

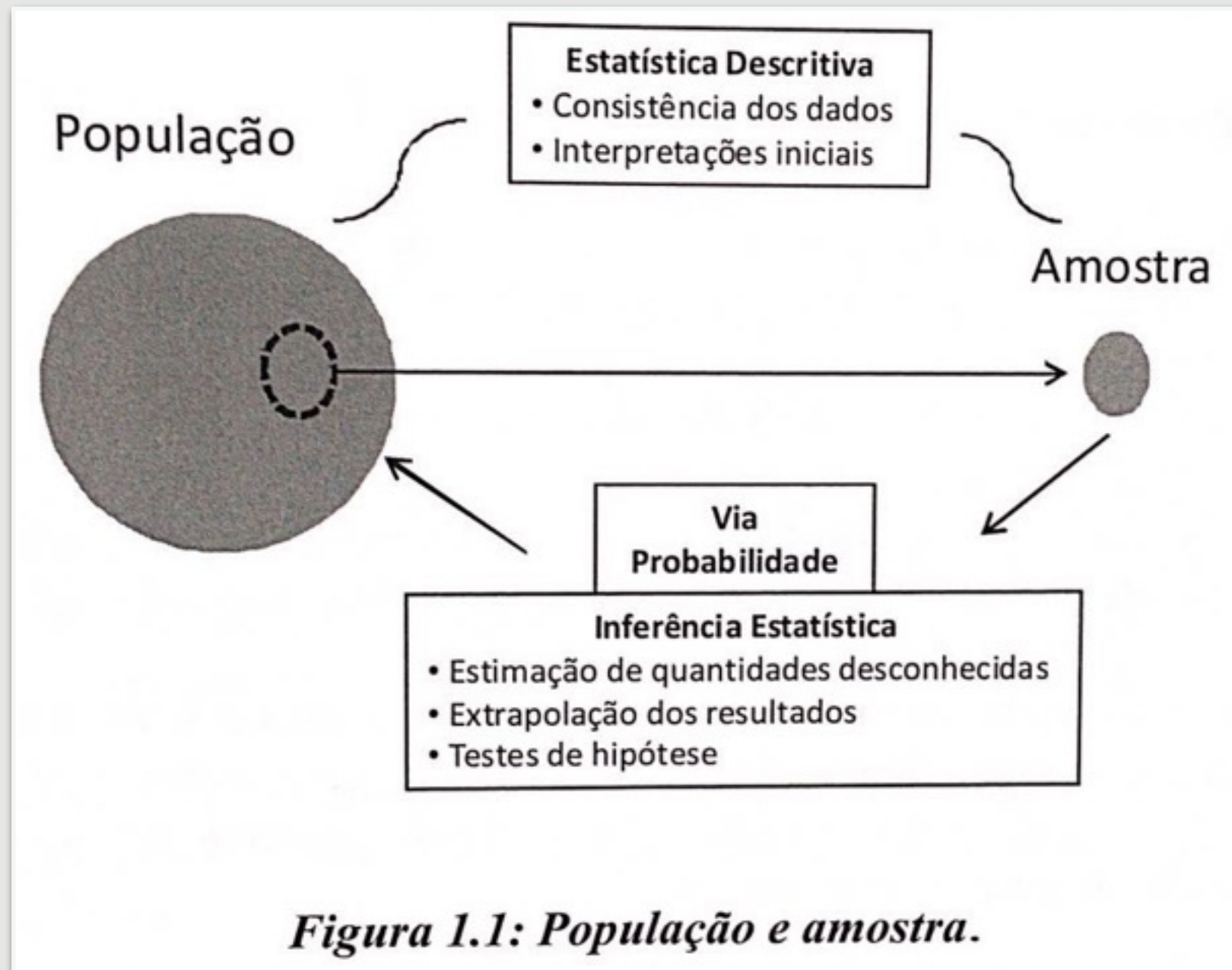
# BioEstat - Parte 1

Murillo F. Rodrigues

# O que é estatística?

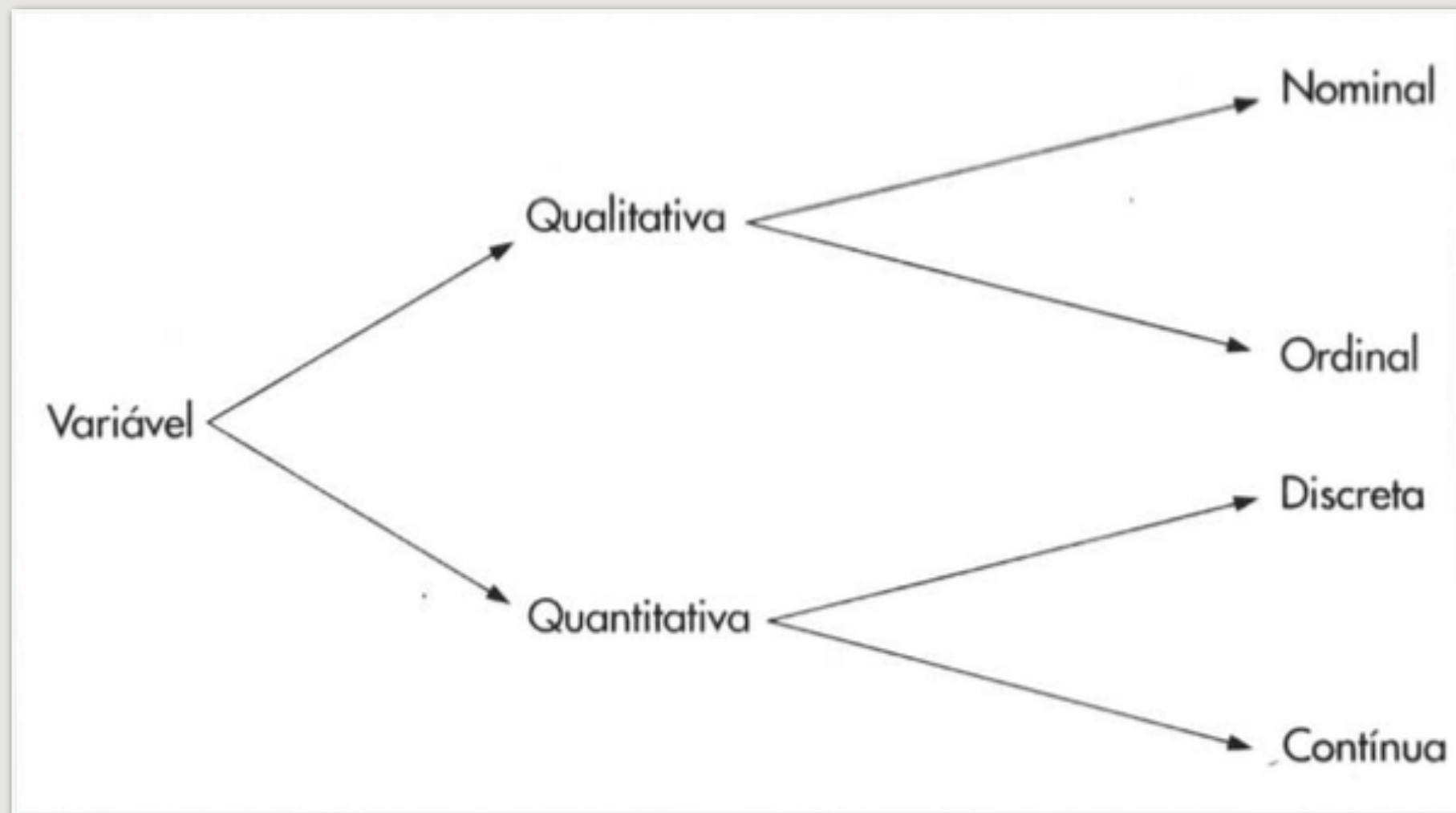
- Análise exploratória
- Probabilidade
- Inferência estatística

# O que é estatística?

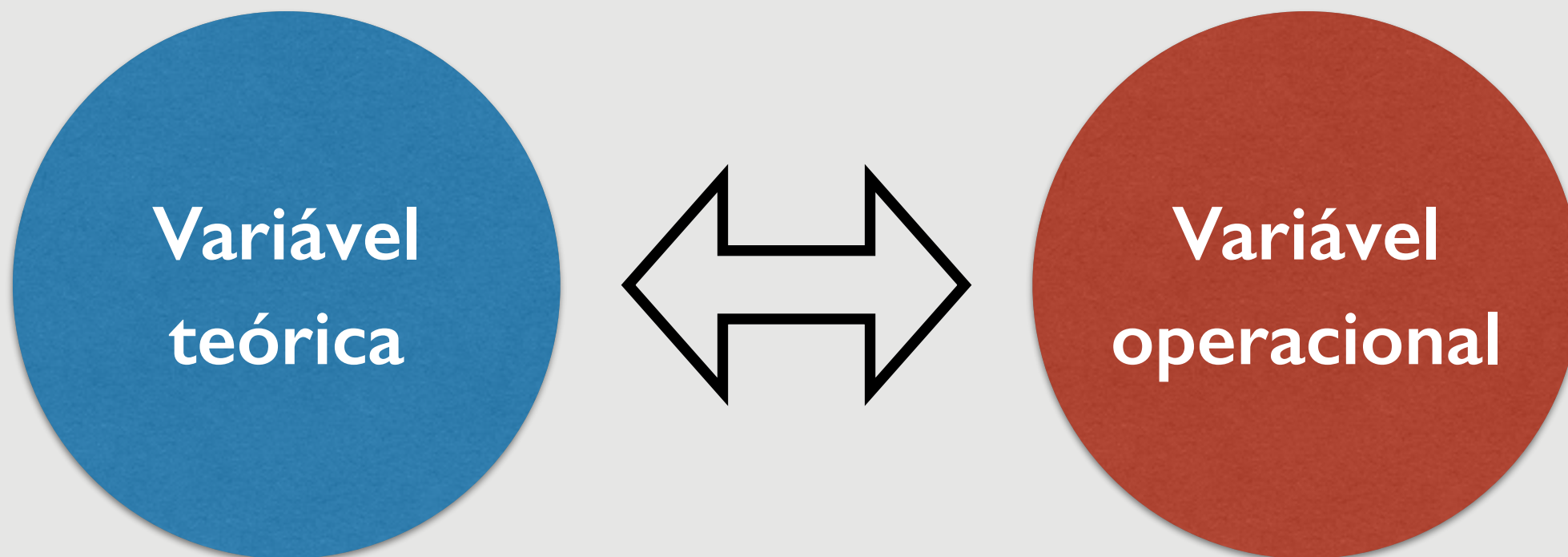


# Análise Exploratória

# Tipos de variáveis



# Tipos de variáveis



# Dados

**Tabela 2.1:** Informações sobre estado civil, grau de instrução, número de filhos, salário (expresso como fração do salário mínimo), idade (medida em anos e meses) e procedência de 36 empregados da seção de orçamentos da Companhia MB.

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (x sal. min.)	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	—	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	—	5,73	20	10	outra
5	solteiro	ensino fundamental	—	6,26	40	07	outra
6	casado	ensino fundamental	0	6,66	28	00	interior
7	solteiro	ensino fundamental	—	6,86	41	00	interior
8	solteiro	ensino fundamental	—	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio	—	7,44	23	06	outra
11	casado	ensino médio	2	8,12	33	06	interior
12	solteiro	ensino fundamental	—	8,46	27	11	capital
13	solteiro	ensino médio	—	8,74	37	05	outra
14	casado	ensino fundamental	3	8,95	44	02	outra
15	casado	ensino médio	0	9,13	30	05	interior
16	solteiro	ensino médio	—	9,35	38	08	outra
17	casado	ensino médio	1	9,77	31	07	capital
18	casado	ensino fundamental	2	9,80	39	07	outra
19	solteiro	superior	—	10,53	25	08	interior
20	solteiro	ensino médio	—	10,76	37	04	interior
21	casado	ensino médio	1	11,06	30	09	outra
22	solteiro	ensino médio	—	11,59	34	02	capital
23	solteiro	ensino fundamental	—	12,00	41	00	outra
24	casado	superior	0	12,79	26	01	outra
25	casado	ensino médio	2	13,23	32	05	interior
26	casado	ensino médio	2	13,60	35	00	outra
27	solteiro	ensino fundamental	—	13,85	46	07	outra
28	casado	ensino médio	0	14,69	29	08	interior
29	casado	ensino médio	5	14,71	40	06	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior	—	16,22	31	05	outra
32	casado	ensino médio	1	16,61	36	04	interior
33	casado	superior	3	17,26	43	07	capital
34	solteiro	superior	—	18,75	33	07	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	02	interior

Fonte: Dados hipotéticos.

# Tabela de frequências

Exemplo 2.2. A Tabela 2.2 apresenta a *distribuição de frequências* da variável grau de instrução, usando os dados da Tabela 2.1.

**Tabela 2.2:** Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução.

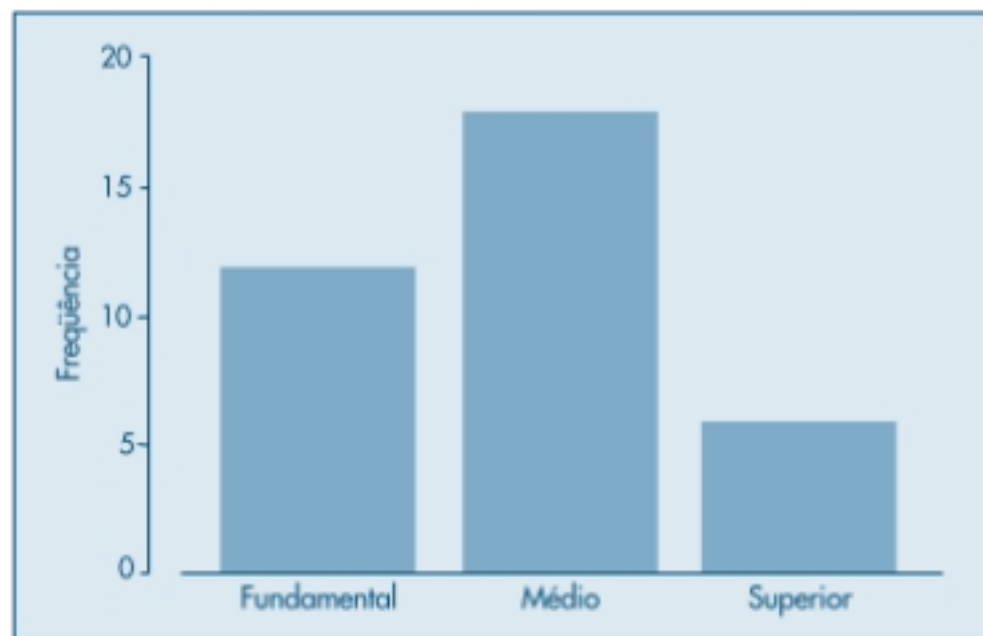
Grau de instrução	Frequência $n_i$	Proporção $f_i$	Porcentagem $100 f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

Fonte: Tabela 2.1.

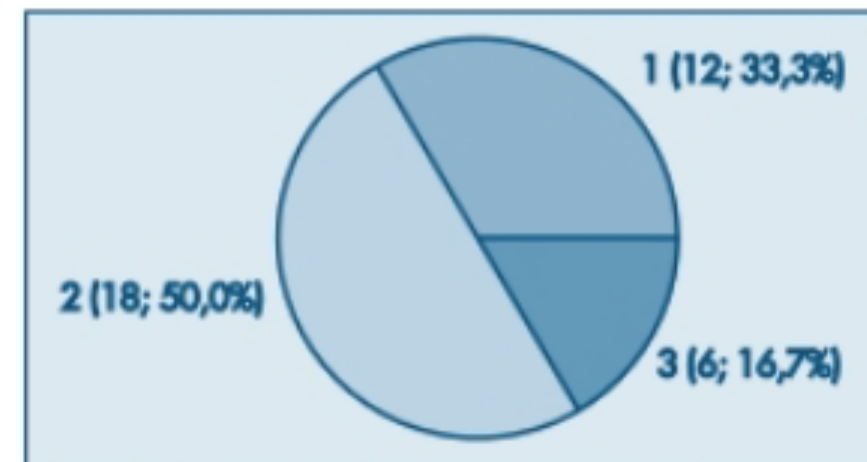


# Gráficos - qualitativa

**Figura 2.2:** Gráfico em barras para a variável Y: grau de instrução.



**Figura 2.3:** Gráfico em setores para a variável Y: grau de instrução.



1 = Fundamental, 2 = Médio e 3 = Superior

# Tabela de frequências

Idade	$n_i$	$f_i$	$n_i^{ac}$	$f_i^{ac}$
17	9	0,18	9	0,18
18	22	0,44	31	0,62
19	7	0,14	38	0,76
20	4	0,08	42	0,84
21	3	0,06	45	0,90
22	0	0,00	45	0,90
23	2	0,04	47	0,94
24	1	0,02	48	0,96
25	2	0,04	50	1,00
Total	$n = 50$			

Exemplo 2.3. A Tabela 2.4 dá a distribuição de frequências dos salários dos 36 empregados da seção de orçamentos da Companhia MB por faixa de salários.

**Tabela 2.4:** Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB por faixa de salário.

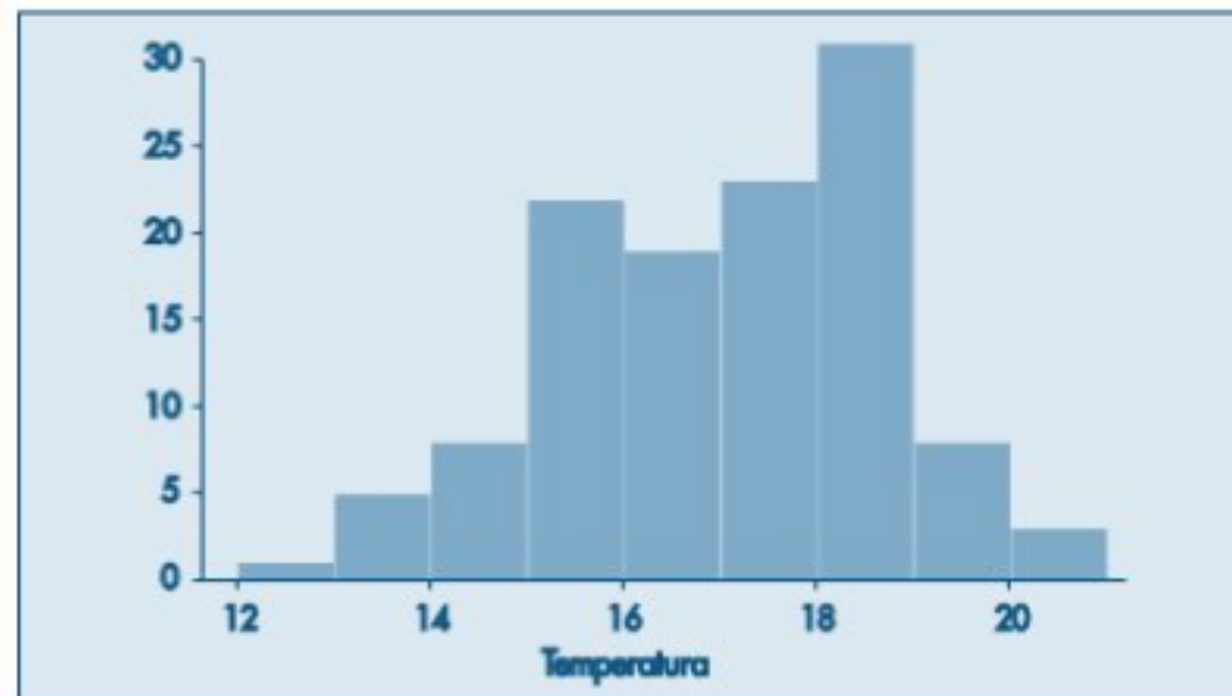
Classe de salários	Frequência $n_i$	Porcentagem $100 f_i$
4,00 – 8,00	10	27,78
8,00 – 12,00	12	33,33
12,00 – 16,00	8	22,22
16,00 – 20,00	5	13,89
20,00 – 24,00	1	2,78
Total	36	100,00

Fonte: Tabela 2.1.

# Histograma

$$d_i = \frac{f_i}{\delta_i}$$

Figura 2.16: Histograma dos dados de temperatura de São Paulo. SPlus.



# Medidas resumo - posição

- Média
- Mediana
- Moda
- Quantis

# Medidas resumo - posição

- Média
- Mediana
- Moda
- Quantis

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

# Medidas resumo - posição

- Média
- Mediana
- Moda
- Quantis

$$md(X) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar} \\ \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right), & \text{se } n \text{ for par} \end{cases}$$

# Medidas resumo - posição

- Média
- Mediana
- Moda
- Quantis

O  $p$ -quantil é o valor que divide o conjunto de dados ordenados em duas partes onde a primeira (à esquerda) concentra  $p\%$  dos dados e a segunda (à direita) concentra  $(1 - p)\%$  dos dados.

# Medidas resumo - dispersão

- Variância
- Desvio padrão
- Intervalo-interquartil
- Coeficiente de variação



# Medidas resumo - dispersão

- Variância
- Desvio padrão
- Intervalo-interquartil
- Coeficiente de variação

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

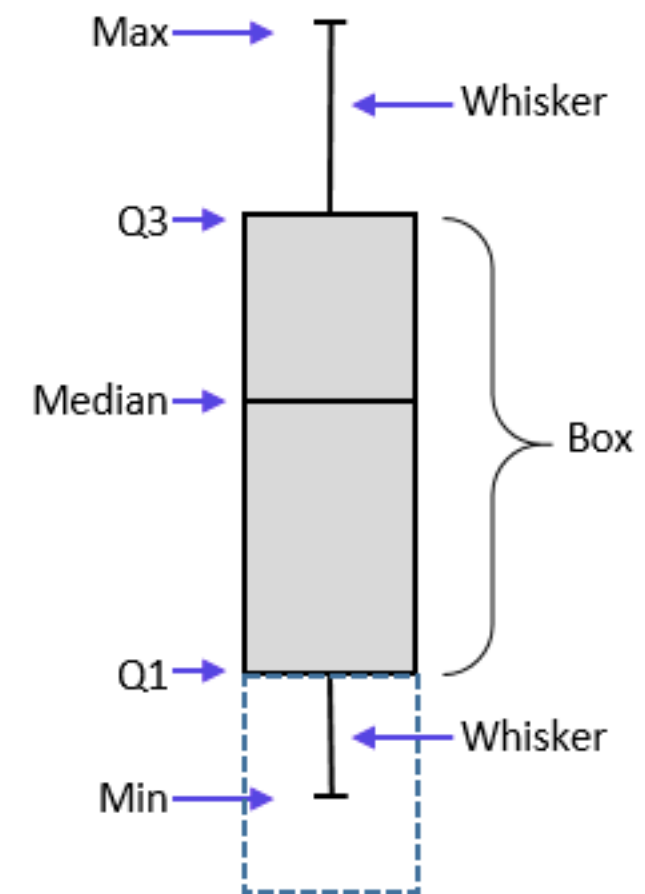
# Medidas resumo - dispersão

- Variância
- Desvio padrão
- Intervalo-interquartil
- Coeficiente de variação

$$CV = \frac{s}{x} \times 100$$

# Boxplot

Ponto de corte superior	$\min\{q_3 + 1,5d_q, x_{(n)}\}$
Terceiro quartil	$q_3$
Mediana	$q_2$
Primeiro quartil	$q_1$
Ponto de corte inferior	$\max\{q_1 - 1,5d_q, x_{(1)}\}$



# Assimetria

Assimetria à esquerda . . . . .

Simétrica . . . . .

Assimetria à direita . . . . .

- (1)  $\text{média} < \text{mediana} < \text{moda} \Rightarrow$  provável assimetria à esquerda.
- (2)  $\text{moda} < \text{mediana} < \text{média} \Rightarrow$  provável assimetria à direita.
- (3)  $\text{moda} = \text{mediana} = \text{média} \Rightarrow$  provável simetria.

# Análise bivariada

# Análise bivariada

- Duas variáveis qualitativas
- Duas variáveis quantitativas
- Uma variável quantitativa e outra qualitativa

# Análise bivariada

- Duas variáveis qualitativas
- Duas variáveis quantitativas
- Uma variável quantitativa e outra qualitativa

# Tabela de dupla entrada

$X Y$	$B_1$	$B_2$	$\dots$	$B_I$	Total
$A_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1I}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2I}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$A_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kI}$	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.I}$	$n_{..}$

- Tabela de frequências relativas: (i) em relação ao total, (ii) em relação ao total de cada coluna, (iii) em relação ao total de cada linha



# Exemplo: tabela de contingência

Exemplo: Considere o seguinte conjunto de dados

<i>X</i> :	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>
<i>Y</i> :	<i>S</i>	<i>S</i>	<i>N</i>	<i>N</i>	<i>S</i>	<i>S</i>	<i>N</i>	<i>S</i>	<i>N</i>

Podemos criar uma tabelas de dupla entrada de frequências

<i>X Y</i>	<i>S</i>	<i>N</i>	Total
<i>A</i>	2	2	4
<i>B</i>	3	2	5
Total	5	4	9

# Exemplo: frequências relativas

Em relação ao total geral

$X Y$	$S$	$N$	Total
$A$	22,22%	22,22%	44,44%
$B$	33,33%	22,22%	55,56%
Total	55,56%	44,44%	100%

Em relação ao total de  $X$  (total das linhas)

$X Y$	$S$	$N$	Total
$A$	50%	50%	100%
$B$	60%	40%	100%

Em relação ao total de  $Y$  (total das colunas)

$X Y$	$S$	$N$
$A$	40%	50%
$B$	60%	50%
Total	100%	100%

# Estatística $\chi^2$

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}.$$

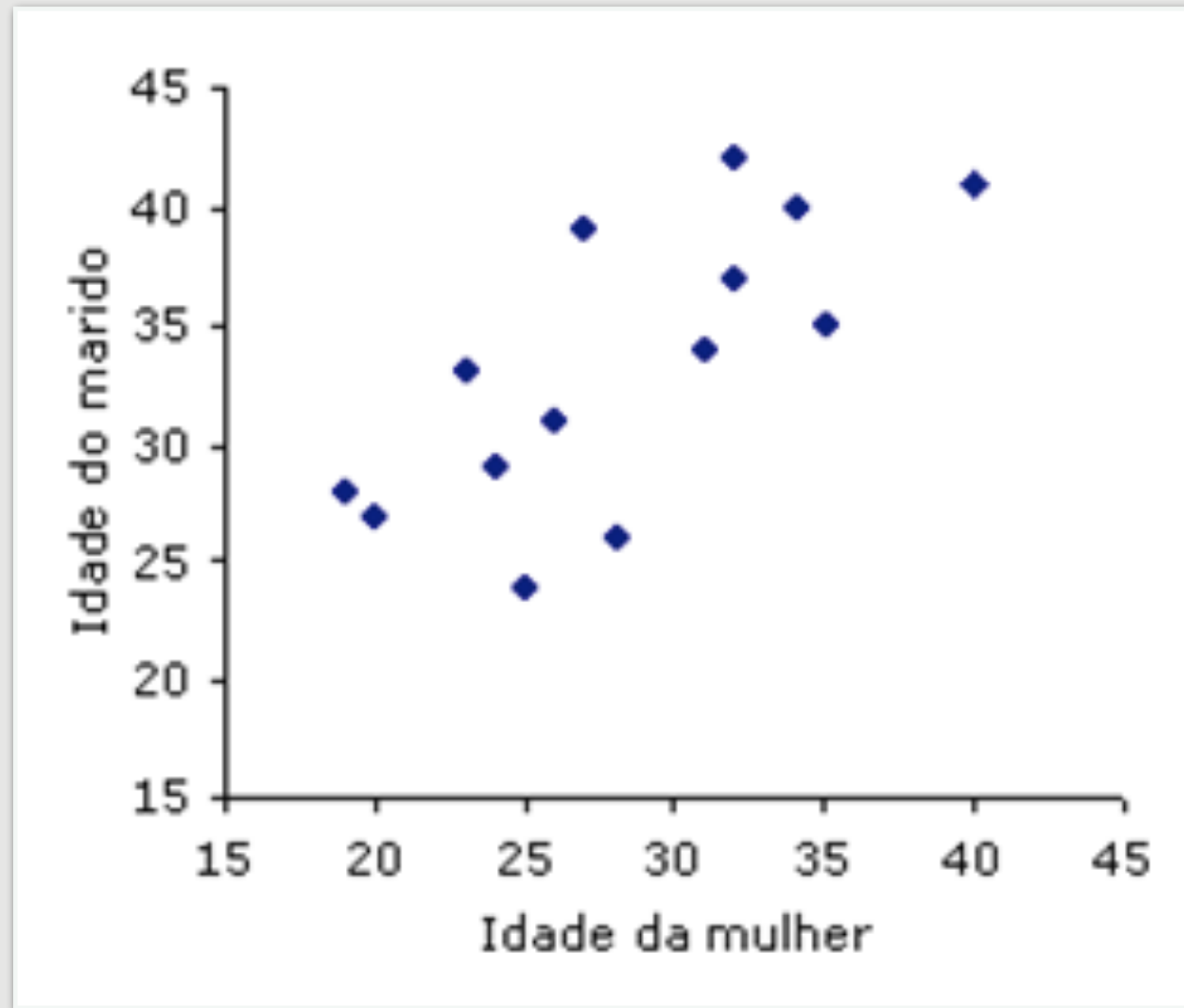
$$n_{ij}^* = \frac{n_{i.} n_{.j}}{n_{..}}$$

$X Y$	$B_1$	$B_2$	$\dots$	$B_l$	Total
$A_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$A_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

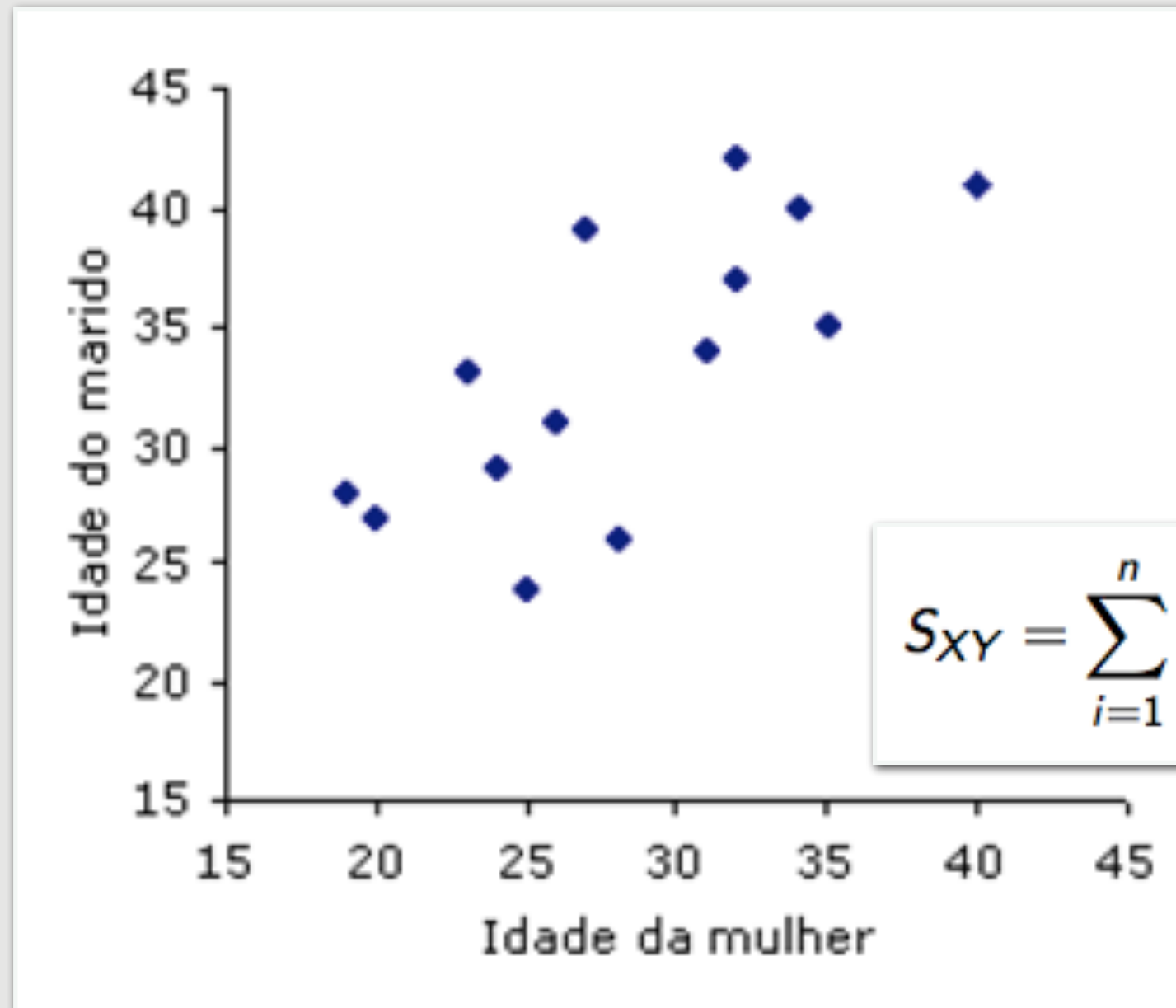
# Análise bivariada

- Duas variáveis qualitativas
- **Duas variáveis quantitativas**
- Uma variável quantitativa e outra qualitativa

# Diagrama de dispersão

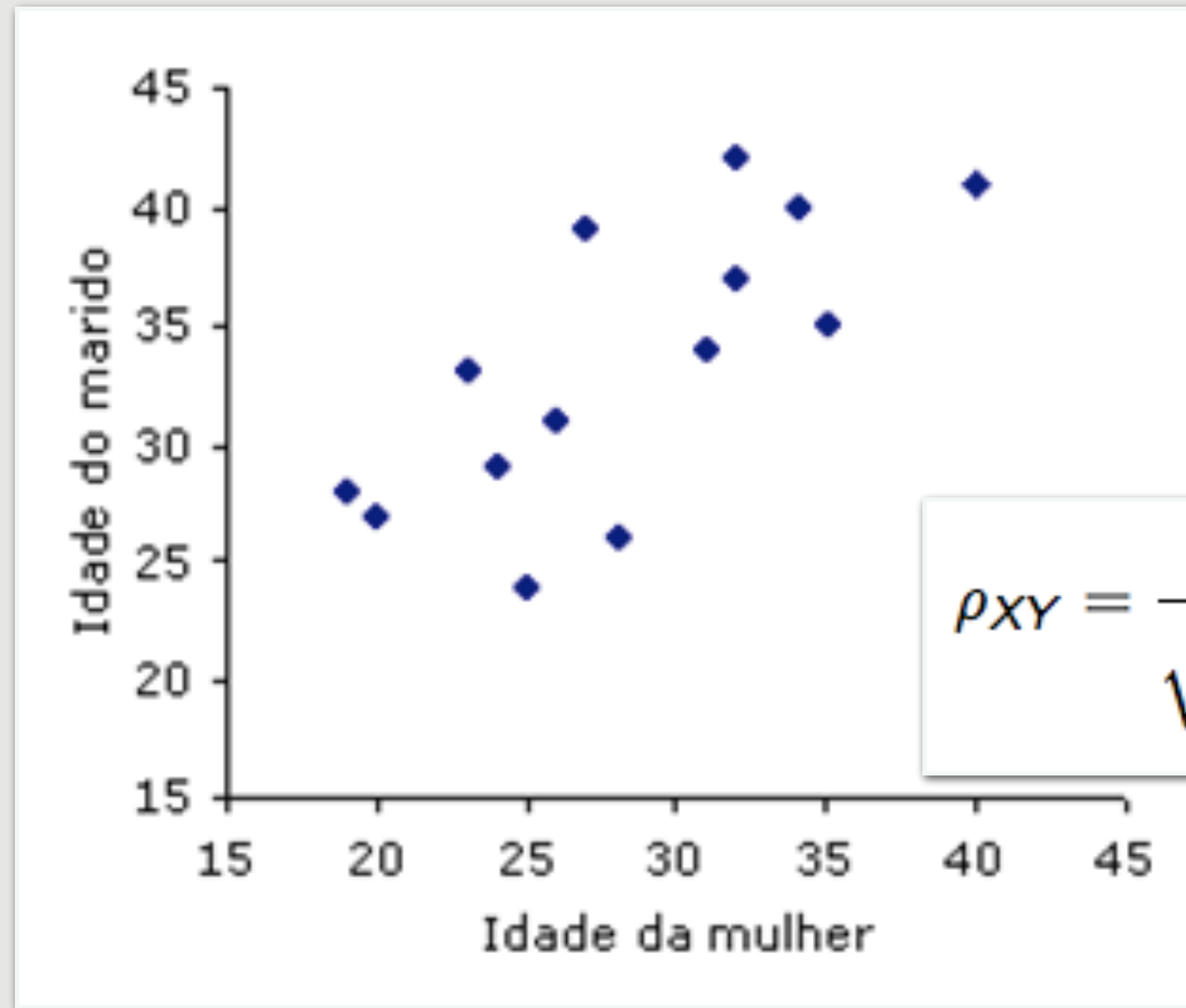


# Covariância



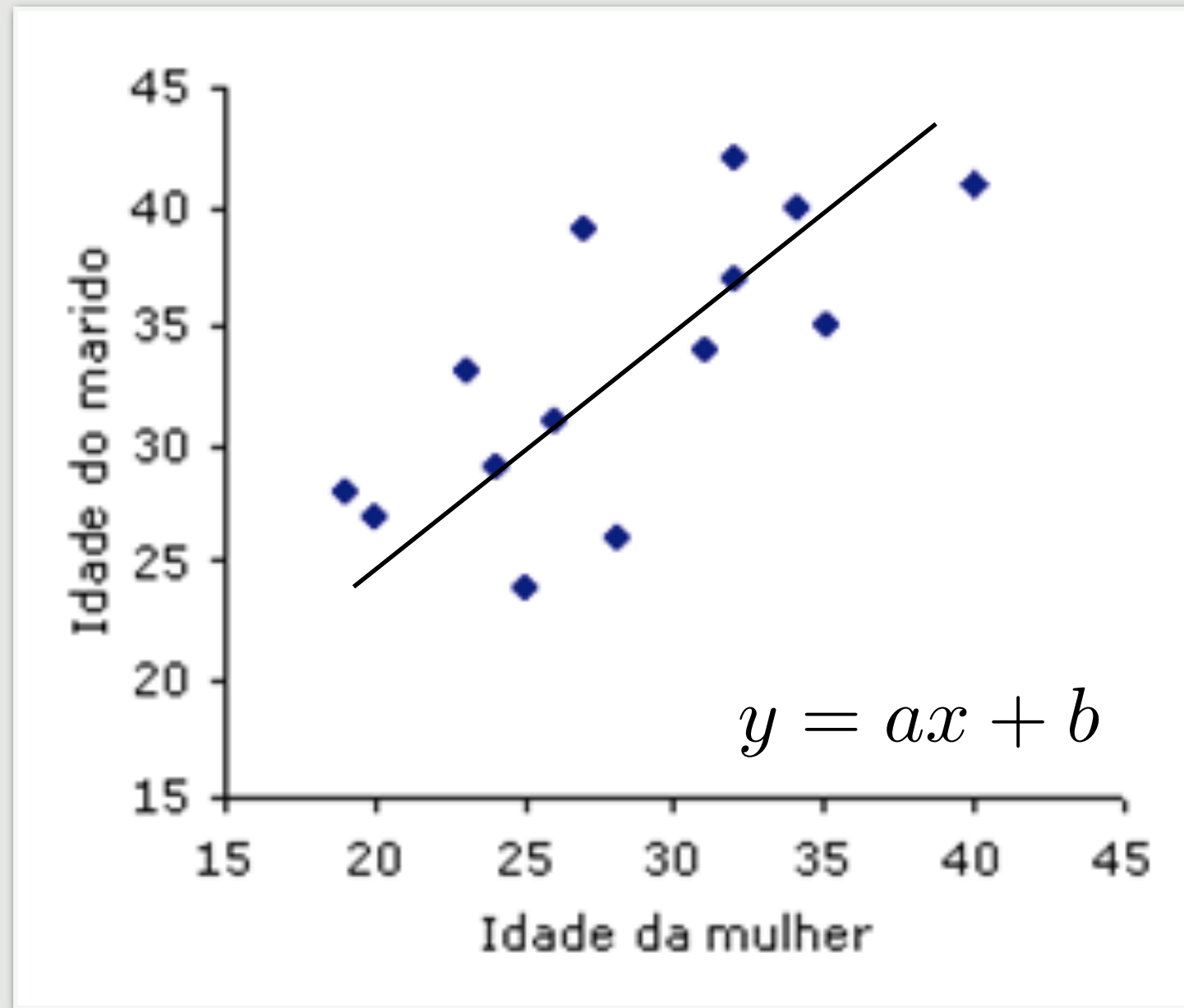
$$S_{XY} = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

# Coeficiente de correlação



$$\rho_{XY} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$$

# Regressão linear





# Análise bivariada

- Duas variáveis qualitativas
- Duas variáveis quantitativas
- Uma variável quantitativa e outra qualitativa

# Dividir em 'classes'

