

## DESCRIÇÃO DE UMA NUMÉRICA, SEGUNDO UMA CATEGÓRICA

<b>1. Antes da Análise: pergunta e dados .....</b>	<b>2</b>
<i>Pergunta de investigação</i> .....	2
<i>Dados</i> .....	2
<b>2. Dados 'crus' (desagregados) .....</b>	<b>3</b>
<b>3. Análise .....</b>	<b>4</b>
<i>Boxplot, por categorias</i> .....	4
<b>4. Teste Estatístico .....</b>	<b>5</b>
<b>5. EXERCÍCIOS .....</b>	<b>7</b>
<b>6. Referencias .....</b>	<b>8</b>
<b>APÊNDICE (R) .....</b>	<b>9</b>

Nesta secção exploramos a relação entre uma variável independente categórica e uma variável dependente numérica.

C -> N

Como veremos, o procedimento consiste em focar na variável dependente aplicando as mesmas ferramentas da análise univariada, mas dividendo a amostra segundo os diferentes grupos da variável categórica.

## 1. Antes da Análise: pergunta e dados

### *Pergunta de investigação*

*As ministras mulheres duram menos que os ministros homem?*

A pergunta inclui 1 variável independente categórica (variável sexo, com 2 categorias) e uma variável dependente numérica (número de dias no cargo de ministro/a).

### *Dados*

Continuamos a trabalhar com a base “BaseTutorial\_1”, agora focando-nos nas variáveis “sexo” e “duração”.

Controlamos a classe:

```
```\n\nclass(B$sexo)\n```\n[1] "factor"
```

Revisamos as categorias:

```
```\n\ntable(B$sexo)\n```\n\nfemale    male\n      8      8
```

*A variável regista 8 casos do sexo feminino e 8 do sexo masculino.*

Controlamos pela existência de valores em falta:

```
```\n\ntable (is.na(B$sexo))\n```\n\nFALSE
```

## 2. Dados 'crus' (desagregados)

Description: df [16 × 2]

sexo <fctr>	duração <dbl>
female	200
male	240
male	140
female	120
female	150
male	230
female	190
male	140
male	80
male	210
female	200
female	80
male	80
male	110
female	190
female	110

16 rows

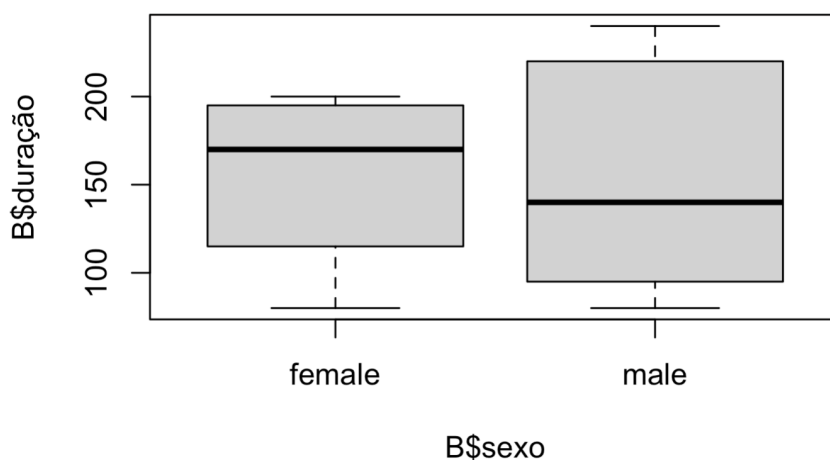
*Os dados aparecem por sujeito. No primeiro caso temos uma ministra, do sexo feminino, com uma duração em funções de 200 dias, enquanto o indivíduo 2 é do sexo masculino e tem uma duração em funções de 240 dias.*

### 3. Análise

#### *Boxplot, por categorias*

A caixa de bigodes (ou “boxplot”) permite-nos realizar uma primeira exploração visual sobre os dados, obtendo informação sobre a mediana amostral, os quartis, o valor mínimo e o valor máximo (i.e., o valor mais reduzido e o mais elevado da amostra), medidas de dispersão e a existência de valores atípicos. Neste caso, obtemos esta informação para cada sexo separadamente.

```
```\npar (mfrow =c(1,1))\nboxplot(B$duração~B$sexo)\n```\n
```



Sumário por categorias :

```
```\nsummary (B$duração[B$sexo=="female"])\nsummary (B$duração[B$sexo=="male"])\n```\n
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
80.0	117.5	170.0	155.0	192.5	200.0
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
80.0	102.5	140.0	153.8	215.0	240.0

*Ao interpretar a primeira linha do output, percebemos que o valor médio de dias no cargo de ministro/a é de 155 no caso das ministras mulheres e de 153,8 no caso dos ministros homens. Vemos ainda que o valor mínimo (i.e., o valor mais reduzido) é de 80 dias para ambos os sexos. Já o valor máximo (i.e., o valor mais elevado da amostra) é de 200 dias no caso das mulheres e de 240 no caso dos homens. Além disso, podemos observar os valores do Quartil 1, da mediana (i.e., Quartil 2) e do Quartil 3, que nos informam sobre que valores ocupam a posição localizada nos 25%, 50% e 75% da amostra ordenada. Neste caso, vemos que a mediana para as ministras mulheres é de 170, ou seja, após ordenar os valores da subamostra feminina o valor que ocupa a posição central é de 170 dias no cargo. Este valor para os ministros homens é de 140 dias. O primeiro quartil para as ministras mulheres é de 117,5, ou seja, pelo menos 25% das ministras mulheres durou até 117,5 dias em funções. No caso dos ministros homens, 25% durou até 102,5 dias. Por fim, o terceiro quartil mostra-nos que 75% das mulheres durou até 192,5 dias no cargo, enquanto 75% dos ministros homens durou até 215 dias.*

## 4. Teste Estatístico

```
```\nt.test(B$duração~B$sexo, B, alternative = "greater") #\n"two.sided", "less"\n```\n
```

Welch Two Sample t-test

data: B\$duração by B\$sexo

t = 0.044102, df = 12.734, p-value = 0.4828

alternative hypothesis: true difference in means between group female and group male is greater than 0

95 percent confidence interval:

-49.0247      Inf

sample estimates:

mean in group female	mean in group male
155.00	153.75

Neste teste estatístico, estamos a comparar a duração média de homens e de mulheres no cargo de ministro/a. O teste- $t$  ajuda-nos a testar duas hipóteses: a hipótese nula ( $H_0$ ), que diz que não há uma diferença significativa entre as médias dos dois grupos (i.e., a diferença de médias é igual a zero), contra a hipótese alternativa ( $H_a$ ), que diz que existe uma diferença entre as médias dos dois grupos (i.e., a diferença de médias não é igual a zero).

O valor- $p$  associado ao teste é de 0,4828, pelo que não podemos rejeitar a hipótese nula ( $H_0$ ) que diz que não há diferença entre as médias dos dois grupos, visto que o valor- $p$  é superior ao valor de significância estatística comumente utilizado ( $p = 0,05$ ). Isto significa que não foi detetada uma diferença significativa entre as médias dos dois grupos.

Concluimos que os ministros homens e as ministras mulheres duram, em média, o mesmo tempo em funções.

## 5. EXERCÍCIOS

### Secção PANORAMA

- 1) Identificar e escolher 5 artigos vinculados a um tópico do seu interesse, que utilizem análise estatística.
- 2) Para cada um dos 5 artigos escolhidos, identificar que tipo de análise aplicam (descritivo, explicativo) e que técnicas e ferramentas (coeficientes, regressões, boxplots, etc.).
- 3) Para 2 dos 5 artigos escolhidos,
  - a) identificar e descrever a base de dados utilizada (unidade de observação, número de observações, variáveis) e
  - b) reconstruir o codebook incluindo pelo menos 5 variáveis.

### Secção UMA QUANTITATIVA

- 4) Para 1 dos 5 artigos escolhidos, identificar a descrição de uma variável numérica. Transcrever essa descrição.

### Secção UMA QUALITATIVA

- 5) Para 1 dos 5 artigos escolhidos, identificar a descrição de uma variável qualitativa. Transcrever essa descrição.

### Secção ANALISE BIVARIADA

- 6) Para 1 dos 5 artigos escolhidos, identifique a variável independente e a variável dependente. Classifique-as (e.g., categórica, numérica).
- 7) Agora escolha um artigo com uma variável independente categórica e uma variável dependente numérica. Defina os grupos da variável categórica.
- 8) Faça uma caixa de bigodes (ou “boxplot”) e, de seguida, interprete o gráfico.
- 9) Por fim, peça um sumário por categorias e calcule o teste estatístico de comparação de médias.

**6. Referencias**

Openintro  
Bryman  
xxxx



## **APÊNDICE (R)**

[Ver markdown]