

Estatística Descritiva II

Ciências Contábeis - FEA - Noturno

1º Semestre 2025

Profs. Leonardo T. Rolla e Nikolai Kolev

(baseado em material previamente desenvolvido
pelo Prof. Gilberto Alvarenga Paula – CC BY-SA)

- 1 Informações sobre esta unidade
- 2 Dados agrupados
- 3 Boxplot
- 4 Boxplot comparado
- 5 Exercícios feitos em classe

Compreender e aplicar técnicas de organização e representação de dados estatísticos, incluindo tabelas de frequência, histogramas e boxplots, para facilitar a interpretação e comparação de distribuições.

Depois desta unidade, o aluno será capaz de:

- 1 Construir histogramas a partir de tabelas de frequência.
- 2 Estimar quartis da população a partir de tabelas de frequência.
- 3 Estimar média e variância a partir de tabelas de frequência.
- 4 Construir e interpretar boxplots.
- 5 Comparar distribuições de diferentes grupos por meio do boxplot.

- 1 Interpretação de tabelas de frequência
- 2 Histogramas: definição, construção e análise
- 3 Cálculo aproximado de medidas resumo a partir do histograma
- 4 Boxplot: conceito, componentes e interpretação
- 5 Boxplot comparado: análise e comparação de distribuições

O aluno pode fazer os exercícios postados no Moodle e posteriormente comparar suas respostas com as soluções apresentadas, identificando possíveis erros, e eventualmente refazer os exercícios.

Além disso, é recomendável refletir sobre as dificuldades encontradas e revisar os conceitos conforme necessário.

- 1 Informações sobre esta unidade
- 2 Dados agrupados**
- 3 Boxplot
- 4 Boxplot comparado
- 5 Exercícios feitos em classe

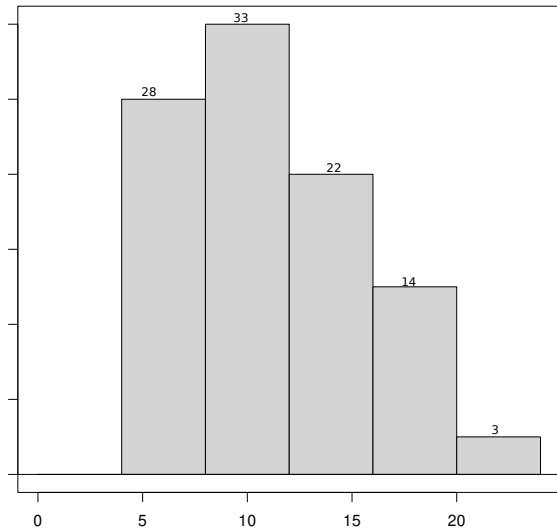
Salário	n_i	f_i (%)
[4, 8)	10	28
[8, 12)	12	33
[12, 16)	8	22
[16, 20)	5	14
[20, 24)	1	3
Total	36	100

Tabela: Salários de 36 empregados de um dado departamento de uma dada empresa, medidos em múltiplos do salário mínimo

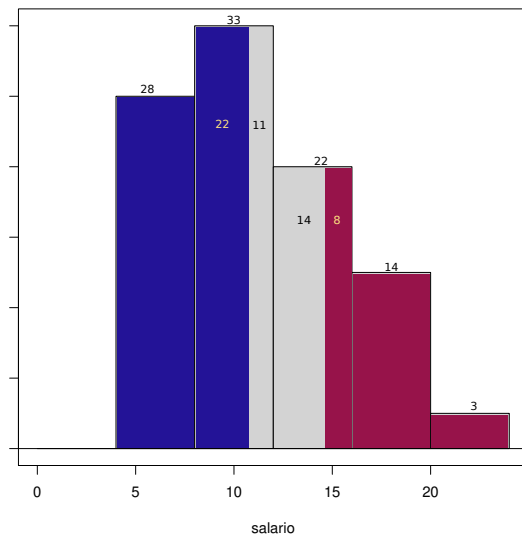
- Gráfico por setores contíguos, onde a altura é proporcional a f_i e a base é constituída pelos segmentos.
- Com pequeno número de classes, perde-se muita informação.
- Com grande número de classes, o objetivo de resumir os dados fica prejudicado.

Histograma

Salário	n_i	f_i (%)
[4, 8)	10	28
[8, 12)	12	33
[12, 16)	8	22
[16, 20)	5	14
[20, 24)	1	3
Total	36	100



Cálculo aproximado de quartis



Cálculo aproximado dos quartis

Salário	f_i (%)	F_i (%)
[4, 8)	28	28
[8, 12)	33	61
[12, 16)	22	83
[16, 20)	14	97
[20, 24)	3	100

$$Q1 = 4 + \frac{25}{28}(8 - 4) = 7,57$$

$$Q2 = 8 + \frac{50 - 28}{61 - 28}(12 - 8) = 10,67$$

$$Q3 = 12 + \frac{75 - 61}{83 - 61}(16 - 12) = 14,55$$

Média Aproximada

No caso de dados agrupados podemos obter uma média aproximada (ponderada)

$$\begin{aligned}\bar{x}^* &= \frac{n_1 x_1^* + \cdots + n_k x_k^*}{n} \\ &= f_1 x_1^* + \cdots + f_k x_k^* \\ \bar{x}^* &= \sum_{i=1}^k f_i x_i^*,\end{aligned}$$

em que k é o número de classes, n_i é o número de indivíduos da i -ésima classe, x_i^* é o ponto médio da i -ésima classe e $f_i = \frac{n_i}{n}$.

Considerando o exemplo dos salários

Tabela: Frequências dos intervalos dos salários.

Salário	Freq. rel f_i	Ponto médio x_i^*	Produto $f_i \cdot x_i^*$
[4, 8)	0.28	6	1.68
[8, 12)	0.33	10	3.30
[12, 16)	0.22	14	3.08
[16, 20)	0.14	18	2.52
[20, 24)	0.03	22	0.66
Total	1	—	11.24

Ou seja,

$$\bar{x}^* = 11.24.$$

Variância Aproximada

No caso de dados agrupados podemos obter uma variância aproximada (ponderada)

$$(s^*)^2 = \sum_{i=1}^k f_i \cdot (x_i^* - \bar{x}^*)^2.$$

Fórmula alternativa

$$(s^*)^2 = (\sum_{i=1}^k f_i \cdot (x_i^*)^2) - (\bar{x}^*)^2.$$

Desvio Padrão Aproximado

O desvio padrão aproximado fica dado por

$$s^* = \sqrt{(s^*)^2}.$$

Considere novamente o exemplo dos salários

Tabela: Frequências dos intervalos dos salários.

Salário	Freq. rel f_i	Ponto médio x_i^*	Produto $f_i \cdot (x_i^*)^2$
[4, 8)	0.28	6	10.08
[8, 12)	0.33	10	33.00
[12, 16)	0.22	14	43.12
[16, 20)	0.14	18	45.36
[20, 24)	0.03	22	14.52
Total	1	—	146.08

Ou seja,

$$s^{*2} = 146.08 - 11.24^2 = 19.74 \rightarrow s^* = 4.443.$$

- 1 Informações sobre esta unidade
- 2 Dados agrupados
- 3 Boxplot**
- 4 Boxplot comparado
- 5 Exercícios feitos em classe

É um tipo de gráfico para descrever a distribuição dos dados que é construído com os quartis $Q1$ e $Q3$, a mediana e valores extremos.

Supondo um conjunto de dados com as seguintes características:

máximo = 22

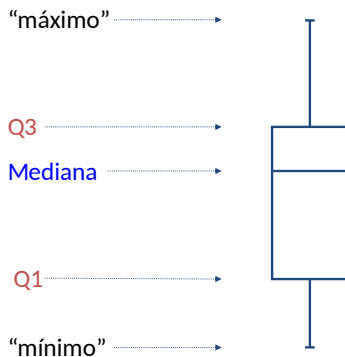
Q3 = 14,55

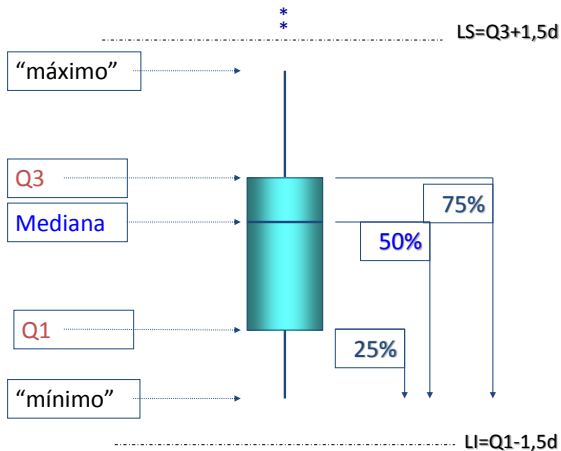
Q2 = 10,67

Q1 = 7,57

mínimo = 4

Como fazer um desenho que descreva todas essas características ao mesmo tempo?





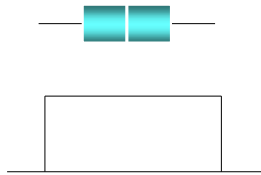
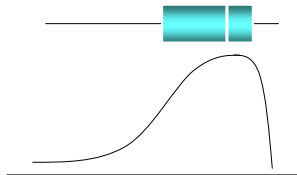
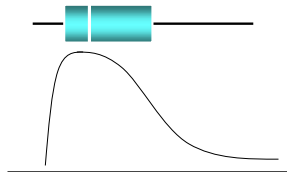
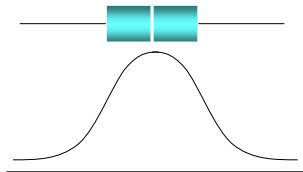
"máximo" é o maior valor menor do que LS

"mínimo" é o menor valor maior do que LI

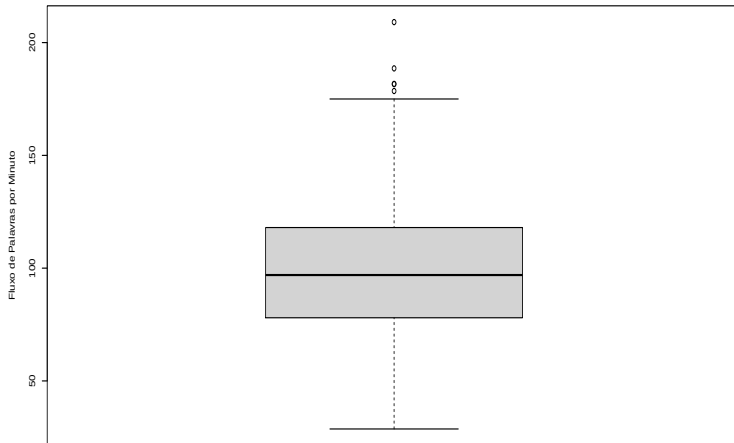
Construção

- o corpo do gráfico consiste de uma caixa que é construída do quartil Q1 até o quartil Q3. Altura da caixa: $d = Q3 - Q1$
- traça-se dentro da caixa uma linha que corresponde à mediana
- traçam-se duas linhas horizontais auxiliares, respectivamente LI e LS, a alturas $LI = Q1 - 1,5d$ e $LS = Q3 + 1,5d$
- linhas verticais são traçadas da caixa até o maior ou menor valor antes da linha horizontal correspondente
- observações que estão acima ou abaixo das linhas horizontais são destacadas

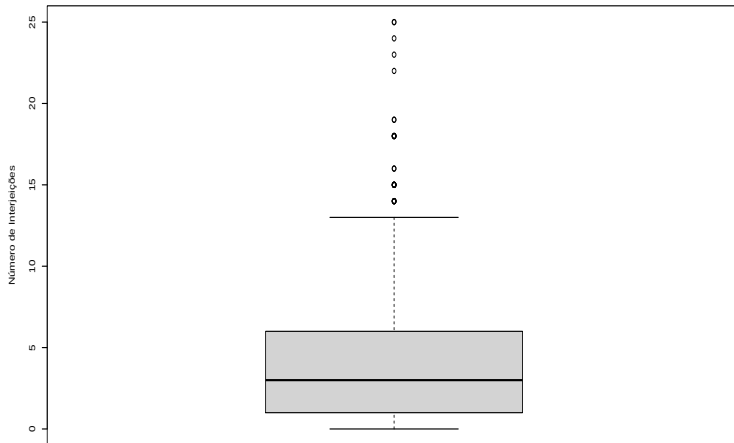
Formas da Distribuição



Exemplo de Boxplot Simétrico: FPM



Exemplo de Boxplot Assimétrico: INTERJ



Previsão PIB de 2018

Supor que $n = 9$ Economistas brasileiros, escolhidos aleatoriamente, foram consultados a respeito do crescimento (em %) previsto para o PIB brasileiro em 2018.

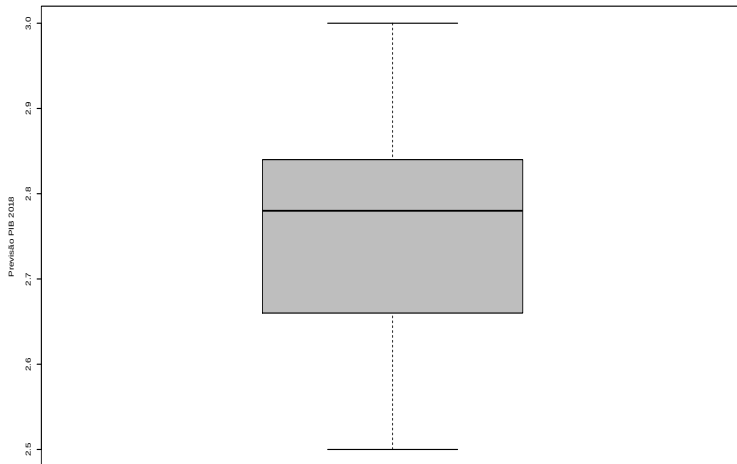
Dados: 2,50; 2,60; 2,66; 2,70; 2,78; 2,80; 2,84; 2,90; 3,00.

Boxplot

Com base nessa amostra, podemos construir o boxplot.

Usar $Q1=2,63$, $Q2=2,78$ e $Q3=2,87$.

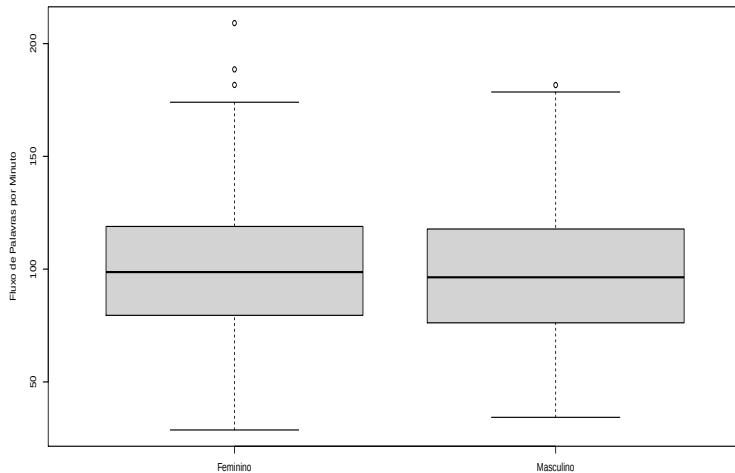
Boxplot Simétrico: PIB 2018



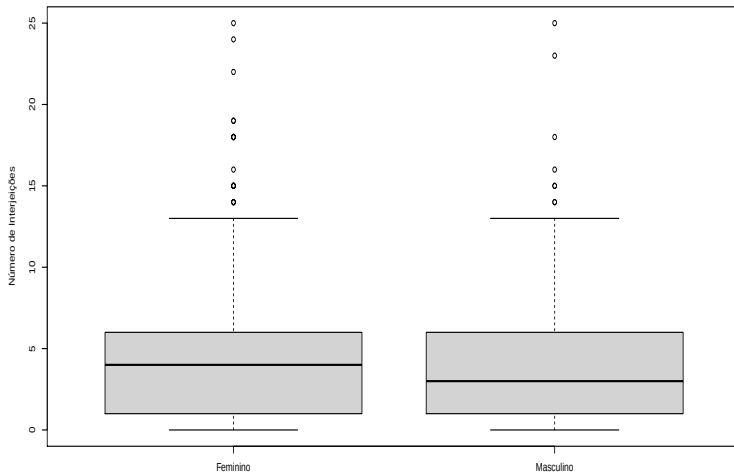
- 1 Informações sobre esta unidade
- 2 Dados agrupados
- 3 Boxplot
- 4 Boxplot comparado**
- 5 Exercícios feitos em classe

Para analisar a associação entre **uma variável qualitativa e uma quantitativa**, podemos usar gráficos de barras lado a lado, comparar as medidas resumo, ou, para ter uma comparação mais visual, exibir os respectivos boxplots lado a lado.

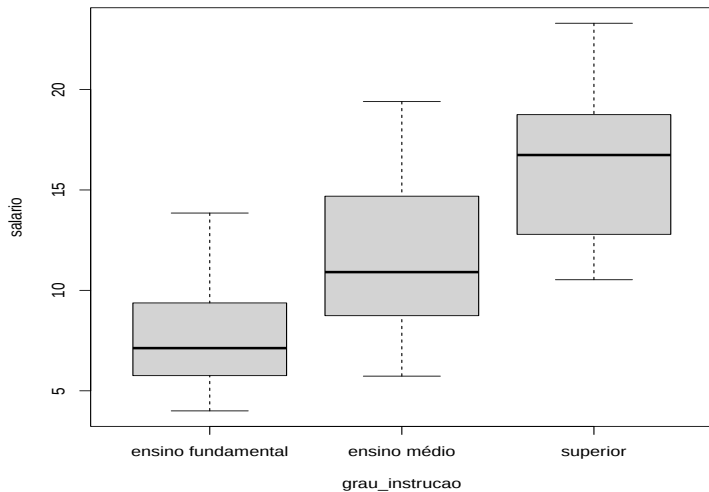
FPM por Gênero



INTERJ por Gênero



Salário por grau de instrução



Comente sobre a diferença na distribuição dos salários segundo o grau de instrução.

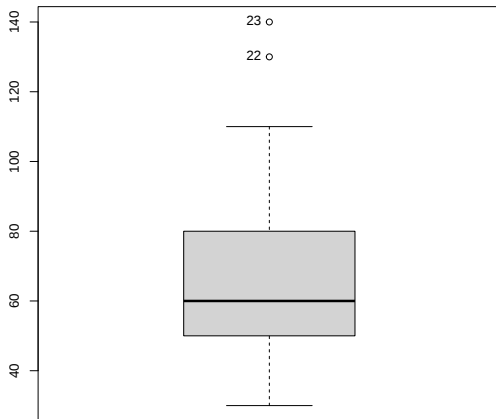
- 1 Informações sobre esta unidade
- 2 Dados agrupados
- 3 Boxplot
- 4 Boxplot comparado
- 5 Exercícios feitos em classe**

Desenhe o Boxplot para os dados abaixo:

2,50; 2,60; 2,66; 2,70; 2,78; 2,80; 2,84; 2,90; 3,00

Resposta: $Q1 = 2,63$; $Q2 = 2,78$; $Q3 = 2,87$

Invente uma base de dados que tenha o seguinte boxplot:



Resposta: 30 31 32 33 34 50 50 55 55 55 55 60 65 65 65 65 80 80 80 90 110 130 140

A distribuição de frequências absolutas, a seguir, mostra os valores de venda para uma amostra de 10 observações.

Tabela: Frequências dos intervalos dos valores de venda.

Vendas	Ponto médio x_i^*	Freq. abs n_i	Freq. rel f_i
[20, 40)	30	1	0.1
[40, 60)	50	2	0.2
[60, 80)	70	4	0.4
[80, 100)	90	2	0.2
[100, 120)	110	1	0.1
Total	—	10	1

Determine a média, o desvio padrão e o coeficiente de variação da amostra.

Resposta: 70; 21,91; 31,3%

Considerando a tabela abaixo:

Altura	f_i (%)
150 ┤ 160	11
160 ┤ 170	21
170 ┤ 180	30
180 ┤ 190	18
190 ┤ 200	13
200 ┤ 210	7
Total	100

Tabela: Altura dos alunos homens de uma faculdade, em cm.

- 1 Faça à mão o gráfico do histograma.
- 2 Estime Q1, Q2 e Q3.

Resposta: Q1 = 166,7; Q2 = 176; Q3 = 187,2