



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA

**Alunos**

Gleyce Alves Pereira da Silva  
Ivanildo Batista da Silva Júnior  
Jaine de Moura Carvalho  
Taciana Araújo da Silva

**Professor**

Dr. Lucian Bogdan Bejan

**Resolução da primeira lista de Estatística Aplicada**

Recife-PE, 31 de março de 2021

# Sumário

<b>1</b>	<b>Questão 1</b>	<b>1</b>
1.1	Resolução da questão 1	2
1.1.1	letra a)	2
1.1.2	letra b)	2
1.1.3	letra c)	2
1.1.4	letra d)	3
1.1.5	letra e)	4
1.1.6	letra f)	4
1.1.7	letra g)	5
<b>2</b>	<b>Questão 2</b>	<b>6</b>
2.1	Resolução da questão 2	6
2.1.1	letra a)	6
2.1.2	letra b)	6
<b>3</b>	<b>Questão 3</b>	<b>7</b>
3.1	Resolução da questão 3	8
3.1.1	letra a)	8
3.1.2	letra b)	8
3.1.3	letra c)	9
3.1.4	letra d)	10
<b>4</b>	<b>Questão 4</b>	<b>11</b>
4.1	Resolução da questão 4	12
4.1.1	letra a)	12
4.1.2	letra b)	13
4.1.3	letra c)	13
4.1.4	letra d)	13
4.1.5	letra e)	14
4.1.6	letra f)	14
<b>5</b>	<b>Questão 5</b>	<b>15</b>
5.1	Resolução da questão 5	15
5.1.1	letra a)	15
5.1.2	letra b)	16
5.1.3	letra c)	16
5.1.4	letra d)	16
<b>6</b>	<b>Questão 6</b>	<b>18</b>
6.1	Resolução da questão 6	18
6.1.1	letra a)	18
6.1.2	letra b)	18
<b>7</b>	<b>Questão 7</b>	<b>19</b>

7.1	Resolução da questão 7 . . . . .	19
7.1.1	letra a) . . . . .	19
7.1.2	letra b) . . . . .	20
7.1.3	letra c) . . . . .	20
<b>8</b>	<b>Questão 8 . . . . .</b>	<b>21</b>
8.1	Resolução da questão 8 . . . . .	21
<b>9</b>	<b>Questão 9 . . . . .</b>	<b>23</b>
9.1	Resolução da questão 9 . . . . .	23
<b>10</b>	<b>Questão 10 . . . . .</b>	<b>24</b>
10.1	Resolução da questão 10 . . . . .	25
10.1.1	letra a) . . . . .	25
10.1.2	letra b) . . . . .	25
10.1.3	letra c) . . . . .	25
10.1.4	letra d) . . . . .	26
10.1.5	letra e) . . . . .	26

# 1 Questão 1

(Cap. 2: ex. 9) A MB Indústria e Comércio, desejando melhorar o nível de seus funcionários em cargos de chefia, montou um curso experimental e indicou 25 funcionários para a primeira turma. Os dados referentes à seção a que pertencem, notas e graus obtidos no curso estão na tabela a seguir. Como havia dúvidas quanto à adoção de um único critério de avaliação, cada instrutor adotou seu próprio sistema de aferição. Usando dados daquela tabela, responda às questões:

- Após observar atentamente cada variável, e com o intuito de resumi-las, como você identificaria (qualitativa ordinal ou nominal e quantitativa discreta ou contínua) cada uma das 9 variáveis listadas?
- Compare e indique as diferenças existentes entre as distribuições das variáveis Direito, Política e Estatística.
- Construa o histograma para as notas da variável Redação.
- Construa a distribuição de frequências da variável Metodologia e faça um gráfico para indicar essa distribuição.
- Sorteado ao acaso um dos 25 funcionários, qual a probabilidade de que ele tenha obtido grau A em Metodologia?
- Se, em vez de um, sorteássemos dois, a probabilidade de que ambos tivessem tido A em Metodologia é maior ou menor do que a resposta dada em (e)?
- Como é o aproveitamento dos funcionários na disciplina Estatística, segundo a seção a que eles pertencem?

Func.	Seção (*)	Administr.	Direito	Redação	Estatíst.	Inglês	Metodologia	Política	Economia
1	P	8,0	9,0	8,6	9,0	B	A	9,0	8,5
2	P	8,0	9,0	7,0	9,0	B	C	6,5	8,0
3	P	8,0	9,0	8,0	8,0	D	B	9,0	8,5
4	P	6,0	9,0	8,6	8,0	D	C	6,0	8,5
5	P	8,0	9,0	8,0	9,0	A	A	6,5	9,0
6	P	8,0	9,0	8,5	10,0	B	A	6,5	9,5
7	P	8,0	9,0	8,2	8,0	D	C	9,0	7,0
8	T	10,0	9,0	7,5	8,0	B	C	6,0	8,5
9	T	8,0	9,0	9,4	9,0	B	B	10,0	8,0
10	T	10,0	9,0	7,9	8,0	B	C	9,0	7,5
11	T	8,0	9,0	8,6	10,0	C	B	10,0	8,5
12	T	8,0	9,0	8,3	7,0	D	B	6,5	8,0
13	T	6,0	9,0	7,0	7,0	B	C	6,0	8,5
14	T	10,0	9,0	8,6	9,0	A	B	10,0	7,5
15	V	8,0	9,0	8,6	9,0	C	B	10,0	7,0
16	V	8,0	9,0	9,5	7,0	A	A	9,0	7,5
17	V	8,0	9,0	6,3	8,0	D	C	10,0	7,5
18	V	6,0	9,0	7,6	9,0	C	C	6,0	8,5
19	V	6,0	9,0	6,8	4,0	D	C	6,0	9,5
20	V	6,0	9,0	7,5	7,0	C	B	6,0	8,5
21	V	8,0	9,0	7,7	7,0	D	B	6,5	8,0
22	V	6,0	9,0	8,7	8,0	C	A	6,0	9,0
23	V	8,0	9,0	7,3	10,0	C	C	9,0	7,0
24	V	8,0	9,0	8,5	9,0	A	A	6,5	9,0
25	V	8,0	9,0	7,0	9,0	B	A	9,0	8,5

(\*) (P = departamento pessoal, T = seção técnica e V = seção de vendas)

## 1.1 Resolução da questão 1

Questão resolvida em *Python*.

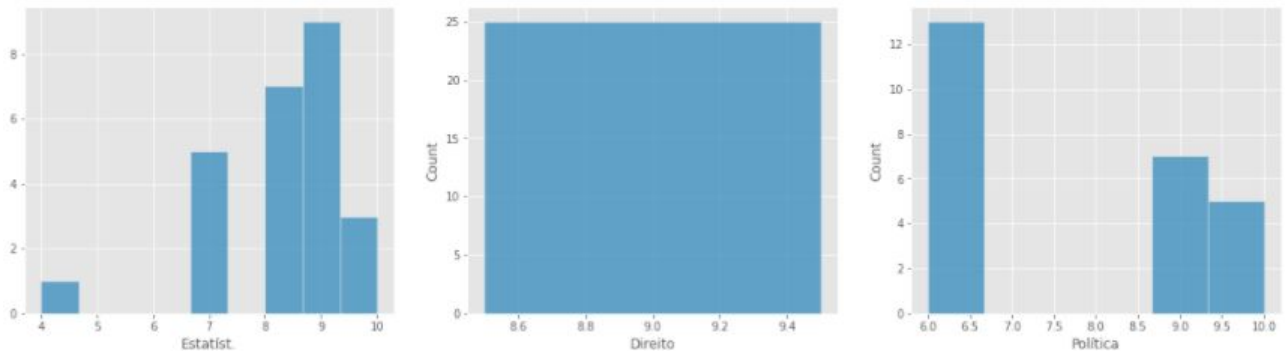
### 1.1.1 letra a)

- **Variáveis qualitativas ordinais:** Inglês e Metodologia;
- **Variável qualitativa nominal:** Seção;
- **Variáveis quantitativas contínuas:** Administração, Direito, Redação, Estatística, Política e Economia.

### 1.1.2 letra b)

```
fig,ax = plt.subplots(1,3, figsize=(20,5))
sns.histplot(x='Estatíst.', data=df5, ax=ax[0])
sns.histplot(x='Direito', data=df5, ax=ax[1])
sns.histplot(x='Política', data=df5, ax=ax[2]);
```

Nos histogramas gerados, respectivamente, de Estatística, Direito e Política, vemos que para a matéria de Estatística as notas estão mais concentradas entre em valores altos (entre 8 e 10), Direito não há variação na nota; e em Política, as notas estão mais distribuídas em notas medianas e notas altas.



### 1.1.3 letra c)

```
plt.figure(figsize=(20,5))
sns.histplot(x='Redação', data=df5)
plt.xlabel('Redação', size=15)
plt.ylabel('')
plt.title('Histograma das notas de Redação',size=15);
```

continuação letra c) : histograma da nota de redação.



#### 1.1.4 letra d)

```
[42] #tabela de frequência
freq = pd.DataFrame(dfs['Metodologia'].value_counts(sort=False))
#frequência acumulada
freq['Met_acumulada'] = freq['Metodologia'].cumsum()
#frequência relativa
freq['Met_relativa'] = freq['Metodologia']/freq['Metodologia'].sum()
#frequência relativa acumulada
freq['Met_rel_acumulada'] = freq['Met_relativa'].cumsum()

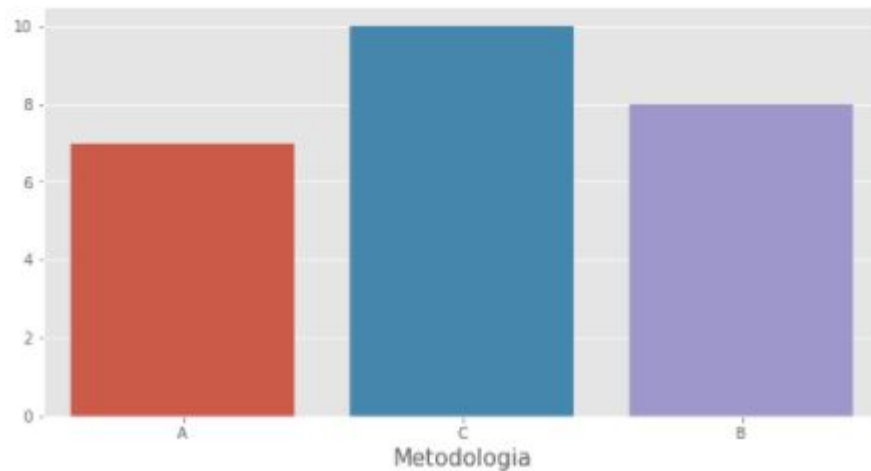
freq
```

Tabela de frequência gerado:

	Metodologia	Met_acumulada	Met_relativa	Met_rel_acumulada
A	7	7	0.28	0.28
C	10	17	0.40	0.68
B	8	25	0.32	1.00

```
plt.figure(figsize=(20,5))
sns.countplot(x='Metodologia', data=df5)
plt.xlabel('Metodologia', size=15)
plt.ylabel("");
```

Gráfico da distribuição:



### 1.1.5 letra e)

Calculando a probabilidade de que um dos 25 funcionários tenha obtido grau A em Metodologia é dado por  $\frac{7}{25} = 0.28$ , que calculei abaixo, em *Python*.

```
[48] print('A probabilidade de um dos 25 funcionários tenha obtido grau A em Metodologia é ',
        freq.T['A'][0]/freq.T['B'][1])
```

A probabilidade de um dos 25 funcionários tenha obtido grau A em Metodologia é 0.28

### 1.1.6 letra f)

Essa resposta depende se o sorteio é com ou sem reposição. Vamos calcular as duas formas:

**com reposição:** Vamos sortear o funcionário e em seguida recolocá-lo novamente entre os funcionários para realizar um novo sorteio. Então na primeira eu terei  $\frac{7}{25}$  e na segunda  $\frac{7}{25}$ , logo a probabilidade será  $\frac{7}{25} * \frac{7}{25}$ .

```
[50] print('A probabilidade de dois dos 25 funcionários tenha obtido grau A em Metodologia é ',
        round((freq.T['A'][0]/freq.T['B'][1])*(freq.T['A'][0]/freq.T['B'][1]),4))
```

A probabilidade de dois dos 25 funcionários tenha obtido grau A em Metodologia é 0.0784

**sem reposição:** Vamos sortear o funcionário, mas ele não será recolocado novamente entre os funcionários para realizar o novo sorteio. Então na primeira eu terei  $\frac{7}{25}$  e na segunda  $\frac{6}{24}$ , logo a probabilidade será  $\frac{7}{25} * \frac{6}{24}$ .  
Continuação letra f)

```
[51] print('A probabilidade de dois dos 25 funcionários tenha obtido grau A em Metodologia é ',
        round((freq.T['A'][0]/freq.T['B'][1])*((freq.T['A'][0]-1)/(freq.T['B'][1]-1)),4))
```

A probabilidade de dois dos 25 funcionários tenha obtido grau A em Metodologia é 0.07

### 1.1.7 letra g)

```
fig,ax = plt.subplots(1,3, figsize=(20,5))
ax[0].title.set_text('Histograma Estatísticas da Seção Pessoal')
ax[1].title.set_text('Histograma Estatísticas da Seção Técnica')
ax[2].title.set_text('Histograma Estatísticas da Seção Vendas')
sns.histplot(df5[df5['Seção (*)']=='P']['Estatíst.'], ax=ax[0], color='red')
sns.histplot(df5[df5['Seção (*)']=='T']['Estatíst.'], ax=ax[1])
sns.histplot(df5[df5['Seção (*)']=='V']['Estatíst.'], ax=ax[2], color='green');
```

Conforme gráficos gerados abaixo, vê-se que a seção de **Pessoal** possui um desempenho melhor em comparação as demais seções. A notas dessa seção estão mais concentradas em valores acima de 8, diferentes das outras seções.





## 2 Questão 2

(Cap. 2: ex. 11) Dispomos de uma relação de 200 aluguéis de imóveis urbanos e uma relação de 100 aluguéis rurais.

- Construa os histogramas das duas distribuições.
- Com base nos histogramas, discuta e compare as duas distribuições.

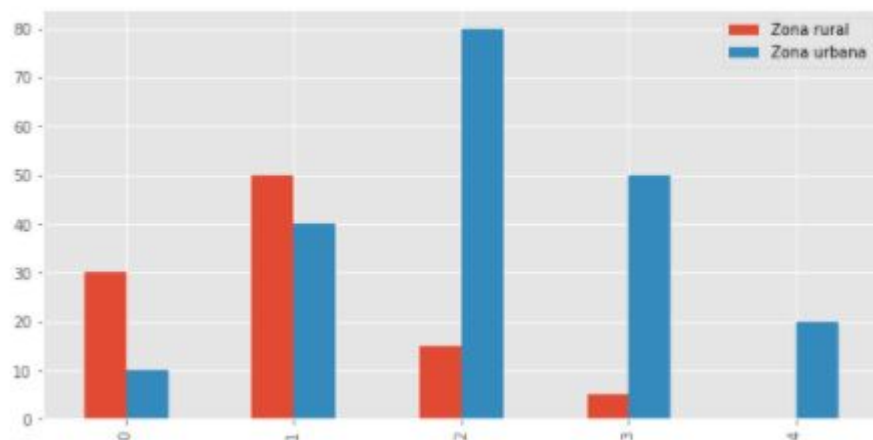
Classes de aluguéis (codificados)	Zona urbana	Zona rural
2├ 3	10	30
3├ 5	40	50
5├ 7	80	15
7├ 10	50	5
10├ 15	20	0
Total	200	100

### 2.1 Resolução da questão 2

Questão resolvida em *Python*.

#### 2.1.1 letra a)

```
df6[['Zona rural', 'Zona urbana']].plot.bar(figsize=(10,5));
```



#### 2.1.2 letra b)

Pelos histogramas vemos que os valores de aluguéis em zonas urbanas são maiores que os aluguéis de zonas rurais. O número de aluguéis na classe 5 - 7 é bem maior em zonas urbanas do que nas rurais.

### 3 Questão 3

Um artigo retirado da revista *Technometrics* (Vol. 19, 1977, p. 425) apresenta os seguintes dados sobre a taxa de octanagem de várias misturas de gasolina:

88,5	87,7	83,4	86,7	87,5	91,5	88,6	100,3	96,5	93,3	94,7
91,1	91,0	94,2	87,8	89,9	88,3	87,6	84,3	86,7	84,3	86,7
88,2	90,8	88,3	98,8	94,2	92,7	93,2	91,0	90,1	93,4	88,5
90,1	89,2	88,3	85,3	87,9	88,6	90,9	89,0	96,1	93,3	91,8
92,3	90,4	90,1	93,0	88,7	89,9	89,8	89,6	87,4	88,4	88,9
91,2	89,3	94,4	92,7	91,8	91,6	90,4	91,1	92,6	89,8	90,6
91,1	90,4	89,3	89,7	90,3	91,6	90,5	93,7	92,7	92,2	92,2
91,2	91,0	92,2	90,0	90,7						

- Construa o diagrama de folhas-e-ramos para esses dados
- Construa a distribuição de frequência e o histograma. Use 8 intervalos de classe.
- Construa a distribuição de frequência e o histograma, agora com 16 intervalos de classe.
- Compare a forma dos dois histogramas em b e c. Ambos os histogramas mostram informações similares?

### 3.1 Resolução da questão 3

Questão resolvida no *Excel*.

#### 3.1.1 letra a)

Diagrama de Ramos-e-folha gerado em *Excel*:

```

83 | 0,4
84 | 0,3 0,3
85 | 0,3
86 | 0,7 0,7 0,7
87 | 0,4 0,5 0,6 0,7 0,8 0,9
88 | 0,2 0,3 0,3 0,3 0,4 0,5 0,5 0,6 0,6 0,7 0,9
89 | 0 0,2 0,3 0,3 0,6 0,7 0,8 0,8 0,9 0,9
90 | 0 0,1 0,1 0,1 0,3 0,4 0,4 0,4 0,5 0,6 0,7 0,8 0,9
91 | 0 0 0 0,1 0,1 0,1 0,2 0,2 0,5 0,6 0,6 0,8 0,8
92 | 0,2 0,2 0,2 0,3 0,6 0,7 0,7 0,7
93 | 0 0,2 0,3 0,3 0,4 0,7
94 | 0,2 0,2 0,4 0,7
95 |
96 | 0,1 0,5
97 |
98 | 0,8
99 |
100 | 0,3

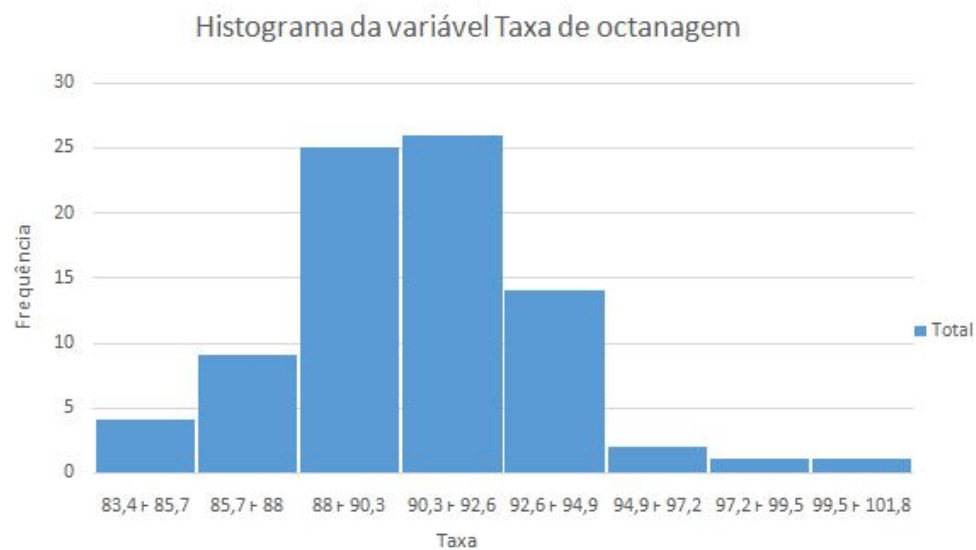
```

#### 3.1.2 letra b)

Tabela de frequência da variável Octanagem com 8 classes (feito em *Excel*):

Intervalo de taxa	Frequência Relativa
83,4 † 85,7	4
85,7 † 88	9
88 † 90,3	25
90,3 † 92,6	26
92,6 † 94,9	14
94,9 † 97,2	2
97,2 † 99,5	1
99,5 † 101,8	1
<b>Total</b>	<b>82</b>

Histograma da variável Octanagem com 8 classes (feito em *Excel*):

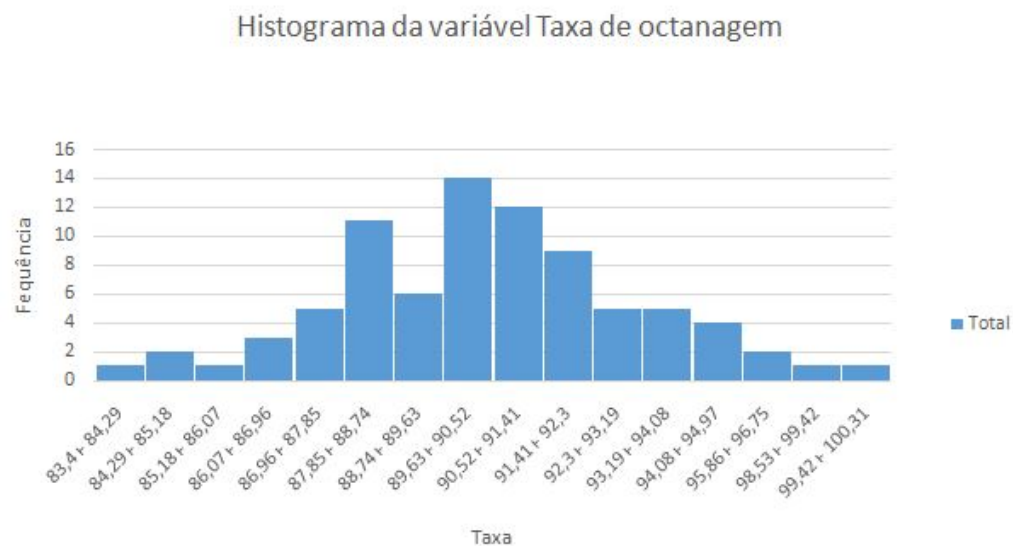


### 3.1.3 letra c)

Tabela de frequência da variável Octanagem com 16 classes (feito em *Excel*):

Intervalo de taxa	Frequência Relativa
83,4 + 84,29	1
84,29 + 85,18	2
85,18 + 86,07	1
86,07 + 86,96	3
86,96 + 87,85	5
87,85 + 88,74	11
88,74 + 89,63	6
89,63 + 90,52	14
90,52 + 91,41	12
91,41 + 92,3	9
92,3 + 93,19	5
93,19 + 94,08	5
94,08 + 94,97	4
95,86 + 96,75	2
98,53 + 99,42	1
99,42 + 100,31	1
<b>Total</b>	<b>82</b>

Histograma da variável Octanagem com 8 classes (feito em *Excel*):



#### 3.1.4 letra d)

Quando construído o histograma com 8 classes, observamos uma maior concentração de dados centralizados e quando analisamos o histograma com 16 classes, nota-se que com o aumento do número de classes temos um histograma mais irregular.

## 4 Questão 4

O seguinte conjunto de dados representa as “vidas” de 40 baterias de carro da mesma marca e mesmas características com aproximação até décimos do ano. As baterias tinham garantia para 3 anos.

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

- Construa a distribuição de frequência e o histograma;
- Faça o gráfico da distribuição de frequências relativas acumuladas.
- Calcule a média aritmética dos dados originais;
- Usando a distribuição de frequência conforme obtido em a calcule a média novamente. Para tal, considere os pontos médios de cada classe (média entre os dois limites de cada classe) para serem os valores da variável no cálculo da média.
- Obtenha a variância para os dados originais conforme feito para a média em c.
- Obtenha a variância a partir da distribuição de frequência conforme feito para a média no item d.

**Obs.: use 7 intervalos de classe; a amplitude da classe igual a 0,5; o início do intervalo mais baixo em 1,5.**

## 4.1 Resolução da questão 4

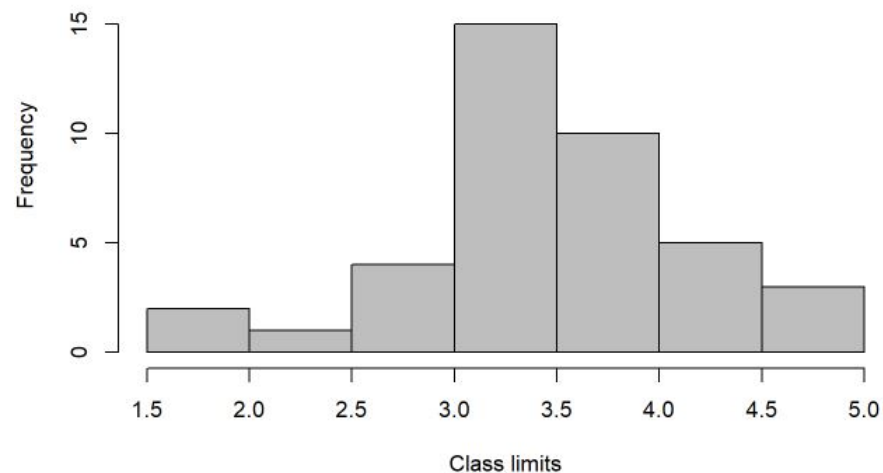
Questão resolvida em R.

### 4.1.1 letra a)

```
d <- fdt(x ,start = 1.5 , end = 5 , h = 0.5)
print(d ,format = TRUE, col = 1:4 , pattern = "%.2f")
```

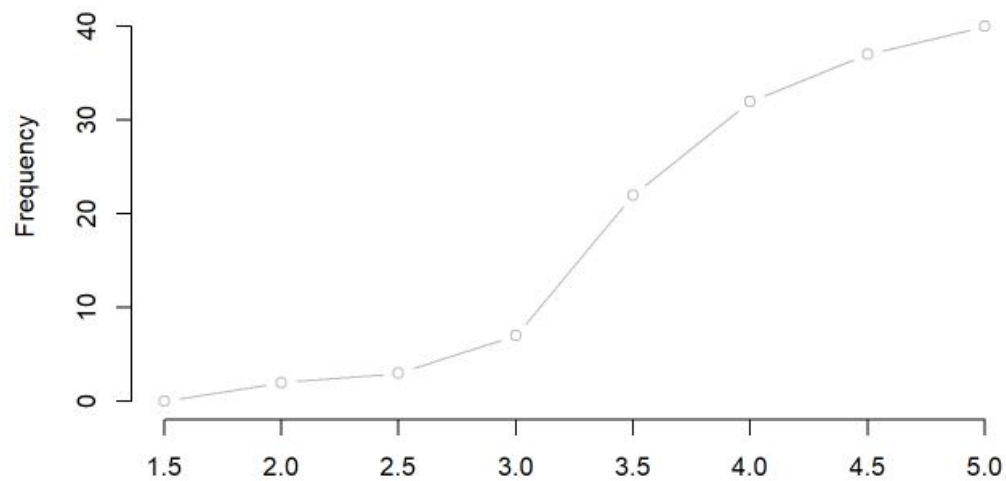
##	Class limits	f	rf	rf(%)
##	[1.50, 2.00)	2	0.05	5.0
##	[2.00, 2.50)	1	0.03	2.5
##	[2.50, 3.00)	4	0.10	10.0
##	[3.00, 3.50)	15	0.38	37.5
##	[3.50, 4.00)	10	0.25	25.0
##	[4.00, 4.50)	5	0.12	12.5
##	[4.50, 5.00)	3	0.07	7.5

```
#plotando o histograma
plot(d)
```



### 4.1.2 letra b)

```
plot(d, type='cfp')
```



### 4.1.3 letra c)

Média dos dados originais.

```
mean(x)
```

```
## [1] 3.4125
```

### 4.1.4 letra d)

Média da distribuição de frequência.

```
mean(d)
```

```
## [1] 3.4625
```



#### 4.1.5 letra e)

Variância dos dados originais.

```
var(x)
```

---

```
## [1] 0.4939423
```

#### 4.1.6 letra f)

Variância da distribuição de frequência.

```
var(d)
```

---

```
## [1] 0.4857372
```

## 5 Questão 5

A média de aprovação na disciplina de Estatística é 6 ou mais. Durante um período letivo foram realizadas quatro provas, sendo que a primeira prova teve peso dois, a segunda e a terceira o dobro do peso da primeira e a última igual ao peso da primeira. Os resultados, incluindo os de uma prova de substituição optativa, foram os seguintes:

Estudantes	1a	2a	3a	4a	Optativa
1	2,5	4,5	5,0	6,0	7,0
2	2,0	8,5	7,0	3,0	5,0
3	8,5	10,0	9,0	8,5	nc
4	3,5	5,5	8,5	7,5	6,5
5	3,0	5,0	6,0	4,5	5,0
6	6,0	3,0	4,0	5,0	2,0
7	8,0	1,5	2,0	9,0	5,0
8	1,5	2,0	1,0	2,5	nc
9	7,5	8,0	8,5	10,0	nc
10	5,5	4,5	5,0	4,5	2,5

Sabendo-se que a nota da prova optativa substitui a menor nota das provas precedentes, determine:

- Média de cada estudante;
- Para cada prova: média, moda, mediana, variância, desvio-padrão, erro-padrão da média e CV;
- Para o período: média, variância, desvio-padrão, erro-padrão da média, CV;
- O box-plot para cada prova e comente as diferenças ou as semelhanças.

### 5.1 Resolução da questão 5

Questão resolvida em *Python* e *Excel*.

#### 5.1.1 letra a)

Estudante	1	2	3	4	5	6	7	8	9	10
Média	5.33	6.50	9.17	7.00	5.25	3.83	5.17	1.67	8.42	4.50

### 5.1.2 letra b)

	1°	2°	3°	4°
Média	6.05	5.50	5.60	5.85
Moda	5.00	2; 4.5; 5	5; 8.5	2.5
Mediana	6.25	5.00	5.5	5.5
Variância	4.02	6.94	7.54	7.78
Desvio-padrão	2.01	2.64	2.75	2.79
Erro padrão da média	0.63	0.83	0.87	0.88
CV(%)	33.16%	47.91%	49.05%	47.68%

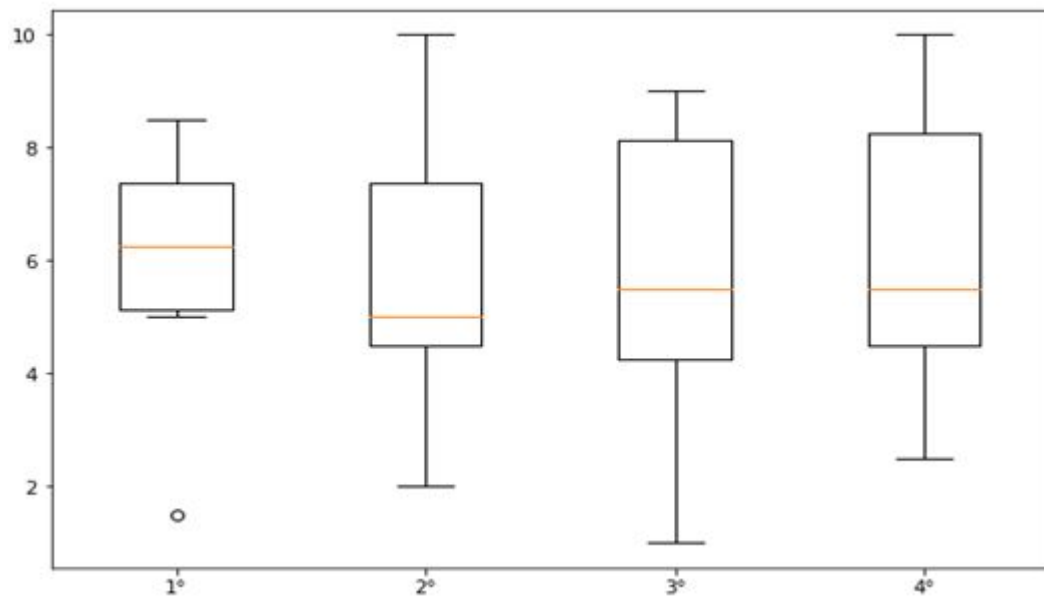
### 5.1.3 letra c)

Para o período:

Média ( $\bar{X}$ )	5.6833
Variância ( $S^2$ )	6.2098
Desvio padrão ( $s$ )	2.4919
Erro padrão da média $s(\bar{X})$	0.2275
CV (%)	43.85%

### 5.1.4 letra d)

```
data=dados.Primeira, dados.Segunda, dados.Terceira, dados.Quarta
fig=plt.figure(figsize=(9,6))
ax=fig.add_subplot(111)
plt.boxplot(data)
ax.set_xticklabels(['1°', '2°', '3°', '4°'])
plt.show()
```



Gráficos boxplot (diagrama de caixa) demonstram a concentração da distribuição dos dados observados. O conjunto de dados utilizados são as 4 maiores notas de cada um dos 10 estudantes (a prova optativa, caso seja realizada, substituirá a menor nota de uma das 4 primeiras provas). A linha laranja denota a mediana, já a linha vertical denota os valores mínimo e máximo da nossa amostra (valor máximo e mínimo não significa, necessariamente, que seja o maior e menor valor da sua amostra, apenas os valores dentro de um intervalo de confiança). Observando os gráficos plotados vemos que a 1ª prova possui o valor mínimo muito próximo do 1º quartil e possui um outlier (o aluno 8 não realizou a prova optativa e tirou nota 1,5 na 1ª prova). A 2ª e 4ª prova possuem valor máximo parecidos, e a 3ª prova possui o menor valor mínimo.

## 6 Questão 6

(Cap. 3: ex. 23) Estamos interessados em estudar a idade dos 12.325 funcionários da Cia. Distribuidora de Leite Teco, e isso será feito por meio de uma amostra. Para determinar que tamanho deverá ter essa amostra, foi colhida uma amostra-piloto. As idades observadas foram: 42, 35, 27, 21, 55, 18, 27, 30, 21, 24.

- (a) Determine as medidas descritivas dos dados que você conhece.
- (b) Qual dessas medidas você acredita que será a mais importante para julgar o tamanho final da amostra? Por quê?

### 6.1 Resolução da questão 6

Questão resolvida em *Python*.

#### 6.1.1 letra a)

*Script* e resultados gerados em *Python*:

```
print('Medidas descritivas obtidas na amostra-piloto')
print('*254')
print('Média :',df4['dados'].mean())
print('Mediana :',df4['dados'].median())
print('Moda :',df4['dados'].mode()[0])
print('Variância :',round(df4['dados'].var(),3))
print('Desvio padrão :',round(df4['dados'].std(),3))
```

Medidas descritivas obtidas na amostra-piloto

Média : 30.0  
Mediana : 27.0  
Moda : 21  
Variância : 128.222  
Desvio padrão : 11.324

#### 6.1.2 letra b)

Das medidas acima, a mais importante para a determinação do tamanho da amostra final é a variância, pois fornece informação a respeito da variabilidade da variável Idade.

## 7 Questão 7

(Cap. 3: ex. 28) A idade média dos candidatos a um determinado curso de aperfeiçoamento sempre foi baixa, da ordem de 22 anos. Como esse curso foi planejado para atender a todas as idades, decidiu-se fazer uma campanha de divulgação. Para se verificar se a campanha foi ou não eficiente, fez-se um levantamento da idade dos candidatos à última promoção, e os resultados estão na tabela a seguir.

Idade	Frequência	Porcentagem
18├ 20	18	36
20├ 22	12	24
22├ 26	10	20
26├ 30	8	16
30├ 36	2	4
Total	50	100

- Baseando-se nesses resultados, você diria que a campanha produziu algum efeito (isto é, aumentou a idade média)?
- Um outro pesquisador decidiu usar a seguinte regra: se a diferença  $\bar{x} - 22$  fosse maior que o valor  $2dp(X)/\sqrt{n}$ , então a campanha teria surtido efeito. Qual a conclusão dele, baseada nos dados?
- Faça o histograma da distribuição.

### 7.1 Resolução da questão 7

Questão resolvida em *Python*

#### 7.1.1 letra a)

Com base nos cálculos realizados abaixo, vê-se que a diferença da média antes e após a campanha é muito pequena, podendo-se concluir que a campanha não surtiu efeito.

```
media = (1/df7['Freq'].sum())*(df7['Média'][0]*df7['Freq'][0] +
                                df7['Média'][1]*df7['Freq'][1] +
                                df7['Média'][2]*df7['Freq'][2] +
                                df7['Média'][3]*df7['Freq'][3] +
                                df7['Média'][4]*df7['Freq'][4])
```

```
print("Diferença entre a média após campanha e média anterior",
      round(media-22,2))
```

Diferença entre a média após campanha e média anterior 0.48

### 7.1.2 letra b)

Calculando o desvio padrão em *Python*:

```
dp = ((df7['Freq']*((df7['Média']-media)**2)).sum()/50)**0.5
dp
3.8274795884498194
```

Comparando o resultado da diferença da média com o resultado de  $\frac{2dp(X)}{\sqrt{n}}$

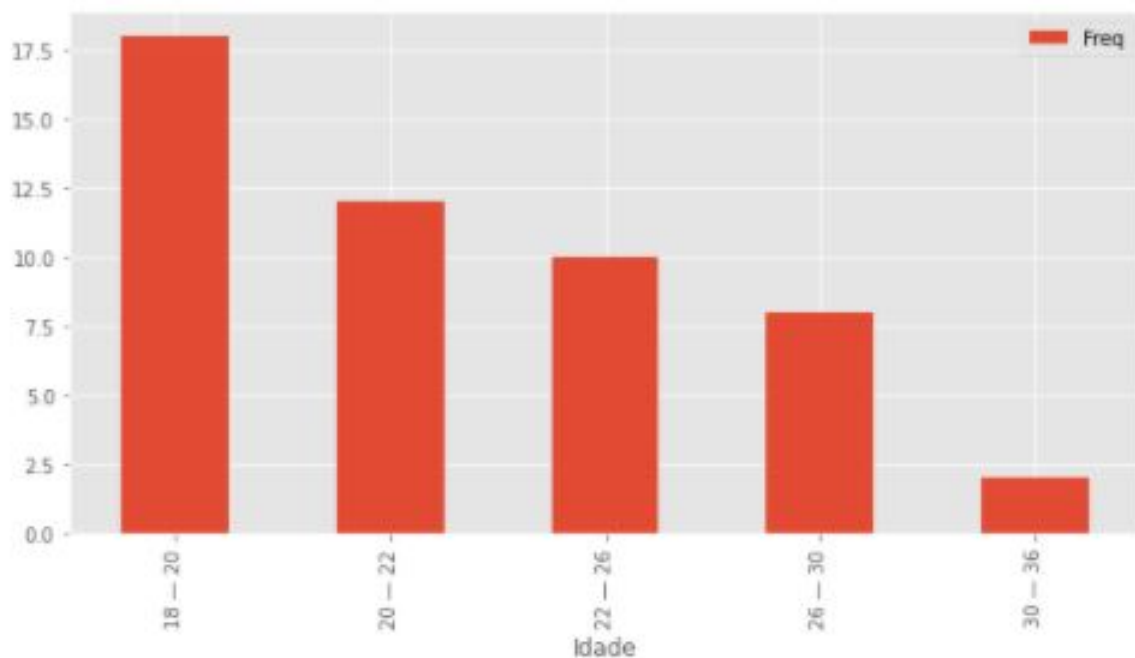
```
if (media-22) > 2*dp/np.sqrt(50):
    print('A campanha surtiu efeito')
else:
    print('A campanha não surtiu efeito')

A campanha não surtiu efeito
```

Conforme imagem acima a campanha não surtiu efeito.

### 7.1.3 letra c)

```
df7.plot.bar(x='Idade', y='Freq', figsize=(10,5));
```



## 8 Questão 8

(Cap. 3: ex. 29) Para se estudar o desempenho de duas corretoras de ações, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada, computou-se a porcentagem de lucro apresentada durante um período fixado de tempo. Os dados estão a seguir.

Corretora A			Corretora B		
45	60	54	57	55	58
62	55	70	50	52	59
38	48	64	59	55	56
55	56	55	61	52	53
54	59	48	57	57	50
65	55	60	55	58	54
			59	51	56

Que tipo de informação revelam esses dados? (**Sugestão:** use a análise proposta nas Seções 3.3 e 3.4.)

### 8.1 Resolução da questão 8

Resumo de cinco número de cada uma dos dados de corretoras:

```
print('O resumo de cinco números dos dados da corretora A')
print('*254')
print('Mínimo :', df9.min()[0])
print('Quantil 0.25 :', df9[0].quantile(0.25))
print('Quantil 0.5 (Mediana) :', df9[0].quantile(0.5))
print('Quantil 0.75 :', df9[0].quantile(0.75))
print('Máximo :', df9.max()[0])
print('*254')
print('O resumo de cinco números dos dados da corretora B')
print('*254')
print('Mínimo :', df10.min()[0])
print('Quantil 0.25 :', df10[0].quantile(0.25))
print('Quantil 0.5 (Mediana) :', df10[0].quantile(0.5))
print('Quantil 0.75 :', df10[0].quantile(0.75))
print('Máximo :', df10.max()[0])
```

O resumo de cinco números dos dados da corretora A

```
Mínimo : 38.0
Quantil 0.25 : 54.0
Quantil 0.5 (Mediana) : 55.0
Quantil 0.75 : 60.0
Máximo : 70.0
```

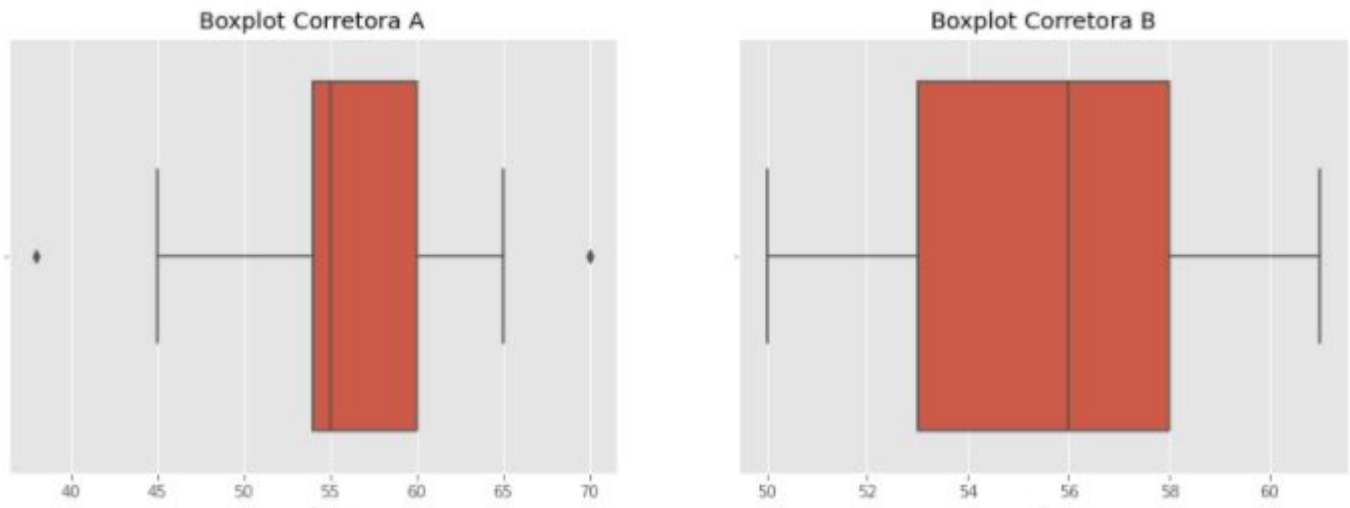
O resumo de cinco números dos dados da corretora B

```
Mínimo : 50
Quantil 0.25 : 53.0
Quantil 0.5 (Mediana) : 56.0
Quantil 0.75 : 58.0
Máximo : 61
```



*Boxplots* dos dados de cada corretora:

```
fig, ax = plt.subplots(1,2, figsize=(15,5))
ax[0].title.set_text('Boxplot Corretora A')
ax[1].title.set_text('Boxplot Corretora B')
sns.boxplot(x=0, data = df9, ax=ax[0])
sns.boxplot(x=0, data = df10, ax=ax[1]);
```



As medidas e a figura acima indicam que, a despeito do fato de o máximo lucro observado ser proveniente da corretora A, é a corretora B que apresenta menor variabilidade nos lucros proporcionados. As medianas das duas empresas estão bastante próximas. Estes elementos permitem acreditar que é mais vantajoso ter o dinheiro investido pela corretora B.

## 9 Questão 9

(Cap. 3: ex. 32) Para decidir se o desempenho das duas corretoras do exercício 29 são semelhantes ou não, adotou-se o seguinte teste: sejam

$$t = \frac{\bar{X}_A - \bar{X}_B}{S_*^2 \sqrt{1/n_A + 1/n_B}}, \quad S_*^2 = \frac{(n_A - 1) \text{var}(X/A) + (n_B - 1) \text{var}(X/B)}{n_A + n_B - 2}.$$

Caso  $|t| < 2$ , os desempenhos são semelhantes, caso contrário, são diferentes. Qual seria a sua conclusão? Aqui,  $n_A$  é o número de ações selecionadas da corretora A e nomenclatura análoga para  $n_B$ .

### 9.1 Resolução da questão 9

*Script Python* para resolução da questão:

Calculando  $S_*^2$

```
[101] s = ((len(df9)-1)*df9.var() + (len(df10)-1)*(df10.var()))/(len(df10) + len(df9) - 2)
      s[0]
32.50686400686401
```

Calculando  $t$ .

```
[106] t = (df9.mean() - df10.mean())/(s[0]*(np.sqrt(1/len(df9) + 1/len(df10))))
      t[0]
0.028123521553959883
```

Conforme enunciado: "Caso  $|t| < 2$ , os desempenhos são semelhantes, caso contrário, são diferentes"

```
[112] if abs(t[0]) < 2:
      print('Os desempenhos das corretoras são semelhantes')
      else:
      print('Os desempenhos das corretoras não são semelhantes')

Os desempenhos das corretoras são semelhantes
```

## 10 Questão 10

(Cap. 3: ex. 38) No Problema 9, do Capítulo 2, temos os resultados de 25 funcionários em vários exames a que se submeteram. Sabe-se agora que os critérios adotados em cada exame não são comparáveis, por isso decidiu-se usar o desempenho relativo em cada exame. Essa medida será obtida do seguinte modo:

- (I) Para cada exame serão calculados a média  $\bar{x}$  e o desvio padrão  $dp(X)$ .
- (II) A nota  $X$  de cada aluno será padronizada do seguinte modo:

$$Z = \frac{X - \bar{x}}{dp(X)}$$

- (a) Interprete o significado de  $Z$ .
- (b) Calcule as notas padronizadas dos funcionários para o exame de Estatística.
- (c) Com os resultados obtidos em (b), calcule  $\bar{z}$  e  $dp(Z)$ .
- (d) Se alguma das notas padronizadas estiver acima de  $2dp(Z)$  ou abaixo de  $-2dp(Z)$ , esse funcionário deve ser considerado um caso atípico. Existe algum nessa situação?
- (e) O funcionário 1 obteve 9,0 em Direito, em Estatística e em Política. Em que disciplina o seu desempenho relativo foi melhor?

## 10.1 Resolução da questão 10

### 10.1.1 letra a)

Esse valor  $Z$  é a nota padronizada, onde o valor 0 indica que o indivíduo em questão obteve a nota média. A nota  $Z$  também fornece idéia sobre o desempenho de cada elemento com relação a todo o grupo.

### 10.1.2 letra b)

Padronizando as notas:

```
df5['Estat_z'] = (df5['Estatíst.'].mean())/df5['Estatíst.'].std()
df5['Direito_z'] = (df5['Direito'].mean())/df5['Direito'].std()
df5['Redação_z'] = (df5['Redação'].mean())/df5['Redação'].std()
df5['Política_z'] = (df5['Política'].mean())/df5['Política'].std()
```

Notas de Estatística padronizadas:

df5['Estat\_z']

0	0.584615
1	0.584615
2	-0.184615
3	-0.184615
4	0.584615
5	1.353846
6	-0.184615
7	-0.184615
8	0.584615
9	-0.184615
10	1.353846
11	-0.953846
12	-0.953846
13	0.584615
14	0.584615
15	-0.953846
16	-0.184615
17	0.584615
18	-3.261538
19	-0.953846
20	-0.953846
21	-0.184615
22	1.353846
23	0.584615
24	0.584615

### 10.1.3 letra c)

Como as notas foram padronizadas pela subtração da média e divisão pelo desvio-padrão, tem-se (Problema 21) que  $\bar{z} = 0$  e  $dp(Z) = 1$ .

### 10.1.4 letra d)

O valor do desvio padrão é igual a 1, assim  $2dp(Z) = 2$  e  $-2dp(Z) = -2$ . Para que uma observação seja considerada atípica, ela deve estar fora desse intervalo e nesse caso a observação de valor  $Z = -3,26$ ; portanto, é uma observação atípica ou um *outlier*.

### 10.1.5 letra e)

Para avaliar o seu desempenho relativo, é necessário comparar as notas padronizadas nas três disciplinas:

- 1) Em Direito, todos obtiveram 9,0, logo a média padronizada tem valor 0.
- 2) Em Política, a média das notas foi 7,76 e o desvio padrão, 1,67. Com isso, a nota padronizada do funcionário 1 é 0,74.
- 3) Em Estatística a média antes foi 8,24 e desvio padrão 1,3, mas agora com a padronização a nota do funcionário 1 foi de 0,58.

Com isso, concluímos, que seu desempenho relativo foi melhor em Política.

```
print('Médias e desvio padrão antes da padronização')
print('*508')
print('DIREITO - ', 'Média :', df5['Direito'].mean(), '|', 'desvio padrão :',
      df5['Direito'].std())
print('ESTATÍSTICA - ', 'Média :', df5['Estatíst.'].mean(), '|', 'desvio padrão :',
      round(df5['Estatíst.'].std(), 2))
print('POLÍTICA - ', 'Média :', df5['Política'].mean(), '|', 'desvio padrão :',
      round(df5['Política'].std(), 2))

print('*508')
print('Notas do funcionário 1 depois da padronização')
print('*508')
print('DIREITO - ', 'Nota :', df5['Direito_z'][0], '|', 'desvio padrão :',
      df5['Direito_z'].std())
print('ESTATÍSTICA - ', 'Nota :', df5['Estat_z'][0], '|', 'desvio padrão :',
      round(df5['Estat_z'].std(), 2))
print('POLÍTICA - ', 'Nota :', df5['Política_z'][0], '|', 'desvio padrão :',
      round(df5['Política_z'].std(), 2))
```

Médias e desvio padrão antes da padronização

DIREITO - Média : 9.0 | desvio padrão : 0.0  
 ESTATÍSTICA - Média : 8.24 | desvio padrão : 1.3  
 POLÍTICA - Média : 7.76 | desvio padrão : 1.67

Notas do funcionário 1 depois da padronização

DIREITO - Nota : nan | desvio padrão : nan  
 ESTATÍSTICA - Nota : 0.5846153846153845 | desvio padrão : 1.0  
 POLÍTICA - Nota : 0.7418148644845932 | desvio padrão : 1.0