

### Exercício 1

Considere os dados abaixo sobre distribuição de salário (em reais) num grupo de 1000 indivíduos.

Tabela 1: Distribuição de frequências da variável salário

Classe de Salário	$n_i$	$f_i$
500—1000	400	0,40
1000—1500	200	0,20
1500—2000	150	0,15
2000—3000	150	0,15
3000—5000	100	0,10

Classifique a variável Classe de Salário. Construa os histogramas pelos métodos de frequência e densidade. Qual deve ser utilizado? Justifique. Calcule aproximadamente do histograma apropriado a média, o desvio padrão e a mediana salarial para essa amostra

## Solução

A variável classe de salário é quantitativa contínua. Para construir os histogramas solicitados observe que,

Tabela 2: Distribuição de frequências da variável salário

Classe de Salário	$n_i$	$f_i$	Amplitude $\Delta_i$	Densidade $f_i/\Delta_i$
500—1000	400	0,40	500	$0,8 \times 10^{-3}$
1000—1500	200	0,20	500	$0,4 \times 10^{-3}$
1500—2000	150	0,15	500	$0,3 \times 10^{-3}$
2000—3000	150	0,15	1000	$0,15 \times 10^{-3}$
3000—5000	100	0,10	2000	$0,05 \times 10^{-3}$

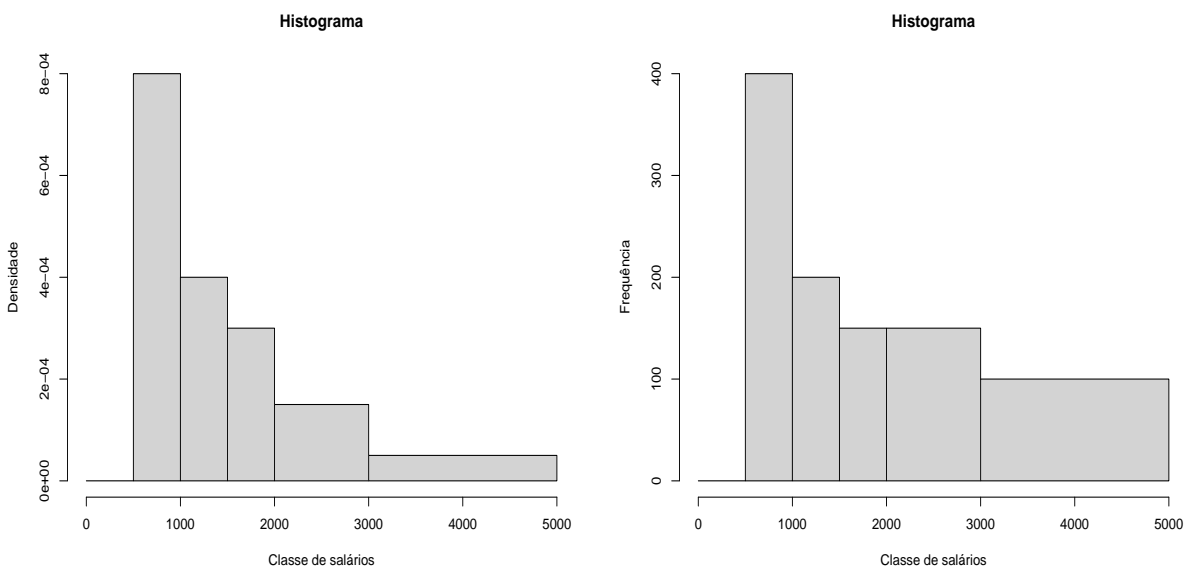


Figura 1: Histogramas para a variável salário usando métodos de densidade e frequência respectivamente.

O gráfico que deve ser utilizado é o histograma obtido pelo método de densidade, visto que as

classes apresentam tamanhos diferentes. A média aproximada fica dada por,

$$\bar{x}^* = 0,4 \times 750 + 1250 \times 0,2 + 1750 \times 0,15 + 2500 \times 0,15400 \times 0,1 = 1587,5.$$

A variância aproximada,

$$\begin{aligned} s^{*2} &= 0,4 \times (750 - 1587,5)^2 + (1250 - 1587,5)^2 \times 0,2 + (1750 - 1587,5)^2 \times 0,15 \\ &\quad + (2500 - 1587,5)^2 \times 0,15 + (4000 - 1587,5)^2 \times 0,1 = 1014219, \end{aligned}$$

portanto, o desvio padrão é

$$s^* = \sqrt{1014219} = 1007,084.$$

A mediana encontra-se na segunda classe (de mil a mil e quinhentos reais) uma vez que a primeira classe possui 40% das observações (menos da metade) e as duas primeiras classes acumulam sessenta por cento das observações. E temos,

$$\frac{md_{obs} - 1000}{0.10} = \frac{1500 - 1000}{0.200} \Rightarrow md_{obs} = 1250.$$

*Comandos utilizados para obtenção dos histogramas no software estatístico R.*

```
f=c(0,400,200,150,150,100)
```

```
y=c(0,500,1000,1500,2000,3000,5000)
```

```
xm=(y[-length(y)]+y[-1])/2
```

```
w=rep(xm,f)
```

```
hist(w,breaks=y,col = "lightgray", main="Histograma", xlab="Classe de salários",
      ylab="Densidade")
```

```
hist(w,breaks=y,prob=FALSE,col = "lightgray", main="Histograma",
     xlab="Classe de salários",
     ylab="Frequência")
```

**Observação:** Ao produzir o histograma usando método de frequência será gerado um aviso de que a área do gráfico está errada e que preferencialmente deve-se utilizar `freq=FALSE`, ou seja, deve-se usar a densidade e não a frequência para este caso (devido ao fato dos tamanhos das classes serem diferentes).

## Exercício 2

Os dados descritos a seguir referem-se a despesas fixas e despesas com pessoal (incluindo encargos) dos departamentos de contabilidade e pessoal de uma empresa durante 15 meses. Os valores estão em mil reais.

Tabela 3: Valores Ordenados das despesas fixas e pessoais em cada departamento.

	Depto. Contabilidade		Depto. Pessoal	
	D. Fixas	D. Pessoal	D. Fixas	D. Pessoal
1	0.91	14.00	1.30	13.63
2	1.80	15.82	1.64	14.12
3	2.31	18.10	2.26	15.07
4	2.36	18.35	2.26	15.37
5	3.00	18.37	2.38	15.38
6	3.44	18.47	2.39	15.39
7	3.46	20.95	2.42	17.18
<b>8</b>	<b>3.54</b>	<b>21.20</b>	<b>2.57</b>	<b>17.70</b>
9	3.60	21.93	2.70	17.78
10	3.68	22.00	3.04	17.78
11	3.75	24.11	3.08	18.00
12	3.88	24.95	3.18	18.54
13	3.89	28.08	3.65	22.47
14	4.00	28.90	4.03	22.56
15	4.45	29.43	4.10	24.61

Construa, num mesmo gráfico, os boxplots das despesas fixas e num outro gráfico os boxplots das despesas com pessoal para os dois departamentos. Comente sobre a dispersão, pontos extremos, mediana e simetria dos dados, para cada departamento. Recomendado uso de computador.

## Solução

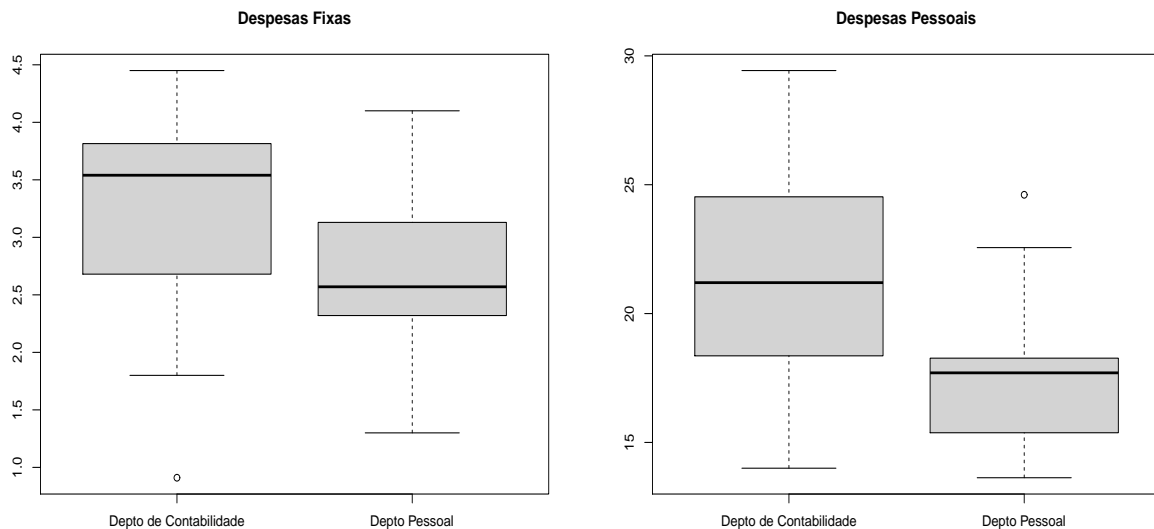


Figura 2: Boxplots das despesas fixas e pessoais para os departamentos de contabilidade e pessoal obtidos no R.

No que se refere as despesas fixas temos,

- i) Os dados são claramente assimétricos em ambos os departamentos. No departamento de contabilidade vemos que a mediana está bem mais próxima do terceiro quartil do que do primeiro o que indica assimetria a esquerda. Já no departamento pessoal a mediana está bem mais próxima do primeiro quartil do que do terceiro o que indica assimetria a direita. No departamento de contabilidade foi detectado um ponto discrepante (*outlier*), correspondente

ao valor 0,91 (mil reais), muito abaixo dos demais.

- ii) Comparando os boxplots das despesas fixas vemos que a distribuição de valores (medidas de posição) é bem diferente em cada departamento. Há evidência que o departamento de contabilidade tem mais valores altos de despesas fixas durante os quinze meses em estudo do que o departamento pessoal.

No que se refere as despesas pessoais temos,

- iii) O departamento de contabilidade no que refere a despesas pessoais parece apresentar dados mais simétricos (levemente simétrico). Já no departamento pessoal constatamos agora assimetria a esquerda. Podemos visualizar a presença de uma observação discrepante no departamento pessoal (consideravelmente mais alta que os demais valores).
- iv) Comparando os boxplots das despesas pessoais temos que a distribuição de valores (medidas de posição) é bem diferente em cada departamento. Agora podemos constatar que o departamento de contabilidade apresenta no geral valores mais altos de despesas pessoais durante os quinze meses em estudo do que o departamento pessoal. Observe inclusive que o primeiro quartil do departamento de contabilidade é realmente próximo do terceiro quartil das observações do departamento pessoal o que colabora com a conclusão que os gastos no departamento de contabilidade são mais elevados. É válido salientar ainda que o gasto de 24,61 (mil reais) (muito alto) é considerado discrepante no departamento pessoal, enquanto que no departamento de contabilidade cerca de 25% dos gastos são mais altos do que isso.

*Os comandos utilizados para obter os gráficos no software estatístico R são,*

```
despesas<-read.table('C:/despesas.txt',head=T)
despesasfixas<-matrix(c(despesas$Fixas.C,despesas$Fixas.P), nrow = 15,
ncol = 2, byrow = FALSE,
```

```

dimnames = list(c(),c("Depto de Contabilidade", "Depto Pessoal"))
boxplot(as.data.frame(despesasfixas), main="Despesas Fixas",col="lightgray")

despesaspessoal<-matrix(c(despesas$Pessoal.C,despesas$Pessoal.P), nrow = 15,
ncol = 2, byrow = FALSE,
dimnames = list(c(),c("Depto de Contabilidade", "Depto Pessoal")))
boxplot(as.data.frame(despesaspessoal),main="Despesas Pessoais",col="lightgray")

```

### Exercício 3

Numa pesquisa sobre rotatividade de mão-de-obra, para uma amostra de 40 pessoas foram observadas duas variáveis: número de empregos nos últimos dois anos (X) e salário mais recente, em número de salários mínimos (Y). Os resultados são descritos abaixo.

X	1	3	2	3	2	2	3	1	2	3	2	3	1	2	3	4	1	2	2	2
Y	6	2	4	1	4	1	3	5	2	2	5	2	6	6	2	2	5	5	1	1
X	2	3	4	1	2	3	4	1	4	3	2	1	1	2	4	3	1	3	2	2
Y	6	2	1	5	4	2	1	5	4	3	2	1	1	6	2	1	4	2	3	5

Responda aos itens abaixo:

- Usando a mediana, classifique os indivíduos em dois níveis, alto e baixo, para cada uma das variáveis, e construa a distribuição de frequências conjunta das duas classificações.
- Qual a porcentagem das pessoas com baixa rotatividade e ganhando pouco?
- Qual a porcentagem das pessoas que ganham pouco?
- Entre as pessoas com baixa rotatividade, qual a porcentagem das que ganham pouco?

- e) A informação adicional dada em (d) mudou muito a porcentagem observada em (c)? O que isso significa?

### Solução

- a) Como temos 40 observações a mediana é o valor que está na posição  $\frac{40+1}{2} = 20,5$ .

Os valores da variável X ordenados são:

1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4,

por outro lado os valores ordenados de Y são,

1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 6.

Portanto,

$$Md_X = \frac{2+2}{2} = 2. \text{ e } Md_Y = \frac{2+3}{2} = 2,5.$$

Definimos então uma pessoa com baixa rotatividade aquela que possui dois empregos ou menos nos últimos dois anos, isto é  $X \leq 2$ , e com alta rotatividade quando  $X > 2$ . E definimos uma pessoa ganhando pouco quando  $Y \leq 2,5$ , e ganhando muito quando  $Y > 2,5$ .

Tabela 4: Distribuição conjunta das duas classificações

	baixa rotatividade	alta rotatividade	total
ganha pouco	7	13	20
ganha muito	17	3	20
total	24	16	

- b) A porcentagem de pessoas com baixa rotatividade e ganhando pouco fica dada por,

$$\frac{7}{40} \times 100 = 17,5\%.$$



c) A porcentagem de pessoas que ganham pouco é dada por,

$$\frac{20}{40} \times 100 = 50\%.$$

d) Entre as pessoas com baixa rotatividade, a porcentagem das que ganham pouco é,

$$\frac{7}{24} \times 100 = 29,1667\%.$$

e) Note que quando restringimos ao grupo de pessoas com baixa rotatividade a porcentagem de pessoas que ganham pouco reduz consideravelmente. Isto significa que as duas variáveis, X e Y, são dependentes e que a porcentagem de pessoas que ganham pouco é menor entre aquelas com baixa rotatividade do que entre as com alta rotatividade.