

Estatística Descritiva II

Exemplo

Dados do projeto CEA15P01: “Avaliação da fadiga em mulheres com câncer de mama durante o tratamento quimioterápico”

- Estudo realizado pela Escola de Enfermagem da Universidade de São Paulo.
- **Objetivo:** Quantificar uma possível associação entre fadiga, qualidade de vida e nível de hemoglobina em mulheres com câncer de mama em tratamento quimioterápico.
- **Coleta dos dados:** junho/2010 a maio/2011.
- **Análise estatística:** realizada pelo Centro de Estatística Aplicada (CEA), IME-USP, em 2015.

Exemplo

Dados do projeto CEA15P01: “Avaliação da fadiga em mulheres com câncer de mama durante o tratamento quimioterápico”

- **Amostra:** 30 mulheres com idades acima de 19 anos, com diagnóstico de câncer de mama, tratadas no Ambulatório de Mastologia da Faculdade de Medicina da USP-Ribeirão Preto, no período de junho/2010 a maio/2011.
- **Dados coletados:**
 - * Variáveis demográficas e biométricas
 - * Fadiga e qualidade de vida: mensuradas por meio do “Questionário FACIT-F versão 4”
 - * Nível de hemoglobina (g/dL)

Exemplo

Dados do projeto CEA15P01: “Avaliação da fadiga em mulheres com câncer de mama durante o tratamento quimioterápico”

- **Amostra:** 30 mulheres com idades acima de 19 anos, com diagnóstico de câncer de mama, tratadas no Ambulatório de Mastologia da Faculdade de Medicina da USP-Ribeirão Preto, no período de junho/2010 a maio/2011.
- **Dados coletados:**
 - * Variáveis demográficas e biométricas
 - * Fadiga e qualidade de vida: mensuradas por meio do “Questionário FACIT-F versão 4”
 - * Nível de hemoglobina (g/dL)
- Dados coletados nas sessões 1, 5 e 8 de quimioterapia (dados longitudinais)



Variáveis

- **Idade** (em anos)
- **Etnia** (codificação no arquivo de dados):
branca (1), preta (2), parda (3), amarela (4)
- **Estado Civil** (codificação no arquivo de dados):
casada (1), solteira (2), separada (3), viúva (4)
- **Escolaridade** (codificação no arquivo de dados):
ensino fundamental incompleto (1), ensino fundamental completo (2), ensino médio incompleto (3), ensino médio completo (4), ensino superior incompleto (5), ensino superior completo (6)

Variáveis

- **Situação do trabalho** (codificação no arquivo de dados):
ativa (1), está de licença saúde (2), desempregada (3), pensionista (4)
- **Fumante** (codificação no arquivo de dados):
fumante (2), não fumante (1)
- **Peso:** em Kg
- **IMC - Índice de massa corporal:** em Kg/m²

Variáveis

- **Situação do trabalho** (codificação no arquivo de dados):
ativa (1), está de licença saúde (2), desempregada (3), pensionista (4)
- **Fumante** (codificação no arquivo de dados):
fumante (2), não fumante (1)
- **Peso:** em Kg
- **IMC - Índice de massa corporal:** em Kg/m^2

Variáveis

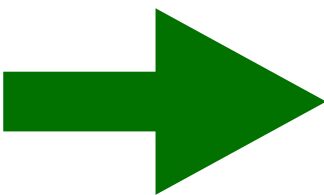
Questionário FACIT-F

- Dividido em 4 domínios: BES, BEFi, BEE, BEF
 - Cada domínio tem 6 ou 7 questões
 - * Cada questão tem pontuação de 0 a 4
- Após o paciente responder, calcula-se o escore de cada um dos domínios
- **Exemplo:** pontuação BEFi
 - 0 a 28 - quanto maior o escore, melhor a sensação de bem-estar

Banco de dados

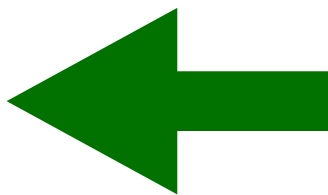
Variáveis demográficas e biométricas da paciente.

Não variam com os tempos de coleta.



indivíduo	Idade	Etnia	EstCivil	Escolar	SitTrab	Fumante	Peso	Altura	IMC
1	49	1	1	1	4	1	70	152	30,3
2	47	1	1	1	4	2	75	162	28,62
3	68	3	4	1	4	2	72,7	146	33,8
4	69	1	1	1	4	1	58	149	26,12
5	58	1	2	1	1	2	77,7	156	31,97
6	63	1	1	1	3	1	59	156	24,27
7	56	2	4	1	5	1	79	161	30,5
8	24	1	2	4	2	1	59,4	160	23,2
9	49	1	1	1	1	2	68	160	26,56
10	41	1	4	1	5	1	65	152	28,13

indivíduo	HB	BEFi	BES	BEE	BEF	SF	FACT.G	Tempo	Idade
1	13,1	28	20	18	18	49	84	1	49
2	12	23	8	10	11	35	52	1	47
3	12,8	24	12	20	15	49	71	1	68
4	13,5	28	17	23	20	48	88	1	69
5	13,7	24	17	19	18	45	78	1	58
6	12,6	22	15	18	18	44	73	1	63
7	14,4	20	14	20	15	39	69	1	56
8	13,9	28	22	19	22	51	91	1	24
.
.
.
1	11,9	28	23	23	22	47	96	5	49
2	10,1	13	9	20	12	23	54	5	47
3	12,4	22	9	20	10	37	61	5	68
4	10,9	21	19	20	17	32	77	5	69

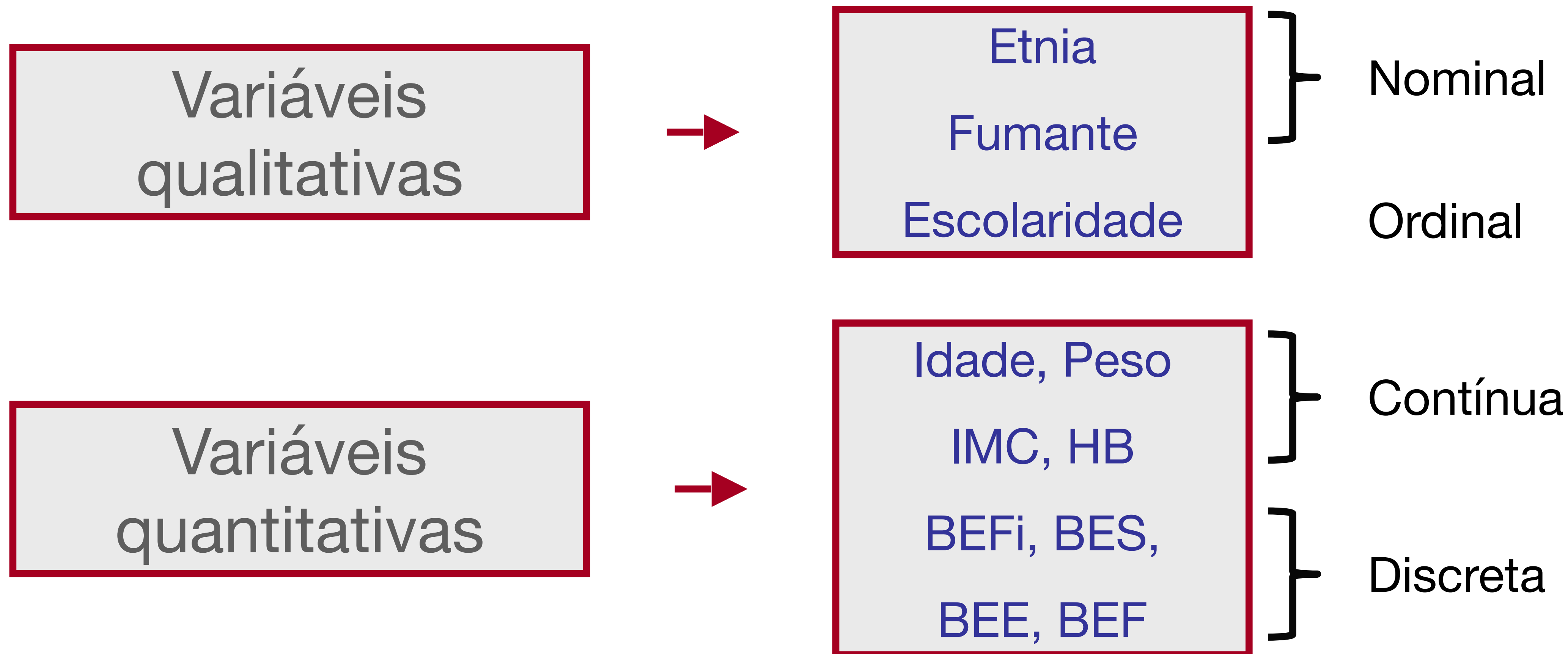


Variáveis do Questionário FACIT-F e do nível de Hemoglobina.

Avaliadas em cada sessão de quimioterapia: sessões 1, 5 e 8.

Banco de dados

Classificação das variáveis



Variáveis quantitativas

Medidas resumo

- Medidas de posição:
 - * Média (\bar{x})
 - * Mediana (md/ Q_2)
 - * Quartis (Q_1/Q_3)
 - * Máximo (max)
 - * Mínimo (min)
- Medidas de dispersão:
 - * Variância (s^2)
 - * Desvio padrão (s)
 - * Distância interquartil ($DI = Q_3 - Q_1$)
 - * Coeficiente de variação (cv)

Apoio computacional

- Linguagem R

<http://cran.rstudio.com/>

- Ambiente de desenvolvimento RStudio

<http://rstudio.com/>



```
> library(readxl)
> CEA2015P01_QV_Demograficos <- read_excel("CEA2015P01-QV-Demograficos.xlsx")
```

```
> CEA2015P01_QV_Demograficos
```

```
# A tibble: 30 x 10
```

	indivíduo	Idade	Etnia	EstCivil	Escolaridade	SitTrab	Fumante	Peso	Altura	IMC
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	49	1	1	1	4	1	70	152	30.3
2	2	47	1	1	1	4	2	75	162	28.6
3	3	68	3	4	1	4	2	72.7	146	33.8
4	4	69	1	1	1	4	1	58	149	26.1
5	5	58	1	2	1	1	2	77.7	156	32.0
6	6	63	1	1	1	3	1	59	156	24.3
7	7	56	2	4	1	5	1	79	161	30.5
8	8	24	1	2	4	2	1	59.4	160	23.2
9	9	49	1	1	1	1	2	68	160	26.6
10	10	41	1	4	1	5	1	65	152	28.1

```
# ... with 20 more rows
```

```
> |
```



```
> summary(CEA2015P01_QV_Demograficos[,c(2,8,9,10)])
```

Idade		Peso		Altura		IMC	
Min.	:24.00	Min.	: 58.00	Min.	:146.0	Min.	:23.20
1st Qu.	:45.50	1st Qu.	: 68.25	1st Qu.	:152.5	1st Qu.	:26.49
Median	:49.00	Median	: 73.10	Median	:158.0	Median	:30.40
Mean	:49.87	Mean	: 76.64	Mean	:157.8	Mean	:30.84
3rd Qu.	:57.00	3rd Qu.	: 85.90	3rd Qu.	:161.0	3rd Qu.	:34.16
Max.	:69.00	Max.	:111.00	Max.	:175.0	Max.	:41.00

- 50% das pacientes tem idade menor ou igual a 49 anos. A idade mínima é de 24 anos e a idade máxima é de 69 anos
- 25% das pacientes tem IMC inferior a 26,49 kg/m²
- IMC de 75% das pacientes foi igual ou menor a 34,16 kg/m²
- A média da idade das pacientes foi de 49,87 anos
- O IMC médio das pacientes foi de 30,40 kg/m²

```
> sd(CEA2015P01_QV_Demograficos$Idade)
[1] 10.9976
> sd(CEA2015P01_QV_Demograficos$Idade)/mean(CEA2015P01_QV_Demograficos$Idade)*100
[1] 22.054
> sd(CEA2015P01_QV_Demograficos$IMC)
[1] 5.133641
> sd(CEA2015P01_QV_Demograficos$IMC)/mean(CEA2015P01_QV_Demograficos$IMC)*100
[1] 16.64335
```

- O desvio padrão da idade das pacientes foi de 11 anos
- O desvio padrão do IMC foi de 5,13 kg/m²
- As pacientes apresentaram maior variabilidade quanto à idade, quando comparada ao IMC (22% e 16,6%, respectivamente)

```
> by(CEA2015P01_QV_Demograficos$Idade, CEA2015P01_QV_Demograficos$Fumante, summary)
```

```
CEA2015P01_QV_Demograficos$Fumante: 1
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
24.00	42.00	50.00	48.42	57.00	69.00

```
-----  
CEA2015P01_QV_Demograficos$Fumante: 2
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
45.00	47.00	49.00	52.36	53.50	68.00

- A média da idade das pacientes não fumantes é menor que a de fumantes
- Contudo, 75% das pacientes do grupo não fumante tem idades inferiores a 57 anos e no grupo de fumantes este valor é de 53 anos
- A diferença interquartil no grupo de não fumantes é maior.


```

> by(CEA2015P01_QV_Demograficos$Idade, CEA2015P01_QV_Demograficos$Fumante, sd)
CEA2015P01_QV_Demograficos$Fumante: 1
[1] 12.23073
-----
CEA2015P01_QV_Demograficos$Fumante: 2
[1] 8.405626
> cv <- function(x) sd(x)/mean(x)*100
> by(CEA2015P01_QV_Demograficos$Idade, CEA2015P01_QV_Demograficos$Fumante, cv)
CEA2015P01_QV_Demograficos$Fumante: 1
[1] 25.25911
-----
CEA2015P01_QV_Demograficos$Fumante: 2
[1] 16.05241

```

- O desvio padrão e o coeficiente de variação da idade no grupo de não fumantes é maior.

Variáveis quantitativas

Distribuições de frequências

- Os dados também podem ser resumidos por meio de tabelas de distribuição de frequências
- Distribuição de frequências de uma variável é uma lista dos valores individuais ou dos intervalos de valores que a variável pode assumir, com as respectivas frequências de ocorrência

Distribuição de frequências

Variável IMC do projeto “Qualidade de vida”

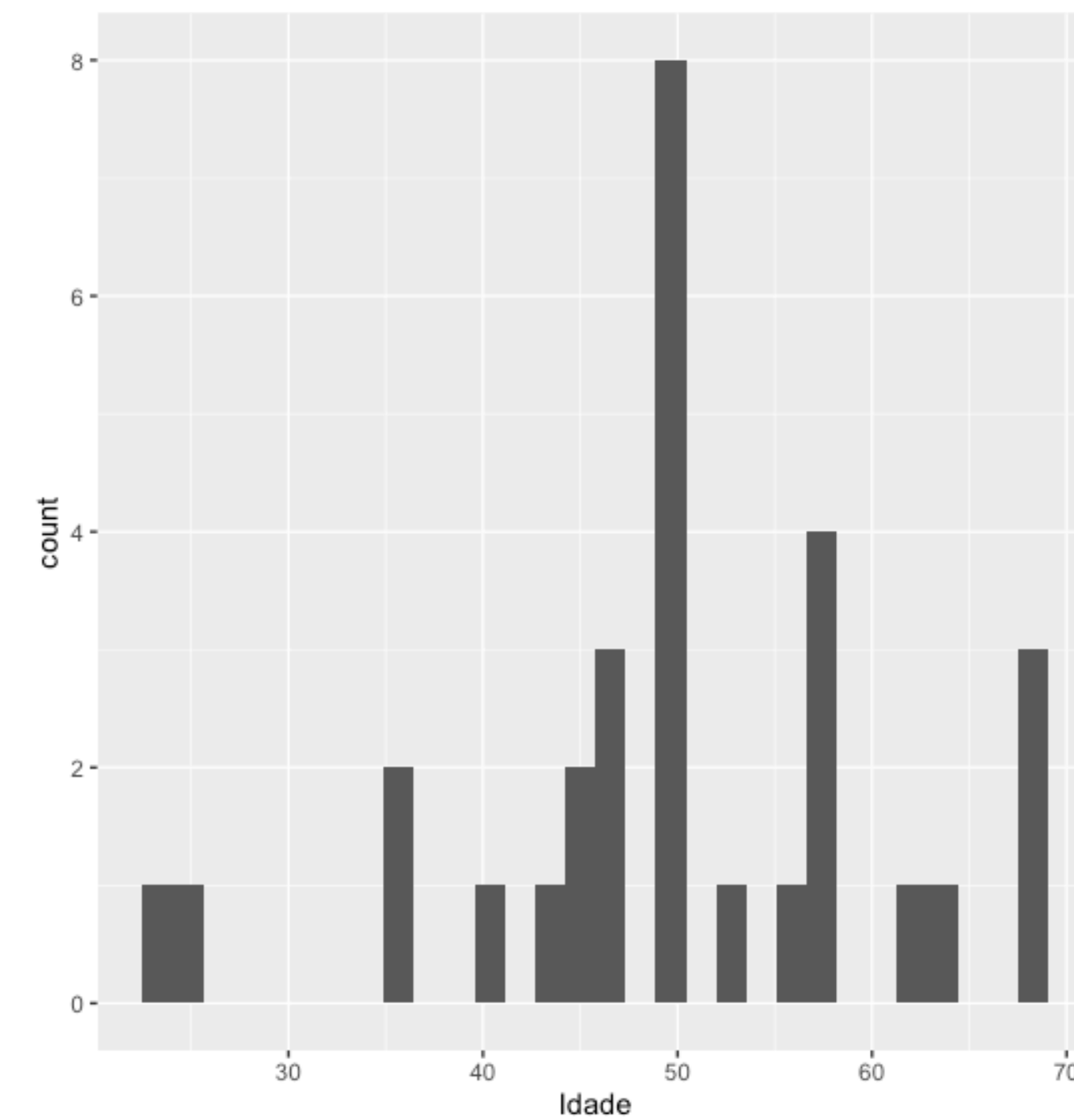
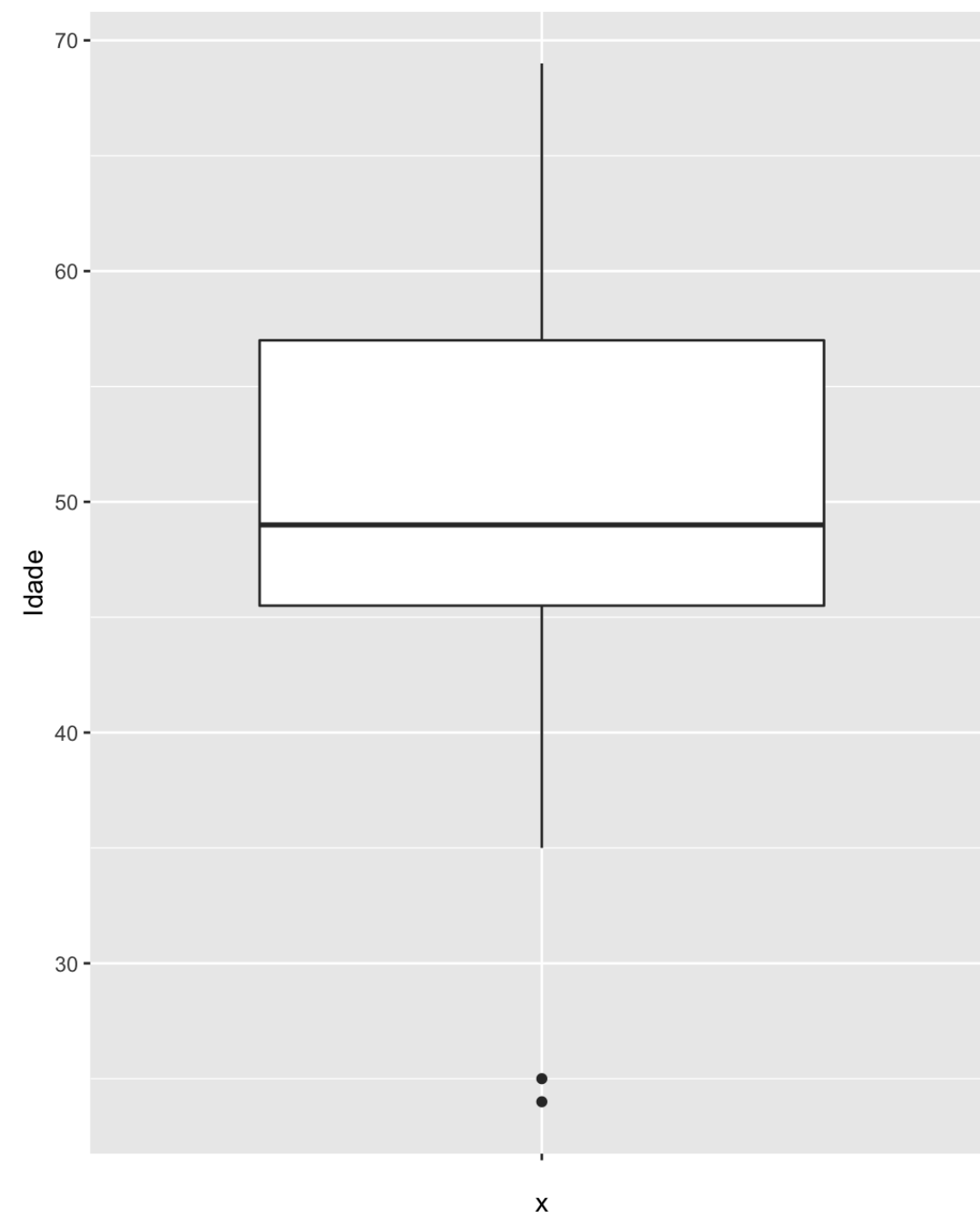
Em geral, para uma variável contínua, o agrupamento em classes leva a uma perda de informação (os dados originais não podem ser recuperados a partir da tabela)

IMC	Frequência	Frequência relativa (%)
(23,0; 26,7]	9	30%
(26,7; 30,3]	6	20%
(30,3; 33,9]	7	23,3%
(33,9; 37,5]	4	13,3%
(37,5; 41]	4	13,3%
Total	30	100%

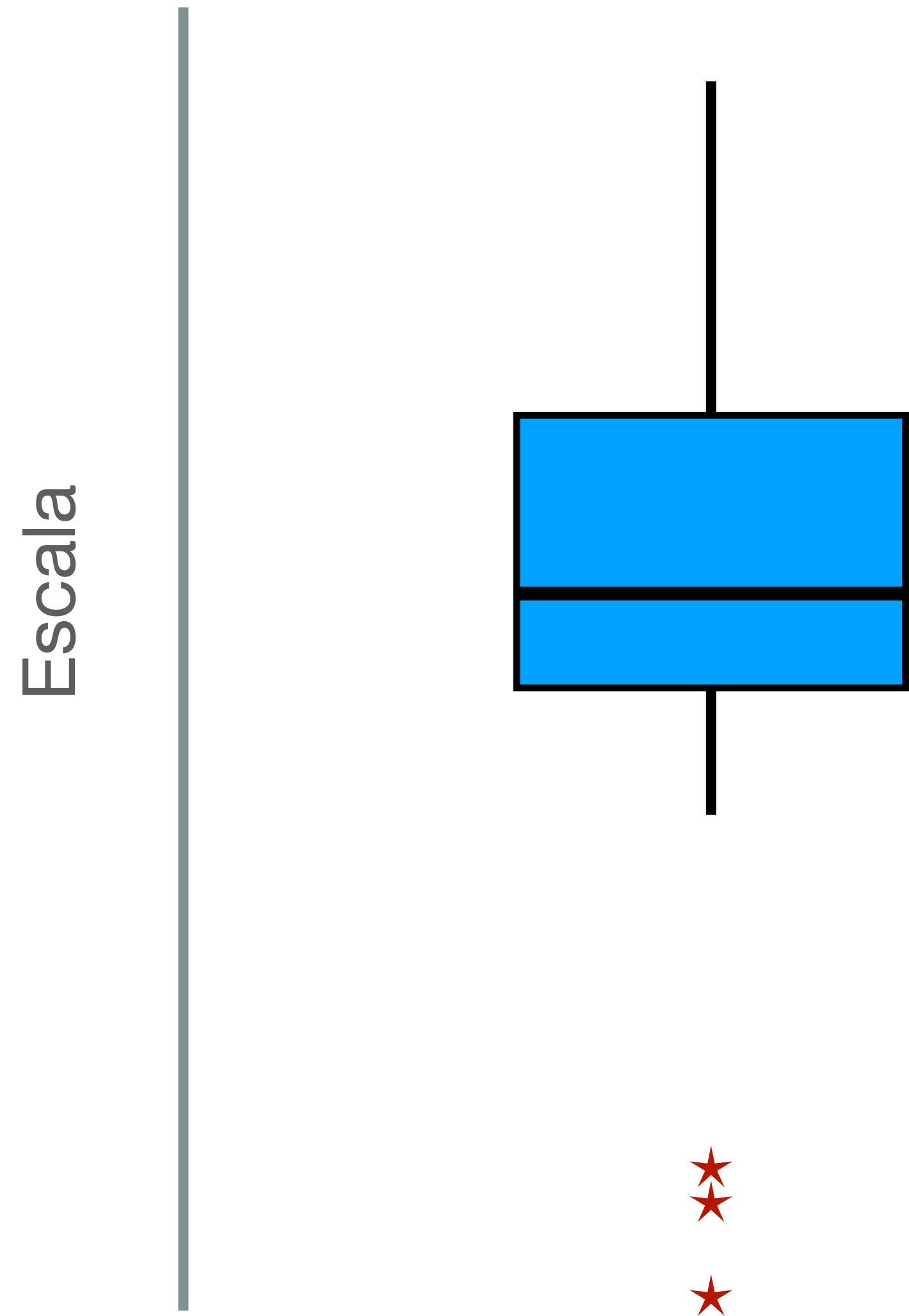
Variáveis quantitativas

Gráficos

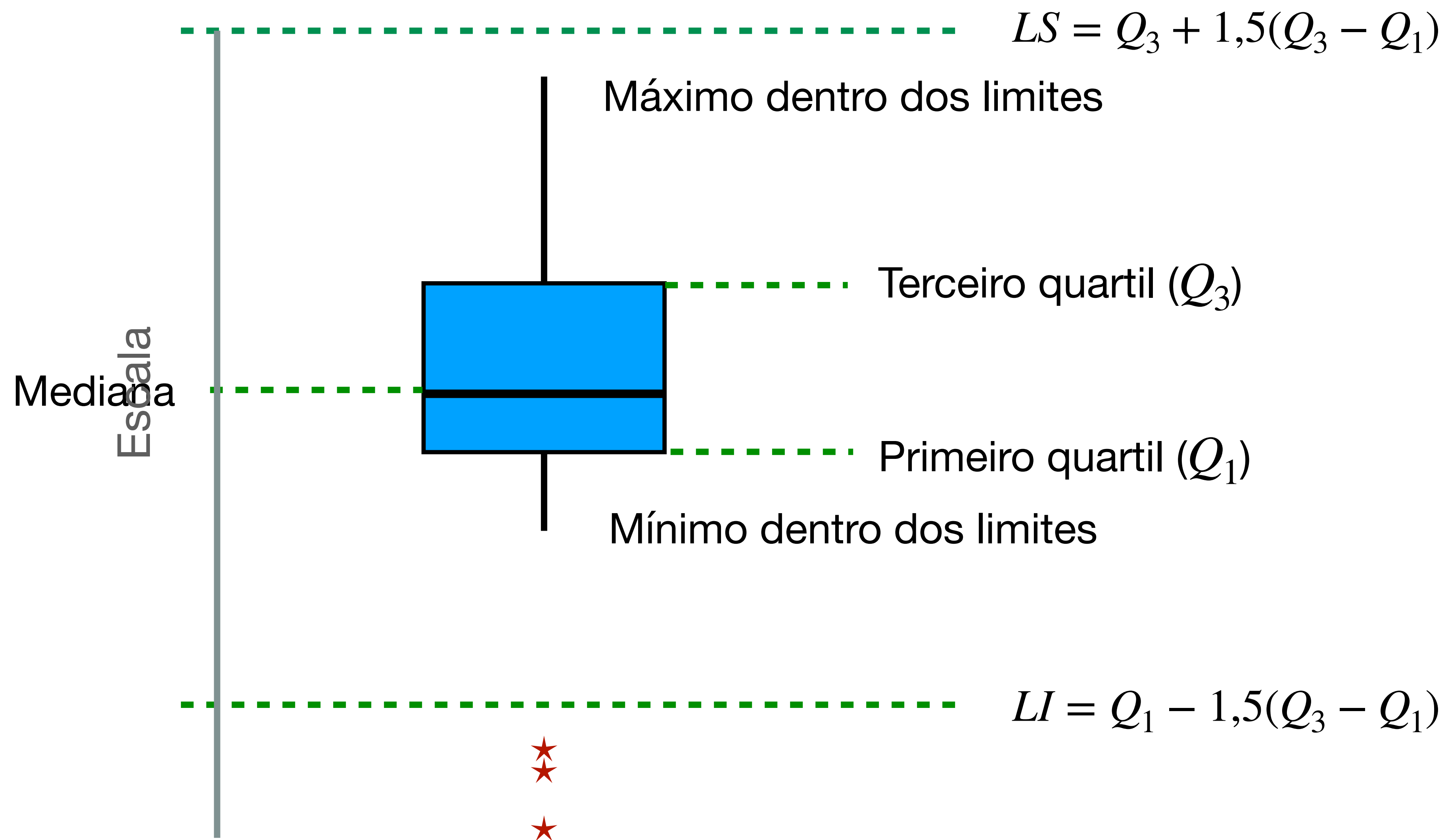
- “Boxplot” (diagrama de caixa)
- Histograma



Boxplot



Boxplot



Boxplot

Exemplo do projeto “Qualidade de vida”

Variável: altura das pacientes

146 148 149 150 150 152

152 153 154 154 156 156

156 156 158 158 159 160

160 160 160 161 161 161

162 164 165 165 172 175

$$\min = 146$$

$$Q_1 = 152,5$$

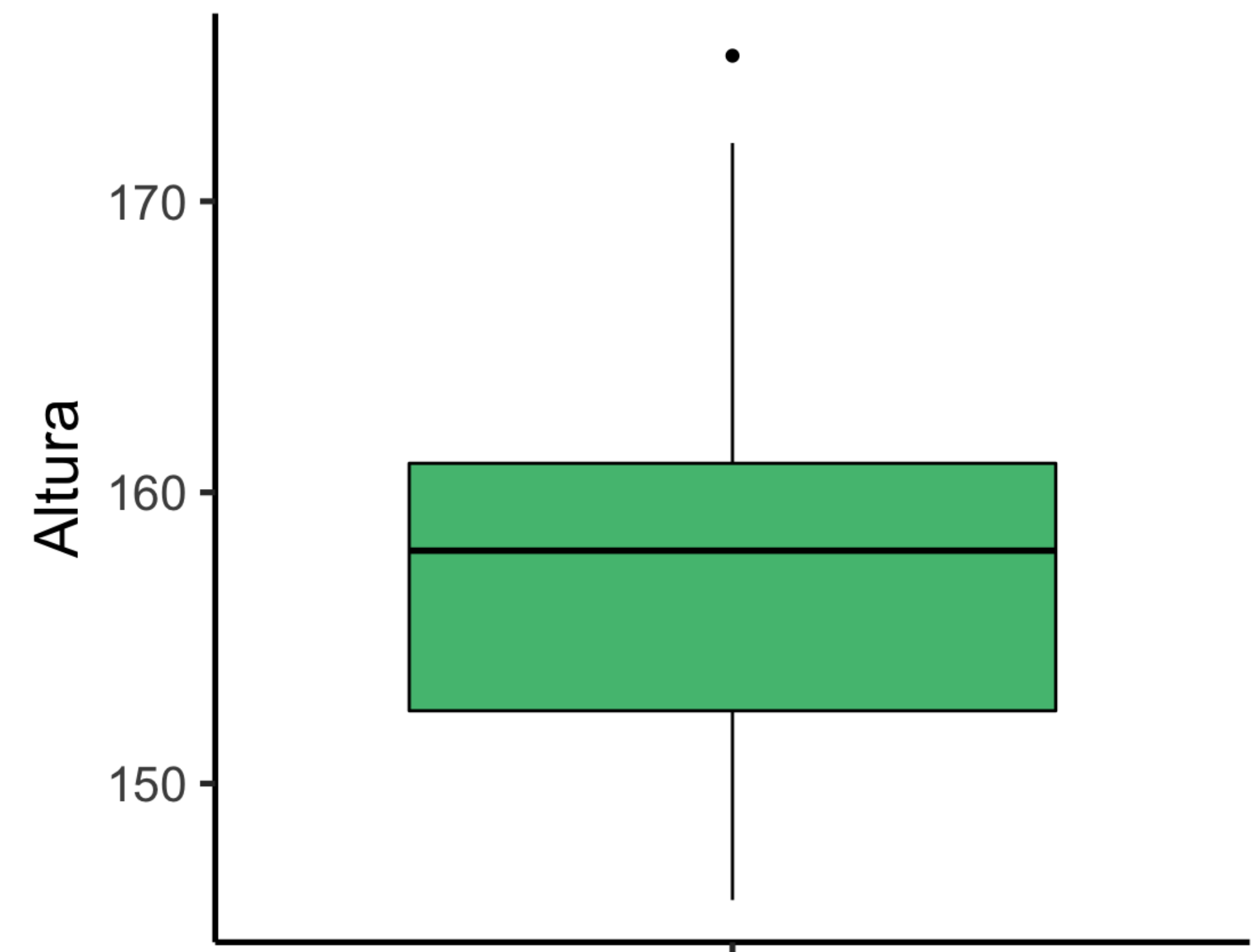
$$Q_2 = 158$$

$$Q_3 = 161$$

$$\max = 175$$

$$LI = Q_1 - 1,5(Q_3 - Q_1) = 139,75$$

$$LS = Q_3 + 1,5(Q_3 - Q_1) = 173,75$$

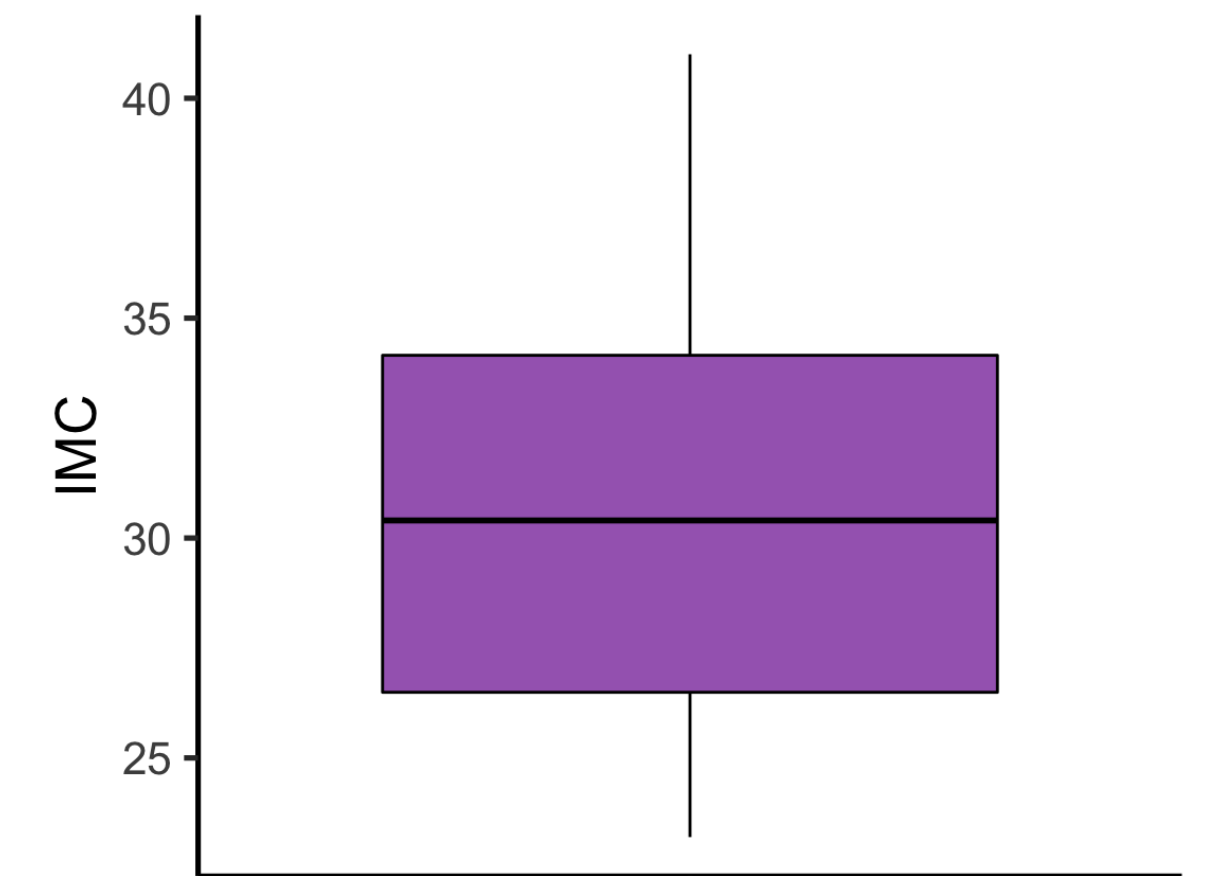
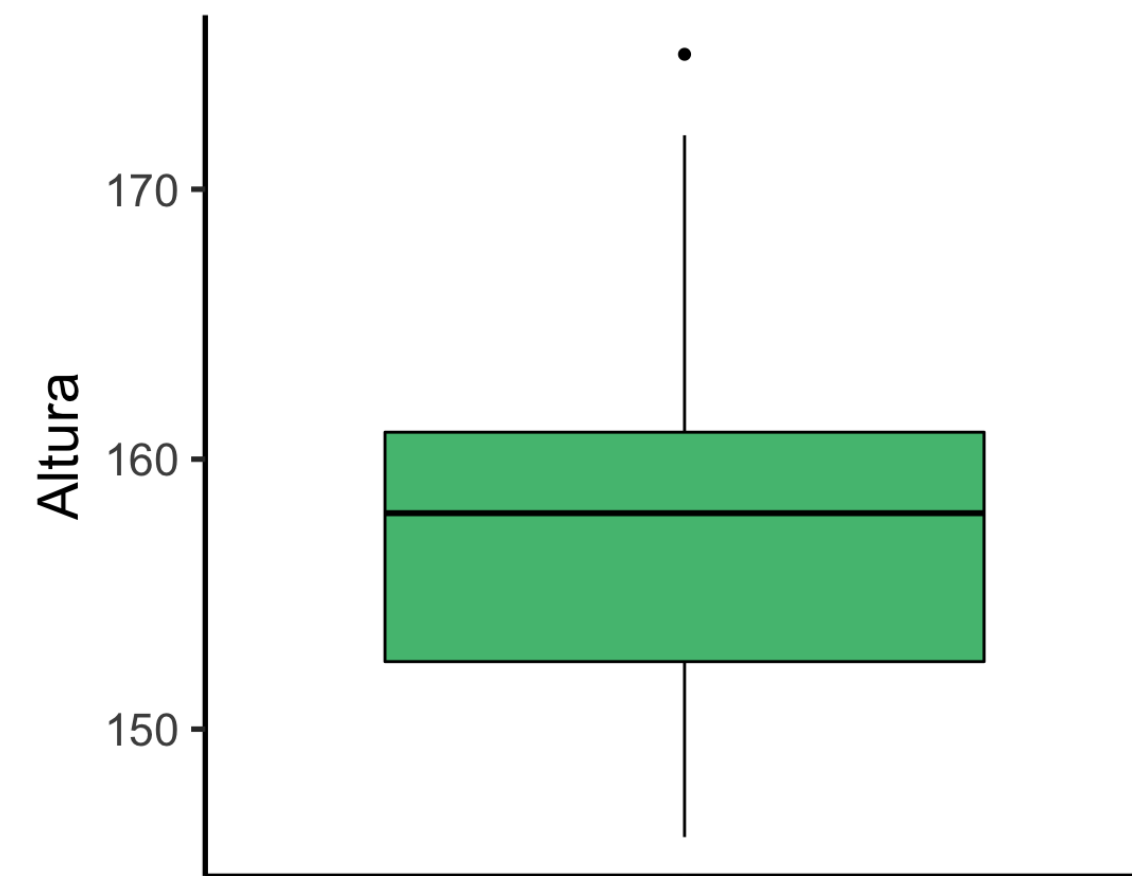
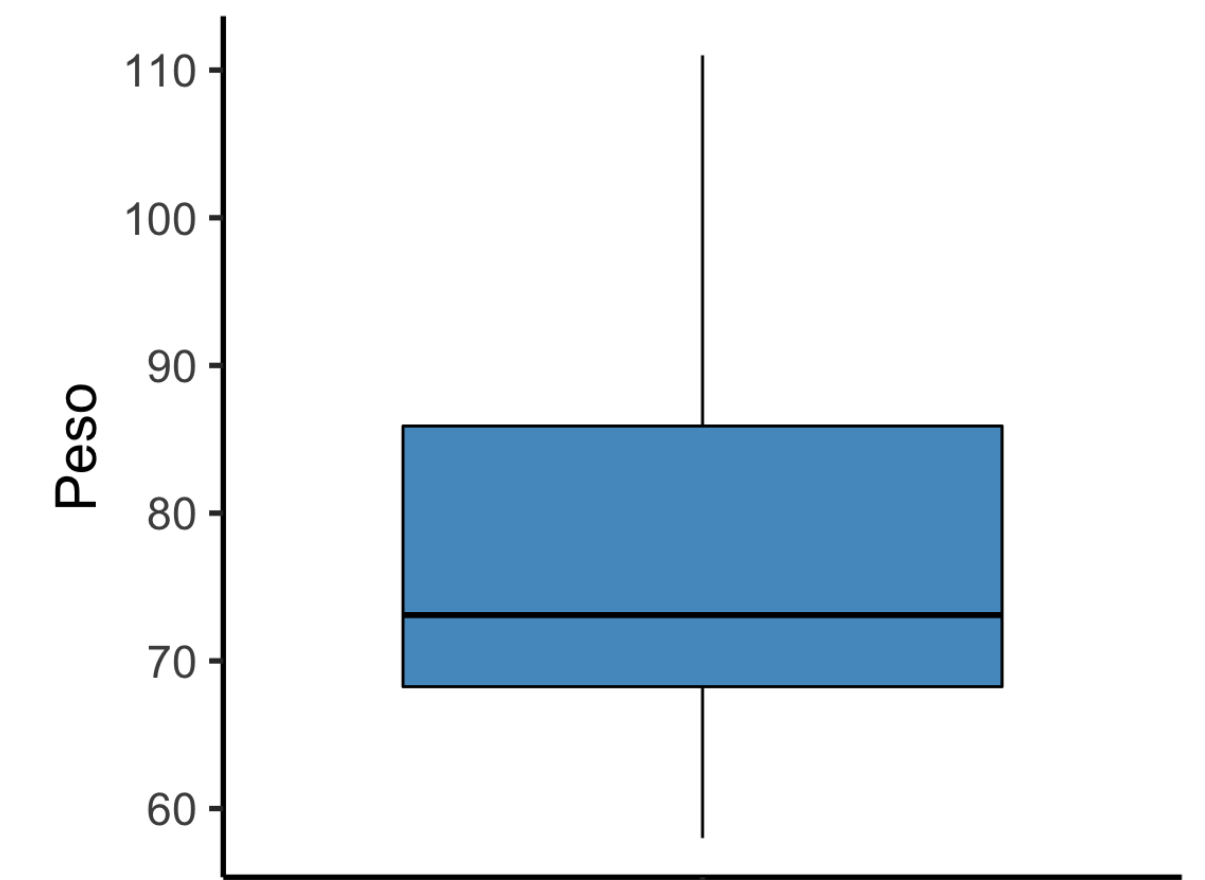
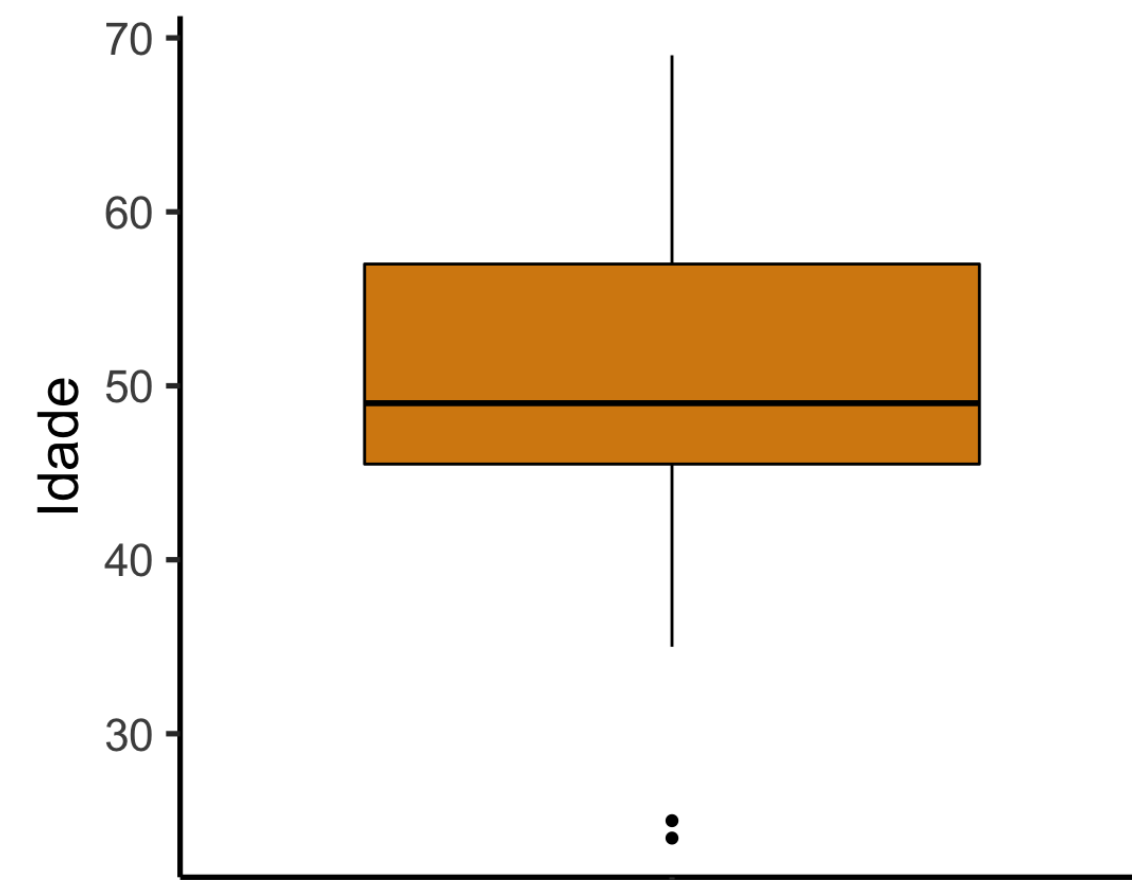


Boxplot

Exemplo do projeto “Qualidade de vida”

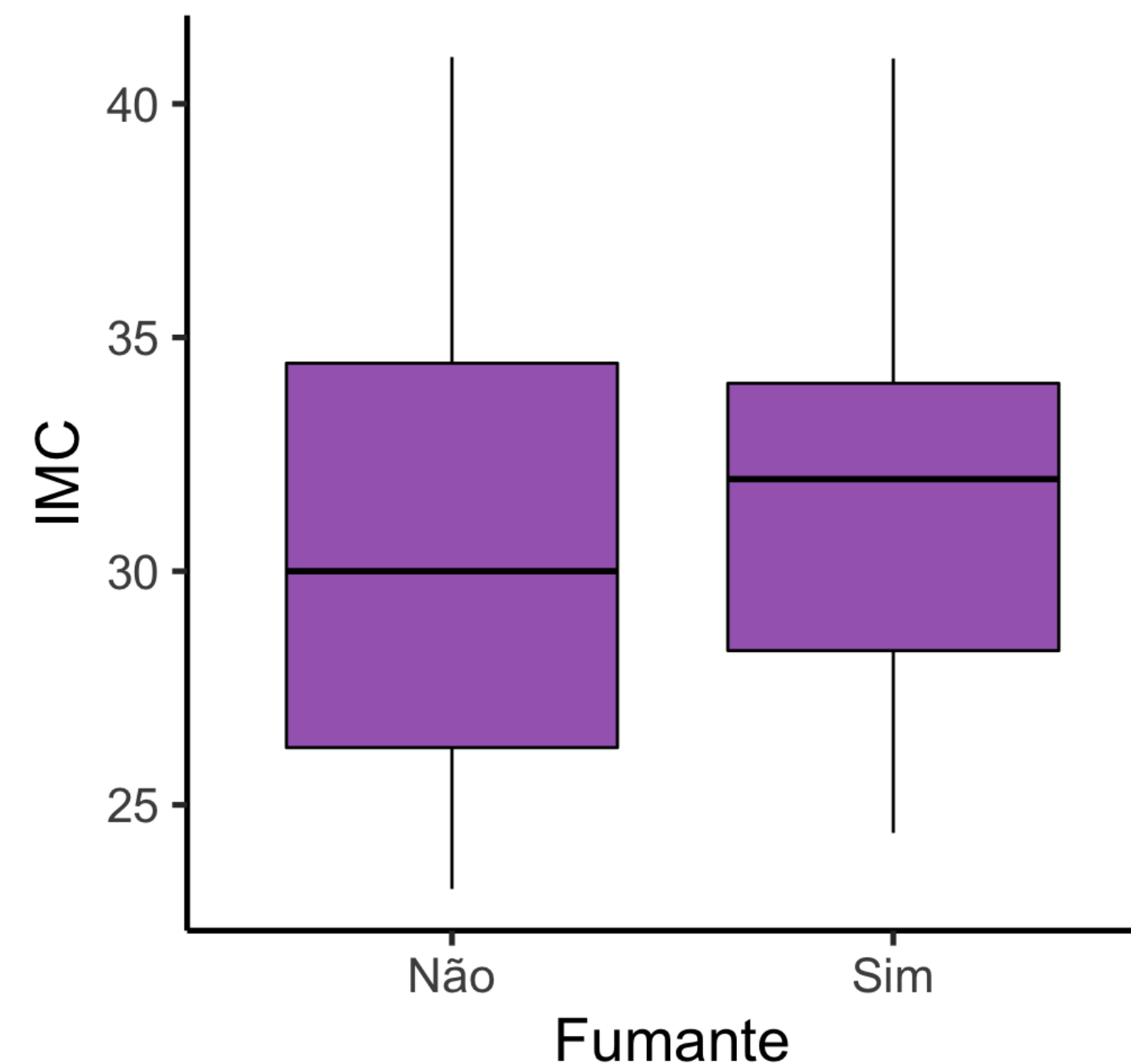
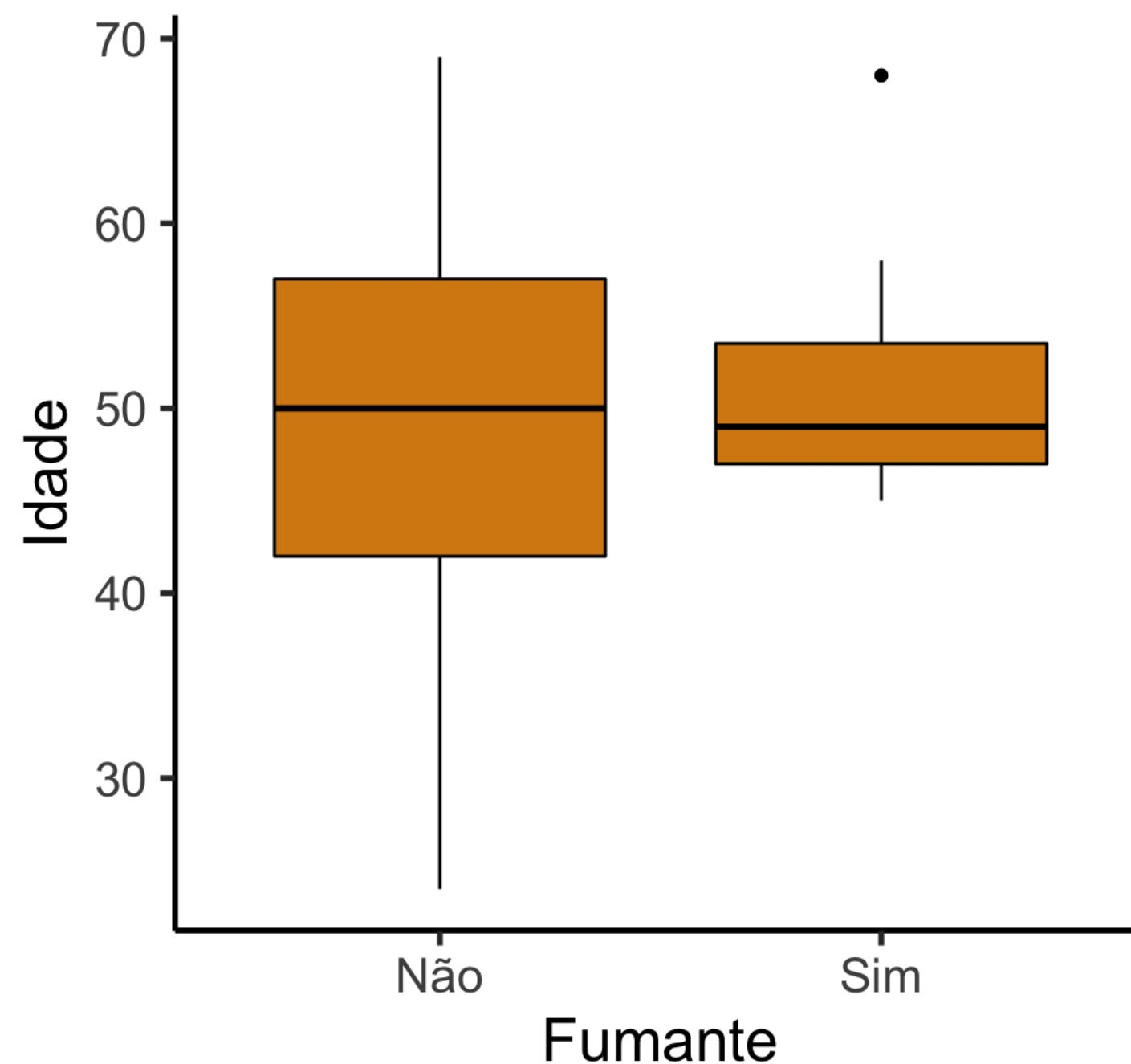
A distribuição do IMC das pacientes é mais simétrica quando comparada às demais variáveis.

Idade e Altura apresentam valores discrepantes.



Boxplot

Exemplo do projeto “Qualidade de vida”



Confirmando as demais análises dos dados, a distribuição da idade é bastante diferente para os dois grupos de pacientes (fumantes e não fumantes).

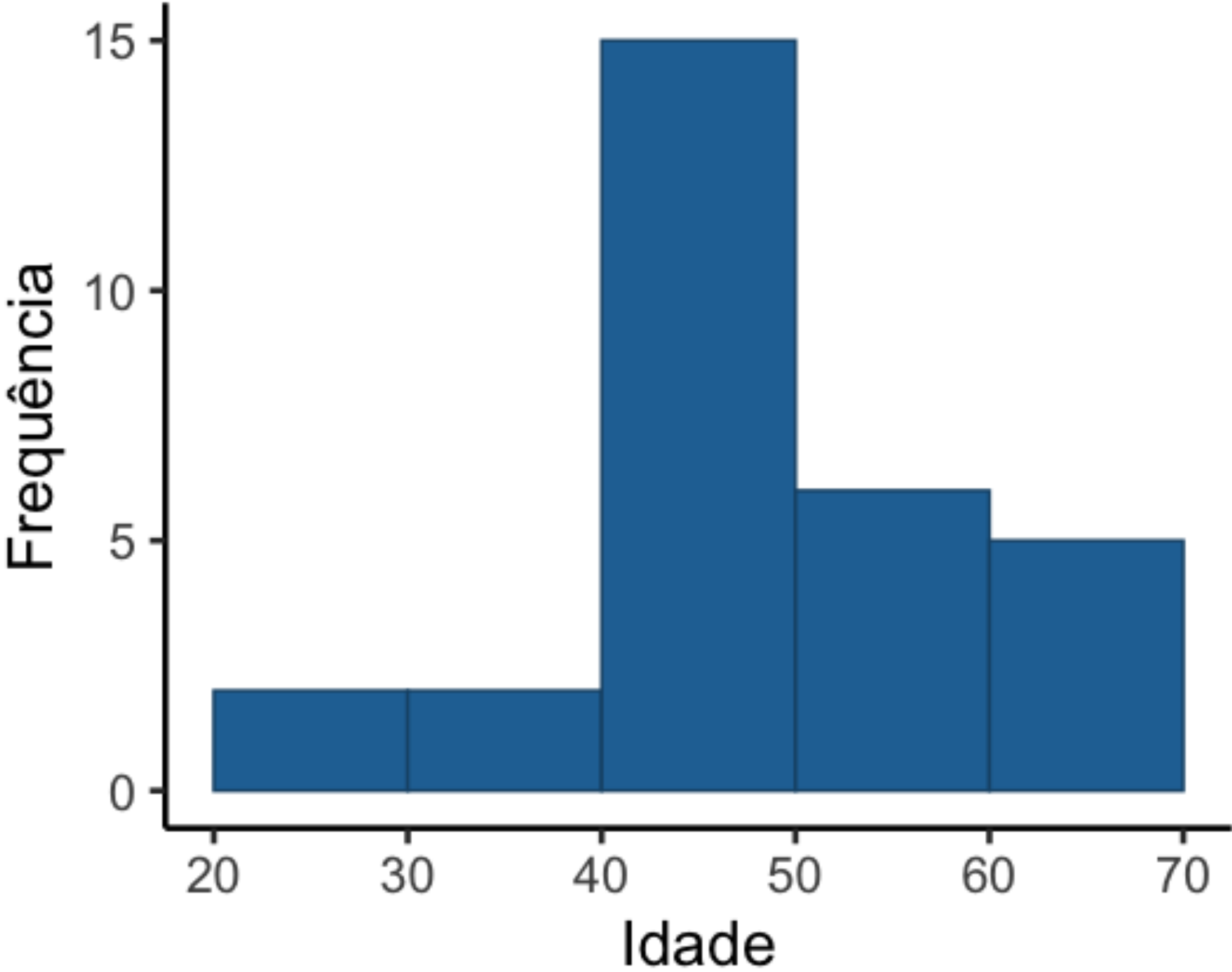
Histograma

- É uma representação gráfica de uma tabela de frequências
- É construído a partir de retângulos, onde a base representa uma classe e a área é proporcional à frequência relativa da classe
- O histograma é um importante indicador da distribuição dos dados
- No caso de variáveis contínuas, quando o volume de dados aumenta e o intervalo de classes tende a zero (o que torna os retângulos cada vez mais finos e altos), a distribuição de frequências se aproxima de uma curva (densidade de probabilidades)

Histograma de frequências

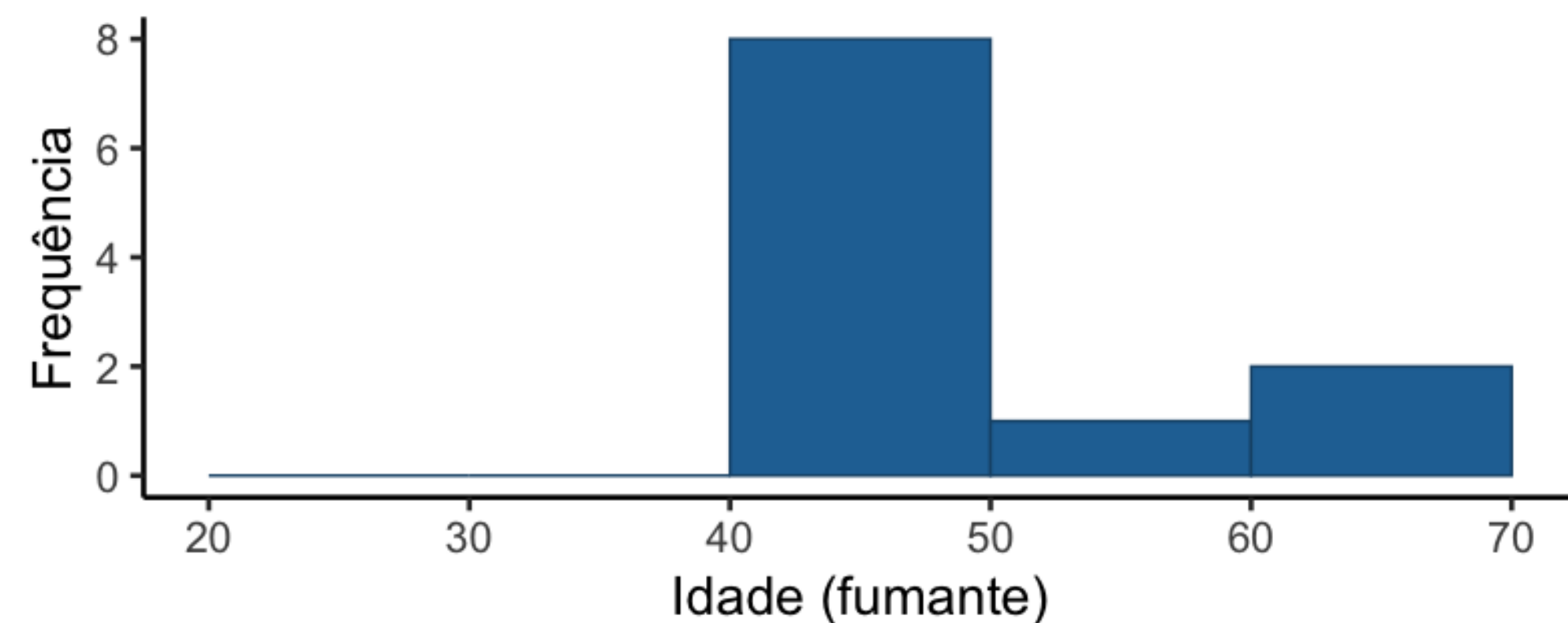
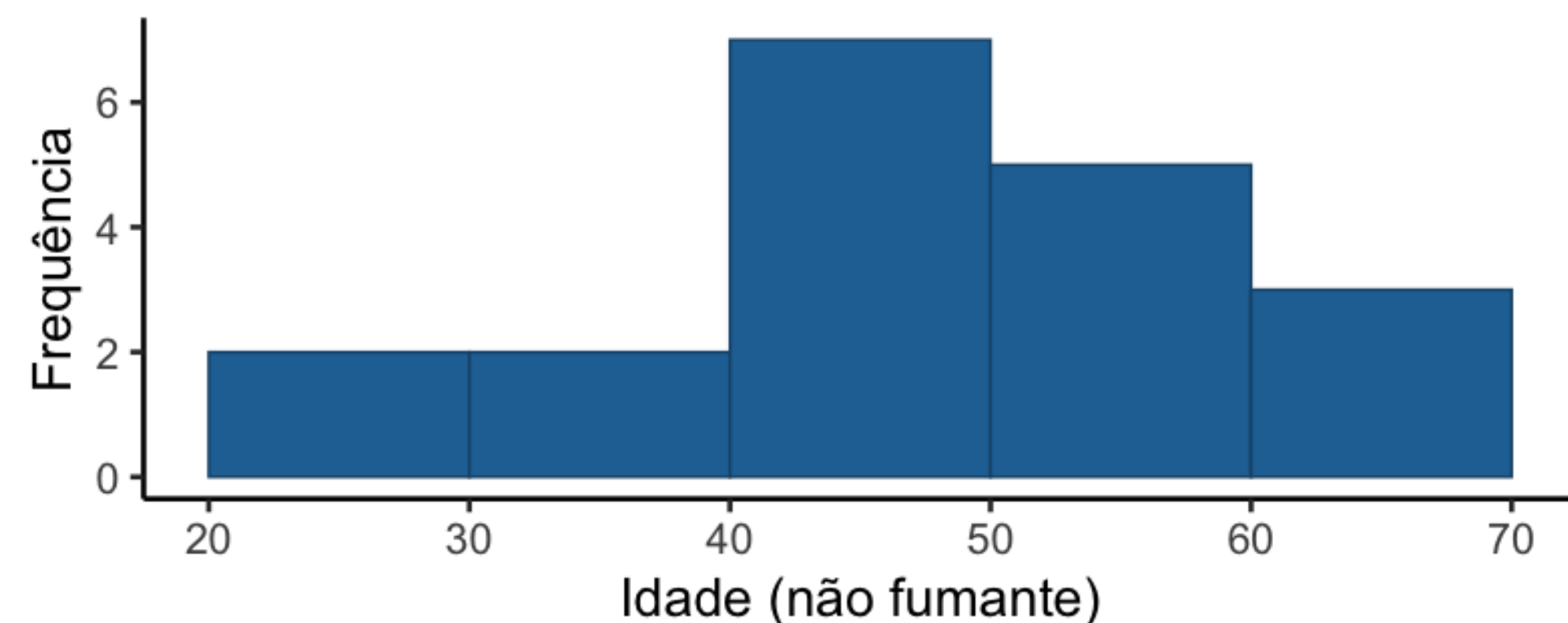
Exemplo do projeto “Qualidade de vida”

Idade	Frequência	Frequência relativa
(20; 30]	2	0,067
(30; 40]	2	0,067
(40; 50]	15	0,500
(50; 60]	6	0,200
(60; 70]	5	0,167
Total	30	1



Histograma de frequências

Exemplo do projeto “Qualidade de vida”



O histograma confirma o padrão já evidenciado por meio do boxplot: a distribuição da idade é diferente para os dois grupos de pacientes (fumantes e não fumantes).

Histograma de densidade

Exemplo do projeto “Qualidade de vida”

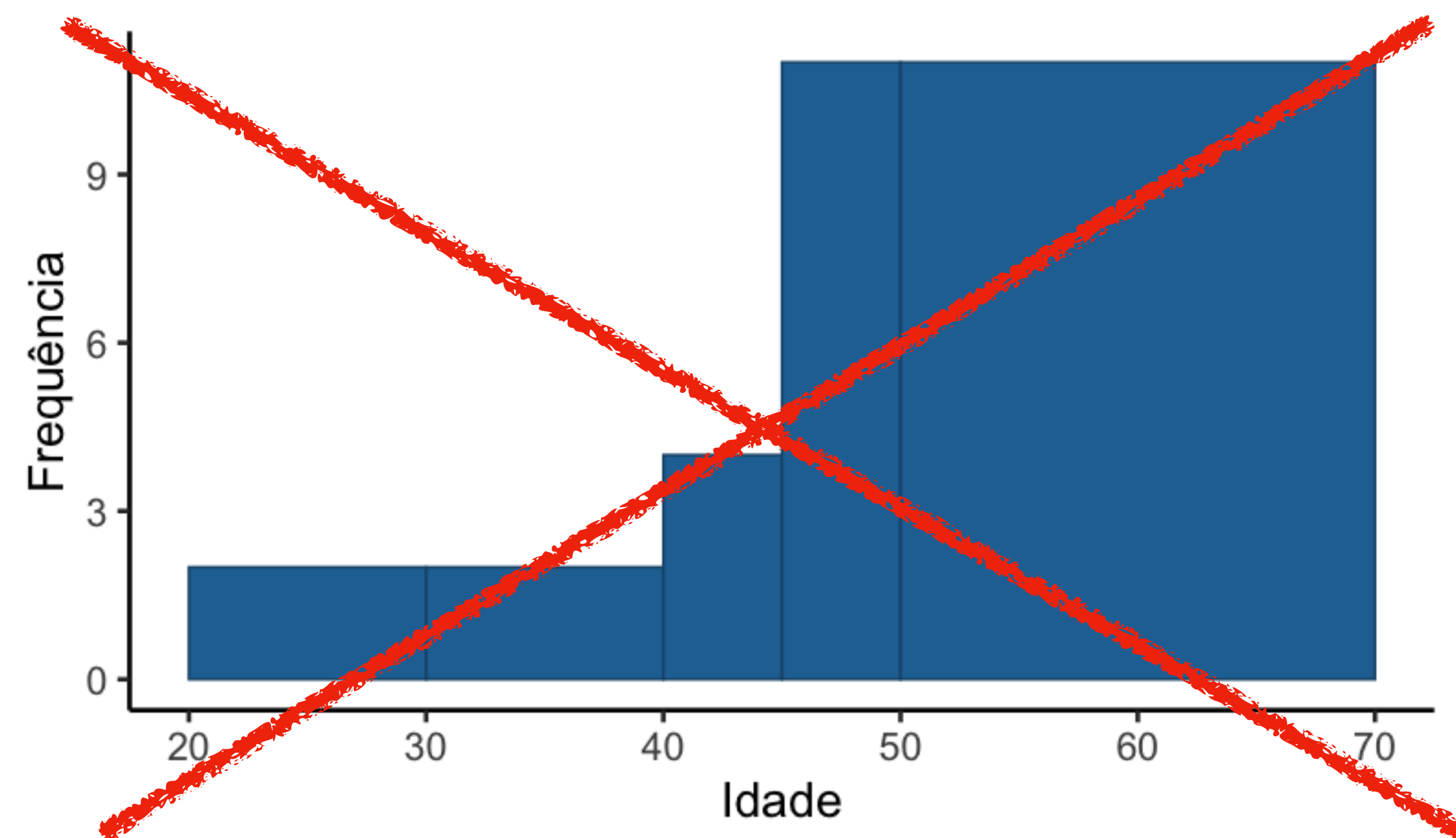
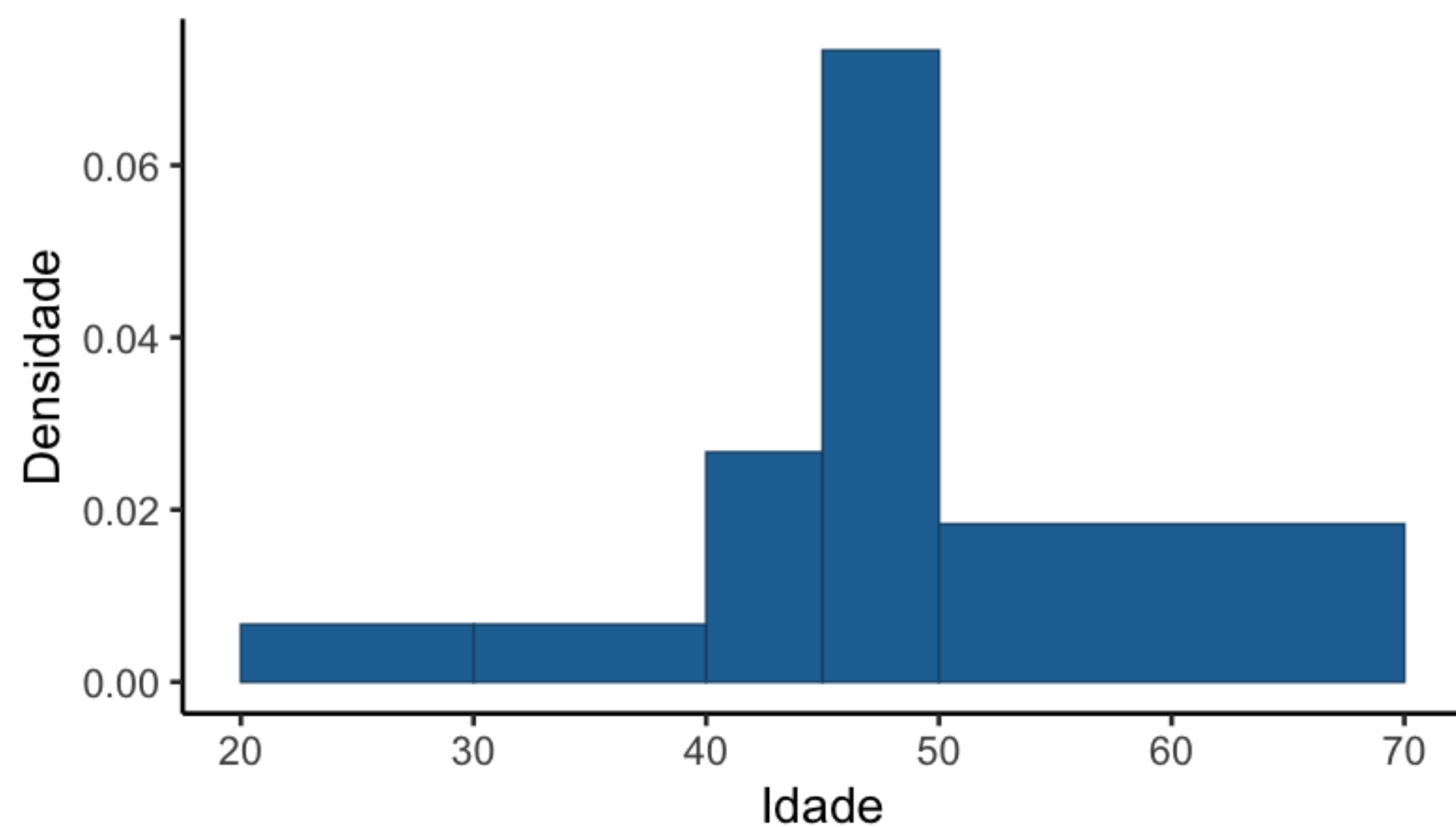
Idade	Frequência	Frequência relativa	Base	Altura (h)
(20; 30]	2	0,067	10	0,0067
(30; 40]	2	0,067	10	0,0067
(40; 45]	4	0,133	5	0,0266
(45; 50]	11	0,367	5	0,0734
(50; 70]	11	0,367	20	0.0184
Total	30	1		

A altura será dada por: $h = fr/base$ (h é a densidade de frequência).

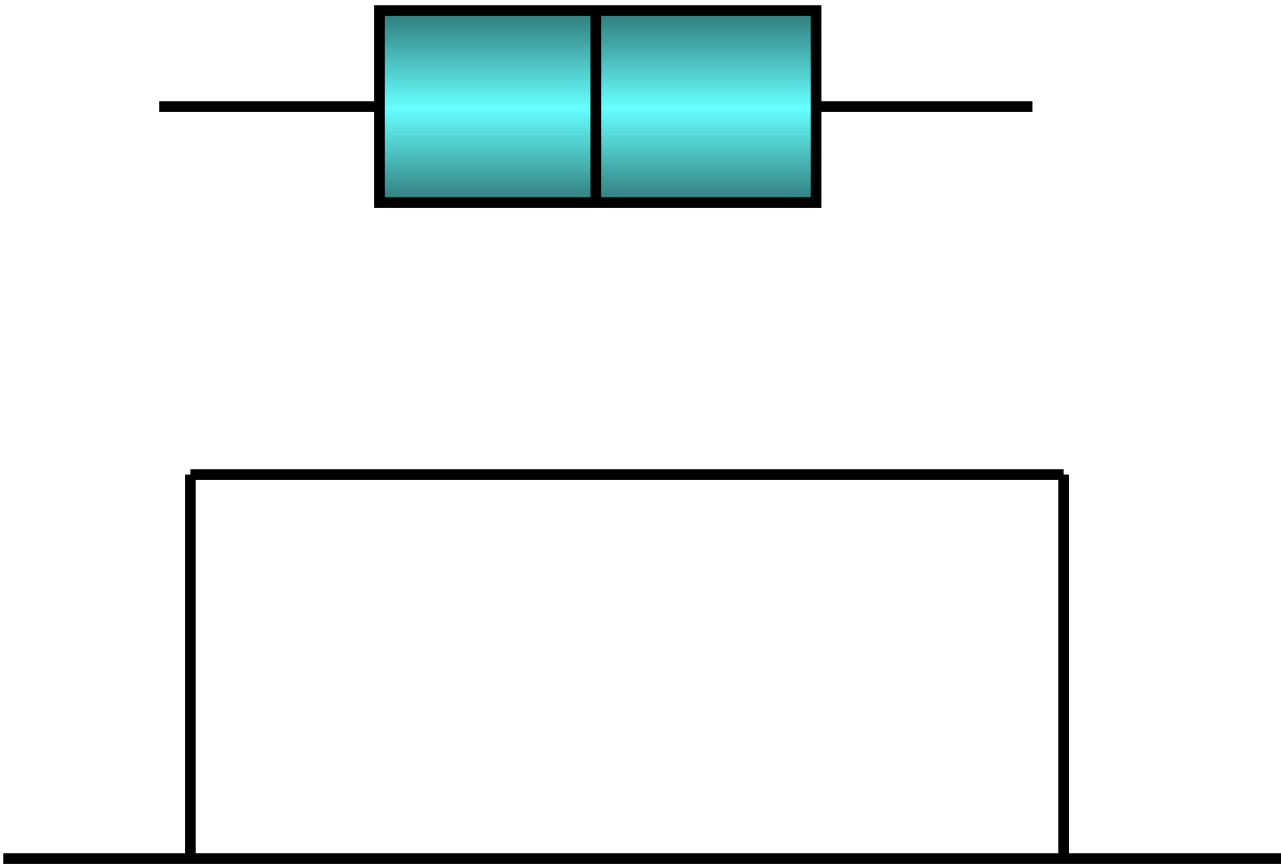
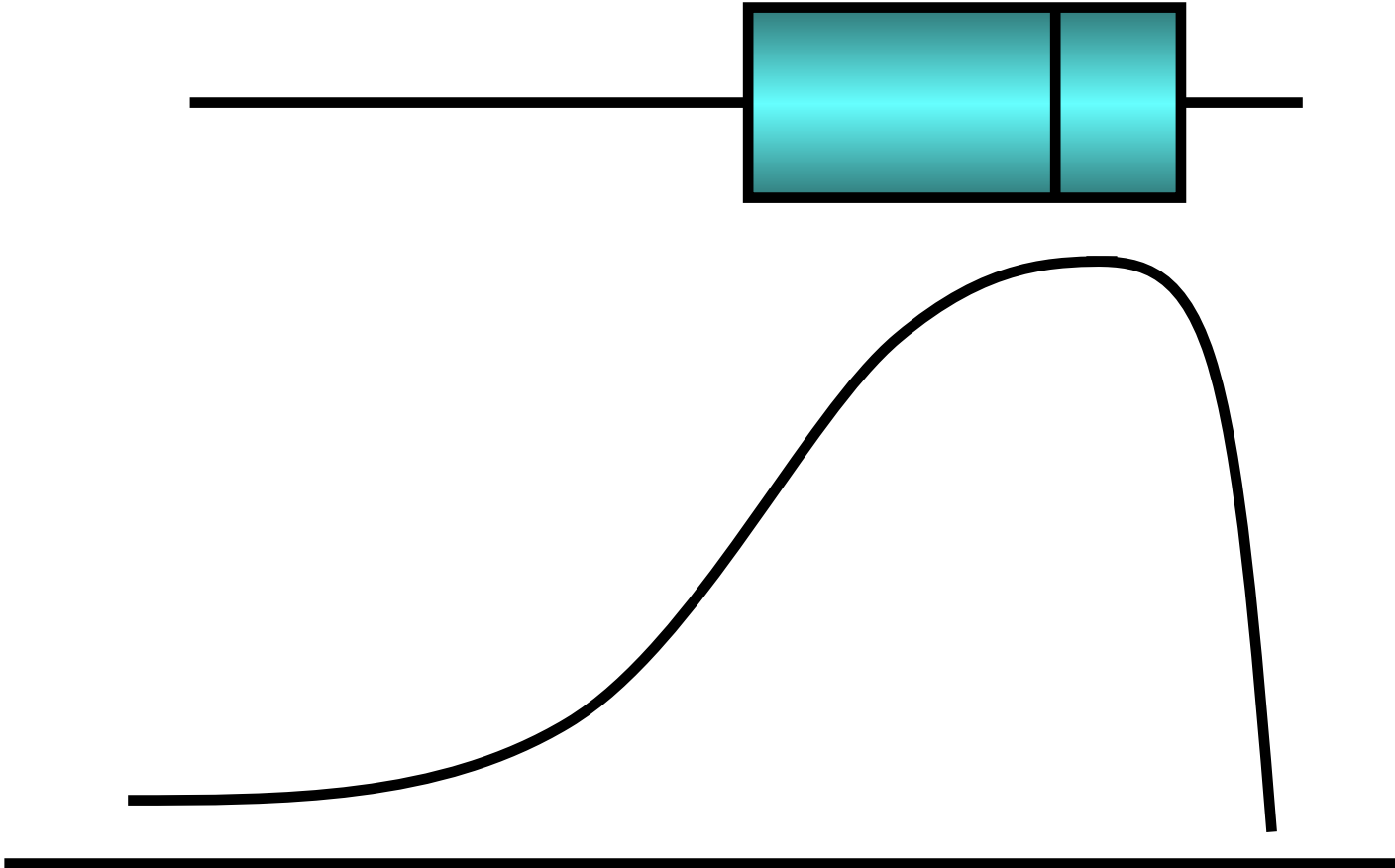
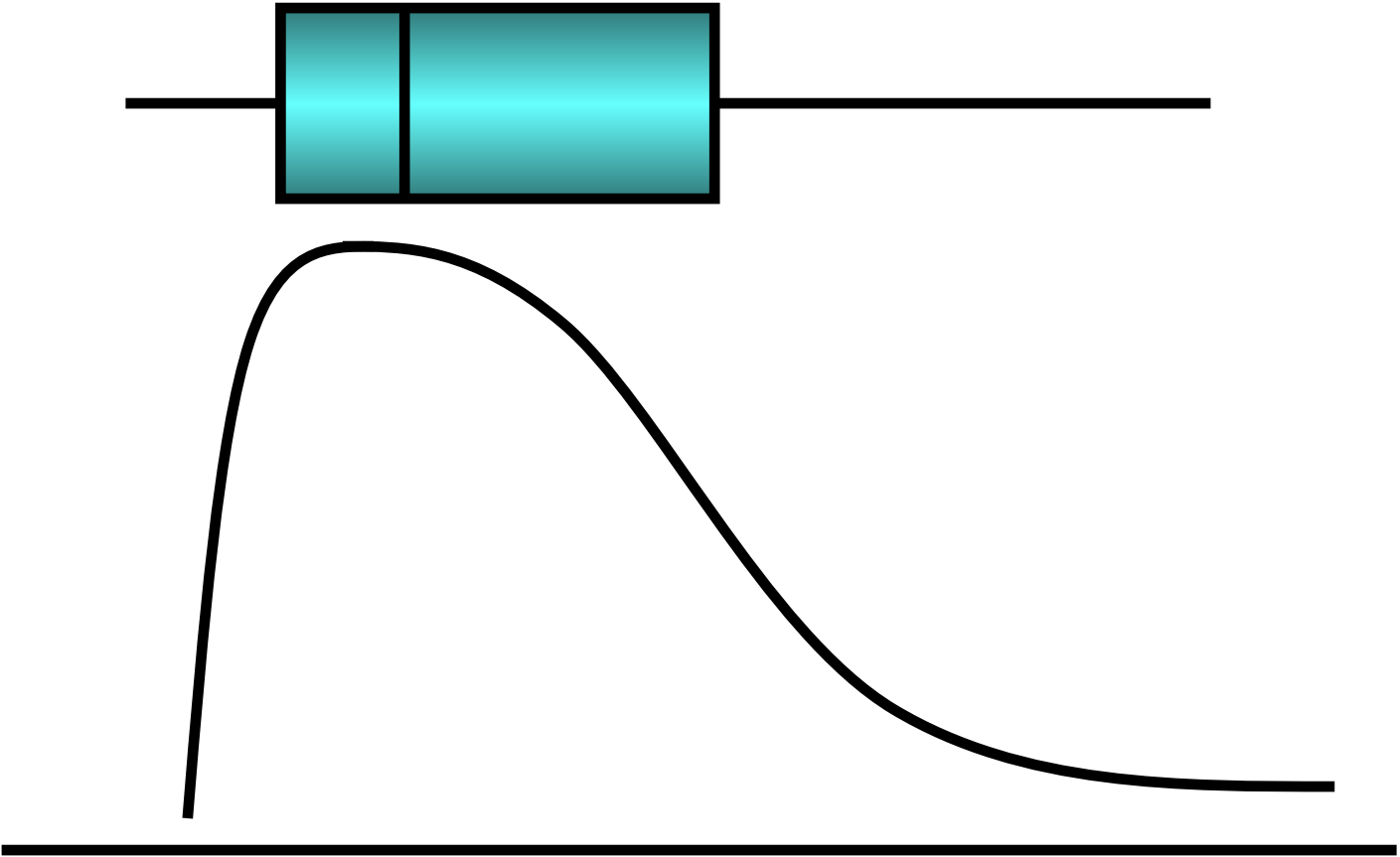
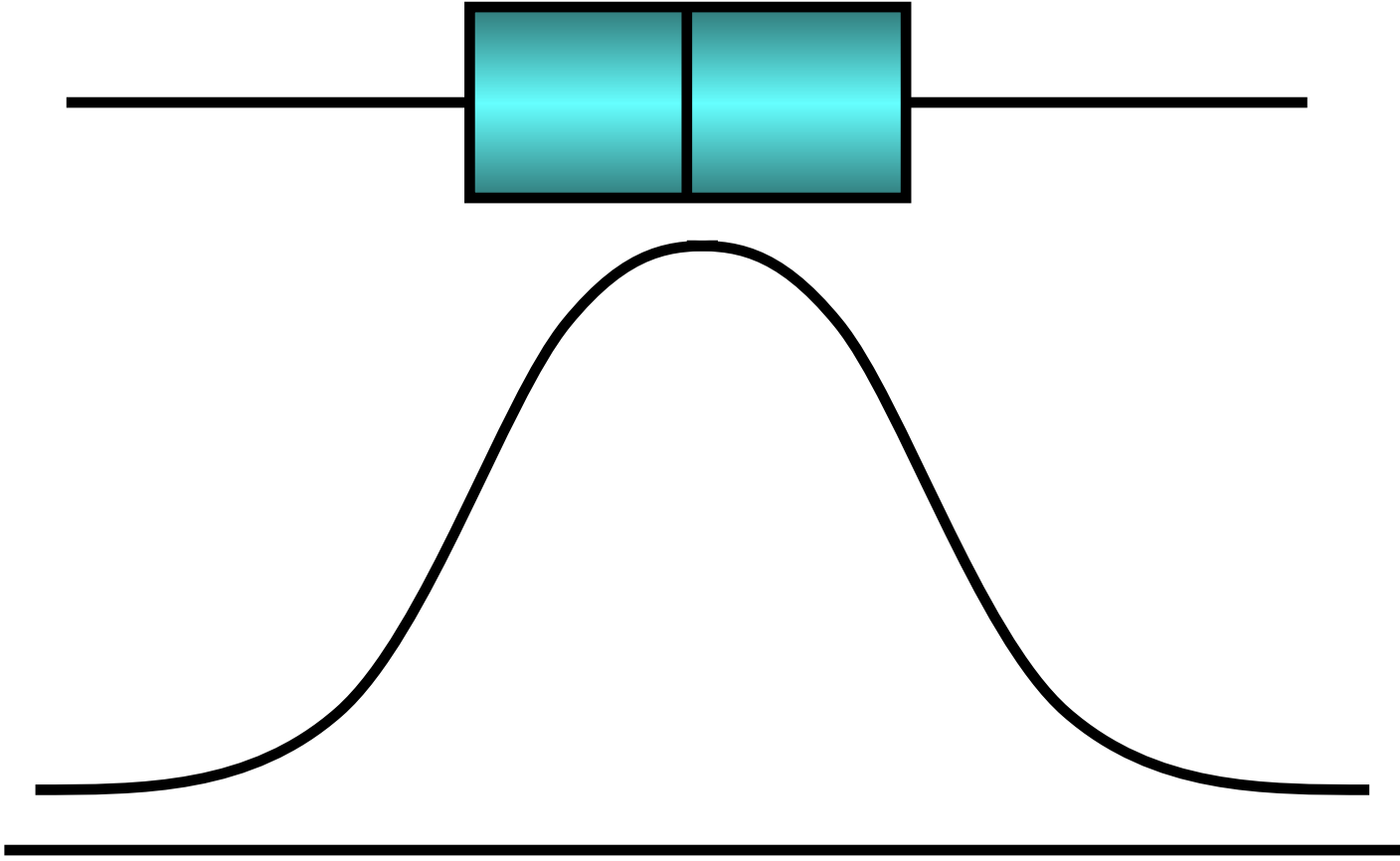
A área do retângulo fica igual à frequência relativa da classe.

Histograma de densidade

Exemplo do projeto “Qualidade de vida”



Forma da distribuição



Variáveis qualitativas

Distribuições de frequências

Os dados podem ser resumidos por meio de tabelas de distribuição de frequências, que quantificam a frequência das distintas categorias.

```
> table(CEA2015P01_QV_Demograficos$Etnia)
```

1	2	3
25	1	4

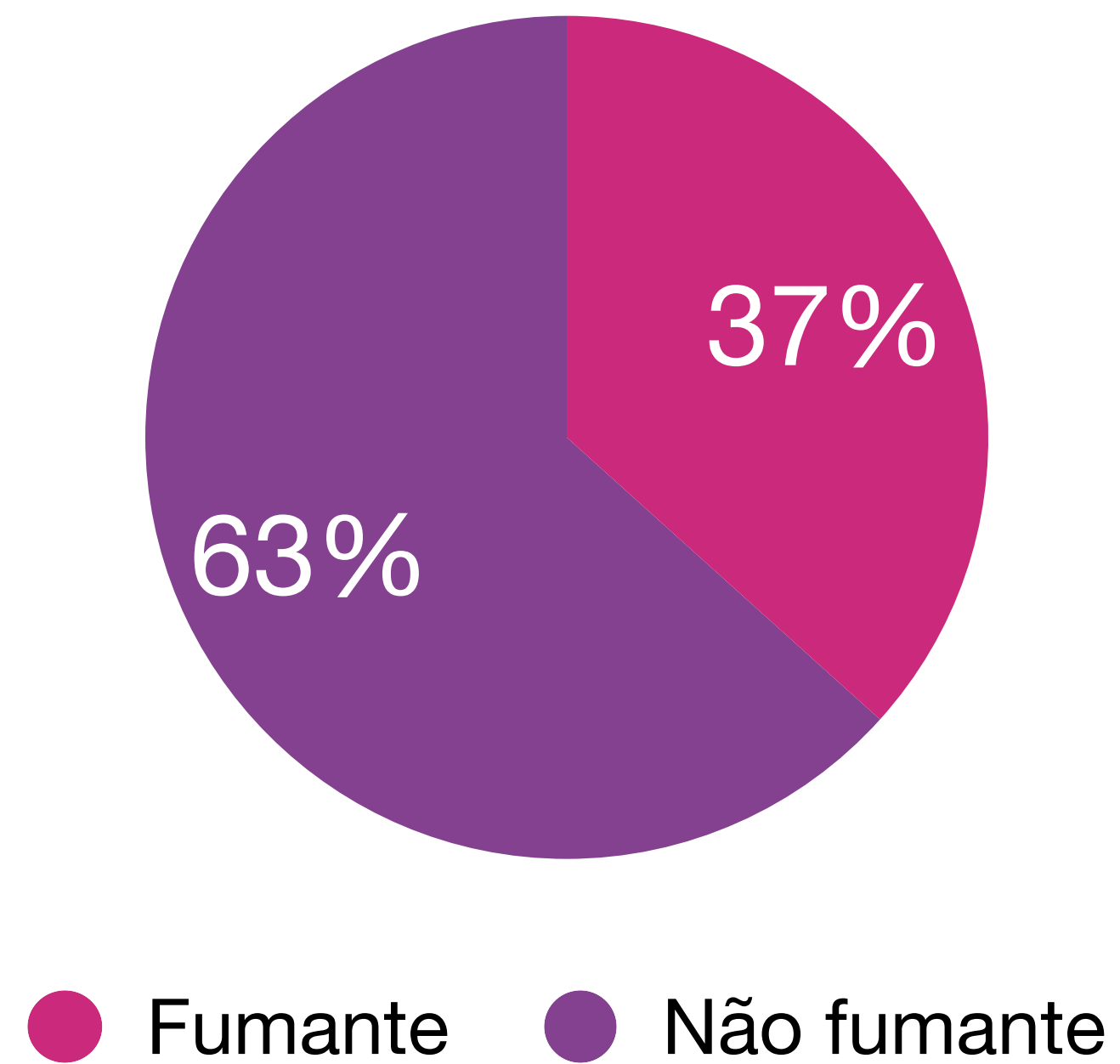
```
> table(CEA2015P01_QV_Demograficos$Escolaridade)
```

1	2	3	4	5	6
20	2	1	4	2	1

Variáveis qualitativas

Gráficos

- Gráficos de setores



- Gráficos de barras

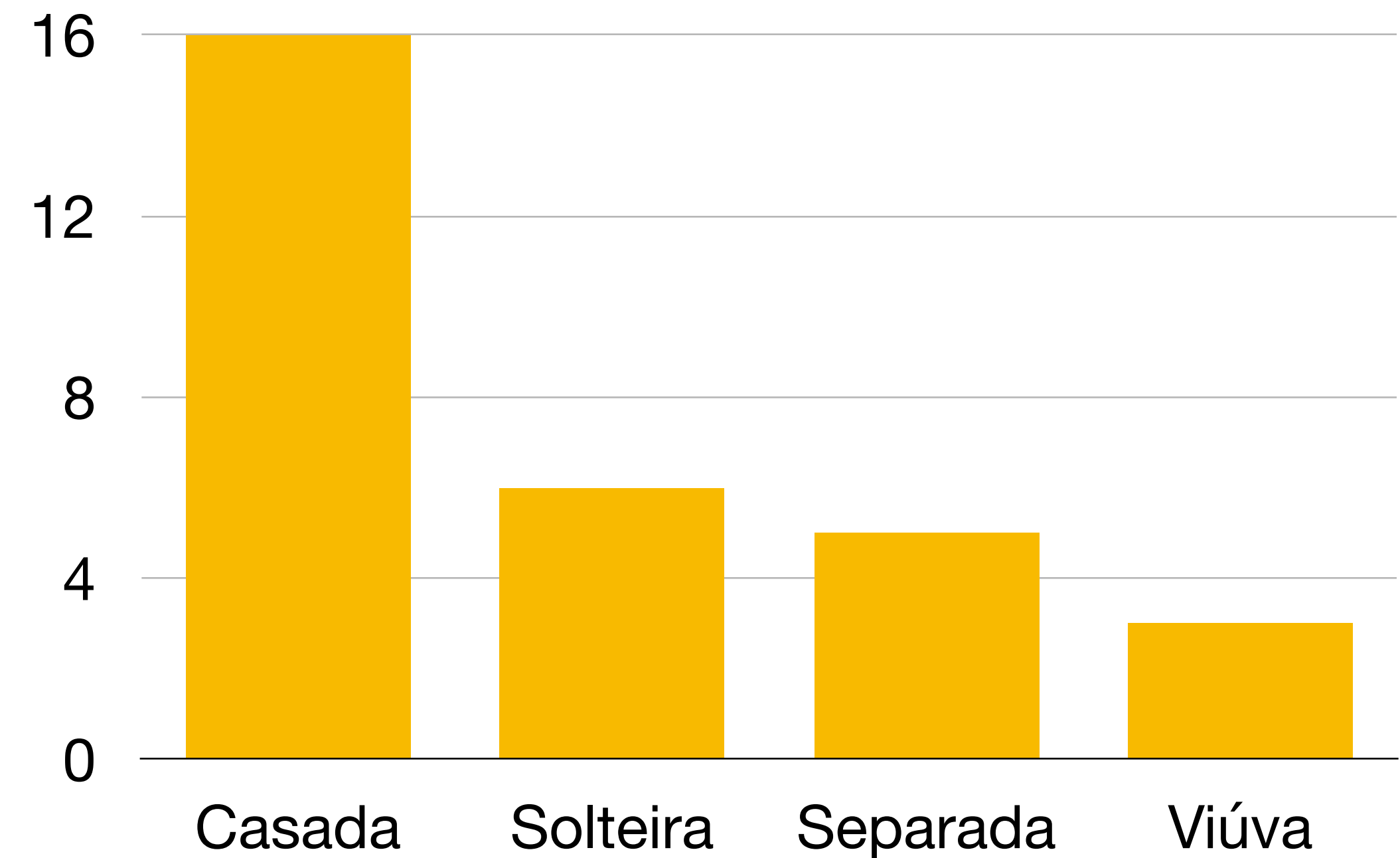
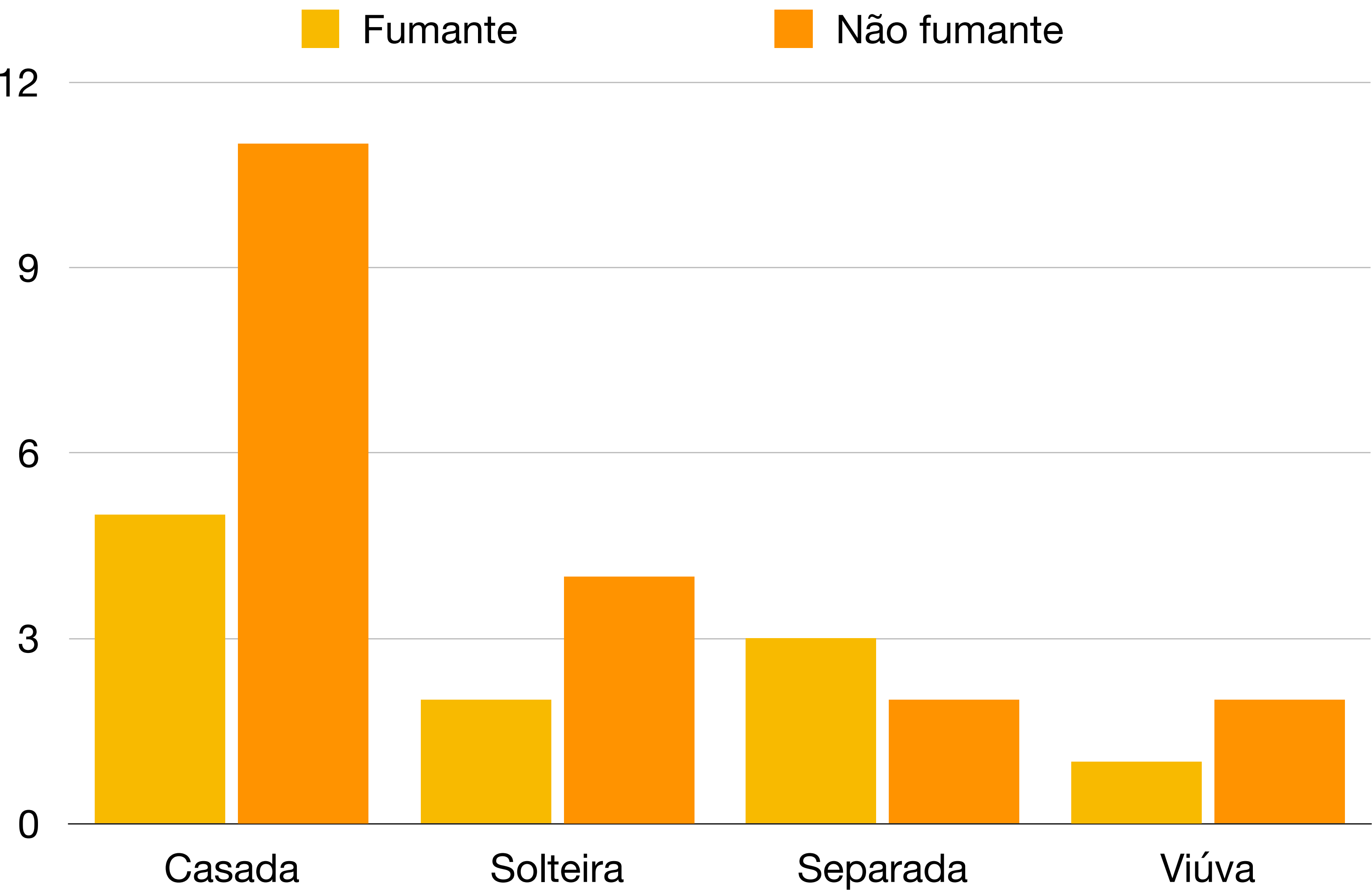


Gráfico de barras



Resumo

**Variáveis
quantitativas**

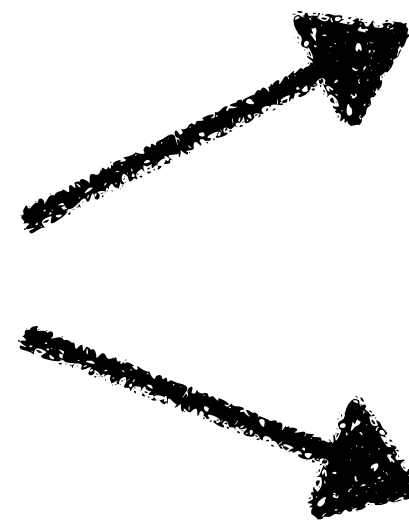


Medidas de posição e dispersão

Gráficos: boxplot, histograma

Distribuições de frequências

**Variáveis
qualitativas**



Gráficos: setores, barras