

## Lista 1 - MAE0127 - Prof Julio Singer

Alexandre Felix da Silva 11221435 Eduardo Janotti Cavalcanti 11221481  
 Emanuel Hark Maciel 11221380 Emerson Silva Aragão 11371583 Guilherme Yukio Iasunaga 11218195  
 Pedro Henrique Gallo Andrade 11221550 William Silva Veçoso 10801513

26/03/2020

## Capítulo 2

### Exercício 4

A Tabela abaixo contém uma planilha encaminhada pelos investigadores responsáveis por um estudo sobre AIDS para análise estatística. Organize-a de forma a permitir sua análise por meio de um pacote computacional como o R.

Grupo I registro	Tempo de Diagnóstico	DST	MAC	Ganho de Peso por Semana
2847111D	pre natal	nao	Pilula	11Kg em 37 semanas
3034048F	6 meses	nao	pilula	?
3244701J	1 ano	nao	Condon	?
2943791B	pre natal	nao	nao	8 Kg em 39 semanas
3000327F	4 anos	condiloma/ sífilis	nao	9Kg em 39 semanas
3232893D	1 ano	nao	DIU	3Kg em 39 semanas
3028772E	3 anos	nao	nao	3 kg em 38 semanas
3240047G	pre natal	nao	pilula	9 Kg em 38 semanas
3017222G		HPV	CONDON	falta exame clinico
3015834J	2 anos	nao	condon	14 Kg em 40 semanas
Grupo I registro	Tempo de Diagnóstico	DST	MAC	Ganho de Peso por Semana
3173611E	3 meses	abcesso ovariano	condon	15 Kg em 40 semanas
3296159D	pre natal	nao	condon	0 Kg em ? semanas
3147820D1	2 anos	nao	sem dados	4 Kg em 37 semanas
3274750K	3 anos	nao	condon	8 Kg em 38 semanas
3274447H	pre natal	sífilis com 3 meses	condon	
2960066D	5 anos	nao	?	13 Kg em 36 semanas
3235727J	7 anos	nao	Condon	(-) 2 Kg em 38 semanas
3264897E		condiloma	condon	nenhum Kg
3044120J	5 anos	HPV		3 Kg em 39 semanas 1

Planilha com dados de um estudo sobre AIDS.

### Resolução

grupo	reg	tempo	dst	mac	gansem	gantot	sem
1	2847111D	37	0	2	0.297	11	37

grupo	reg	tempo	dst	mac	gansem	gantot	sem
1	3034048F	6	0	2	NA	NA	NA
1	3244701J	12	0	1	NA	NA	NA
1	2943791B	39	0	0	0.205	8	39
1	3000327F	48	1	0	0.231	9	39
1	3232893D	12	0	3	0.077	3	39
1	3028772E	36	0	0	0.079	3	38
1	3240047G	38	0	2	0.237	9	38
1	3017222G	NA	1	1	NA	NA	NA
1	3015834J	24	0	1	0.350	14	40
2	3173611E	3	1	1	0.375	15	40
2	3296159D	NA	0	1	0.000	0	NA
2	3147820D1	24	0	NA	0.108	4	37
2	3274750K	3	0	1	0.211	8	38
2	3274447H	NA	1	1	NA	NA	NA
2	2960066D	60	0	NA	0.361	13	36
2	3235727J	72	0	1	-0.053	-2	38
2	3264897E	NA	1	1	0.000	0	NA
2	3044120J	60	1	NA	0.077	3	39

Nós escolhemos classificar pré-natal = 0, a variável dst como binária e nac como categórica. Com os devidos rótulos:

## Exercício 5

Num estudo planejado para avaliar o consumo médio de combustível de veículos em diferentes velocidades foram utilizados 4 automóveis da marca A e 3 automóveis da marca B selecionados ao acaso das respectivas linhas de produção. O consumo (em L/km) de cada um dos 7 automóveis foi observado em 3 velocidades diferentes (40 km/h, 80 km/h e 110 km/h). Construa uma tabela apropriada para a coleta e análise estatística dos dados, rotulando-a adequadamente.

Table 2: Rótulos da planilha de um estudo sobre AIDS

Rótulos	Variável	Unidade de Medida
grupo	Número de grupo em que o paciente foi observado	
reg	Registro da identificação do paciente	
tempo	Tempo de diagnóstico do teste sobre DST	meses
dst	Presença de doenças sexualmente transmissíveis (HPV/AIDS/condiloma/sífilis)	0: Falso 1: Verdadeiro
mac	Método anticonceptivo do paciente	0: Não 1: Condon 2: Pílula 3: DIU
gansem	Ganho de peso por semana	Kg/Semana
gantot	Ganho de peso total	Kg
sem	Semana no qual o ganho do peso foi vistoriado	semana
Obs:	Observações omissas são representadas por 'NA'	

Table 3: Rótulos da planilha de consumo de carros

Rótulos	Variável	Unidade de Medida
marca	identificador da marca do carro	0: Marca A 1: Marca B
carro	identificador do carro medido	
Consumo Pequeno	Consumo de combustível do carro em pequena velocidade (40km/h)	L/Km
Consumo Médio	Consumo de combustível do carro em média velocidade (80Km/h)	L/Km
Consumo Alto	Consumo de combustível do carro em alta velocidade (110 Km/h)	L/Km
Obs:	Observações omissas são representadas por NA	

## Resolução

Marca	Carro	Consumo		
		Pequeno	Médio	Alto
0	1	NA	NA	NA
0	2	NA	NA	NA
0	3	NA	NA	NA
0	4	NA	NA	NA
1	5	NA	NA	NA
1	6	NA	NA	NA
1	7	NA	NA	NA

Com os devidos rótulos

## Exercício 6

A planilha apresentada na Figura 2.7 contém dados de um estudo em que o limiar auditivo foi avaliado nas orelhas direita (OD) e esquerda (OE) de 13 pacientes em 3 ocasiões (Limiar, Teste 1 e Teste 2). Reformate-a segundo as recomendações da Seção 2.3

Limiar	Teste1	Teste2
OD 50 / OE 55	OD/OE 50	OD/OE 80%
OD 41 /OE 40	OD 45/OE 50	OD 68% OE 80%
OD/OE 41,25	OD/OE 45	OD 64% OE 72%
OD 45/OE 43,75	OD 60/OE 50	OD 76%/OE 88%
OD51,25/ OE47,5	OD/OE 50	OD 80%/OE 88%
OD45/ OE 52,5	OD/OE 50	OD 84%/OE 96%
OD 52,5/OE 50	OD55/OE45	OD 40%/OE 28%
OD 42,15/OE48,75	OD 40/OE 50	OD80%/OE76%
OD50/ OE 48,75	OD/OE 50	OD 72%/OE 80%
OD47,5/OE46,25	OD/OE 50	OD/OE 84%
OD55/OE 56,25	OD55/OE60	OD80%/OE 84%
OD/OE 46,25	OD40/OE35	OD72%/OE 84%
OD 50/OE 47,5	OD/OE45	OD/OE 76%

Figure 1: Limiar auditivo de pacientes observados em 3 ocasiões.

## Resolução

id	Olho		Olho		Olho	
	Limiar		Teste 1		Teste2	
	Direito(%)	Esquerdo(%)	Direito(%)	Esquerdo(%)	Direito(%)	Esquerdo(%)
paca	50	55	50	50	80	80
pacb	41	40	45	50	68	80
pacc	41	41	45	45	64	72
pacd	45	44	60	45	76	88
pace	51	48	50	50	80	88
pacf	45	53	50	50	84	96
pacg	53	50	55	45	40	28
pach	42	49	40	50	80	76
paci	50	49	50	50	72	80
pacj	48	46	50	50	84	84
pacl	55	56	55	60	80	84
pacm	46	46	40	35	73	84
pacn	50	48	45	45	76	76

Nós escolhemos arredondar as unidades, em porcentagem, com dois números significativos.

## Capítulo 3

### Exercício 1

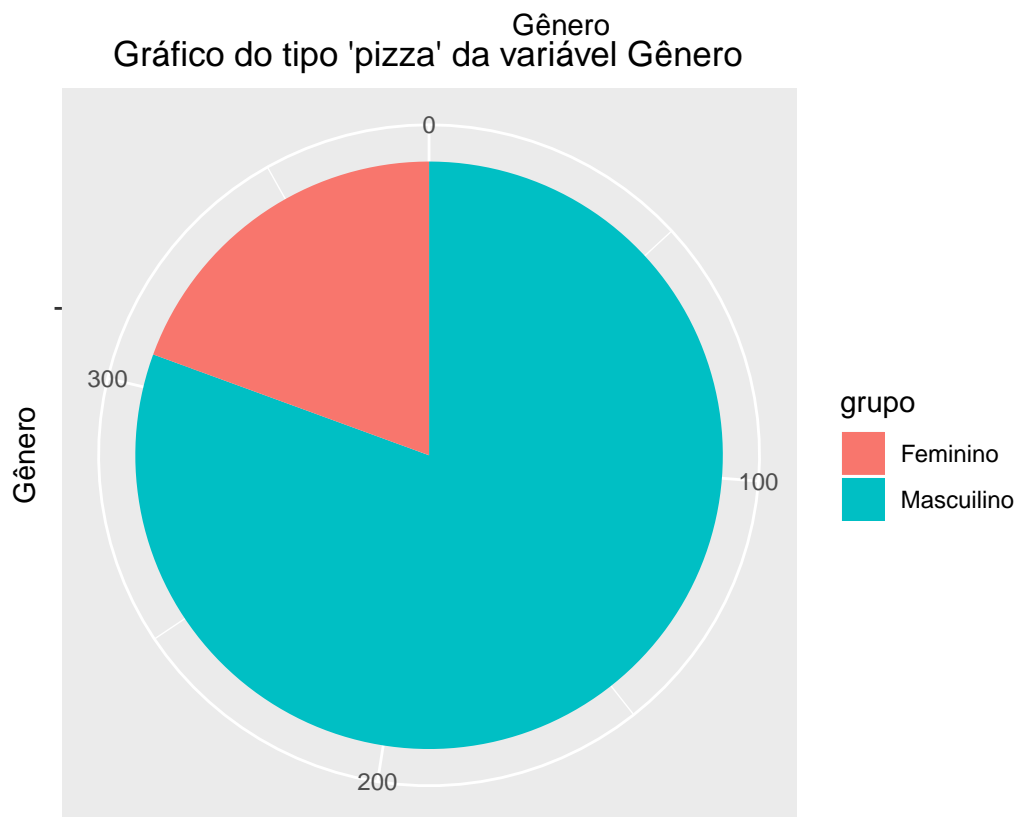
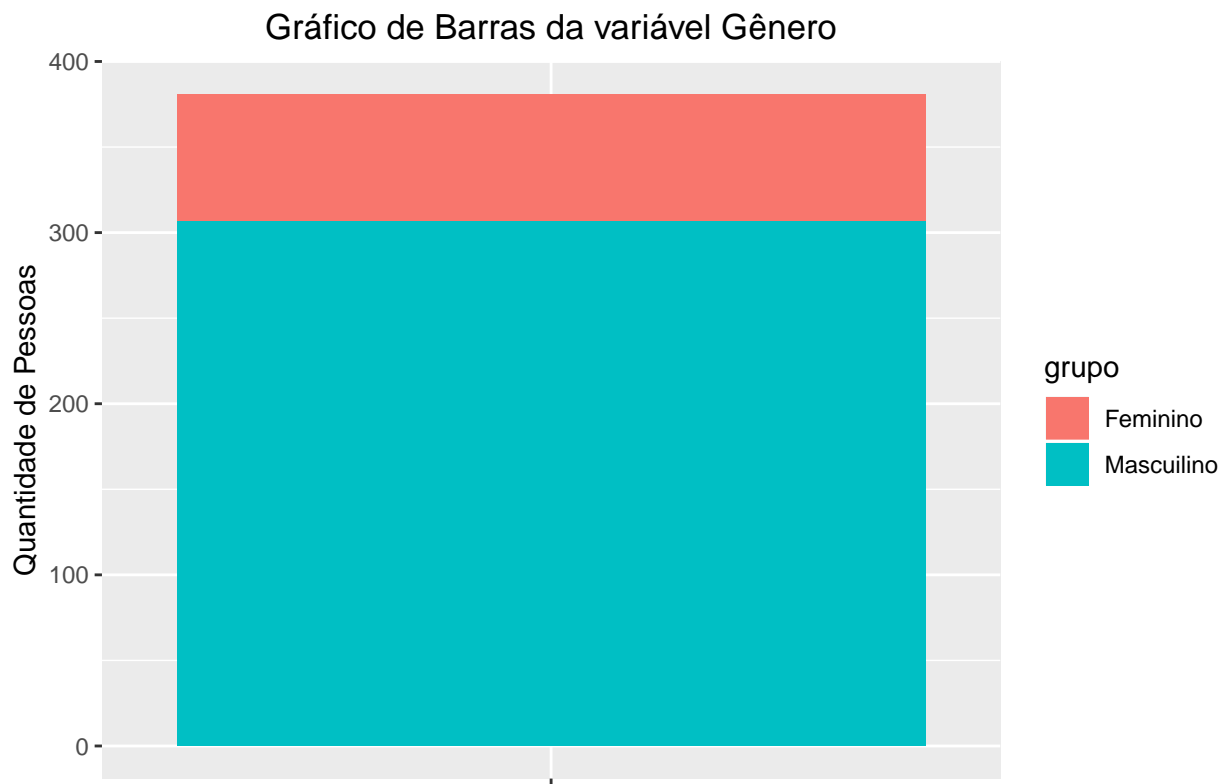
O arquivo *rehabcardio* contém informações sobre um estudo de reabilitação de pacientes cardíacos. Elabore um relatório indicando possíveis inconsistências na matriz de dados e faça uma análise descritiva de todas as variáveis do estudo.

### Resolução

#### Variável qualitativa

Table 4: Distribuição de frequência Gênero

	Gênero	Frêquencia observada	Frêquencia relativa (%)	Frequência acumulada(%)
F	Feminino	74	19	19
M	Masculino	307	81	100



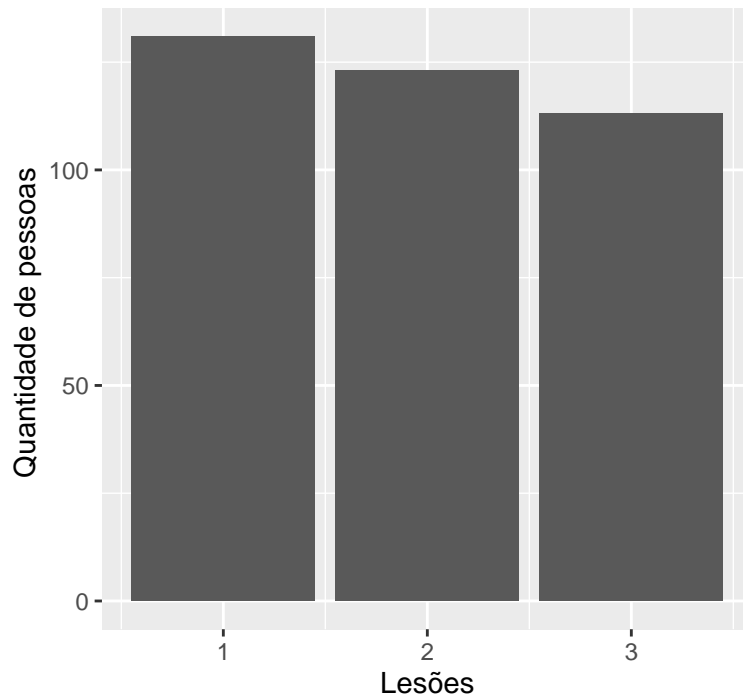
Percebe-se que há preferencialmente membros do sexo masculino dentro dessa amostra.

## Variável quantitativa discreta

Table 5: Número de

Nº de lesões	1	2	3	NA
Nº de pacientes	131	123	113	14

Gráfico de barra da variável Lesões



Como a variável Lesões é discreta, nota-se que os números de pacientes decaem com o aumento do número de lesões, mostrando um comportamento decrescente.

## Variável quantitativa contínua

Table 6: Medidas de posição e de dispersão da variável PESO

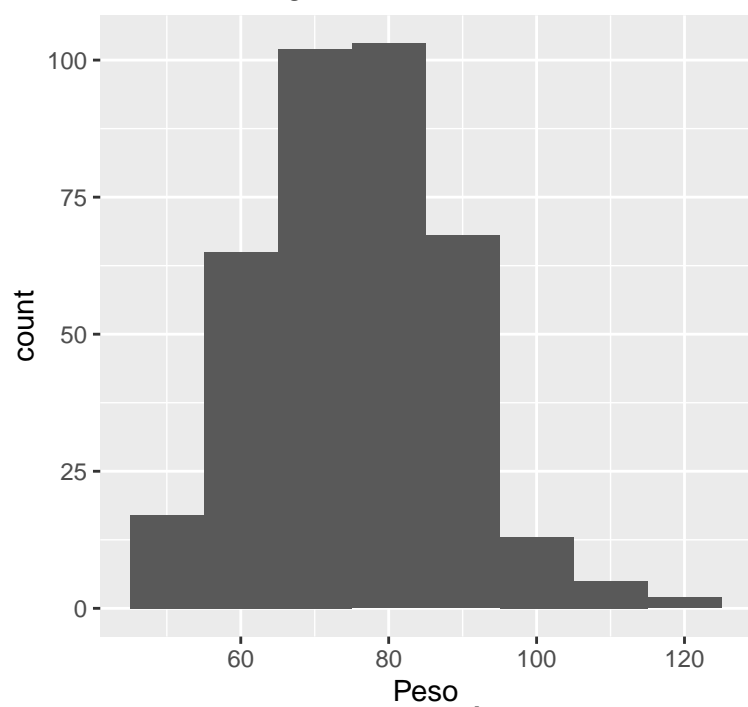
Resumo (Kg)						Dispersão(Kg)				Assimetria				
Quartis						Desvio				Coeficientes de Assimetria				
Mínimo	1º Quar-til	Mediana	3º Quar-til	Média	Máxi-mo	Moda	Padrão	Médio	Inter-quartil	Normal	Ajustado	Pearson	Kurtose	NA
47	67.05	75.5	76.11	84.95	119	78	34.87	NA	17.9	0.286	0.288	0.285	0.287	6



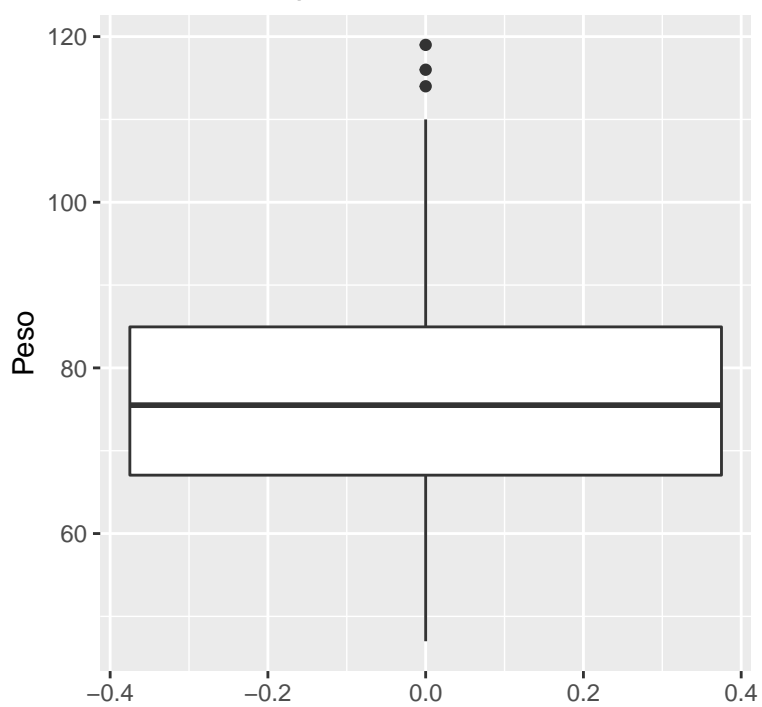
Table 7: Distribuição da frequência da variável PESO

Classes	Frequência	Frequência acumulada	Frequência relativa (%)	Frequência relativa acumulada (%)
40,50	5	5	1	1
50-60	42	47	11	12
60-70	80	127	21	33
70-80	119	246	31	64
80-90	86	332	22	86
90-100	29	361	8	94
100-110	11	372	3	97
110-120	3	375	1	98
NA	6	381	2	100
Total	381	381	100	100

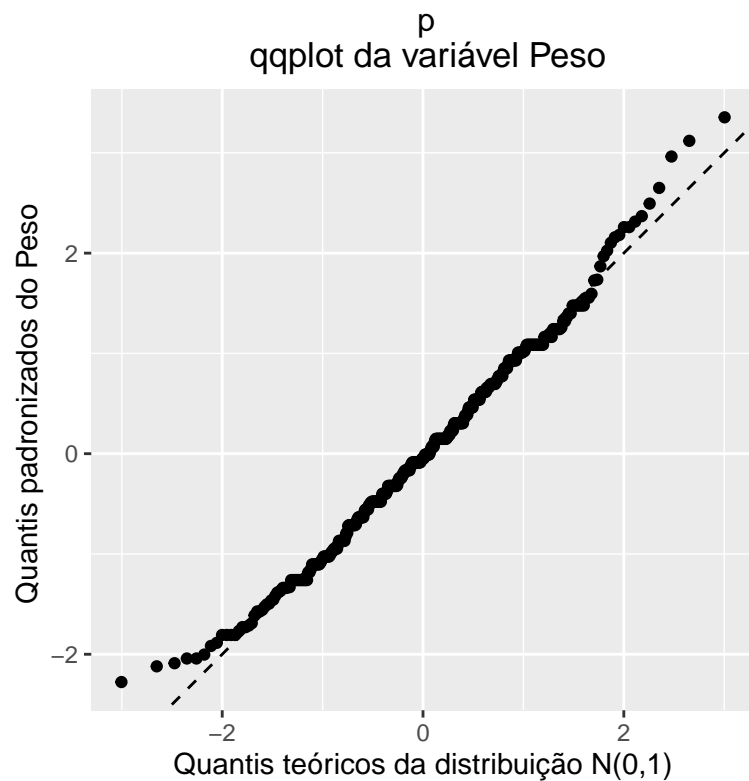
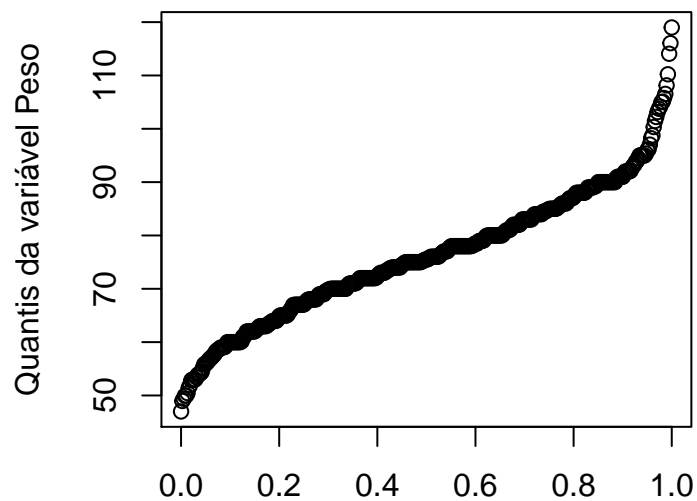
Histograma da variável Peso



Boxplot da variável Peso



## Gráfico de Quantis da variável Peso



Como a variável Peso é contínua, nota-se, pelos gráficos de histograma, de quantis e do qqplot com os quantis teóricos normais, que o comportamento dos dados é aproximadamente normal, tendo três valores discrepantes superiores.

Nós não encontramos variáveis qualitativas ordinais.

## Exercício 6

Considere do arquivo *vento*. Observe o valor atípico 61,1, que na realidade ocorreu devido a forte tempestade no dia 2 de dezembro. Calcule as medidas de posição e dispersão dadas na Seção 3.3. Comente os resultados.

## Resolução

Table 8: Medidas de posição e de dispersão das variáveis vt

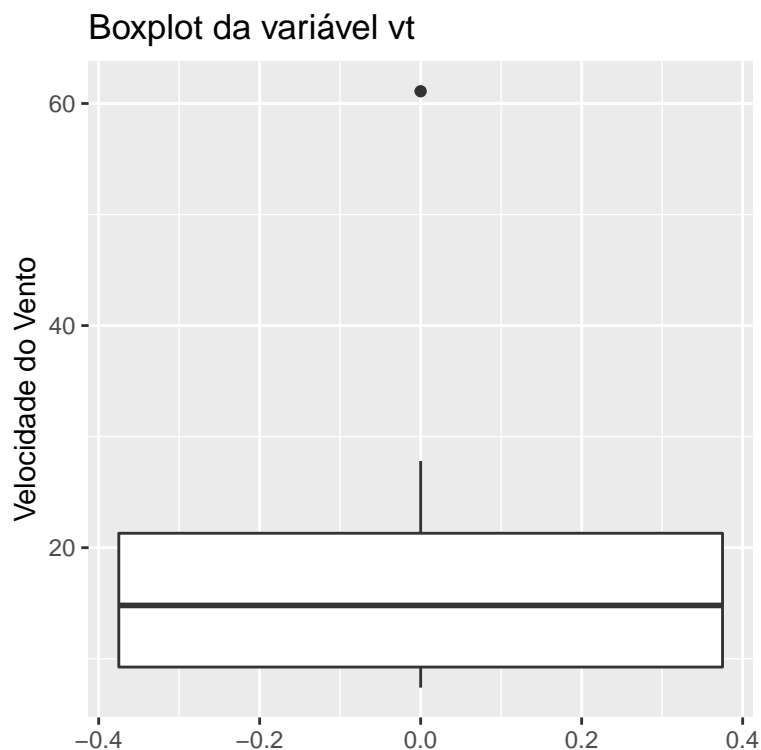
	Média	1º Quartil	Mediana	3º Quartil	Variância	Desvio Padrão	Desvio Médio	Intervalo Interquar- til
Velocidade do Vento	18.4	9.25	14.8	21.3	192.93	13.52	8.76	12.05

Tirando o valor atípico, as medidas resumo dizem que a variável vt pode ser distribuídas uniformemente.

## Exercício 7

Construa gráficos ramo-e-folhas e *boxplot* para os dados do Exercício 6.

## Resolução

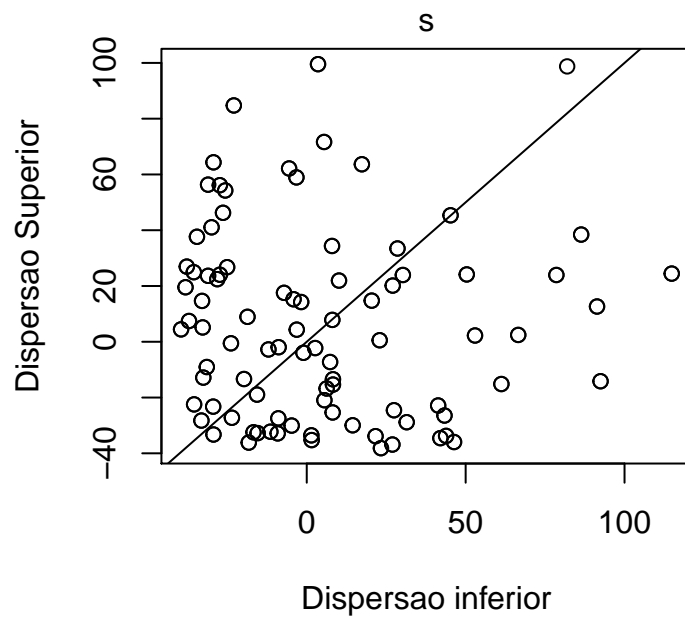
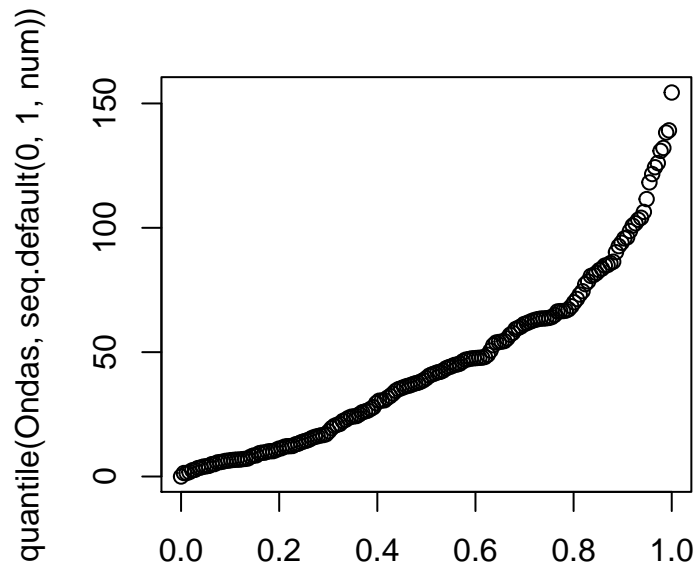


```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 0 | 7777
## 1 | 1335
## 2 | 000228
## 3 |
## 4 |
## 5 |
## 6 | 1
```

## Exercício 12

Construa gráficos de quantis e simetria para os dados de manchas solares disponíveis no arquivo *manchas*.

## Resolução



## Exercício 15

Considere os valores  $X_1, \dots, X_n$  de uma variável  $X$ , com média  $\bar{X}$  e desvio padrão  $S$ . Mostre que a variável  $Z$ , cujos valores são  $\{Z_i = (X_i - \bar{X})/S\}$ ,  $i = 1, \dots, n$  tem média 0 e desvio padrão 1.

## Resolução

Fórmula da média:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i, i = 1, \dots, n$$

Fórmula da variância:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2, i = 1, \dots, n$$

Fórmula do desvio padrão:

$$S = \sqrt{S^2}$$

Calculando a média de Z:

$$\mu_Z = \frac{1}{n} \sum_{i=1}^n (z_i) = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \mu_x}{s} \right) = \frac{1}{n_s} \sum_{i=1}^n (x_i - \mu_x) = \frac{1}{n_s} \left( \left( \sum_{i=1}^n x_i \right) - n\mu_x \right) = \frac{1}{n_s} (n\mu_x - n\mu_x) = \frac{0}{n_s} = 0$$

Calculando a variância de Z, e consequentemente, o desvio padrão:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \mu_z)^2 = \frac{1}{n} \sum_{i=1}^n (Z_i)^2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \mu_x}{S} \right)^2 = \frac{1}{nS^2} \sum_{i=1}^n (x_i - \mu_x)^2 = \frac{1}{nS^2} (nS^2) = 1$$

Então:

$$S[2] = \sqrt{S_z^2} = \sqrt{1} \Rightarrow S_z = 1$$

## Exercício 25

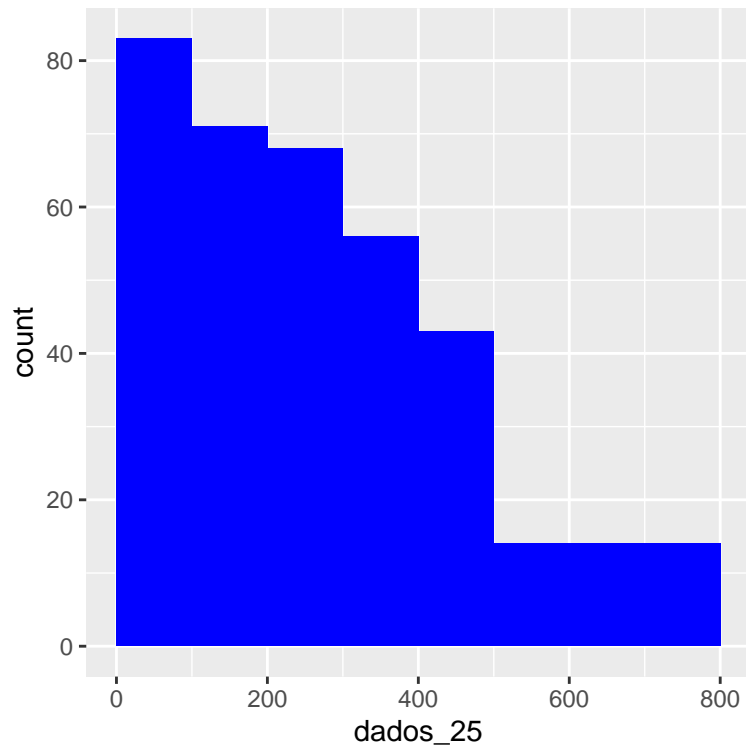
Na tabela abaixo estão indicadas as durações de 335 lâmpadas.

Duração (horas)	Número de lâmpadas
0 - 100	82
100 - 200	71
200 - 300	68
300 - 400	56
400 - 500	43
500 - 800	15

- Esboce o histograma correspondente.
- Calcule os quantis de ordem  $p=0,1, 0,3, 0,5, 0,7$  e  $0,9$ .

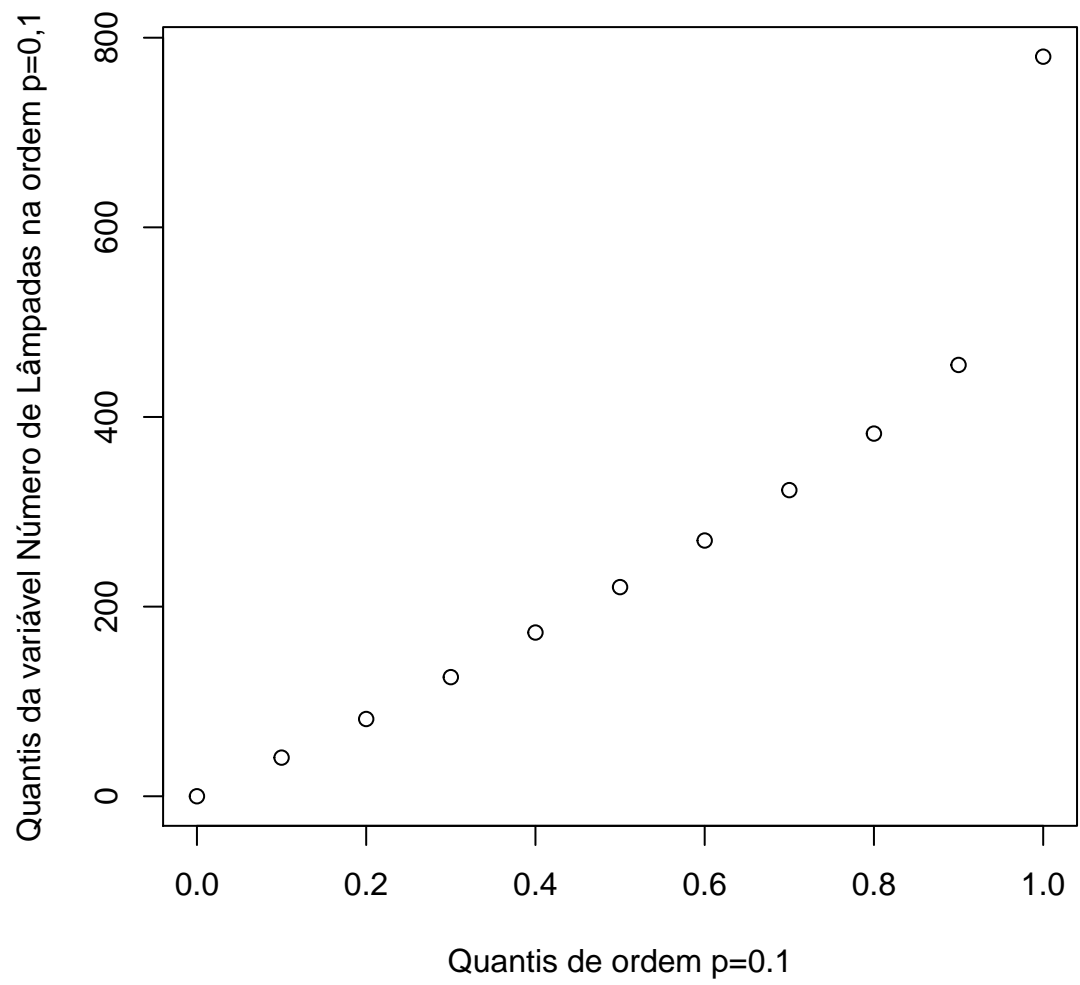
## Resolução

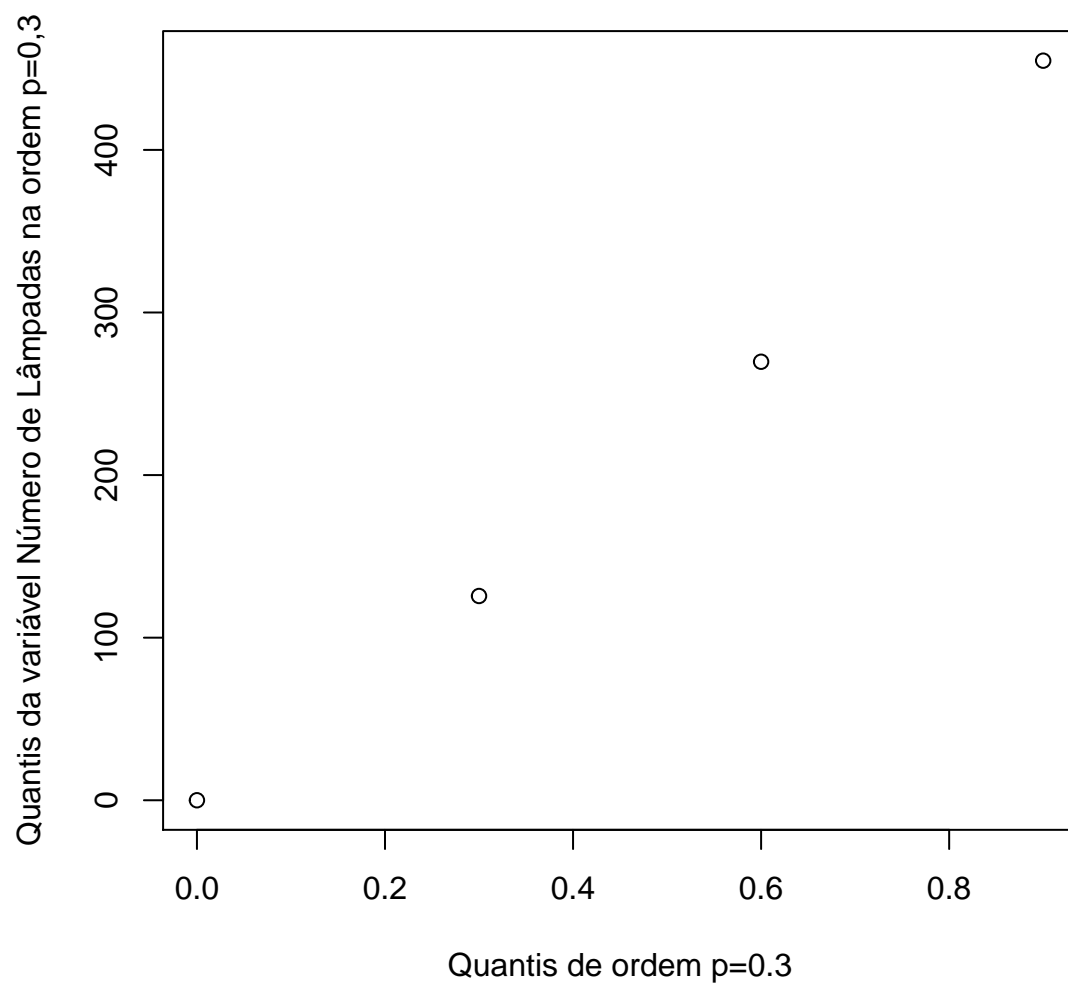
-

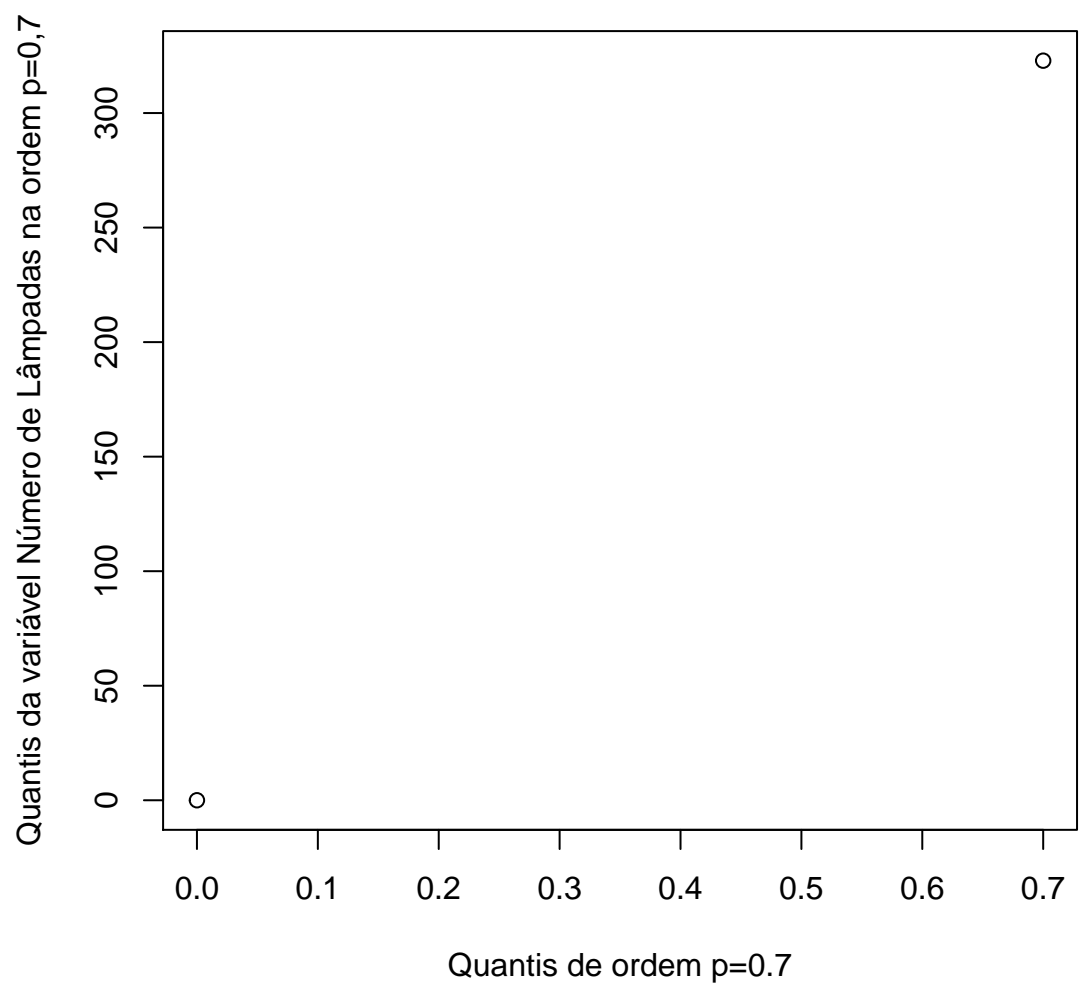


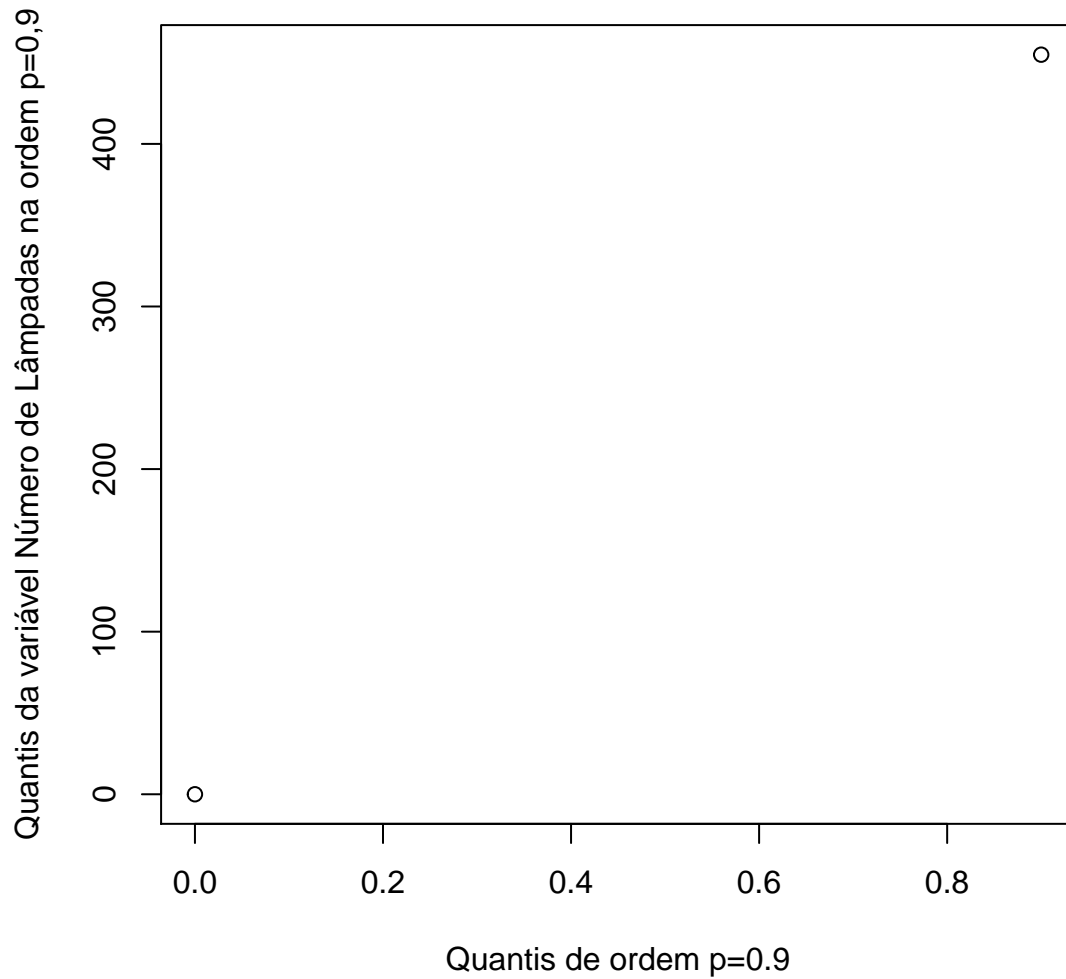
b)











### Exercício 17

Os dados apresentados na figura abaixo referem-se aos instantes nos quais o centro de controle operacional de estradas rodoviárias recebeu chamados solicitando algum tipo de auxílio em duas estradas num determinado dia.

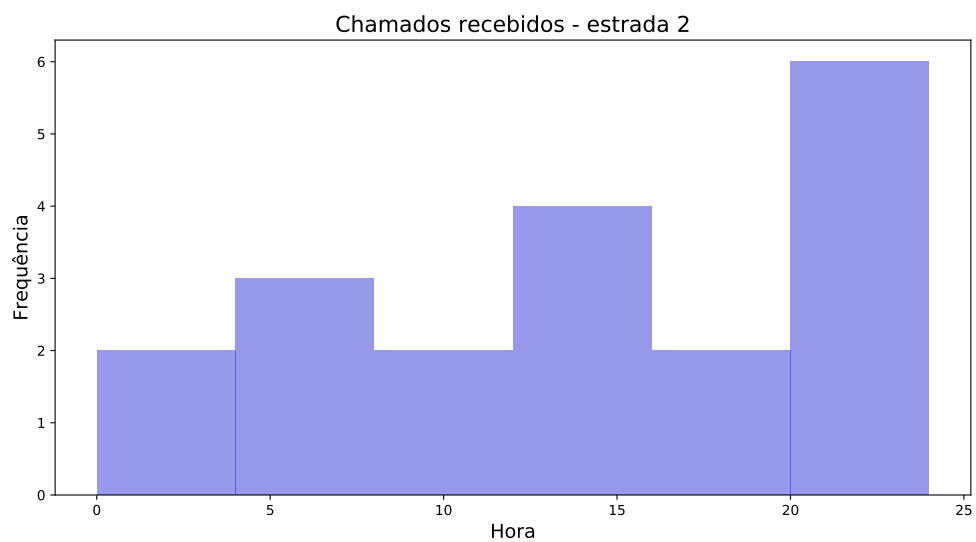
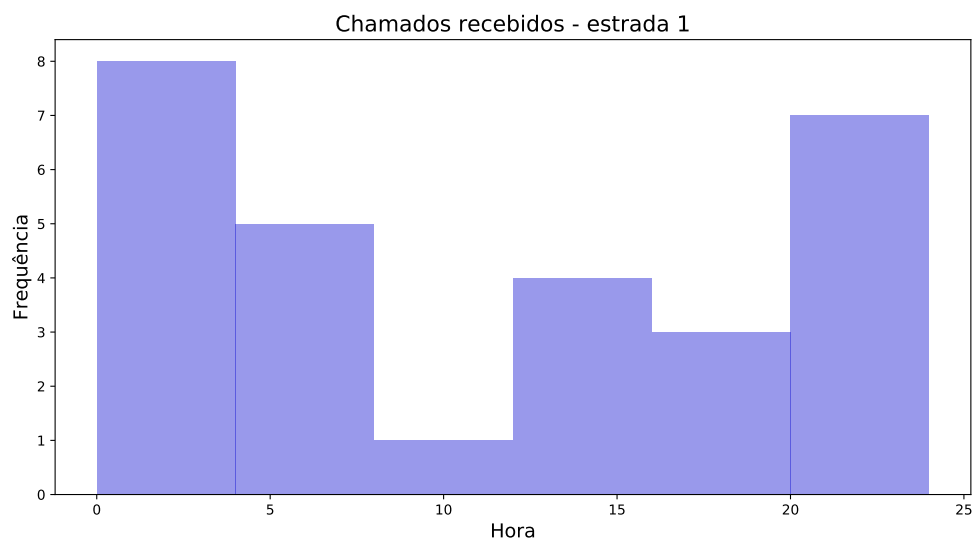
Estadra 1	12:07:00 AM	12:58:00 AM	01:24:00 AM	01:35:00 AM	02:05:00 AM
	03:14:00 AM	03:25:00 AM	03:46:00 AM	05:44:00 AM	05:56:00 AM
	06:36:00 AM	07:26:00 AM	07:48:00 AM	09:13:00 AM	12:05:00 PM
	12:48:00 PM	01:21:00 PM	02:22:00 PM	05:30:00 PM	06:00:00 PM
	07:53:00 PM	09:15:00 PM	09:49:00 PM	09:59:00 PM	10:53:00 PM
	11:27:00 PM	11:49:00 PM	11:57:00 PM		
Estrada 2	12:03:00 AM	01:18:00 AM	04:35:00 AM	06:13:00 AM	06:59:00 AM
	08:03:00 AM	10:07:00 AM	12:24:00 AM	01:45:00 PM	02:07:00 PM
	03:23:00 PM	06:34:00 PM	07:19:00 PM	09:44:00 PM	10:27:00 PM
	10:52:00 PM	11:19:00 PM	11:29:00 PM	11:44:00 PM	

- Construa um histograma para a distribuição de frequências de chamados em cada uma das estradas.
- Calcule os intervalos de tempo entre as sucessivas chamadas e descreva-os, para cada uma das estradas, utilizando medidas resumo gráficos do tipo *boxplot*. Existe alguma relação entre o tipo de estrada e o intervalo de tempo entre as chamadas ?

- c) Por intermédio de um gráfico do tipo QQ, verifique se a distribuição da variável “Intervalo de tempo entre as chamadas” em cada estrada é compátivel com um modelo Normal. Faça o mesmo para um modelo exponencial. Compare as distribuições de frequências correspondentes às duas estradas.

## Resolução

a)



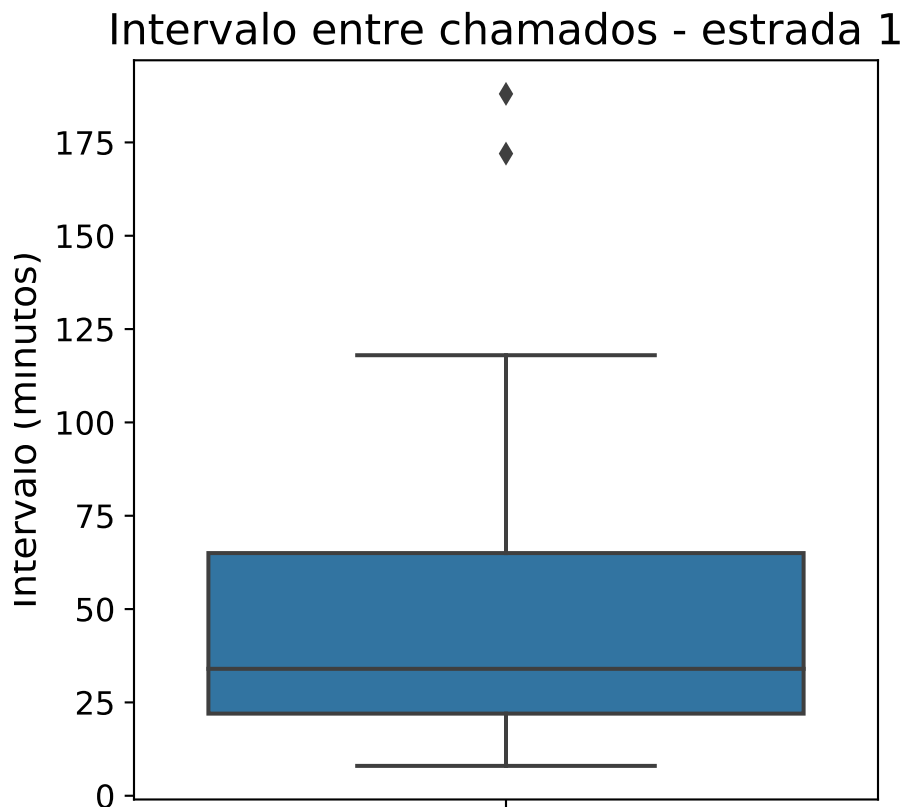
b)

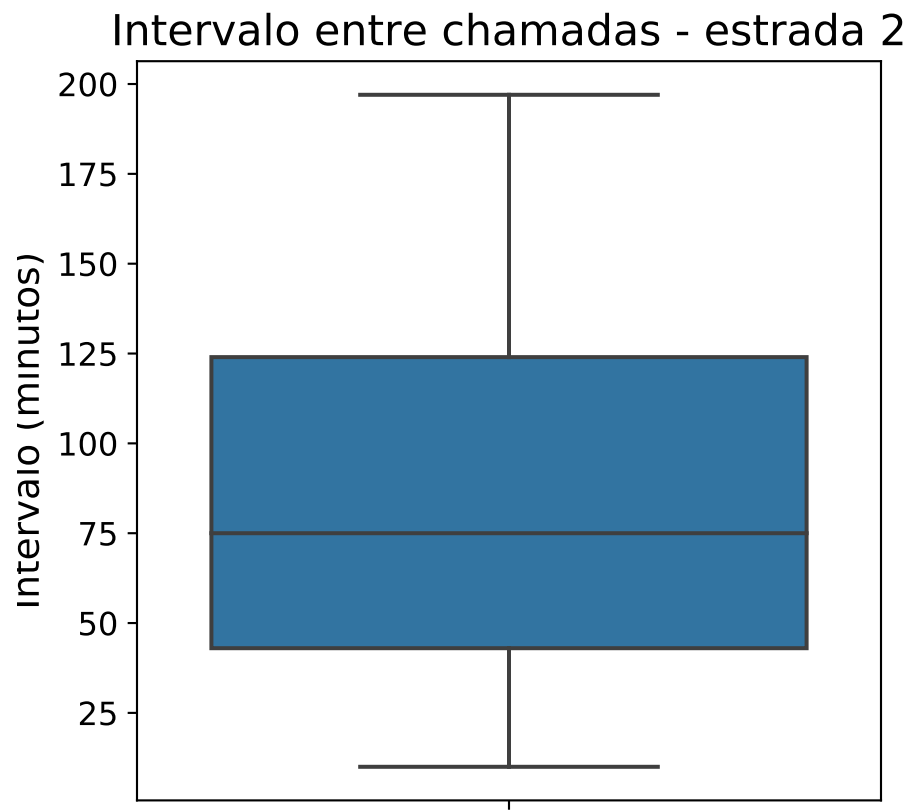
Table 9: Mediadas resumo e de dispersão das variáveis Estrada 1 e Estrada 2

Estrada	N	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão
1	27	8	22	34	53	65	188	47
2	17	10	43	75	83	124	197	58

Table 10: Rótulo das mediadas resumo e de dispersão das variáveis Estrada 1 e Estrada 2

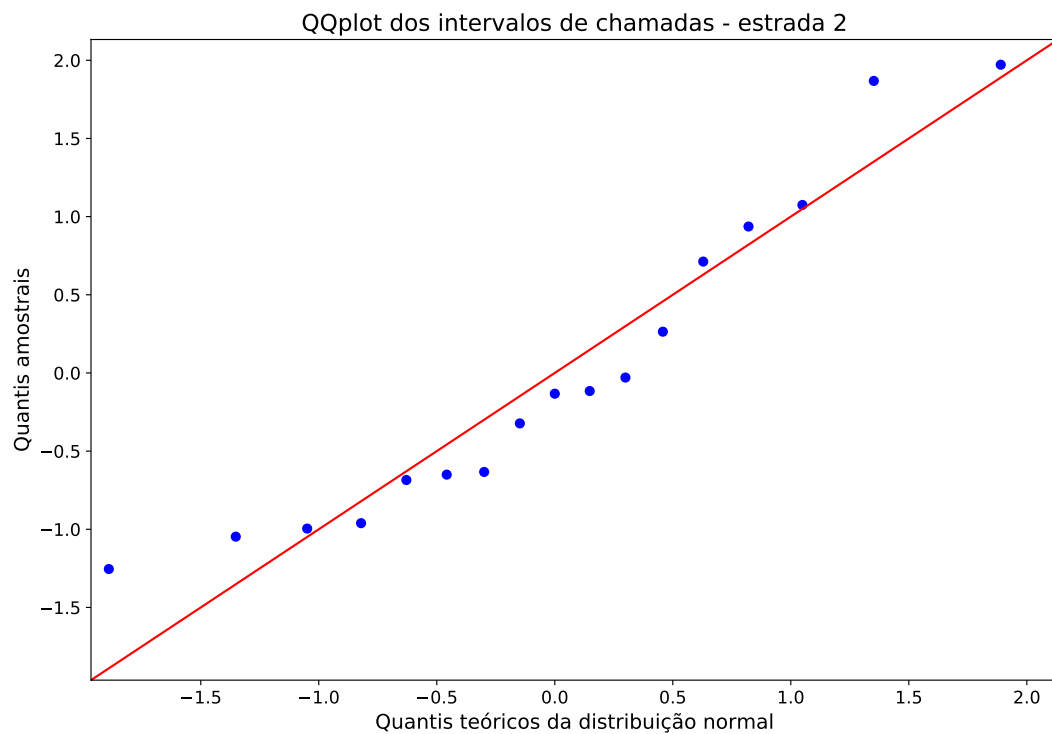
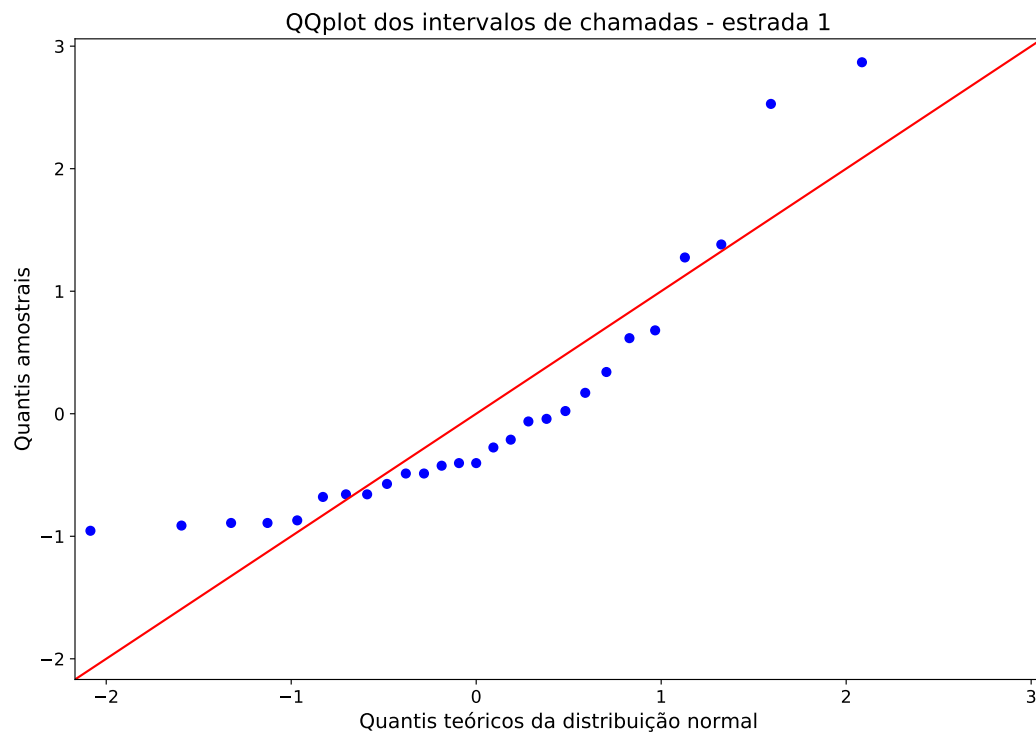
Rótulo	Descrição	Unidade de medida
Estrada	Estrada estudada	
N	Quantidade da amostra	
Mínimo	Valor mínimo da amostra	min
1º Quartil	1º Quartil da amostra	min
Mediana	Mediana da amostra	min
Média	Média da amostra	min
3º Quartil	3º Quartil da amostra	min
Máximo	Valor máximo da amostra	min
Desvio	Desvio padrão da amostra	min



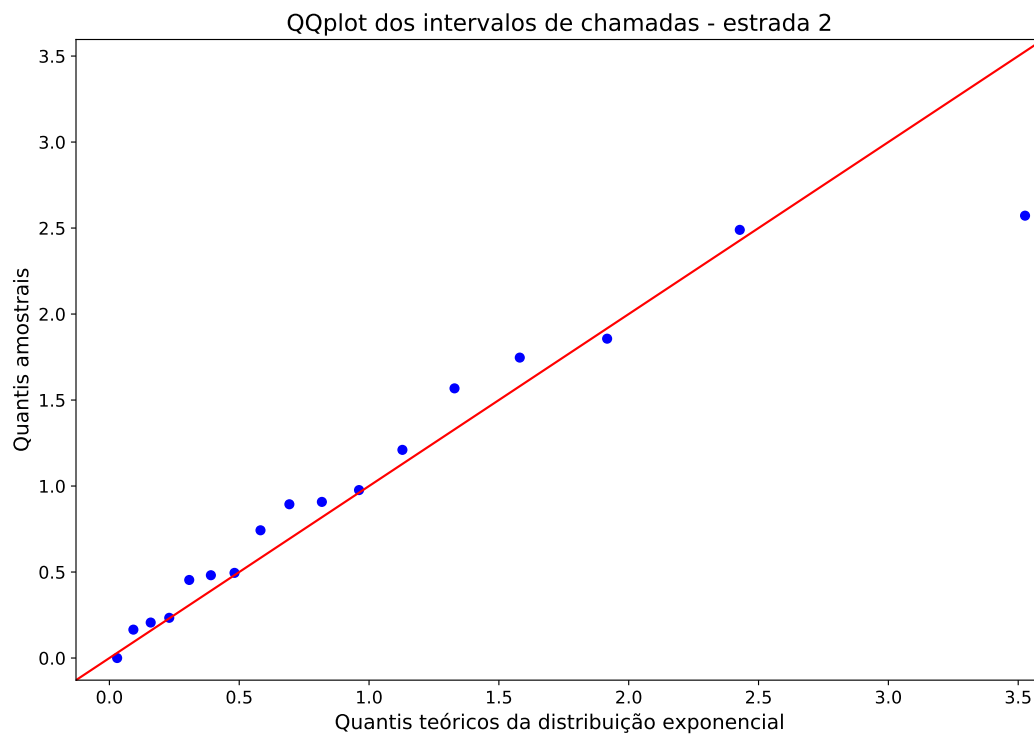
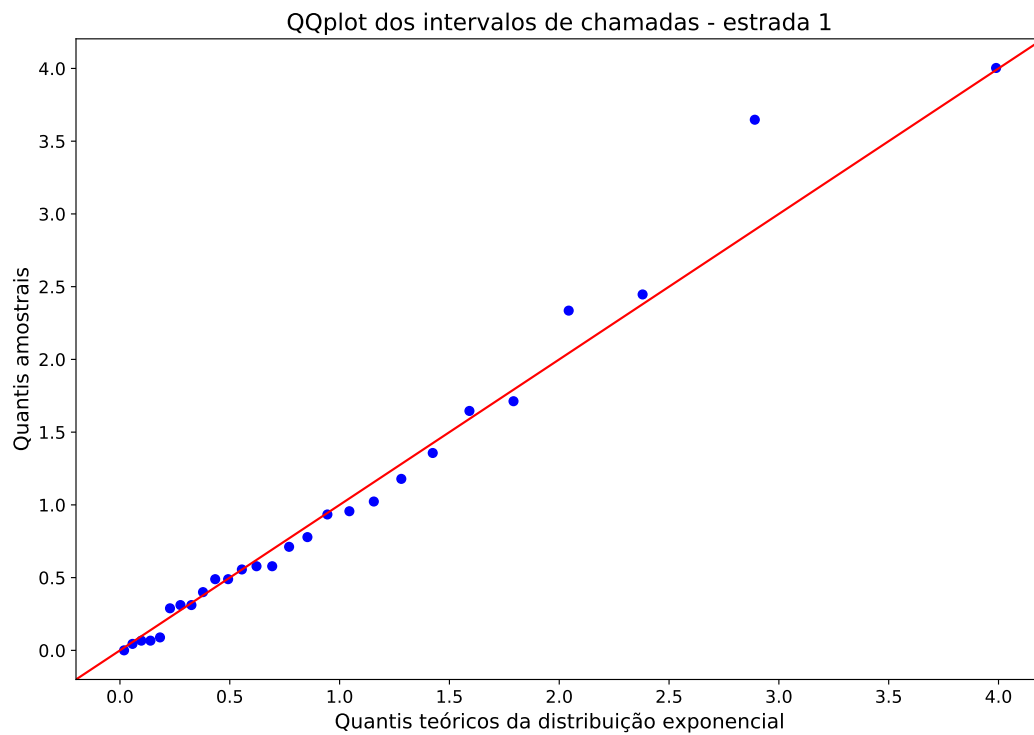


Com base nos boxplots e nas medidas descritivas, que foram transformadas da unidade hora para a unidade minuto, podemos observar que existe uma relação entre o intervalo de chamadas e o tipo de estrada, pois vemos que, na estrada 1, os intervalos entre chamadas são mais baixos que os da estrada 2, analisando de forma geral.

c)







Com base nos gráficos QQ, vemos que a distribuição dos intervalo entre chamadas sucessivas da estrada 1

se ajusta melhor a distribuição exponencial do que a distribuição normal. Já os dados da estrada 2 não se ajustam bem a nenhuma das distribuições, entretanto tem ajuste melhor a distribuição normal se comparado com a dos dados da estrada 1. Ao comparar as duas distribuições pelo qqplot, podemos dizer que as duas distribuições são, de fato, diferentes.

## Exercício 18

As notas finais de um curso de Estatística foram: 7,5,4,5,6,3,8,4,5,4,6,4,5,6,4,6,6,3,8,4,5,4,5,5 e 6.

- Determine a mediana, os quartis e a média.
- Separe o conjunto de dados em dois grupos denominados **aprovados**, com nota pelo menos igual a 5, e **reprovados**. Compare a variância desses dois grupos.

## Resolução

a)

	Média	1º Quartil	Mediana	3º Quartil
Média	5,12	3	4	5

b)

	Alunos	
	Aprovados	Reprovados
Variância	0,1944444	1,05

## Exercício 28

Os dados encontrados no arquivo *esforco* são provenientes de um estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca. As variáveis medidas durante a realização do teste foram observadas em quatro momentos distintos: repouso (REP), limiar anaeróbio (LAN), ponto de compensação respiratório (PCR) e pico (PICO). As demais variáveis são referentes às características demográficas e clínicas dos pacientes e foram registradas uma única vez.

- Descreva a distribuição da variável consumo de oxigênio (VO2) em cada um dos quatro momentos de avaliação utilizando medidas resumo (mínimo, máximo, quartis, mediana, média, desvio padrão, etc), *boxplots* e histogramas. Você identifica algum paciente com valores de consumo de oxigênio discrepantes? Interprete os resultados.
- Descreva a distribuição da classe funcional NYHA por meio de uma tabela de frequências. Utilize um método gráfico para representar essa tabela.

## Resolução

a)

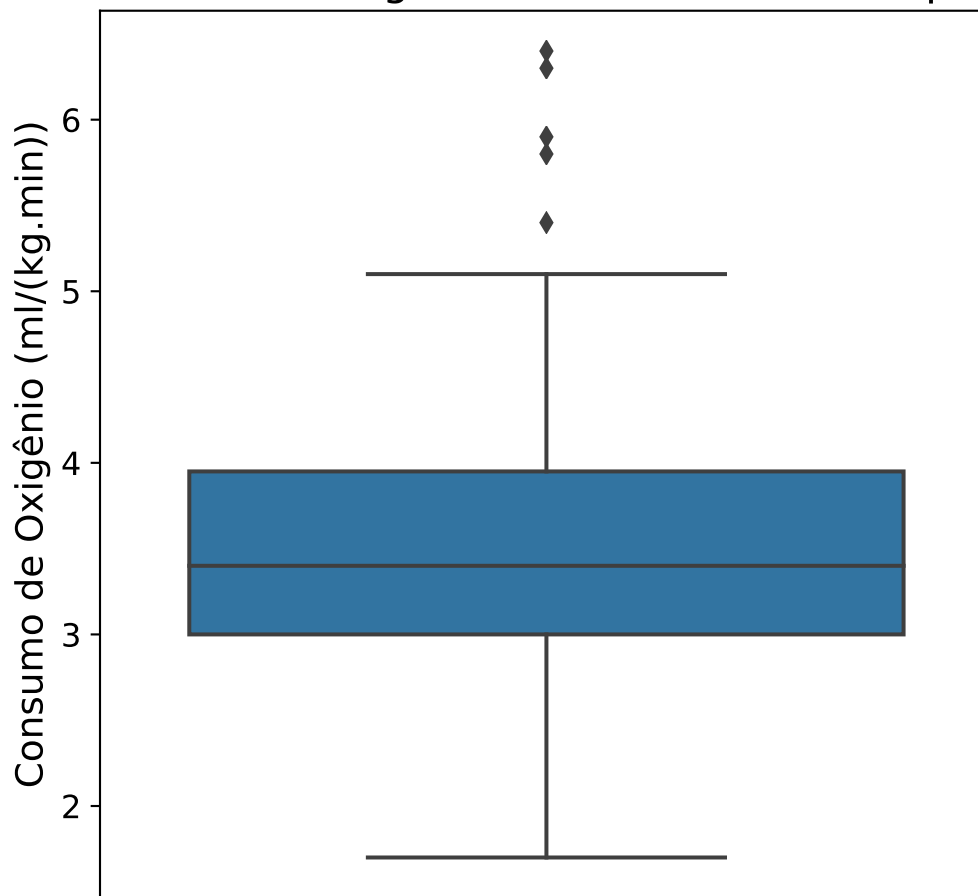
Table 11: Medidas resumo e de dispersão das variáveis Repouso, LAN, PCR e PCIO, em  $\frac{ml}{kg \cdot min}$ .

Variáveis	N	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão
Repouso	127	1.7	3.0	3.4	3.6	4.0	6.4	0.8
LAN	124	5.1	8.6	10.0	10.5	11.7	21.4	2.8
PCR	121	6.9	11.3	13.9	14.8	16.6	29.8	4.7
PICO	127	5.2	13.5	17.1	18.1	21.0	41.0	6.5

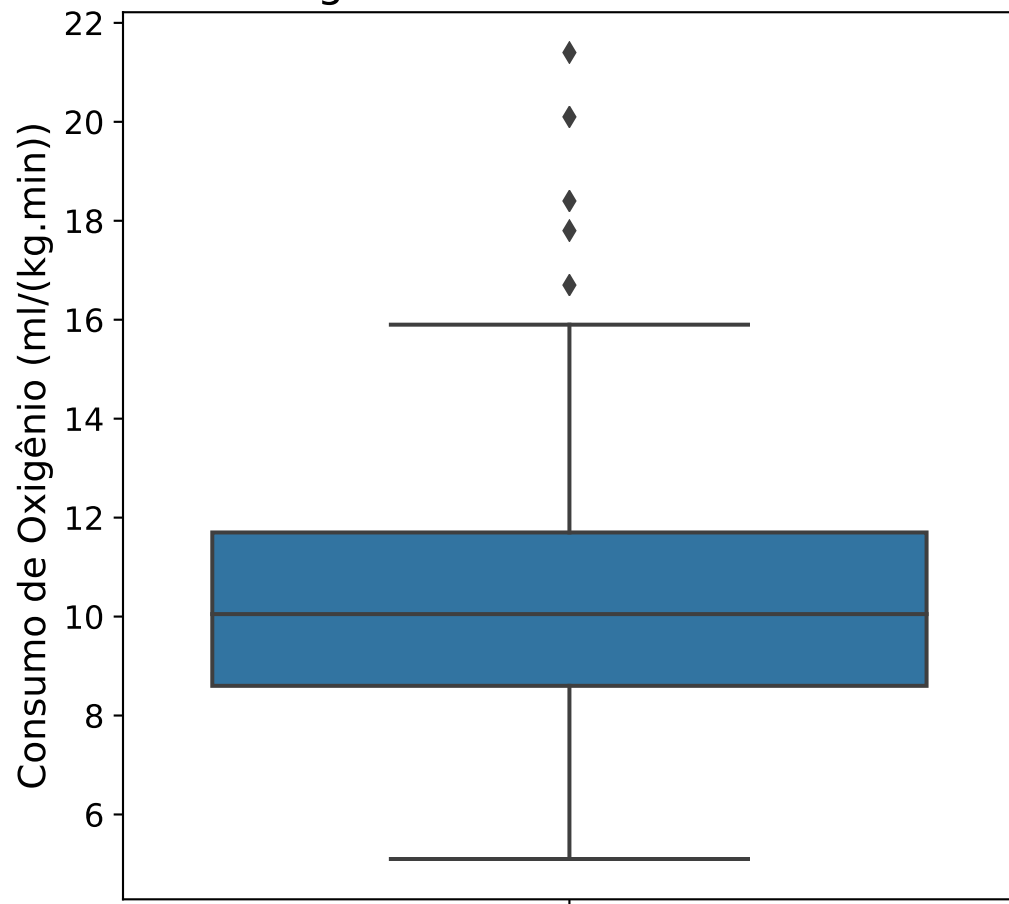
Table 12: Rótulo das mediadas resumo e de dispersão das variáveis Estrada 1 e Estrada 2

Rótulo	Descrição	Unidade de medida
Variável	Variável estudada da variável	
N	Quantidade da amostra da variável	
Mínimo	Valor mínimo da amostra da variável	$\frac{ml}{kg \cdot min}$
1º Quartil	1º Quartil da amostra da variável	$\frac{ml}{kg \cdot min}$
Mediana	Mediana da amostra da variável	$\frac{ml}{kg \cdot min}$
Média	Média da amostra da variável	$\frac{ml}{kg \cdot min}$
3º Quartil	3º Quartil da amostra da variável	$\frac{ml}{kg \cdot min}$
Máximo	Valor máximo da amostra da variável	$\frac{ml}{kg \cdot min}$
Desvio	Desvio padrão da amostra da variável	$\frac{ml}{kg \cdot min}$

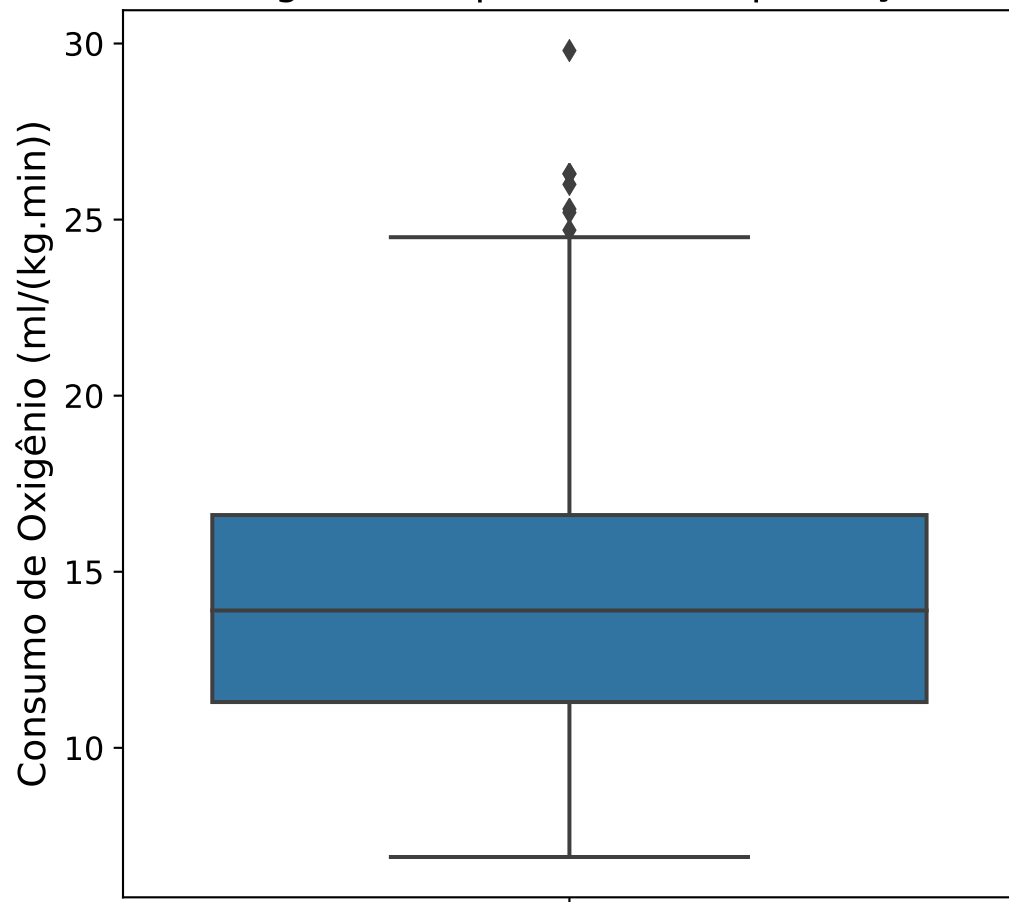
## Consumo de oxigênio em momento de repouso

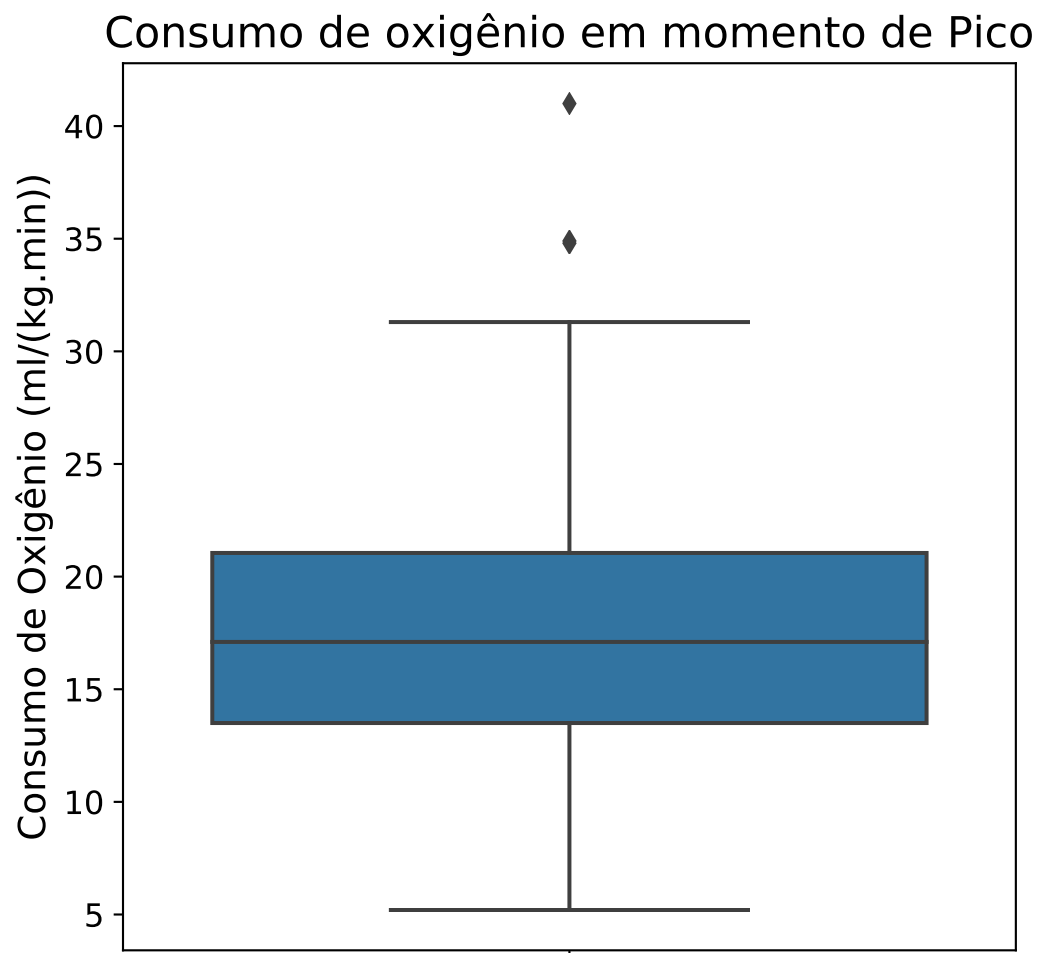


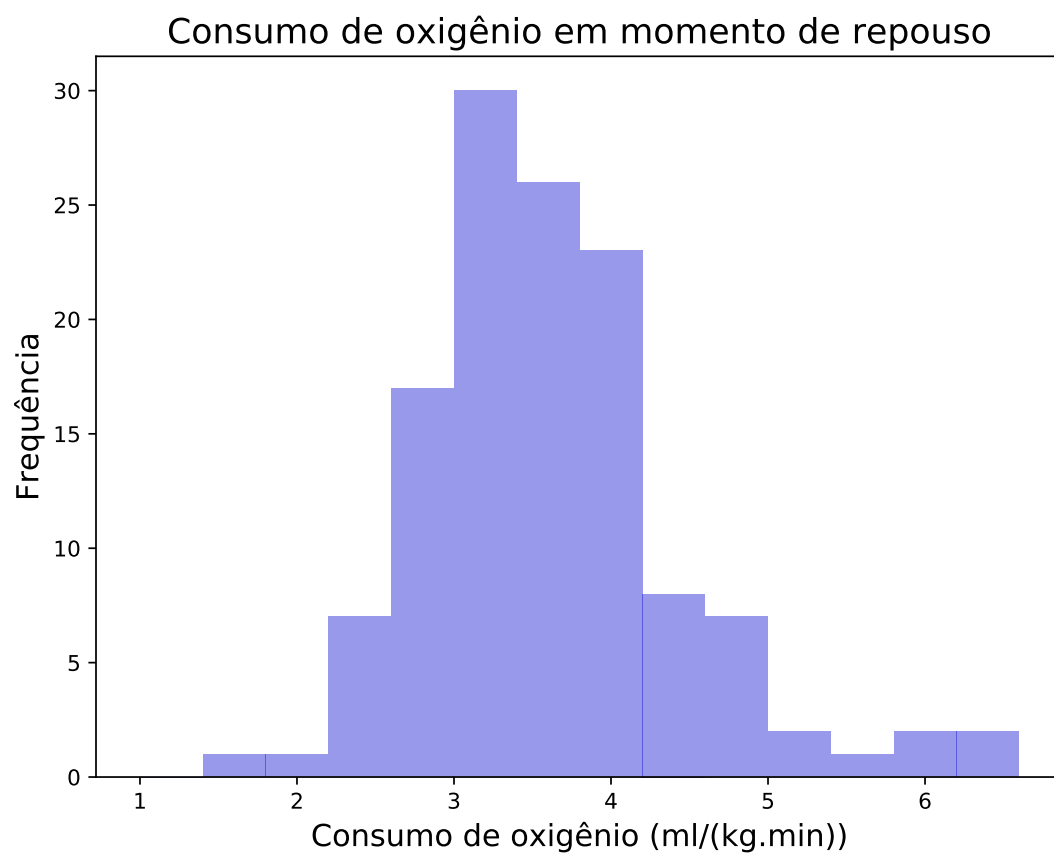
Consumo de oxigênio em momento de limiar anaeróbi

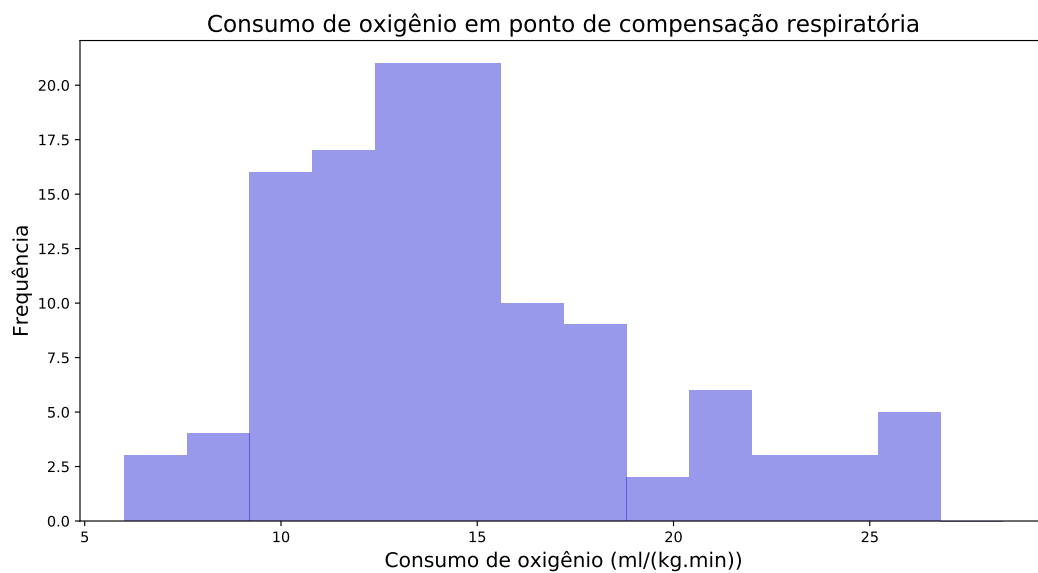
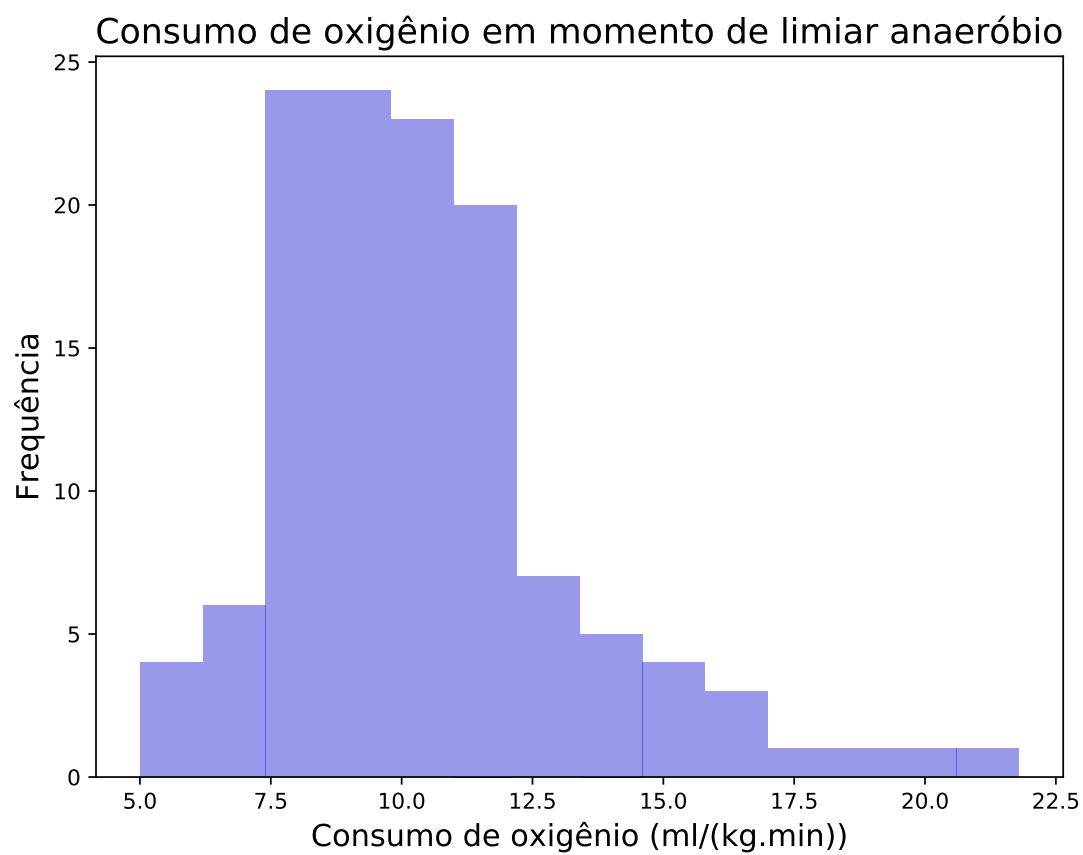


Consumo de oxigênio em ponto de compensação respiratório

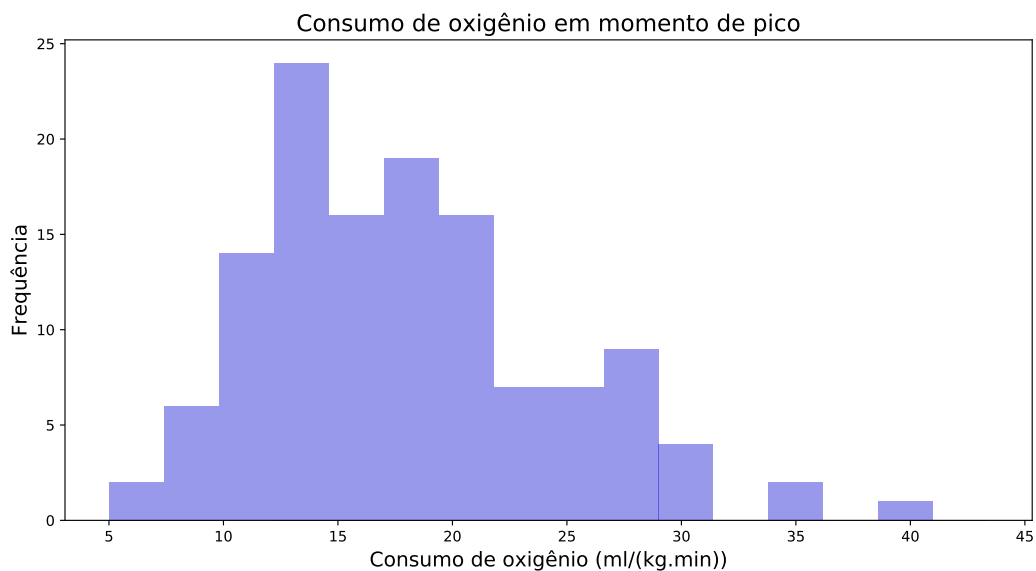










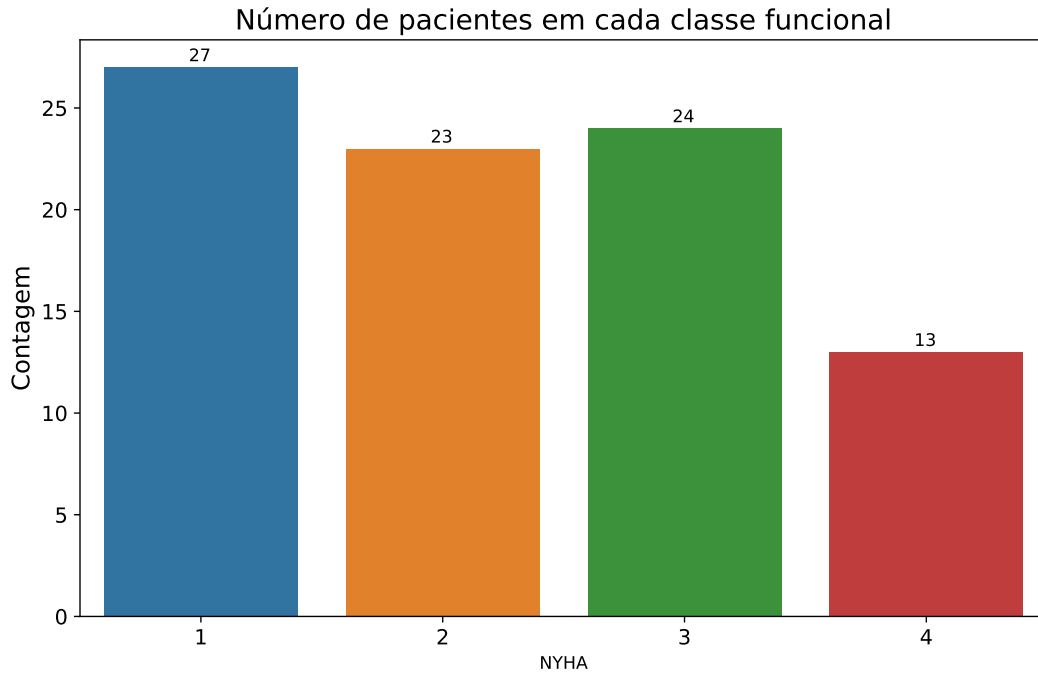


Podemos identificar algumas medidas discrepantes nos quatro gráficos boxplot. Além disso, notamos que a distribuição do consumo de oxigênio é semelhante nos quatro momentos em que foi medido, o consumo de oxigênio tende a subir de acordo com momento, sendo o Pico o momento com valores mais altos e Repouso o com valores mais baixos.

b)

Table 13: Distribuição de frequência do número de pacientes da variável NYHA

Classe Funcional NYHA	Frequência	Frequência acumulada	Frequência relativa (%)	Frequência relativa acumulada (%)
1	27	27	31	31
2	23	50	26	57
3	24	74	28	85
4	13	87	15	100
Total	87	67	100	100



### Exercício 32

Considere uma amostra aleatória simples  $X_1, \dots, X_n$  de uma variável  $X$  que assume o valor 1 com probabilidade  $0 < p < 1$  e o valor 0 com probabilidade  $1 - p$ . Seja  $\hat{p} = n^{-1} \sum_{i=1}^n X_i$ . Mostre que

i)  $E(X_i) = p$  e  $\text{Var}(X_i) = p(1-p)$ .

ii)  $E(\hat{p}) = p$  e  $\text{Var}(\hat{p}) = p(1-p)/n$ .

iii)  $0 < \text{Var}(X_i) < 0,25$ .

Com base nesses resultados, utilize o Teorema Limite Central [ver Sen et al. (2009), por exemplo] para construir um intervalo de confiança aproximado conservador (i.e. com a maior amplitude possível) para  $p$ . Utilize o Teorema de Sverdrup [ver Sen et al. (2009), por exemplo] para construir um intervalo de confiança aproximado para  $p$  com amplitude menor que a do intervalo mencionado acima.

### Resolução

i)

$$E(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$\text{Var}(x) = E(x^2) - [E(x)]^2 = (0^2 \cdot (1 - p) + 1 \cdot 1) - p^2 = p - p^2 = p \cdot (1 - p)$$

ou

$$\text{Var}(x) = (0 \cdot p) \cdot (1 - p) + (1 - p)^2 \cdot p = p^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p \cdot (1 - p) \cdot [p + (1 - p)] = p \cdot (1 - p)$$

ii)

$$E(\hat{p}) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i)\right) = \frac{1}{n} \cdot E\left(\sum_{i=1}^n X_i\right) = \frac{n \cdot p}{n} = p$$

$$Var(\hat{p}) = Var\left(\frac{1}{n} \sum_{i=1}^n (X_i)\right) = \frac{1}{n^2} \cdot E\left(\sum_{i=1}^n X_i\right) = \frac{n \cdot p \cdot (1-p)}{n^2} = \frac{p \cdot (1-p)}{n}$$

iii)

$$\frac{\partial}{\partial p} \cdot p \cdot (1-p) = 1 - 2 \cdot p = 0 \Rightarrow p = \frac{1}{2}$$

Logo, a variância máxima é atingida quando

$$p = \frac{1}{2}$$

, então

$$\frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

.

A variância é mínima quando  $p = 0$ , então

$$0 \cdot (1 - 0) = 0$$

.

Portanto,

$$0 < Var(X) < 0.25$$

.

$$\hat{p} \sim N\left(\hat{p}, \frac{\hat{p} \cdot (1-\hat{p})}{n}\right)$$

$$\begin{aligned} \mathcal{P}\left(-Z_\gamma < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}}\right) &= \mathcal{P}\left(-Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} < \hat{p} - p < Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right) = \\ \mathcal{P}\left(-\hat{p} - Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} < -p < \hat{p} + Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right) &= \mathcal{P}\left(\hat{p} - Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} < -p < \hat{p} + Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right) \end{aligned}$$

Intervalo de confiança aproximadamente conservador ( $\hat{p} \cdot (1-\hat{p}) = \frac{1}{4}$ )

$$IC(p, \gamma) = \left[ \hat{p} \pm Z_\gamma \cdot \sqrt{\frac{1}{4 \cdot n}} \right] = \left[ \hat{p} \pm \frac{Z_\gamma}{2 \cdot \sqrt{n}} \right]$$

Intervalo de confiança aproximado para p com amplitude menor

$$IC(p, \gamma) = \left[ \hat{p} \pm Z_\gamma \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \right]$$