

# Fundamentos de Análise Exploratória de Dados

## Conceitos e Aplicações

Encontro 2  
Exercícios

Prof. Me. Lineu Alberto Cavazani de Freitas



# Exercício 1

## Exercício 1

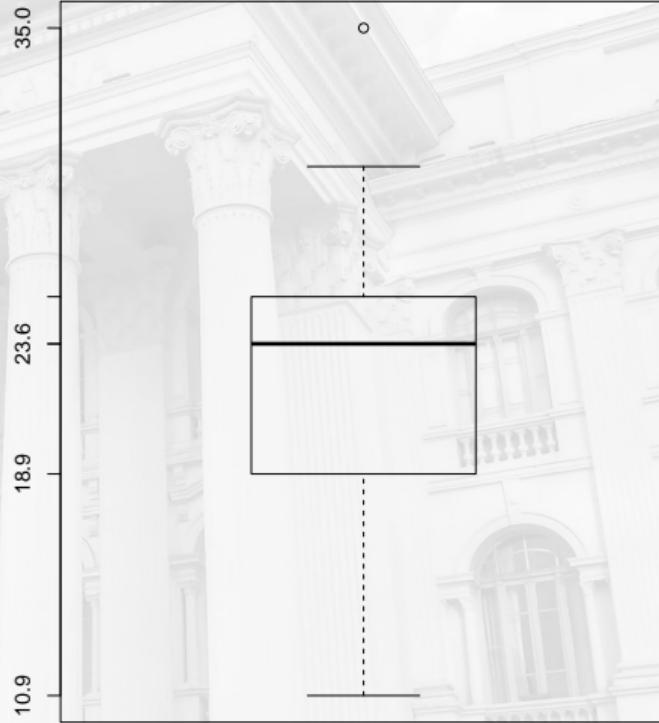
O Índice de Massa Corporal (IMC) é um cálculo que faz uso do peso e da altura de um indivíduo que permite avaliar se este indivíduo está ou não com o peso “ideal”. Para fazer o cálculo, basta dividir o peso pela altura ao quadrado. Uma escola tinha como objetivo avaliar como o IMC dos alunos de uma turma se comportava. Os dados estão na tabela:

18.9	22.9	17.8	30.0	23.6
17.9	24.4	25.7	24.9	20.5
29.6	23.9	18.9	10.9	35.0

- ▶ Obtenha média, mediana, mínimo, máximo, quartis, amplitude interquartílica e esboce o boxplot para os dados fornecidos.
- ▶ Obtenha amplitude, desvio absoluto médio com relação à média e mediana, variância, desvio padrão, coeficiente de variação e z-escore.

# Exercício 1

Média	22.99
Mediana	23.60
Mínimo	10.90
Máximo	35.00
1º Quartil	18.90
2º Quartil	23.60
3º Quartil	25.30
AIQ	6.40



## Exercício 1

i	IMC	Desv. abs. média	Desv. abs. mediana	Desv. quadrático	z-escore
1	10.9	12.09	12.7	146.25	-0.6919
2	17.8	5.19	5.8	26.97	-0.0158
3	17.9	5.09	5.7	25.94	-0.8779
4	18.9	4.09	4.7	16.76	1.1844
5	18.9	4.09	4.7	16.76	0.1026
6	20.5	2.49	3.1	6.22	-0.8610
7	22.9	0.09	0.7	0.01	0.2378
8	23.6	0.61	0.0	0.37	0.4575
9	23.9	0.91	0.3	0.82	0.3223
10	24.4	1.41	0.8	1.98	-0.4215
11	24.9	1.91	1.3	3.64	1.1168
12	25.7	2.71	2.1	7.33	0.1533
13	29.6	6.61	6.0	43.65	-0.6919
14	30.0	7.01	6.4	49.09	-2.0443
15	35.0	12.01	11.4	144.16	2.0296

# Exercício 1

---

Amplitude	24.10
Desvio absoluto médio (média)	4.42
Desvio absoluto médio (mediana)	4.38
Variância	34.99
Desvio padrão	5.92
Coeficiente de variação	25.73

---

## Exercício 2

## Exercício 2

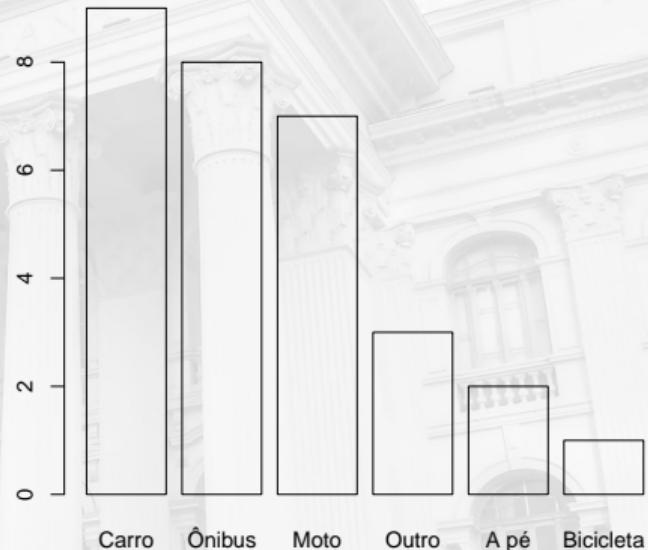
Uma pesquisa tinha como objetivo verificar qual o meio de transporte mais comum entre os alunos de uma turma. As respostas foram:

Carro	Ônibus	Ônibus	Outro	Carro	A pé	Outro	Moto	Moto	Carro
Carro	Carro	Moto	Ônibus	Moto	Ônibus	Moto	Bicicleta	Ônibus	Moto
Outro	Carro	Moto	Carro	Carro	Ônibus	Carro	Ônibus	A pé	Ônibus

Obtenha uma medida de variabilidade adequada.

## Exercício 2

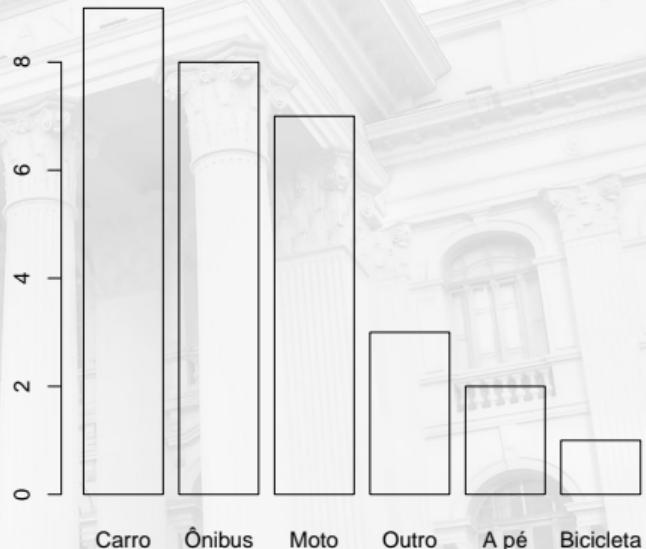
Respostas	$f_a$	$f_r$
Carro	9	0.300
Ônibus	8	0.267
Moto	7	0.233
Outro	3	0.100
A pé	2	0.067
Bicicleta	1	0.033
Total	29	0.967



## Exercício 2

Respostas	$f_a$	$f_r$
Carro	9	0.300
Ônibus	8	0.267
Moto	7	0.233
Outro	3	0.100
A pé	2	0.067
Bicicleta	1	0.033
Total	29	0.967

$$\underline{H = 1.58}$$



# Exercício 3

## Exercício 3

Em um estudo, uma universidade selecionou uma amostra de 10 alunos pertencentes a 2 turmas. Destes alunos, registrou-se se tinha ou não feito um curso pré vestibular e qual a nota obtida nas provas de português e matemática. Os dados coletados estão na tabela abaixo.

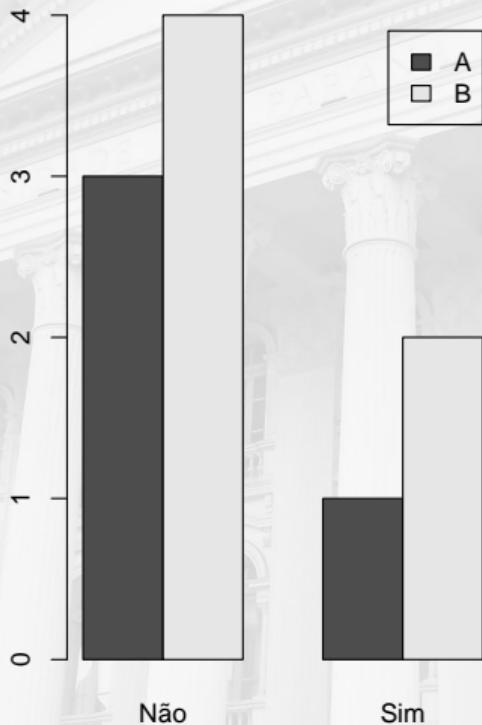
Aluno	Turma	Curso	Português	Matemática
1	A	Não	4	4
2	B	Não	3	3
3	B	Sim	4	5
4	A	Não	7	1
5	B	Não	6	5
6	B	Não	5	4
7	B	Não	3	9
8	A	Sim	4	9
9	A	Não	10	6
10	B	Sim	7	3

## Exercício 3

1. Para turma e curso, obtenha tabelas de dupla entrada usando frequência absoluta e relativa (total, por linha e por coluna) e esboce gráficos adequados para representar as tabelas.
2. Obtenha uma medida de associação para turma e curso.
3. Para as notas, obtenha o coeficiente de correlação de Pearson e o diagrama de dispersão.
4. Obtenha medidas descritivas e box-plots das notas em função do curso e da turma.

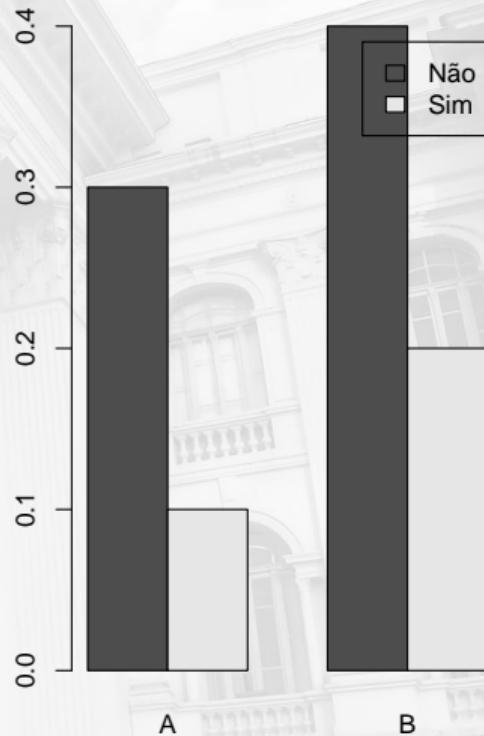
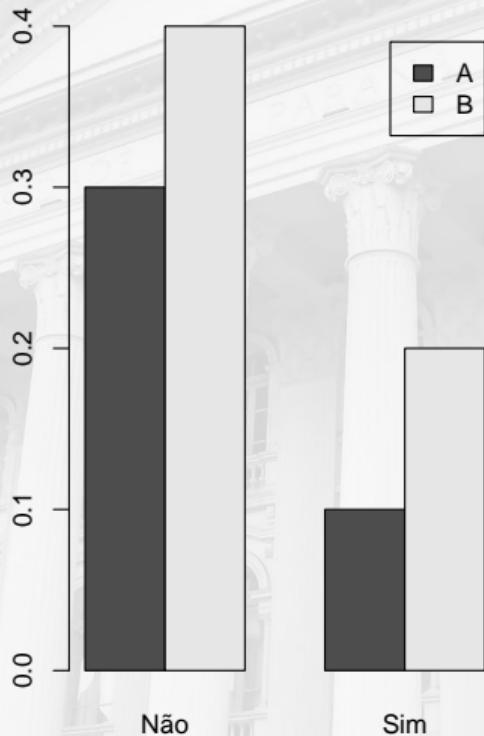
# 1 Tabela de dupla entrada com frequências absolutas

	Não	Sim
A	3	1
B	4	2



# 1 Tabela de dupla entrada com frequências relativas (total geral)

	Não	Sim
A	0.3	0.1
B	0.4	0.2



# 1 Tabela de dupla entrada com frequências relativas (total linha)

	Não	Sim	Sum
A	0.75	0.25	1
B	0.67	0.33	1
Sum	1.42	0.58	2



# 1 Tabela de dupla entrada com frequências relativas (total coluna)

	Não	Sim	Sum
A	0.43	0.33	0.76
B	0.57	0.67	1.24
Sum	1.00	1.00	2.00



## 2 Qui-quadrado para associação

Tabela 14. Valores observados.

	Não	Sim	Total
A	3	1	4
B	4	2	6
Total	7	3	10

Tabela 15. Valores esperados.

	Não	Sim	Total
A	2.8	1.2	4
B	4.2	1.8	6
Total	7.0	3.0	10

Tabela 16.  $\frac{(o-e)^2}{e}$ .

	Não	Sim
A	0.01	0.03
B	0.01	0.02

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \approx 0$$

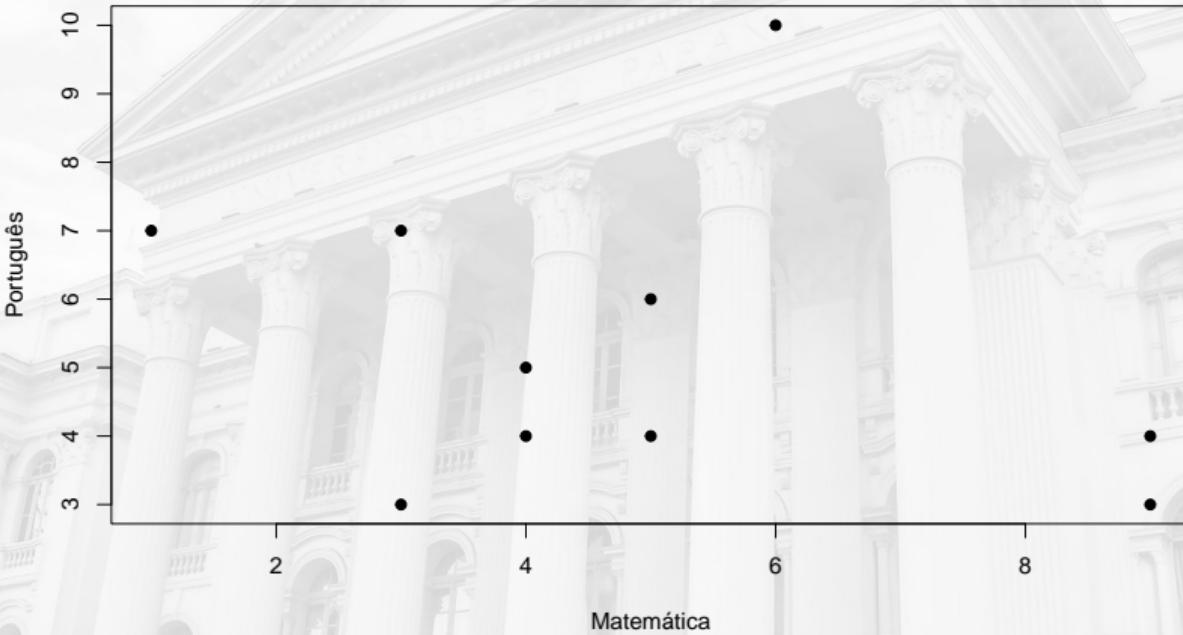
### 3 Correlação

PT	$PT - \bar{PT}$	$(PT - \bar{PT})^2$	MT	$MT - \bar{MT}$	$(MT - \bar{MT})^2$	$(PT - \bar{PT}) \times (MT - \bar{MT})$
4	-1.3	1.69	4	-0.9	0.81	1.17
3	-2.3	5.29	3	-1.9	3.61	4.37
4	-1.3	1.69	5	0.1	0.01	-0.13
7	1.7	2.89	1	-3.9	15.21	-6.63
6	0.7	0.49	5	0.1	0.01	0.07
5	-0.3	0.09	4	-0.9	0.81	0.27
3	-2.3	5.29	9	4.1	16.81	-9.43
4	-1.3	1.69	9	4.1	16.81	-5.33
10	4.7	22.09	6	1.1	1.21	5.17
7	1.7	2.89	3	-1.9	3.61	-3.23

---

$$\bar{PT} = 5.30$$
$$\bar{MT} = 4.90$$
$$V(PT) = 4.90$$
$$V(MT) = 6.54$$
$$COV(PT, MT) = -1.52$$
$$COR(PT, MT) = -0.27$$

### 3 Diagrama de dispersão



## 4 Medidas descritivas por turma e curso

Tabela 19. Notas em português em função do turma. Tabela 21. Notas em português em função do curso.

Turma	Média	Mediana	Desvio padrão
A	6.25	5.5	2.87
B	4.67	4.5	1.63

Curso	Média	Mediana	Desvio padrão
Não	5.43	5	2.51
Sim	5.00	4	1.73

Tabela 20. Notas em matemática em função do turma.Tabela 22. Notas em matemática em função do curso.

Turma	Média	Mediana	Desvio padrão
A	5.00	5.0	3.37
B	4.83	4.5	2.23

Curso	Média	Mediana	Desvio padrão
Não	4.57	4	2.51
Sim	5.67	5	3.06

# 4 box-plots das notas em função de turma e curso

