Resumo de Dados

Tipos de Variáveis

Exemplo 2.1 **Um pesquisador está interessado em fazer um levantamento sobre alguns aspectos socioeconômicos dos empregados da seção de orçamentos da Companhia MB. Usando informações obtidas do departamento pessoal, ele elaborou a Tabela 2.1.**

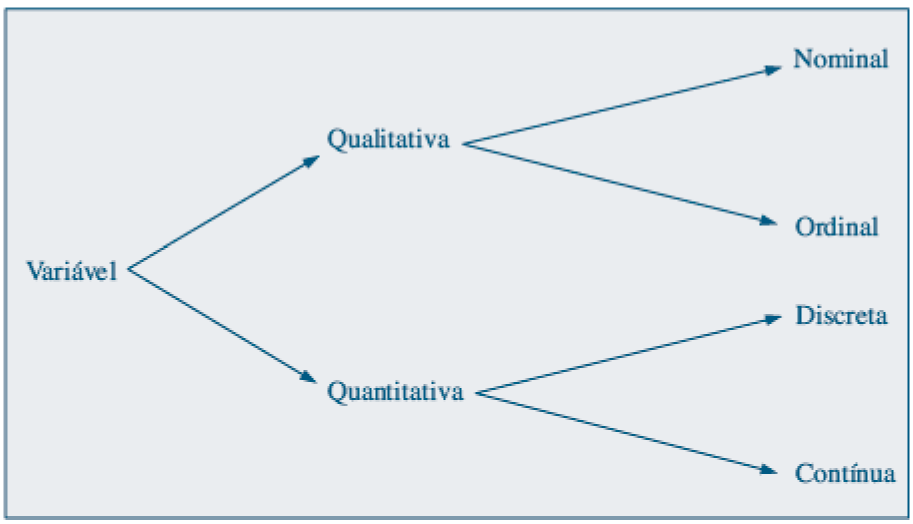
Tabela 2.1 Informações sobre estado civil, grau de instrução, número de filhos, salário (expresso como fração do salário mínimo), idade (medida em anos e meses) e procedência de 36 empregados da seção de orçamentos da Companhia MB.

Nº	Estado civil	Grau de instrução	Nº de filhos	Salário (× sal. mín.)	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

Variável: Qualquer característica associada a uma população.

Variável	Representação
Estado civil	X
Grau de instrução	Y
Número de filhos	Z
Salário	S
Idade	U
Região de procedência	V

Figura 2.1 Classificação de uma variável.



- **Variável qualitativa nominal, para a qual não existe nenhuma ordenação nas possíveis realizações: a região de procedência, sexo, cor dos olhos, estado civil;**

- **Variável qualitativa ordinal, para a qual existe uma ordem nos seus resultados: grau de instrução, classe social;**

- **Variáveis quantitativas discretas, cujos possíveis valores formam um conjunto finito ou enumerável de números, e que resultam, frequentemente, de uma contagem, como número de filhos (0, 1, 2, ...);**

- **Variáveis quantitativas contínuas, cujos possíveis valores pertencem a um intervalo de números reais e que resultam de uma mensuração, como por exemplo peso, altura, salário, idade.**

Distribuições de Frequências

Variáveis qualitativas

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.

Tabela 2.3 Frequências e porcentagens dos 2.000 empregados da Companhia MB, segundo o grau de instrução.

Grau de instrução	Frequência n_i	Porcentagem $100\,f_i$
Fundamental	650	32,50
Médio	1.020	51,00
Superior	330	16,50
Total	2.000	100,00

Fonte: Dados hipotéticos.

Variáveis quantitativas

A construção de tabelas de frequências para variáveis contínuas necessita de certo cuidado.

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 $\vdash$ 8,00	10	27,78
8,00 $\vdash$ 12,00	12	33,33
12,00 $\vdash$ 16,00	8	22,22
16,00 $\vdash$ 20,00	5	13,89
20,00 $\vdash$ 24,00	1	2,78
Total	36	100,00

Fonte: Tabela 2.1.

Procedendo-se desse modo, ao resumir os dados referentes a uma variável contínua, perde-se alguma informação.

A escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas.

Entretanto, deve-se observar que, com um pequeno número de classes, perde-se informação, e com um

número grande de classes, o objetivo de resumir os dados fica prejudicado.

Normalmente, sugere-se o uso de 5 a 15 classes com a mesma amplitude.

Gráficos

Gráficos para Variáveis Qualitativas

Exemplo 2.4 Tomemos como ilustração a variável Y: grau de instrução, exemplificada nas Tabelas 2.2 e 2.3.

O gráfico em barras consiste em construir retângulos ou barras, em que uma das dimensões é proporcional à magnitude a ser representada (n_i ou f_i).

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.

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

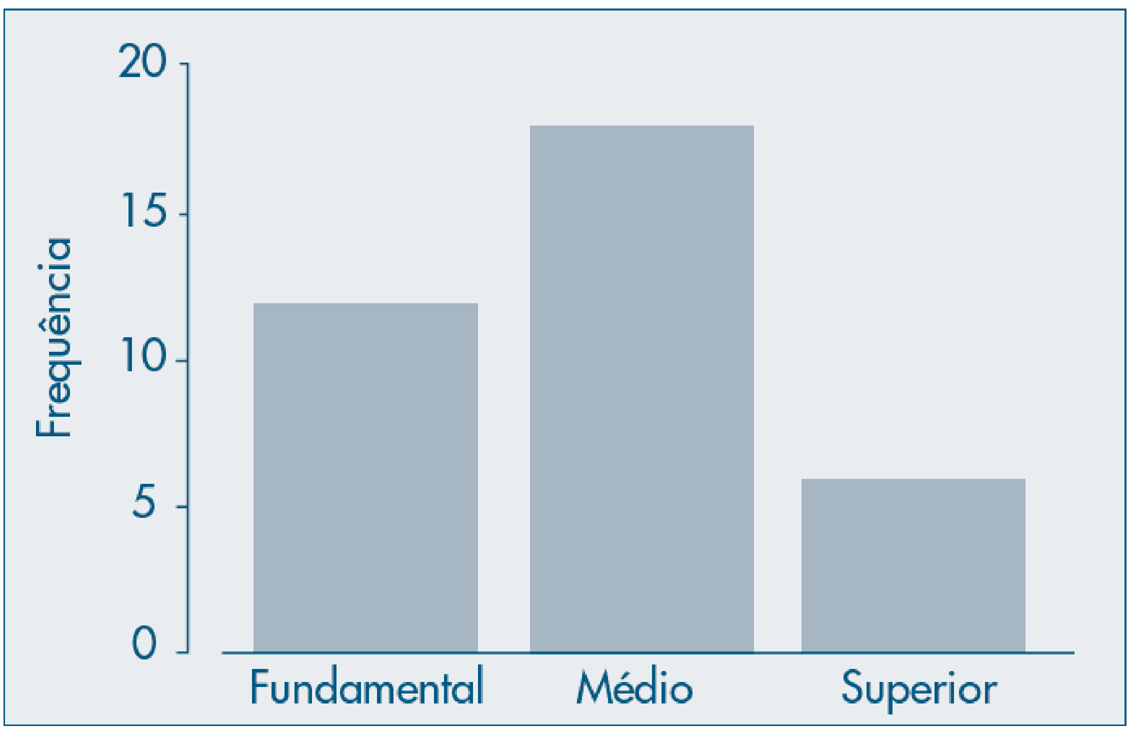
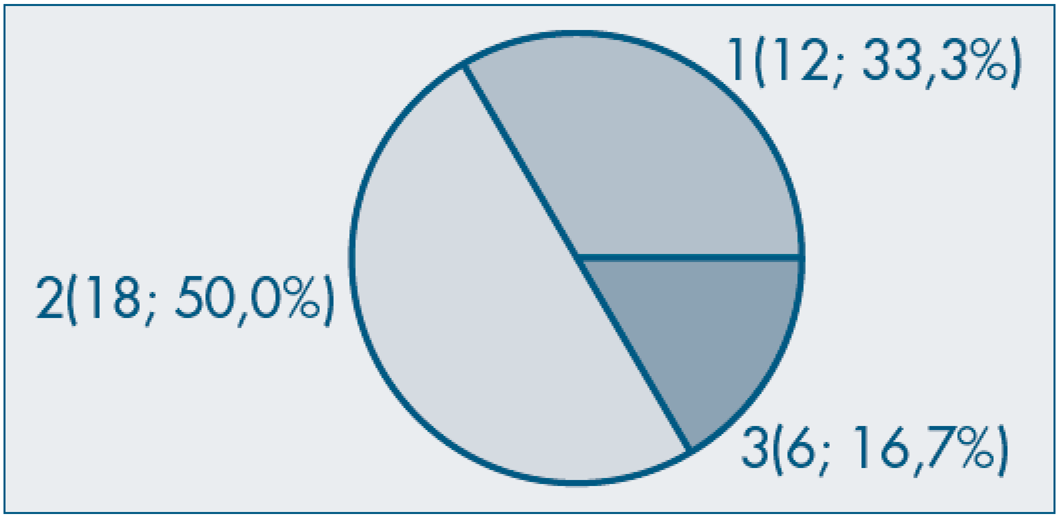


Figura 2.3 Gráfico em setores para a variável Y : grau de instrução.



1= Fundamental, 2 = Médio e 3 = Superior

Gráficos para Variáveis Quantitativas Discretas

Exemplo 2.5 Considere a distribuição da variável Z, número de filhos dos empregados casados da seção de orçamentos da Companhia MB (Tabela 2.1). Na Tabela 2.5, temos as frequências e porcentagens.

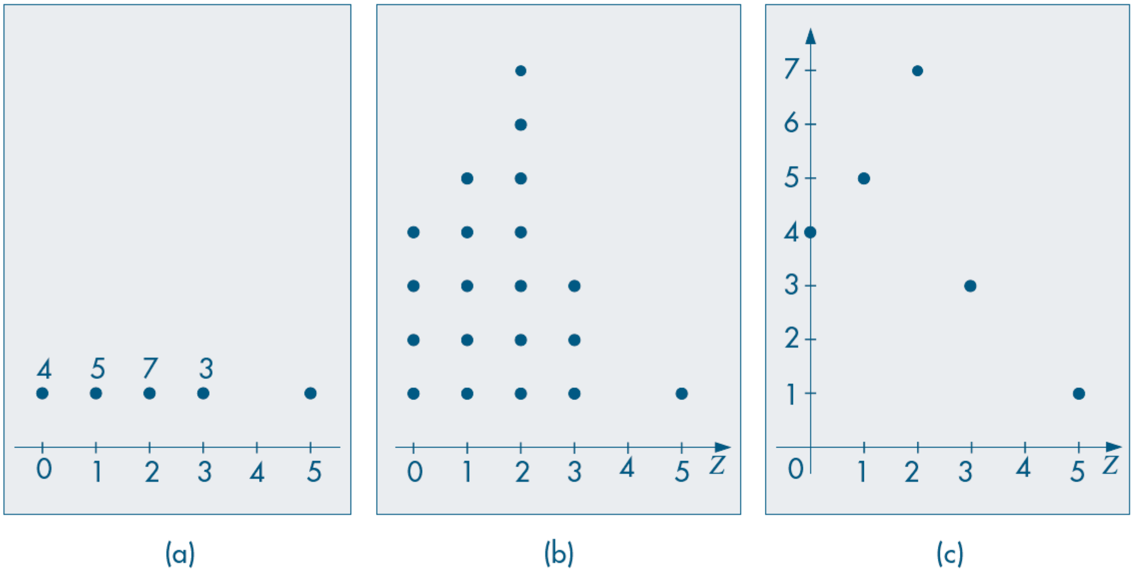
Tabela 2.5 Frequências e porcentagens dos empregados da seção de orçamentos da Companhia MB, segundo o número de filhos.

Nº de filhos z_i	Frequência n_i	Porcentagem $100\,f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Fonte: Tabela 2.1.

Figura 2.4 Gráfico em barras para a variável Z: número de filhos.

Figura 2.5 Gráficos de dispersão unidimensionais para a variável Z: número de filhos.



Gráficos para variáveis quantitativas contínuas

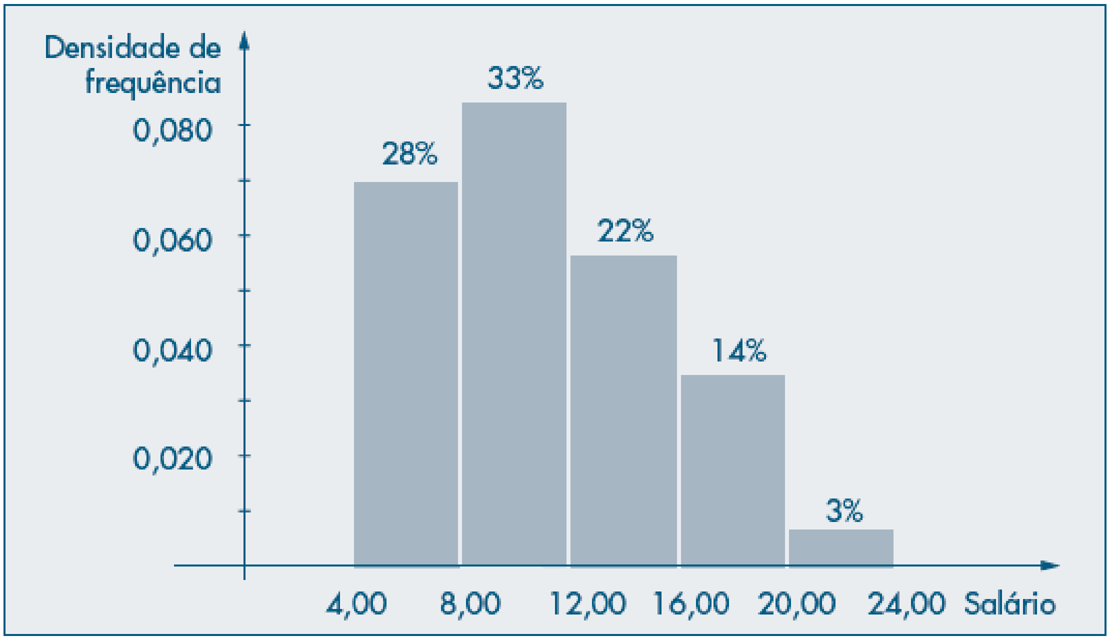
Exemplo 2.3 A Tabela 2.4 dá a distribuição de frequências dos salários dos 36 empregados da seção de orçamentos da Companhia MB por faixa de salários.

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.

Figura 2.7 Histograma da variável S: salários.



O histograma é um gráfico de barras contíguas, com as bases proporcionais aos intervalos das classes e a área de cada retângulo proporcional à respectiva frequência. Pode-se usar tanto a frequência absoluta, n_i, como a relativa, f_i.

Indiquemos a amplitude do i-ésimo intervalo por Δ_i. Para que a área do retângulo respectivo seja proporcional a f_i, a sua altura deve ser proporcional a f_i/Δ_i (ou a n_i/Δ_i), que é chamada densidade de frequência da i-ésima classe.

Com essa convenção, a área total do histograma será igual a um.

Ramo-e-Folhas

Exemplo 2.8 Na Figura 2.9, construímos o ramo-e-folhas dos salários de 36 empregados da Companhia MB (Tabela 2.1). Não existe uma regra fixa para construir o ramo-e-folhas, mas a ideia básica é dividir cada observação em duas partes: a primeira (o ramo) é colocada à esquerda de uma linha vertical, a segunda (a folha) é colocada à direita. Assim, para os salários 4,00 e 4,56, o 4 é o ramo e 00 e 56 são as folhas.

Um ramo com muitas folhas significa maior incidência daquele ramo (realização).

Figura 2.9 Ramo-e-folhas para a variável S: salários.

4	00	56		
5	25	73		
6	26	66	86	
7	39	44	59	
8	12	46	74	95
9	13	35	77	80
10	53	76		
11	06	59		
12	00	79		
13	23	60	85	
14	69	71		
15	99			
16	22	61		
17	26			
18	75			
19	40			
20				
21				
22				
23	30			

Algumas informações que se obtêm deste ramo-e-folhas são:

(a) Há um destaque grande para o valor 23,30.

(b) Os demais valores estão razoavelmente concentrados entre 4,00 e 19,40.

(c) Um valor mais ou menos típico para este conjunto de dados poderia ser, por exemplo, 10,00.

(d) Há uma leve assimetria em direção aos valores grandes; a suposição de que estes dados possam ser considerados como amostra de uma população com

distribuição simétrica, em forma de sino (a chamada distribuição normal), pode ser questionada.

Exemplo 2.9 Os dados abaixo referem-se à dureza de 30 peças de alumínio (Hoaglin; Mosteller; Tukey, 1983, p. 13).

53,0	70,2	84,3	69,5	77,8	87,5	53,4	82,5	67,3	54,1
70,5	71,4	95,4	51,1	74,4	55,7	63,5	85,8	53,5	64,3
82,7	78,5	55,7	69,1	72,3	59,5	55,3	73,0	52,4	50,7

Na Figura 2.10, temos o ramo-e-folhas correspondente. Aqui, optamos por truncar cada valor, omitindo os décimos, de modo que 69,1 e 69,5, por exemplo, tornam-se 69 e 69 e aparecem como 9 na linha que corresponde ao ramo 6.

Figura 2.10 Ramo-e-folhas para os dados de dureza de peças de alumínio.

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

Este é um exemplo em que temos muitas folhas em cada ramo. Uma maneira alternativa é duplicar os ramos. Criamos os ramos 5* e 5•, 6* e 6• etc., nos quais colocamos folhas de 0 a 4 na linha * e folhas de 5 a 9 na linha •. Obtemos o ramo-e-folhas da Figura 2.11.

Um ramo-e-folhas pode ser "adornado" com outras informações, como o número de observações em cada ramo. Para outros exemplos, veja o Problema 19.

Figura 2.11 Ramo-e-folhas para os dados de dureza, com ramos divididos.

5*	0 1 2 3 3 3 4
5•	5 5 5 9
6*	3 4
6•	7 9 9
7*	0 0 1 2 3 4
7•	7 8
8*	2 2 4
8•	5 7
9*	
9•	5