartis?

Cálculo dos quartis

10 19 20 21 25 30 31 33 37 61 77 88 91 → n=13

Posição do Q1 = $(13+1)/4 = 3,5^a$ posição

$$Q1 = 20 + 0,5*(21-20) \longrightarrow Q1 = 20,5$$

Posição do Q2 = $(13+1)/2 = 7^a$ posição

$$Q2 = 31$$

Posição do Q3 = $[3 \cdot (13+1)]/4 = 10,5^a$ posição

$$Q3 = 61 + 0,5*(77-61)$$

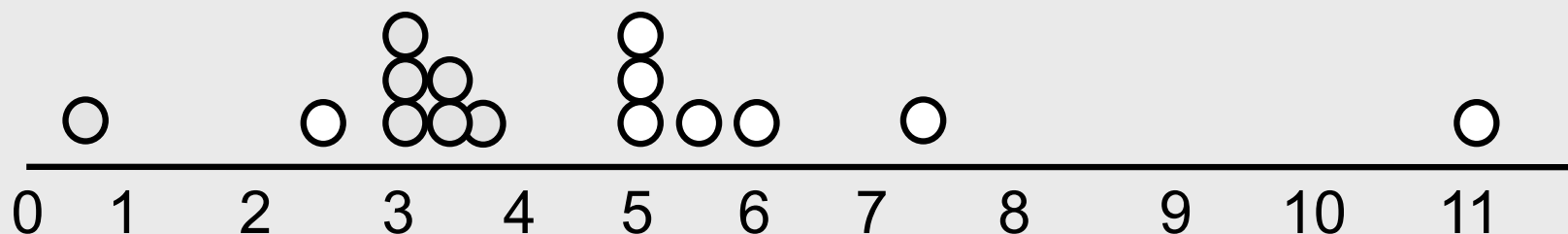


Exercício

- Os dados abaixo referem-se aos salários (milhares de R\$/mês) de 15 pessoas com curso superior.
- Obtenha os quartis deste conjunto de dados.
- Você diria que há algum *outlier*?

11	2,5	5,0	5,0	5,5	3,0	3,5	3,0	0,4	3,2	5,0	3,0	3,2	7,4	6,0
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Gráfico de pontos dos salários das 15 pessoas com curso superior (milhares de R\$/mês):



0,4	2,5	3,0	3,0	3,0	3,2	3,2	3,5	5,0	5,0	5,0	5,5	6,0	7,4	11
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Exercício

- Os dados abaixo referem-se ao tempo de permanência em UTI de bebês nascidos prematuros (dias).
- Obtenha os quartis deste conjunto de dados.
- Você diria que há algum *outlier*?

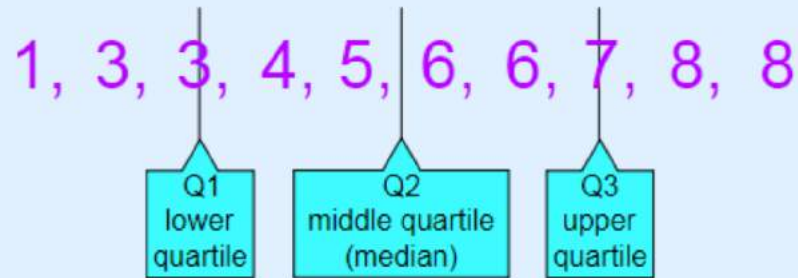
18	21	21	23	23	25
27	29	30	31	32	32
32	34	35	36	38	41
42	42	43	44	45	46
46	47	48	50	54	56
57	58	60	61	98	116

Quartis: exemplo

- Quartis, que dividem os dados em quartas partes

Example: 1, 3, 3, 4, 5, 6, 6, 7, 8, 8

The numbers are in order. Cut the list into quarters:



In this case Quartile 2 is half way between 5 and 6:

$$Q2 = (5+6)/2 = \mathbf{5,5}$$

And the result is:

- Quartile 1 (Q1) = 3
- Quartile 2 (Q2) = 5,5
- Quartile 3 (Q3) = 7

Percentis

- Percentil: valor abaixo do qual está uma porcentagem dos dados

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you:



That means you are at the **80th percentile**.

If your height is 1,85m then "1,85m" is the 80th percentile height in that group.

Percentis para dados agrupados

- Quando os dados estão agrupados: adicione todas as porcentagens **abaixo** da pontuação, então some metade da porcentagem **na** pontuação

Example: You Score a B!

In the test 12% got D, 50% got C, 30% got B and 8% got A

You got a B, so add up

- all the 12% that got D,
- all the 50% that got C,
- half of the 30% that got B,



for a total percentile of $12\% + 50\% + 15\% = 77\%$

In other words you did "as well or better than 77% of the class"

(Why take half of B? Because you shouldn't imagine you got the "Best B", or the "Worst B", just an average B.)

Decis

- Decis são semelhantes a percentis, na medida em que partem os dados em grupos de 10%
 - O 1º decil é o 10º percentil (o valor que divide os dados tal que 10% deles estão abaixo)
 - O 2º decil é o 20º percentil (o valor abaixo do qual 20% dos dados estão)
 - etc.

Example: (continued)

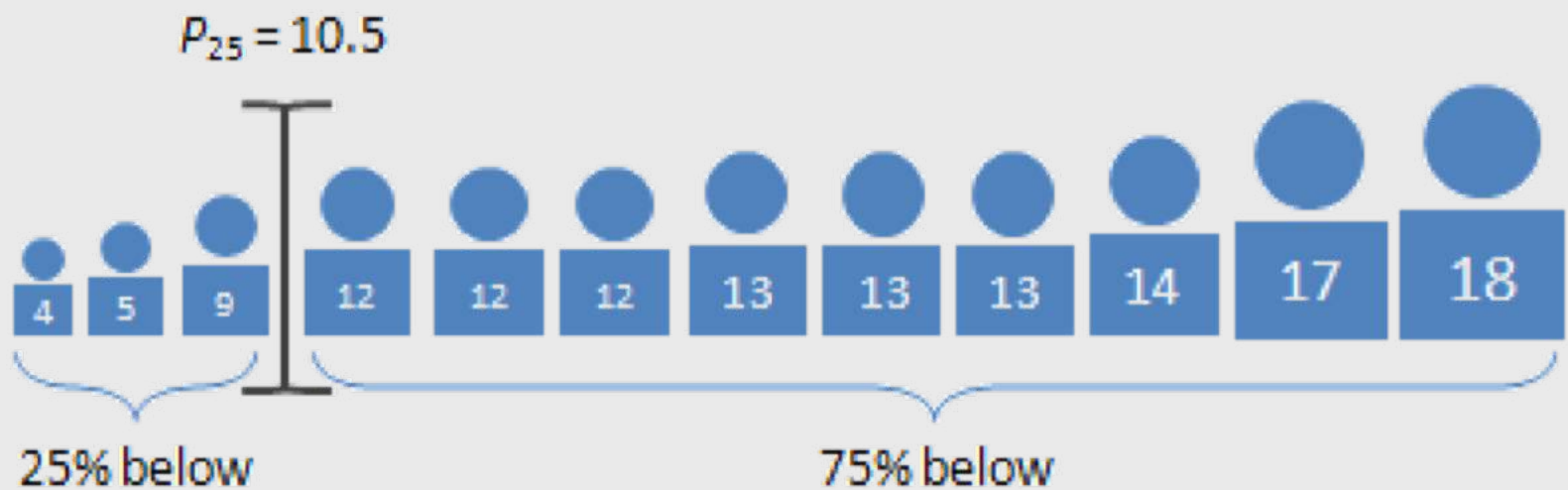


You are at the **8th decile** (the 80th percentile).

Quartis e percentis

- Os quartis também dividem os dados em divisões de 25%, então
 - Quartil 1 (Q1) pode ser chamado de 25º percentil
 - Quartil 2 (Q2) pode ser chamado de 50º percentil
 - Quartil 3 (Q3) pode ser chamado de 75º percentil
- Exemplo (continuado dos slides anteriores):
 - Para: 1, 3, 3, 4, 5, 6, 6, 7, 8, 8:
 - O 25º percentil é 3
 - O 50º percentil é 5,5
 - O 75º percentil é 7

Percentil 25° = Q1



Exercício: óbitos de bebês

- Uma pesquisa foi realizada com o objetivo de verificar se o peso ao nascer está associado à ocorrência de óbito em bebês nascidos com problemas respiratórios.
- Nesta pesquisa foram obtidos os pesos de 50 recém-nascidos, sendo que 23 sobreviveram e 27 foram a óbito (*).
- Qual foi a conclusão da pesquisa?

1.050*	2.500*	1.890*	1.760	2.830
1.175*	1.030*	1.940*	1.930	1.410
1.230*	1.100*	2.200*	2.015	1.715
1.310*	1.185*	2.270*	2.090	1.720
1.500*	1.225*	2.440*	2.600	2.040
1.600*	1.262*	2.560*	2.700	2.200
1.720*	1.295*	2.730*	2.950	2.400
1.750*	1.300*	1.130	2.550	3.160
1.770*	1.550*	1.575	2.570	3.400
2.275*	1.820*	1.680	3.005	3.640

Cálculo dos quartis

- *Recém nascidos que sobreviveram*

Posição Q_1 = $\frac{1}{4} (23+1) = 6^{\circ}\text{elemento} \rightarrow Q_1 = 1720\text{g}$

Posição Q_2 = $\frac{1}{2} (23+1) = 12^{\circ}\text{elemento} \rightarrow Q_2 = 2200\text{g}$

Posição Q_3 = $\frac{3}{4} (23+1) = 18^{\circ}\text{elemento} \rightarrow Q_3 = 2830\text{g}$

Recém nascidos que sobreviveram

1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	
1	4	5	6	7	7	7	9	0	0	0	2	4	5	5	6	7	8	9	0	1	4	6
3	1	7	8	1	2	6	3	1	4	9	0	0	5	7	0	0	3	5	0	6	0	4
0	0	5	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5	0	0	0

Cálculo dos quartis

Recém nascidos que foram a óbito

Posição Q_1 = $\frac{1}{4} (27+1) = 7^{\circ}\text{elemento} \rightarrow Q_1 = 1230\text{g}$

Posição Q_2 = $\frac{1}{2} (27+1) = 14^{\circ}\text{elemento} \rightarrow Q_2 = 1600\text{g}$

Posição Q_3 = $\frac{3}{4} (27+1) = 21^{\circ}\text{elemento} \rightarrow Q_3 = 2200\text{g}$

[illegible]

1	1	1	1	1	1
0	0	1	1	1	2
3	5	0	7	8	2
0	0	0	5	5	5
*	*	*	*	*	*
1	1	1	1	1	1
2	2	2	3	3	5
6	9	0	1	0	5
2	5	0	0	0	0
*	*	*	*	*	*
1	1	1	1	1	1
7	7	7	8	8	9
2	5	7	2	9	4
0	0	0	0	0	0
*	*	*	*	*	*
2	2	2	2	2	2
2	2	4	5	5	7
7	7	4	0	6	3
0	5	0	0	0	0
*	*	*	*	*	*

Bebês que sobreviveram

mínimo= 1130g

$Q_1 = 1720\text{g}$

$Q_2 = 2200\text{g}$

$Q_3 = 2830\text{g}$

máximo= 3640g

Bebês que morreram

mínimo= 1030g

$Q_1 = 1230\text{g}$

$Q_2 = 1600\text{g}$

$Q_3 = 2200\text{g}$

máximo=2730g

BOX-PLOT

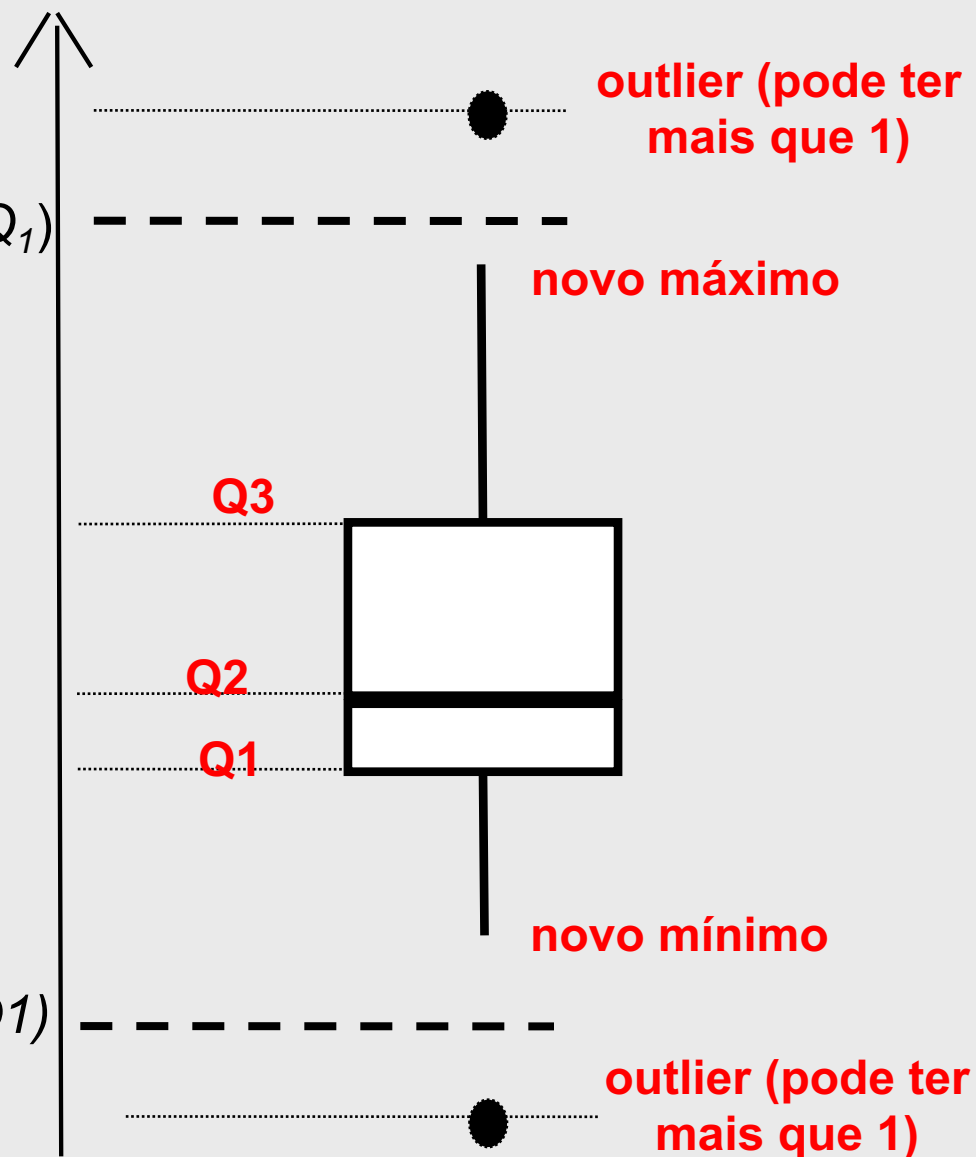
Sessões para estudo no livro texto: 3.3, 3.4

Fazer exercícios correspondentes do Cap. 3

BOX-PLOT

$$\text{Limite superior} = Q_3 + 1,5(Q_3 - Q_1)$$

$$\text{Limite Inferior} = Q_1 - 1,5(Q_3 - Q_1)$$



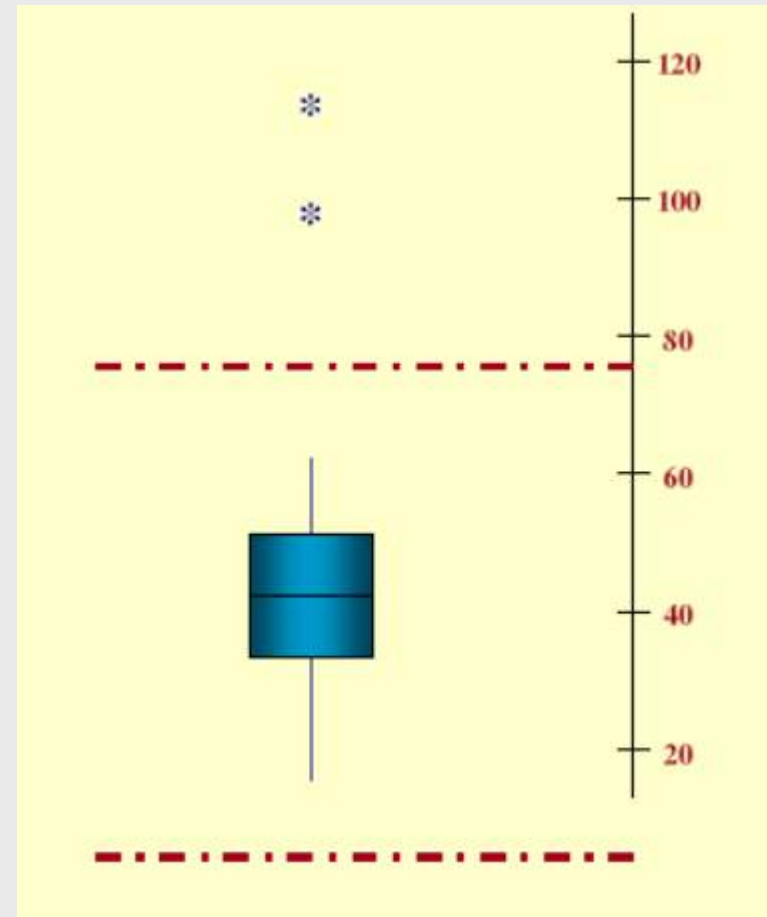
Exemplo: bebês em UTI

Tempo de permanência em UTI de
bebes nascidos prematuros (dias)

18	21	21	23	23	25
27	29	30	31	32	32
32	34	35	36	38	41
42	42	43	44	45	46
46	47	48	50	54	56
57	58	60	61	98	116

$$LI = Q1 - 1,5(Q3 - Q1) = 1,38$$

$$LS = Q3 + 1,5(Q3 - Q1) = 78,38$$



Exemplo: bebês em UTI

Bebês que sobreviveram

mínimo= 1130g

$Q_1 = 1720\text{g}$

$Q_2 = 2200\text{g}$

$Q_3 = 2830\text{g}$

máximo= 3640g

Bebês que morreram

mínimo= 1030g

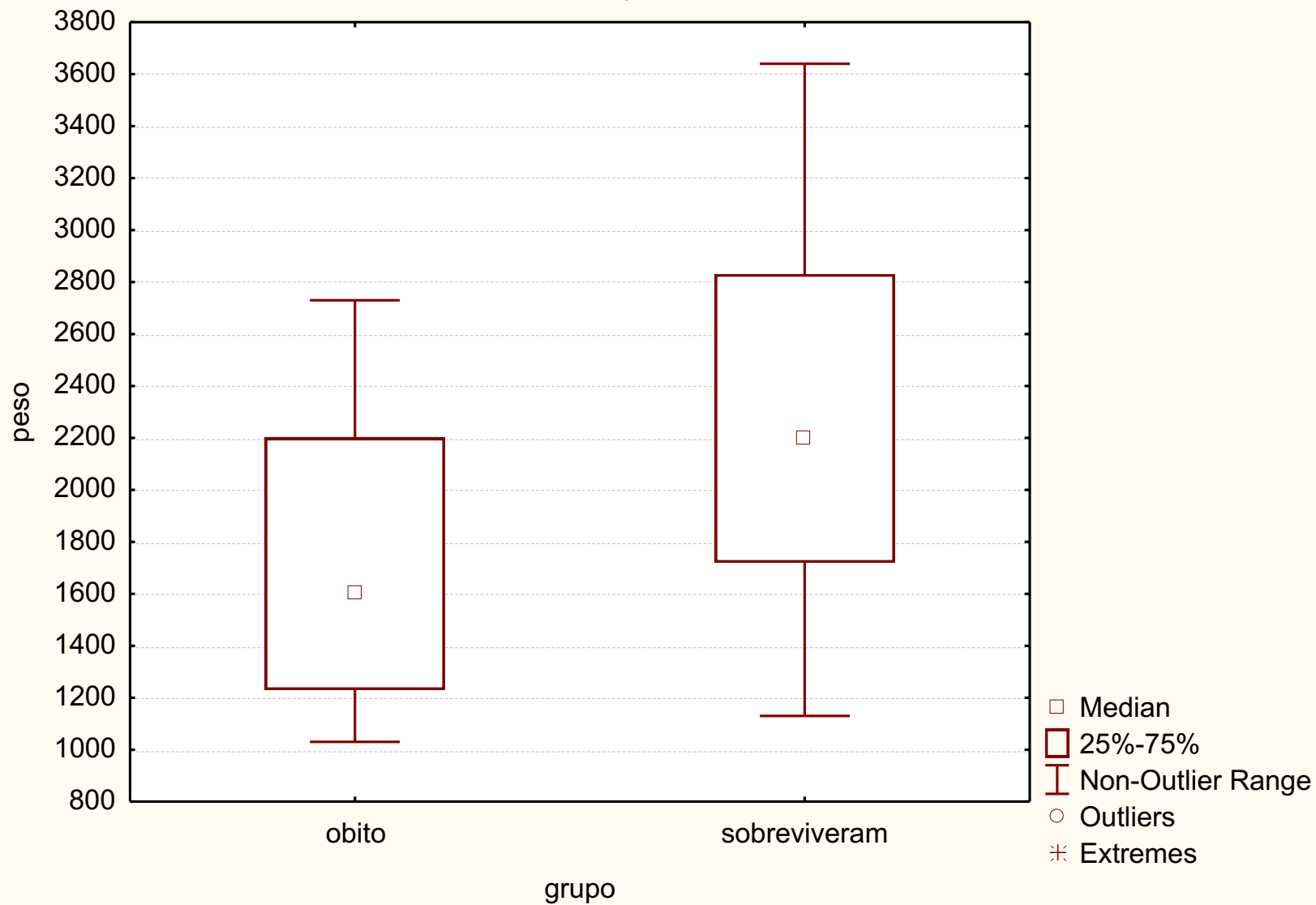
$Q_1 = 1230\text{g}$

$Q_2 = 1600\text{g}$

$Q_3 = 2200\text{g}$

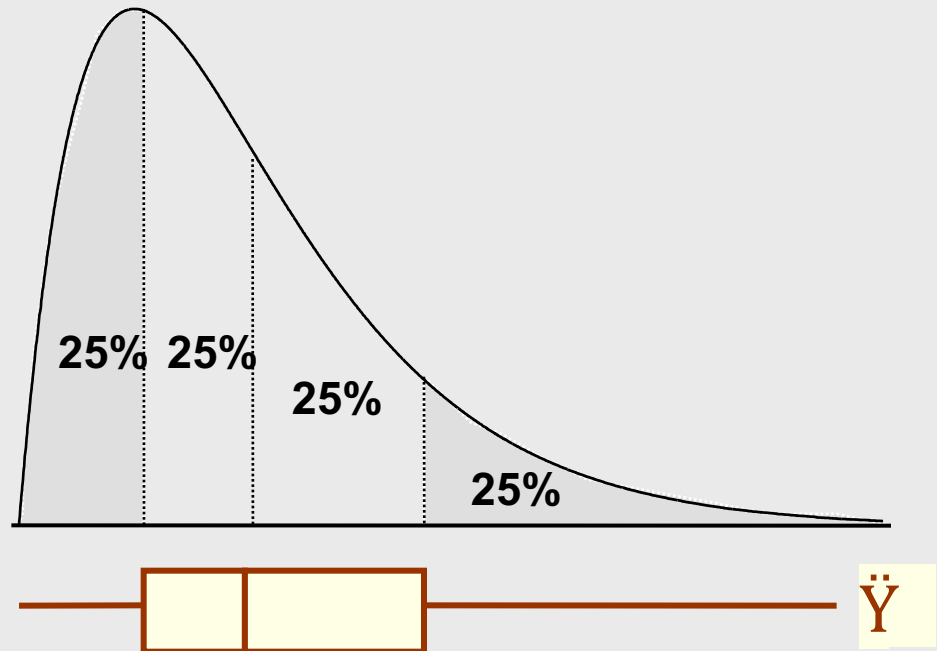
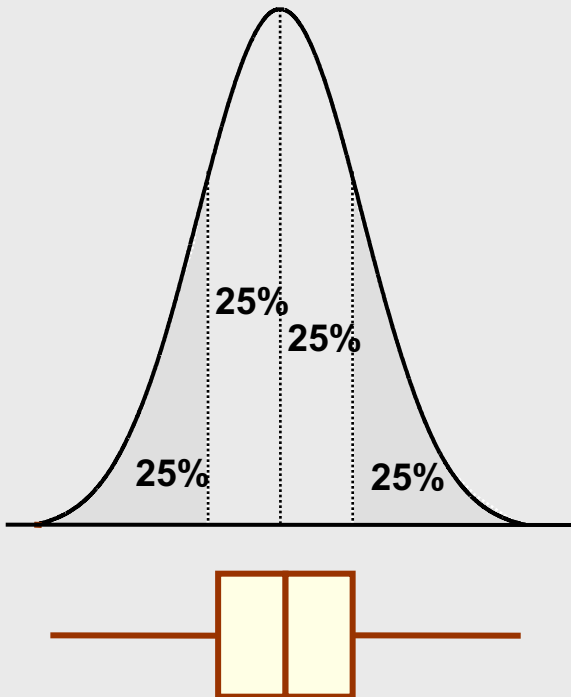
máximo=2730g

BOXPLOT
Pesos separados por grupo

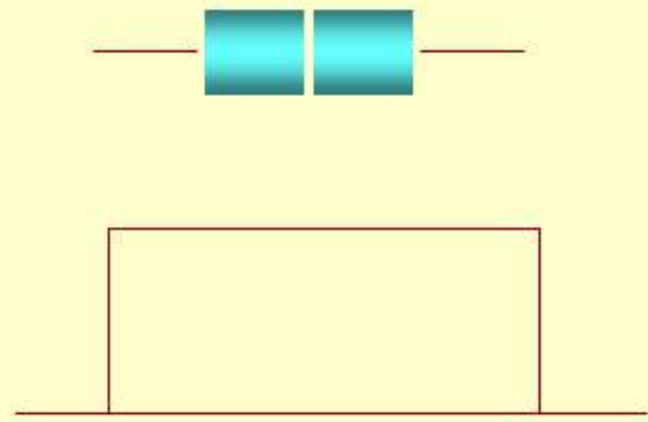
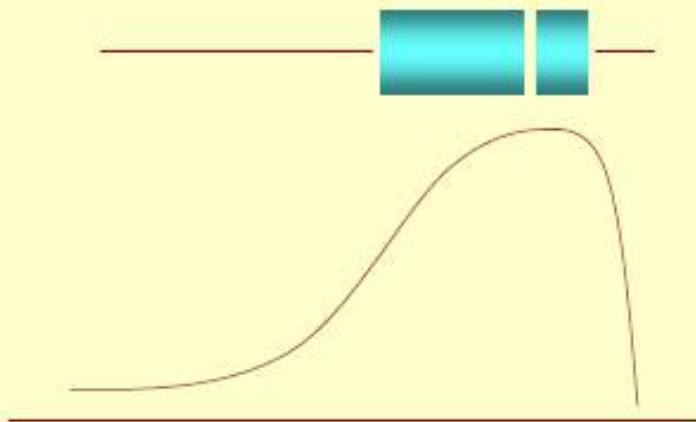
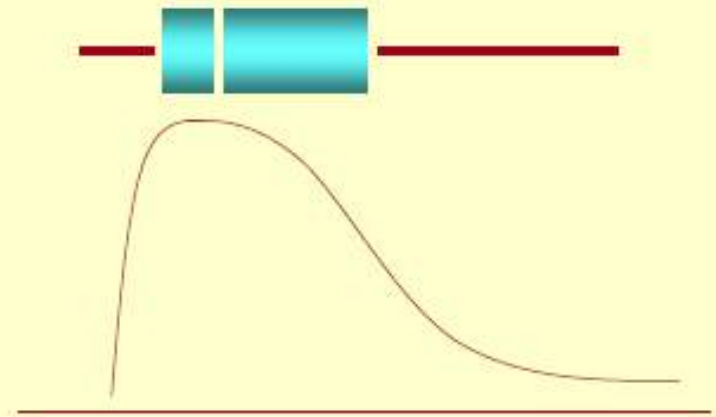
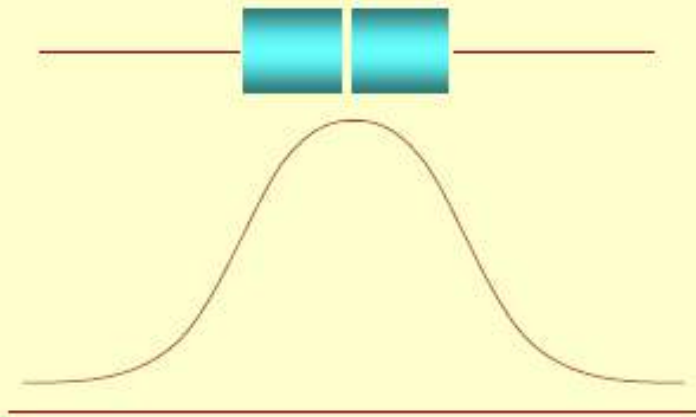


BOX-PLOT

- Veja o box-plot típico para diferentes distribuições de dados

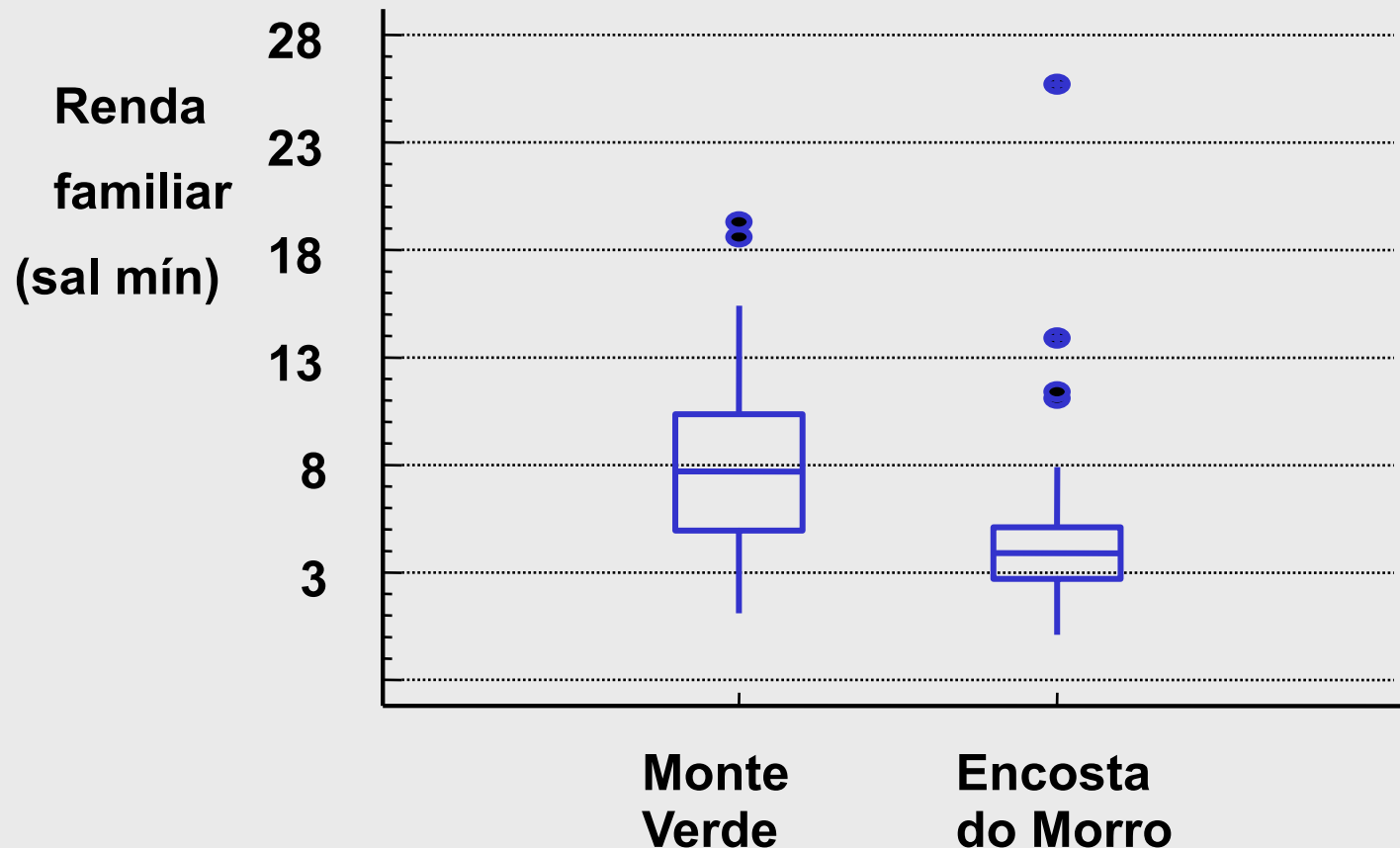


Forma da Distribuição



Comparando box-plots

- Comparando duas distribuições



Quanto ganha quem tem curso superior?

Entrevistamos 15 pessoas com curso superior que informaram seus salários (milhares de R\$/mês):

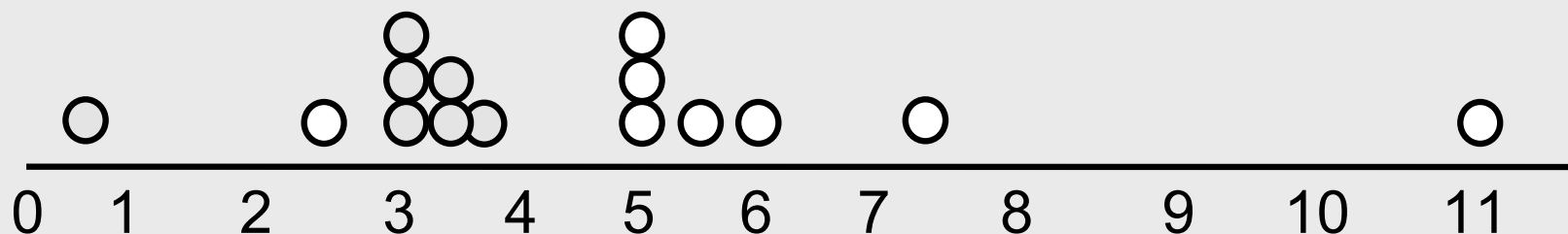
11	2,5	5,0	5,0	5,5	3,0	3,5	3,0	0,4	3,2	5,0	3,0	3,2	7,4	6,0
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

colocando em ordem

0,4	2,5	3,0	3,0	3,0	3,2	3,2	3,5	5,0	5,0	5,0	5,5	6,0	7,4	11
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

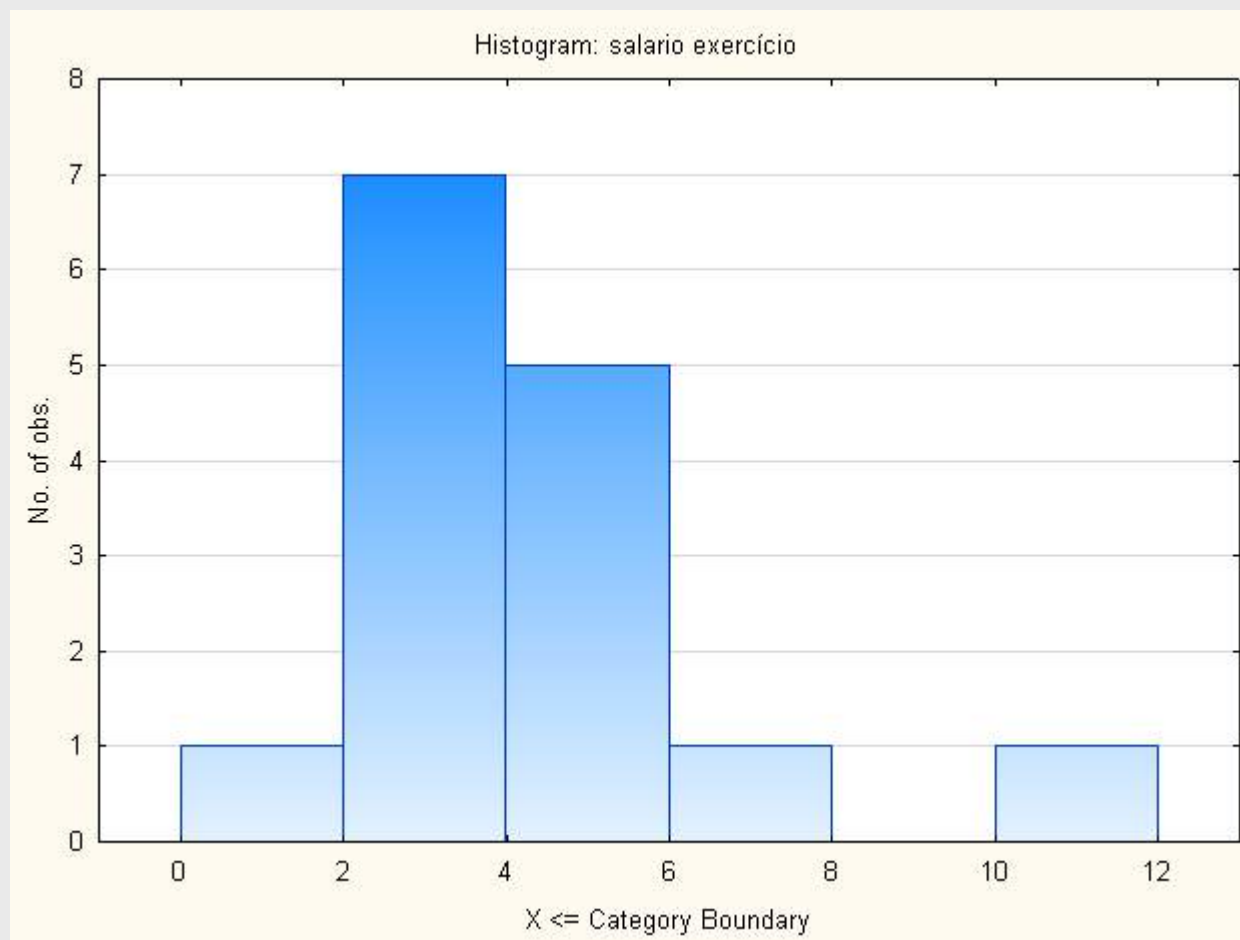
Quanto ganha quem tem curso superior?

Gráfico de pontos dos salários das 15 pessoas com curso superior (milhares de R\$/mês):



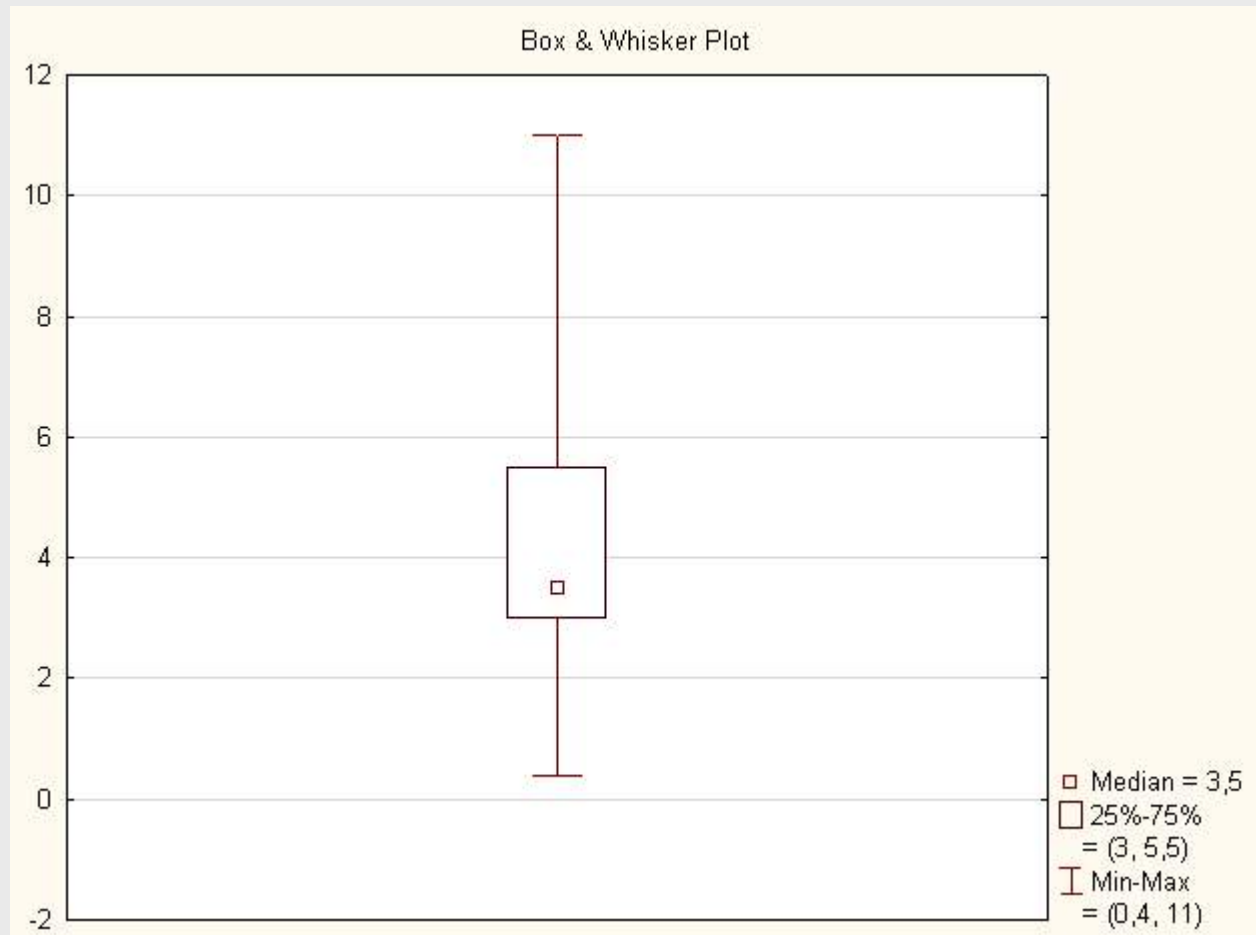
0,4	2,5	3,0	3,0	3,0	3,2	3,2	3,5	5,0	5,0	5,0	5,5	6,0	7,4	11
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Quanto ganha quem tem curso superior?



Quanto ganha quem tem curso superior?

- Box-plot



Box-plot usando o RStudio

- Exemplo dos salários

- Comandos no R:

```
> ganha <- c(0.4,2.5,3.0,3.0,3.0,3.2,  
3.2,3.5,5.0,5.0,5.0,5.5,6,7.4,11)
```

```
> ganha
```

```
[1] 0.4 2.5 3.0 3.0 3.0 3.2 3.2  
3.5 5.0 5.0
```

```
[11] 5.0 5.5 6.0 7.4 11.0
```

```
> boxplot(ganha)
```

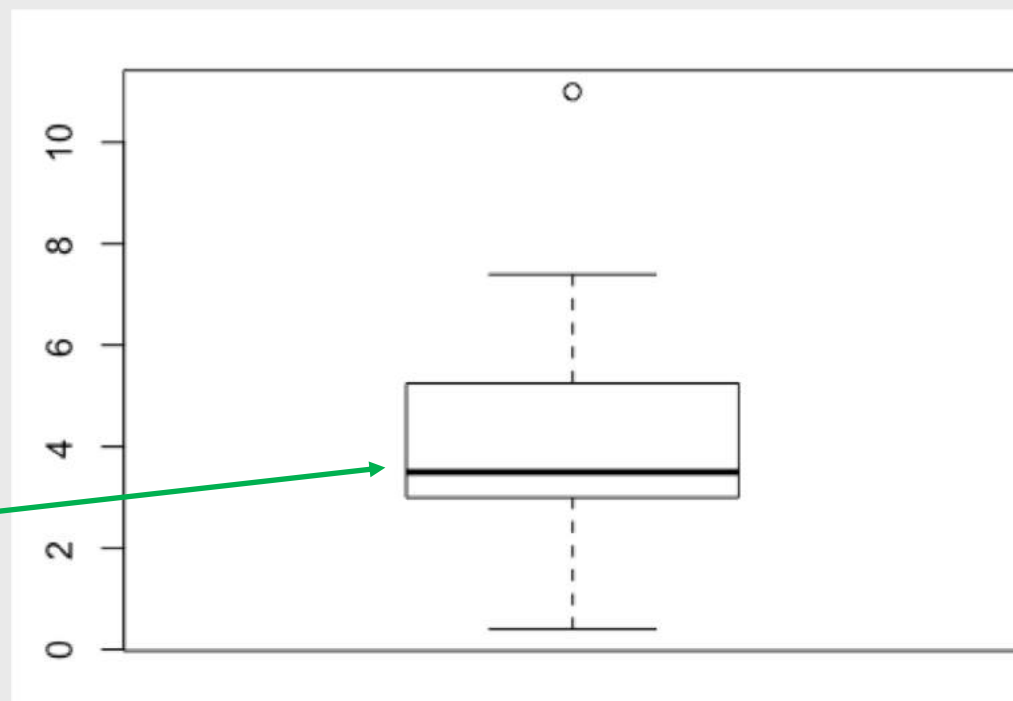
```
> median(ganha)
```

```
[1] 3.5
```

```
> quantile(ganha)
```

```
0% 25% 50% 75% 100%  
0.40 3.00 3.50 5.25 11.00
```

Fabricando o vetor dos dados ("ganha")



Mediana (Q2)

Quartis: min(0%) Q1 Q2 Q3 max (100%)

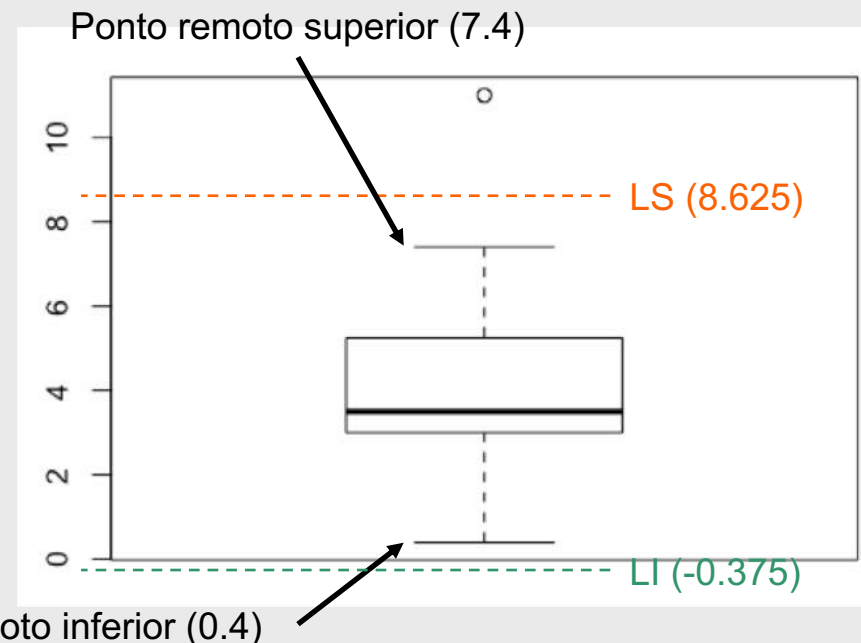
Pontos remotos e outliers

- Note que, a partir do retângulo, para cima, segue uma linha até o **ponto mais remoto** que não exceda o *limite superior* $LS = q3 + (1,5)(q3 - q1)$.
- No exemplo dos salários, $LS = 5.25 + (1,5)(5.25 - 3.0) = 8.625$. O **ponto mais remoto** $< LS = 7.4$
- De modo similar, da parte inferior do retângulo, para baixo, segue uma linha até o **ponto mais remoto** que não seja menor do que o *limite inferior* $LI = q1 - (1,5)(q3 - q1)$.
- No exemplo, $LI = 3.0 - (1,5)(5.25 - 3.0) = -0.375$. O **ponto mais remoto** $> LI = 0.4$ (min. 0%)

Box-plot do exemplo dos salários:

Quantis:

min(0%)	q1(25%)	q2(50%)	q3(75%)	max(100%)
0.40	3.00	3.50	5.25	11.0



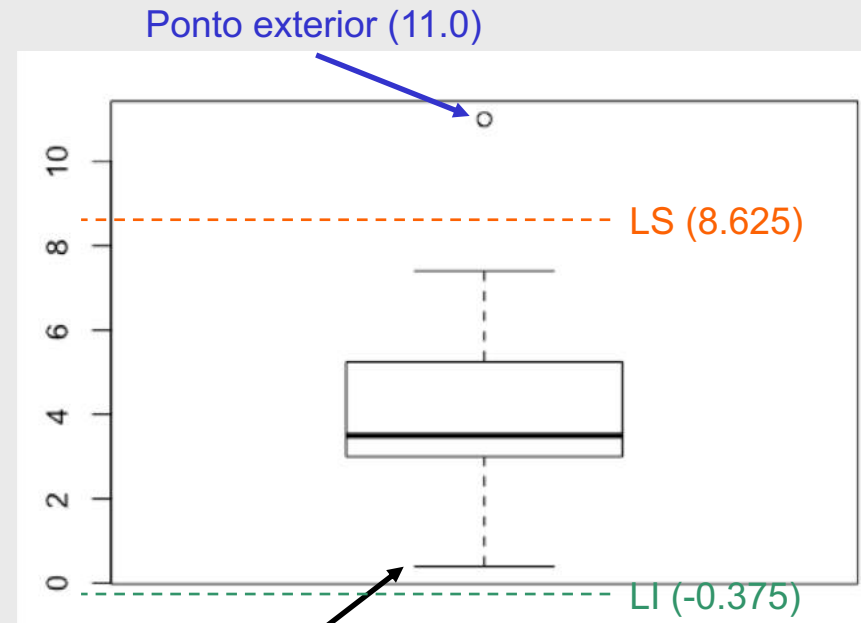
Pontos remotos e outliers

- Os valores compreendidos entre esses dois limites (isto é, entre os pontos remotos superior e inferior) são chamados *valores adjacentes*.
- As observações que estiverem acima do limite superior ou abaixo do limite inferior estabelecidos serão chamadas *pontos exteriores* e representadas por * (asteriscos) ou "o".
- Essas são observações destoantes das demais e podem ou não ser o que chamamos de *outliers* (*valores atípicos*).

Box-plot do exemplo dos salários:

Quantis:

min(0%)	q1(25%)	q2(50%)	q3(75%)	max(100%)
0.40	3.00	3.50	5.25	11.0



Ponto remoto inferior (0.4)

Exercício

- As rendas mensais em reais dos 25 ingressantes num certo curso de pós-graduação em finanças de uma universidade foram as seguintes:

2200, 2200, 2200, 2200, 2300, 2300, 2400, 2400,
2400, 2400, 2500, 2500, 2600, 2600, 2600, 2700,
2700, 2800, 3500, 3600, 3600, 4000, 4200, 4400,
5400.

- Pede-se: Para estes dados: calcule a média, o desvio padrão; determine: quartis (min, q1, q2, q3, max), e construa o “boxplot”.