

BIOESTATÍSTICA

Juliane Silveira
Freire da Silva

Revisão técnica:

Rute Henrique da Silva Ferreira

Licenciada em Matemática

Mestre em Educação Matemática

Doutora em Sensoriamento Remoto



P228b Parenti, Tatiane.

Bioestatística / Tatiane Parenti, Juliane Silveira Freire da Silva, Jamur Silveira; [revisão técnica : Rute Henrique da Silva Ferreira]. – Porto Alegre: SAGAH, 2018.

207 p. il. ; 22,5 cm

ISBN 978-85-9502-362-8

1. Bioestatística. I. Silva, Juliane Silveira Freire da. II. Silveira, Jamur. III. Título.

CDU 311

Organização de dados: tabelas e gráficos

Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Reconhecer por que os dados devem ser organizados em estatística.
- Identificar os principais tipos de tabelas e gráficos.
- Selecionar o tipo de gráfico mais adequado para cada tipo de situação.

Introdução

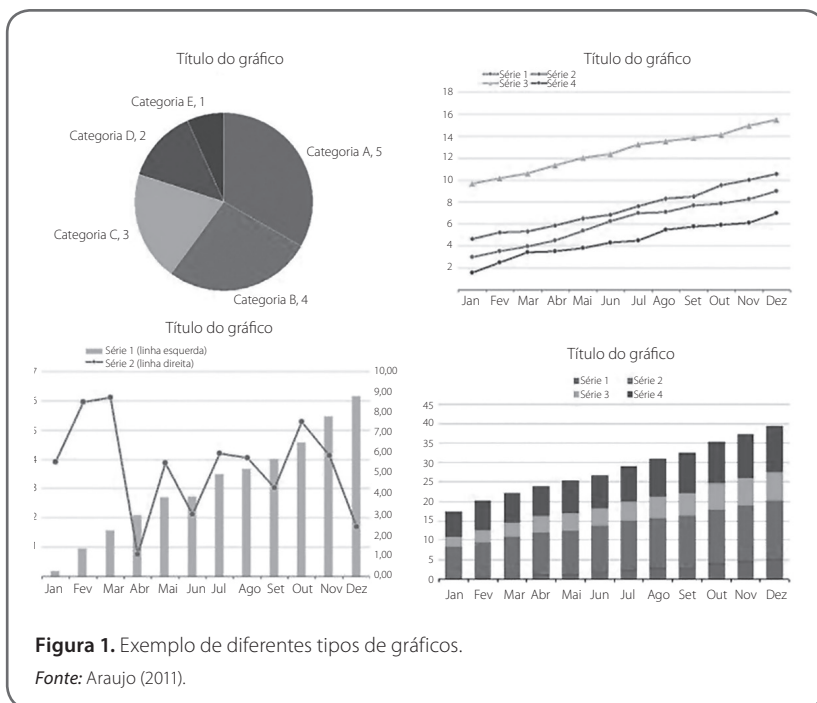
Neste capítulo, vamos calcular e aplicar métodos estatísticos à análise de dados. A partir daí, construir e analisar tabelas e gráficos utilizando as normas científicas.

Por que organizamos os dados em estatística?

Quando estamos coletando os dados, essa coleta ocorre de forma aleatória e, durante esse processo, não temos a capacidade de organizá-los e também não temos condições de tomar alguma decisão com base na coleta, sem o tratamento desses dados.

Por esse motivo, precisamos começar a analisar os dados coletados e, de alguma forma, resumi-los para podermos visualizar os resultados de forma organizada, iniciando, assim, a análise descritiva dos dados.

Primeiramente, resumimos em tabelas de distribuição de frequências e depois podemos fazer gráficos, o que visualmente é melhor para representar os dados (Figura 1). A análise descritiva dos dados ainda dispõe de outras técnicas além dessas, mas, neste capítulo, atentaremos para a análise de tabelas e gráficos.



Fique atento

Sobre a coleta de dados, é importante estarmos sempre atentos à forma como coletamos os dados. Precisamos, antes de qualquer coleta, estabelecer a metodologia para a escolha das unidades amostrais. Muitas vezes, quando coletamos dados, estamos interessados em poder fazer inferência para o restante da população (extrapolar para toda a população). Somente quando temos uma amostra probabilística – ou seja, os elementos da população são escolhidos por sorteio aleatório – que poderemos realizar inferências. Caso a amostra não seja probabilística, poderemos apenas fazer uma análise descritiva dos dados e o resultado dessa análise dirá respeito somente à amostra pesquisada.

Tipos de tabelas e gráficos

Existem tabelas que são para dados qualitativos, que também chamamos de tabelas para dados categóricos (Tabela 1). São tabelas simples em que se anota a frequência que cada uma das opções de resposta aparece na amostra.

Tabela 1. Exemplo de tabela com dados qualitativos.

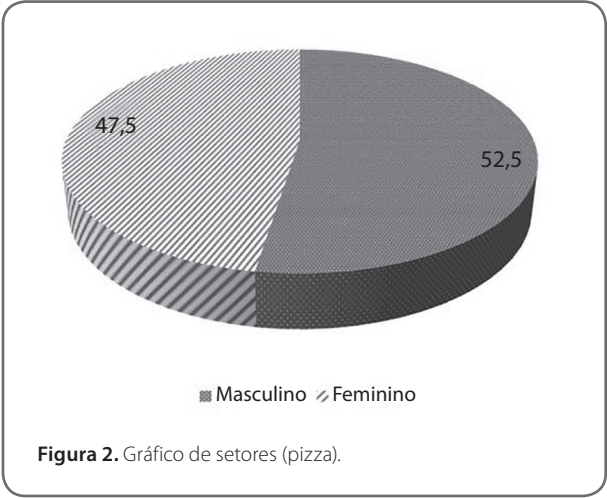
Sexo	F	Fr
Masculino	63	52,5
Feminino	57	47,5
Total	120	100,0

Conforme verificado na Tabela 1, a coluna f (frequência simples absoluta) é resultado da contagem da frequência que cada uma das palavras apareceu na amostra. Ou seja, havia 63 pessoas do sexo masculino e 57 do sexo feminino na amostra.

Para calcularmos a coluna fr , precisamos ver quanto cada uma das frequências tem de proporção no total da amostra. Podemos resolver isso por regra de três.

$$\begin{array}{ccc} 120 & 100\% & \\ 63 & x & \end{array} \rightarrow 120 \cdot x = 63 \cdot 100 \rightarrow x = \frac{63 \cdot 100}{120} = 52,5\%$$

Podemos representar essa tabela com um gráfico de setores, também conhecido como gráfico de pizza, conforme a Figura 2.



Observe que em tabelas para dados de uma variável qualitativa nominal, devemos ordenar do mais frequente para o menos frequente. Já quando temos uma variável qualitativa ordinal, precisamos respeitar a ordem em que a variável é apresentada (Tabela 2).

Tabela 2. Exemplo de tabela com dados qualitativos ordinais sobre a satisfação com o atendimento recebido em uma Unidade de Pronto Atendimento (UPA) de Porto Alegre, RS.

Satisfação	F	fr
Muito satisfeito	12	13,3
Satisfeito	14	15,6
Indiferente	21	23,3
Insatisfeito	19	21,1
Muito insatisfeito	24	26,7
Total	90	100

Para representarmos essa tabela, podemos fazer um gráfico de colunas, conforme a Figura 3.

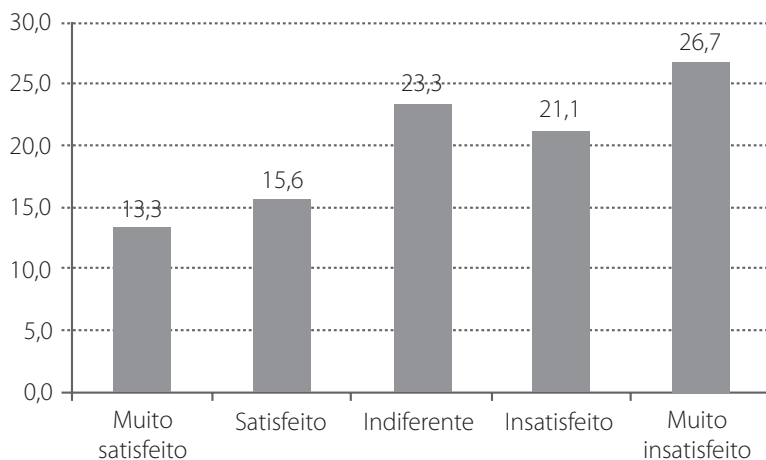


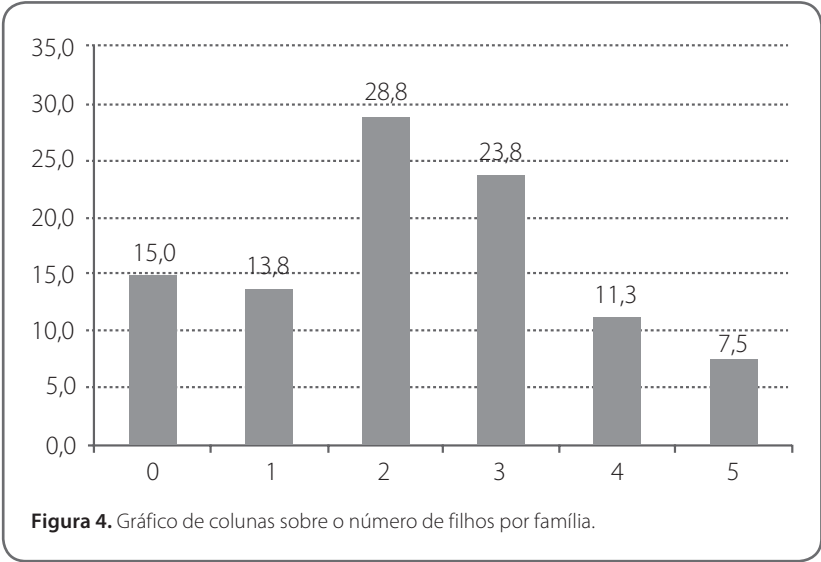
Figura 3. Exemplo de gráfico de colunas.

Podemos também utilizar as tabelas para representar dados quantitativos. Nesse caso, podemos ter tabelas por ponto e tabelas por intervalos (também chamadas de tabelas por classes). Variáveis quantitativas discretas costumam gerar tabelas de distribuição de frequência por ponto (Tabela 3).

Tabela 3. Exemplo de tabela quantitativa sobre o número de filhos por família.

Número de filhos	F	fr
0	12	15,0
1	11	13,8
2	23	28,8
3	19	23,8
4	9	11,3
5	6	7,5
Total	80	100

Também podemos representar esses dados com um gráfico de colunas, conforme a Figura 4.

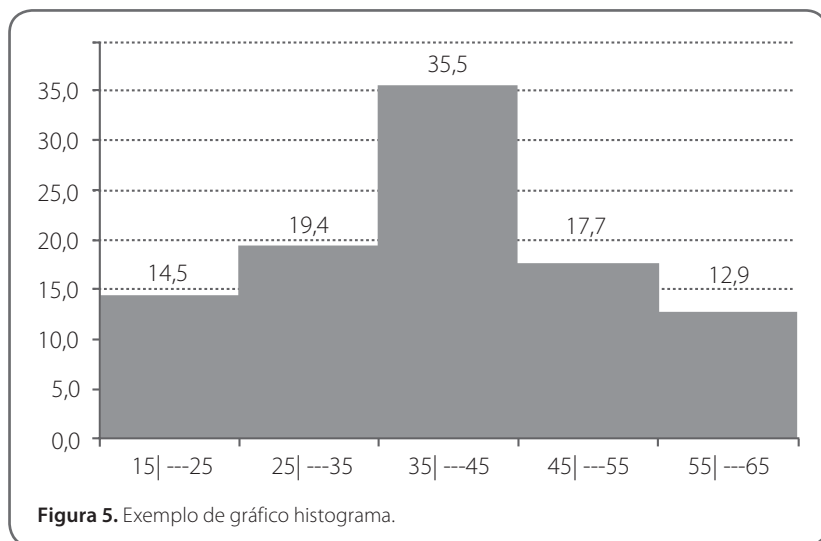


Já as variáveis quantitativas geram tabelas de distribuição de frequências por intervalos (Tabela 4).

Tabela 4. Exemplo de tabela com variáveis quantitativas sobre a faixa etária.

Faixa	F	Fr
15 ---25	9	14,5
25 ---35	12	19,4
35 ---45	22	35,5
45 ---55	11	17,7
55 ---65	8	12,9
Total	62	100,0

Para representarmos essa tabela, precisamos nos dar conta de um fato: entre as faixas, não existe um intervalo numérico, pois chegamos ao limite de um número e na faixa seguinte já iniciamos com ele. Assim, não podemos representar nenhum espaço no eixo do gráfico quando temos um gráfico de colunas. Nesse caso, as colunas estão grudadas umas às outras, e chamamos esse gráfico de histograma (Figura 5).



Saiba mais

Quando temos uma variável quantitativa discreta, pode ser que também precisemos fazer intervalos para melhor representar os dados. Caso existam mais de 10 opções de resposta, já podemos montar os intervalos para poder representar melhor esses dados.

Podemos ainda acrescentar mais colunas a essas tabelas que representam dados quantitativos para utilizarmos para fins de análise (Tabela 5). As colunas que necessariamente precisam aparecer em uma tabela de distribuição de frequências, além da primeira coluna que representa as opções de resposta dos dados coletados, são:

- $f \rightarrow$ frequência simples absoluta (resulta da contagem na amostra).
- $fr \rightarrow$ frequência simples relativa (resulta da regra de três vista anteriormente no capítulo).
- $F \rightarrow$ frequência acumulada absoluta (resulta somando a coluna f).
- $Fr \rightarrow$ frequência acumulada relativa (resulta somando a coluna fr).
- $x' \rightarrow$ ponto médio do intervalo, no caso da tabela de intervalos.

Tabela 5. Exemplo de tabela de faixa etária com demais colunas.

Faixa	f	fr	F	Fr	
15 ---25	9	14,5	9	14,5	$(15+25)/2=20$
25 ---35	12	19,4	$9+12=21$	33,9	$(25+35)/2=30$
35 ---45	22	35,5	$21+22=43$	69,4	$(35+45)/2=40$
45 ---55	11	17,7	$43+11=54$	87,1	$(45+55)/2=50$
55 ---65	8	12,9	$54+8=62$	100,0	$(55+65)/2=60$
Total	62	100,0	-	-	-



Saiba mais

Sobre a nomenclatura para a tabela de distribuição de frequências por intervalos, a barra na vertical (|) indica que o número ao seu lado está contido no intervalo. Quando temos o traço na horizontal, chegamos muito próximo ao número que está ao seu lado, mas não chegamos até ele. Por exemplo:

15|---25 \rightarrow o número 15 está contido nesse intervalo, mas o número 25 não.

15---|25 \rightarrow o número 15 não está contido nesse intervalo e o número 25 sim.

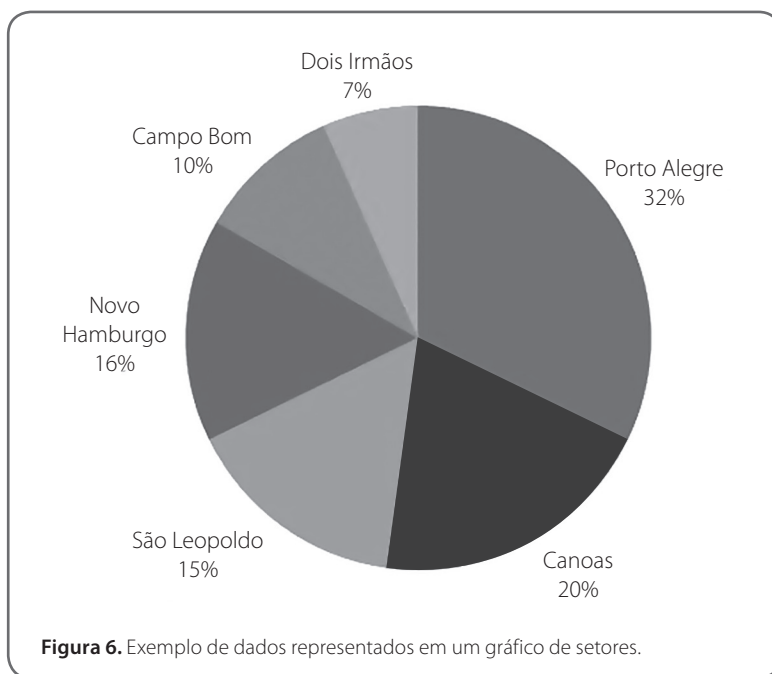
15---25 \rightarrow o número 15 não está contido nesse intervalo e o número 25 também não.

15|---|25 \rightarrow o número 15 está contido nesse intervalo e o número 25 também.

Agora, qual gráfico escolher?

Além dos gráficos apresentados aqui, temos uma grande quantidade de gráficos. Os mais básicos para a análise descritiva de dados são os de setores e os de barras ou colunas, mas não são somente esses que podemos utilizar.

Quando tivermos uma variável qualitativa, tanto nominal quanto ordinal, podemos representar esses dados com um gráfico de setores, de colunas ou barras (Figura 6).



Para os mesmos dados, poderíamos representar em um gráfico de colunas e de barras (Figuras 7 e 8).

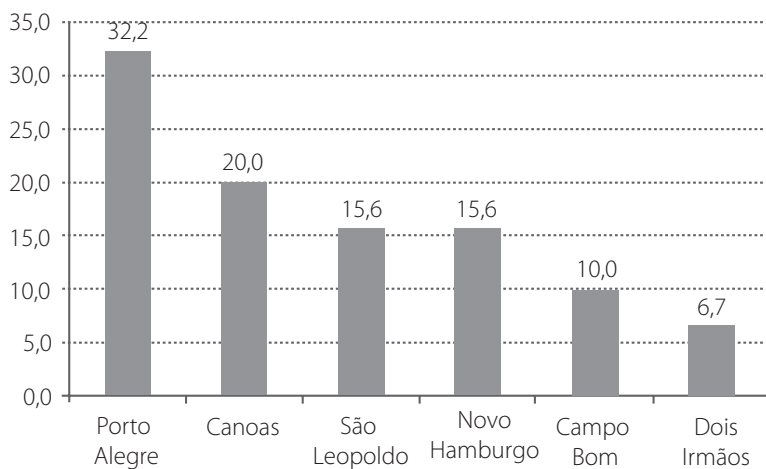


Figura 7. Exemplo de gráfico de colunas utilizando os dados da Figura 6.

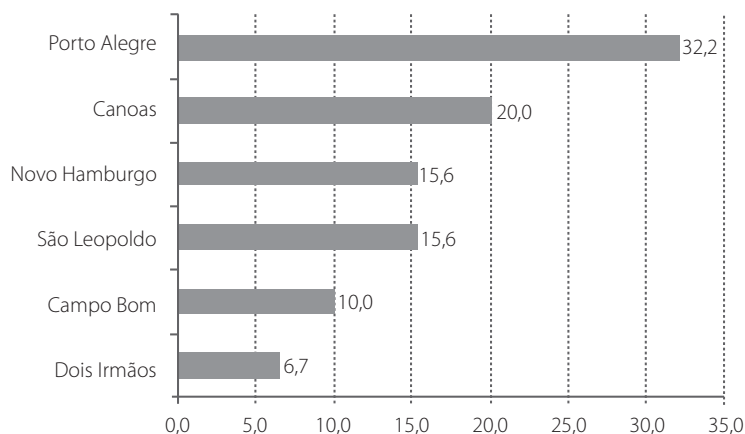


Figura 8. Exemplo de gráfico de barras utilizando os dados da Figura 6.

Agora, para as variáveis quantitativas para tabelas de distribuição de frequências simples ou por intervalos, podemos ter gráficos de colunas para representar as variáveis quantitativas discretas, conforme mostra a Figura 9.

Para os dados de variáveis quantitativas representadas em tabelas de distribuição de frequências por intervalos, representamos graficamente com um histograma, conforme mostra a Figura 10.

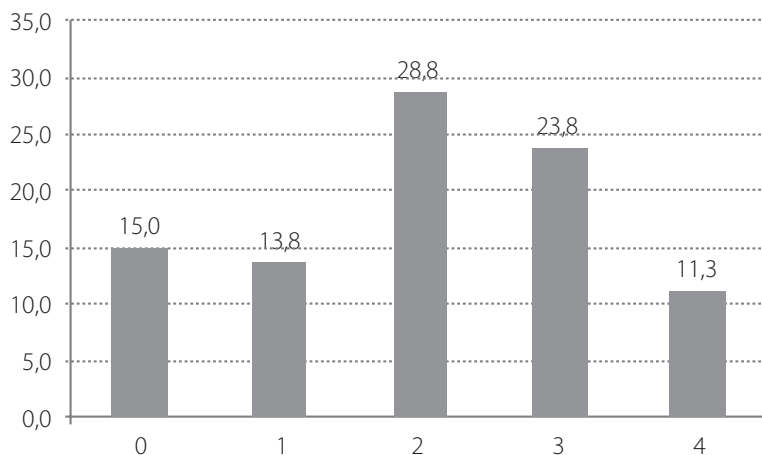


Figura 9. Exemplo de gráfico de colunas com variáveis quantitativas discretas.

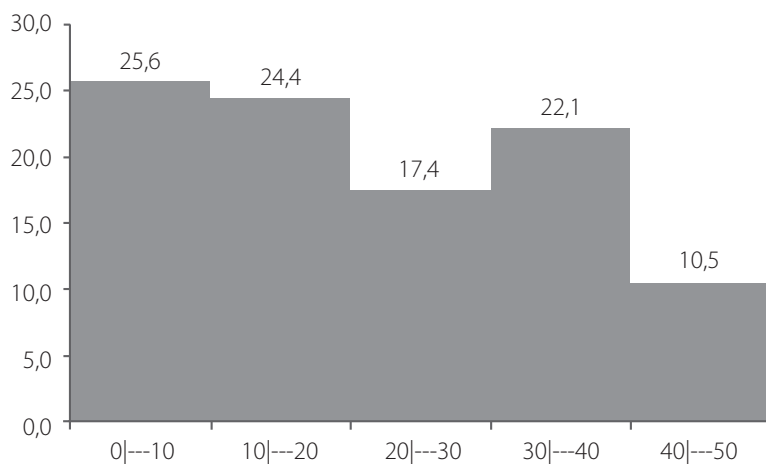


Figura 10. Exemplo de histograma.

Além desses gráficos, podemos citar ainda o gráfico de dispersão, que é utilizado em análise de correlação e regressão, quando temos duas variáveis e verificamos a relação entre elas. Imaginemos duas variáveis, peso e altura. Podemos, com o gráfico de dispersão (Figura 11), verificar a relação entre elas. Cada um dos pontos representa um par de valores (peso no eixo y e altura no eixo x).

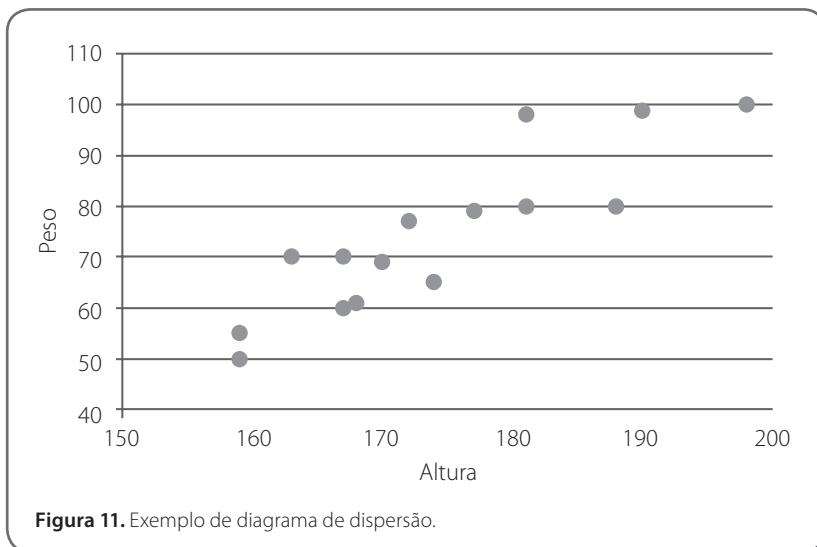


Figura 11. Exemplo de diagrama de dispersão.

O gráfico de linhas é utilizado quando desejamos representar uma variável quantitativa ao longo do tempo (Figura 12). O eixo x sempre será o tempo. Imaginemos acompanhar a evolução do número de nascidos vivos em uma pequena maternidade ao longo dos anos.

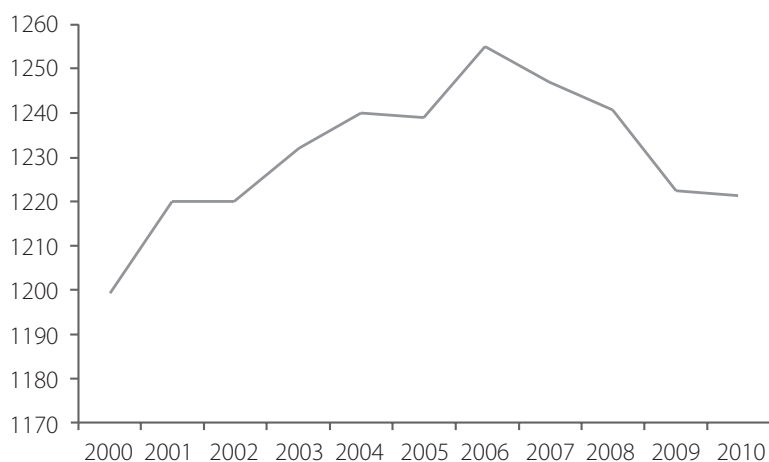


Figura 12. Exemplo de gráfico de linhas.



Link

Nestes endereços eletrônicos, você pode aprender a fazer gráficos utilizando o Excel:

<https://goo.gl/4mQZ0m>

<https://goo.gl/Ek3Ydy>



Exemplo

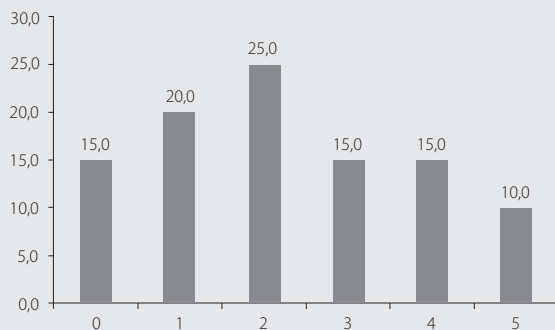
Considere os dados referentes a uma pesquisa com 20 famílias de um bairro pequeno, onde foi perguntado quantas vezes o chefe da família procurou o médico no ano anterior. As respostas da coleta são as seguintes:

1	4	2	0	2
2	2	3	0	4
5	0	1	1	3
3	1	4	2	5

Para representarmos esses dados, o primeiro passo é a montagem da tabela de distribuição de frequências. Precisamos contar quantas vezes cada um dos números apareceu e então fazer os seus percentuais.

nº de visitas	f	fr
0	3	15
1	4	20
2	5	25
3	3	15
4	3	15
5	2	10
total	20	100

A segunda maneira de representarmos esses dados seria por meio de um gráfico.



Concluimos então que o número mais frequente de visitas é igual a 2, representando 25%. Ou seja, mais da metade dos chefes de família foi, no máximo, até duas vezes a uma consulta com um médico no último ano.



Referência

ARAUJO, A. *Gráficos: modelos prontos*. 04 fev. 2011. Disponível em: <<http://geomorfologiacesc.blogspot.com.br/2011/02/graficos-modelos-prontos.html>>. Acesso em: 26 out. 2017.

Leituras recomendadas

CALLEGARI-JACQUES, S. M. *Bioestatística: princípios e aplicações*. Porto Alegre: Artmed, 2007.

FREUND, J. E. *Estatística aplicada economicamente*. 11. ed. Porto Alegre: Bookman, 2007.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.