

Estatística e Probabilidade

Representação gráfica e tabular



Núcleo de Educação a Distância

www.unigranrio.com.br

Rua Prof. José de Souza Herdy, 1.160
25 de Agosto – Duque de Caxias - RJ

Reitor

Arody Cordeiro Herdy

Pró-Reitoria de Programas de Pós-Graduação

Nara Pires

Pró-Reitoria de Programas de Graduação

Lívia Maria Figueiredo Lacerda

Pró-Reitoria Administrativa e Comunitária

Carlos de Oliveira Varella

Núcleo de Educação a Distância (NEAD)

Lúcia Inês Kronemberger Andrade

Produção: Gerência de Desenho Educacional - NEAD

Desenvolvimento do material: Gregório Dalle Vedove Nosaki

1ª Edição

Copyright © 2022, Unigranrio

Nenhuma parte deste material poderá ser reproduzida, transmitida e gravada, por qualquer meio eletrônico, mecânico, por fotocópia e outros, sem a prévia autorização, por escrito, da Unigranrio.

Sumário

Representação gráfica e tabular

Para Início de Conversa	4
Objetivos	4
1. Tipos de gráficos	5
2. Tipos de tabelas	10
3. Distribuição de frequências	12
Referências	17



Para Início de Conversa

Os gráficos e as tabelas podem expressar de maneira concisa e organizada muitas informações relevantes sobre uma pesquisa, um estudo ou simplesmente o levantamento de dados realizado em um trabalho ou publicação. Esses dois tipos de representação de dados são muito importantes na Estatística e podem auxiliar a transmitir uma mensagem de maneira mais impactante ou até mesmo tendenciosa. Estudaremos, neste capítulo, as diferentes formas, gráficos e tabelas e os elementos fundamentais de cada um deles. Além da construção e da elaboração desses dois elementos representativos, é importante trabalhar a interpretação dessas representações, pois essa análise crítica é uma habilidade fundamental a ser desenvolvida dentro da Estatística. Trabalharemos ainda com a distribuição de frequências, que pode auxiliar na compreensão da relação dos dados como um todo e que pode ser apresentada em forma de tabela ou gráfico de diferentes tipos. Estudaremos diversos exemplos para ilustrar todas as definições e os elementos que serão introduzidos neste capítulo.

Objetivos

- Descrever e apresentar os resultados de um conjunto de observações a partir de uma distribuição de frequências.
- Interpretar os resultados apresentados em gráficos e tabelas de forma clara, objetiva e passando o máximo de informações possíveis.



1. Tipos de gráficos

Os gráficos têm um poder de síntese e um apelo visual que os tornam ferramentas muito importantes para a divulgação de dados ou da relação entre esses dados em qualquer texto. Morettin e Bussab (2017) destacam as seguintes finalidades que uma representação gráfica traz:

1. buscar padrões e relações;
2. confirmar (ou não) certas expectativas sobre os dados;
3. descobrir novos fenômenos;
4. confirmar (ou não) suposições feitas sobre os procedimentos estatísticos usados;
5. apresentar resultados de modo mais rápido e fácil.

Um dos gráficos mais utilizados para representar variáveis qualitativas

é o **gráfico de barras**. Os grupos ou as categorias que servem para classificar uma determinada variável são apresentados no eixo horizontal, e cada um deles possui uma barra referente à quantidade de indivíduos que se encaixa naquela categoria. Observe as diferentes variações de um mesmo gráfico de barras apresentadas em Vieira (2019), com base nos dados do censo demográfico de 2010.



Sexo	População recenseada
Brasil	190.755.799
Norte	15.864.454
Nordeste	53.081.950
Sudeste	80.364.410
Sul	27.386.891
Centro-Oeste	14.058.094

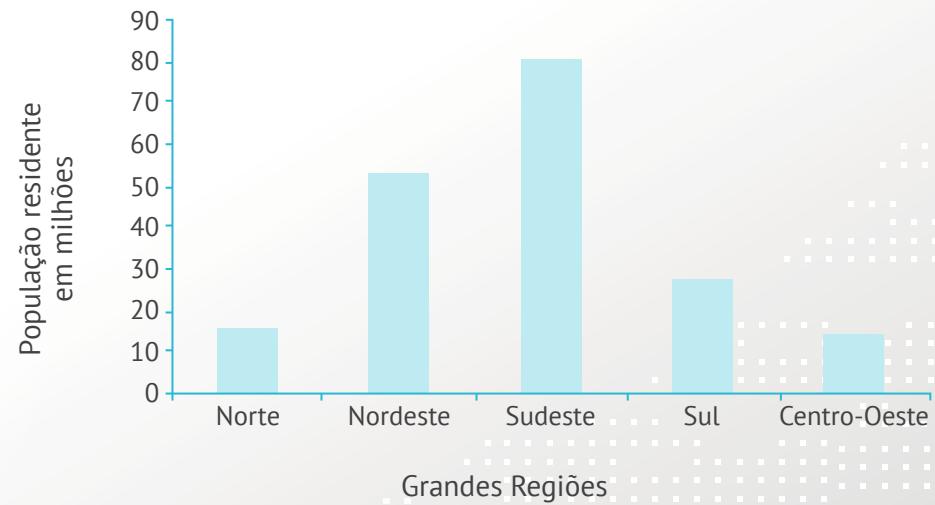


Figura 1: Exemplo de gráfico de barras. Fonte: Adaptado de Vieira (2019).

Neste primeiro gráfico, temos um gráfico de barras bidimensional, no qual o eixo horizontal apresenta as diferentes regiões do Brasil, e as barras são referentes à população total de cada região.

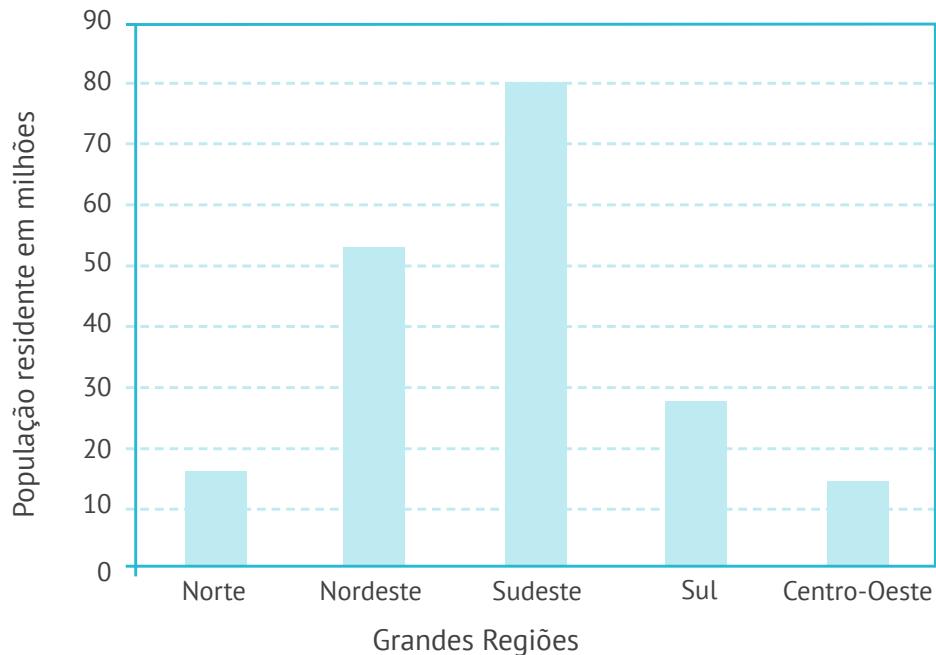


Figura 2: Gráfico de barras com linhas horizontais. Fonte: Adaptado de Vieira (2019).

Às vezes, é conveniente desenhar as linhas horizontais da escala vertical para que fique mais fácil a leitura e a compreensão dos dados apresentados no gráfico de barras.

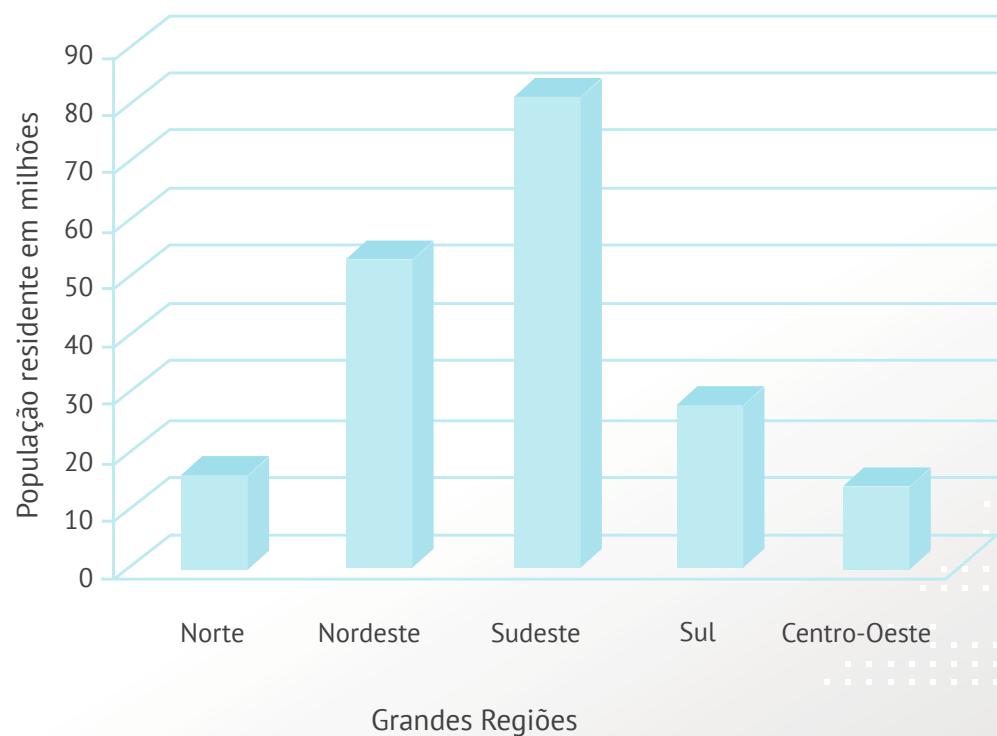


Figura 3: Gráfico de barras tridimensional. Fonte: Adaptado de Vieira (2019).

Uma variação do gráfico de barras é o gráfico de barras tridimensional, em que cada categoria apresentada no eixo horizontal apresenta um prisma de base idêntica, mas de altura variável de acordo com os dados de cada categoria.

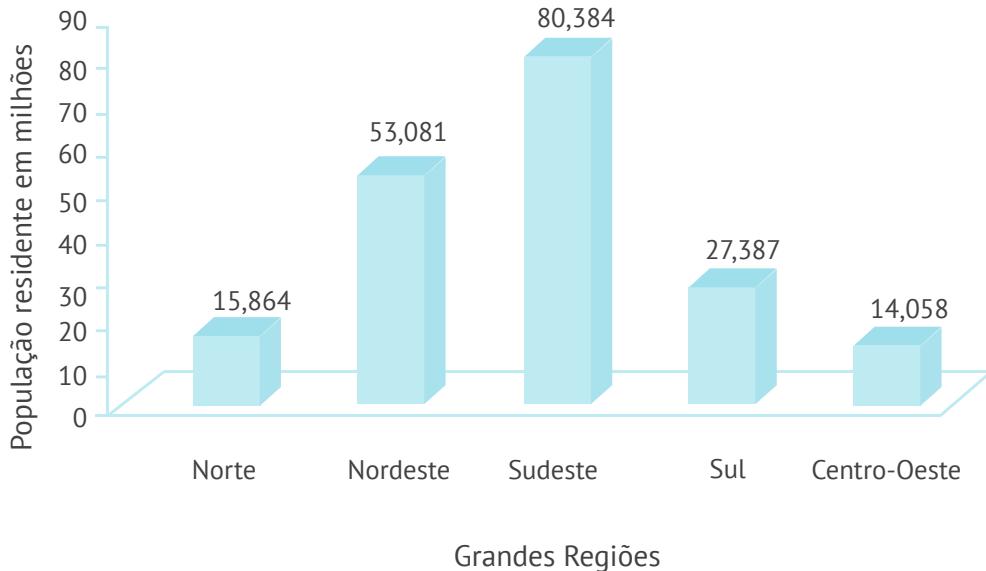
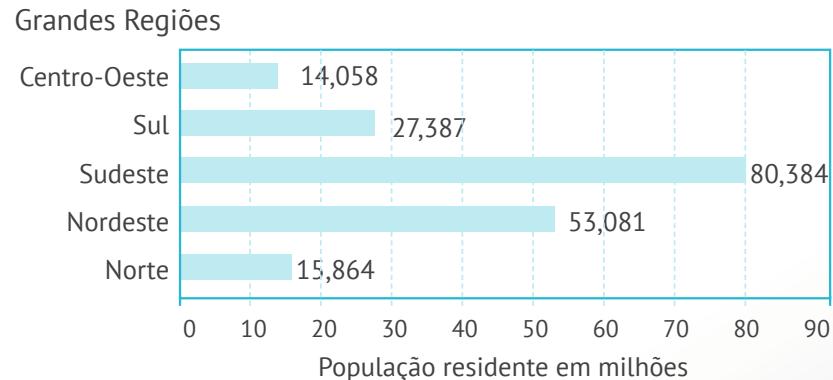


Figura 4: Gráfico de barras tridimensional com dados no topo do prisma. Fonte: Adaptado de Vieira (2019).

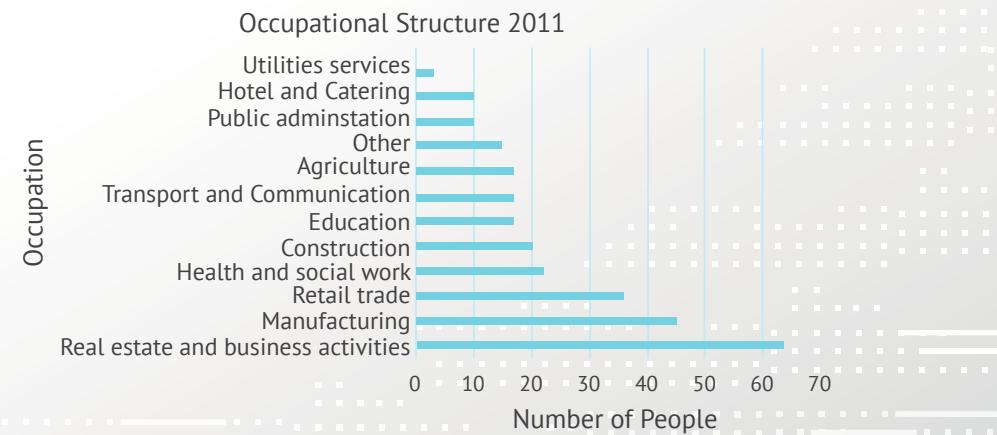
É comum encontrar também gráficos de barras que apresentam os dados exatos no topo de cada prisma para facilitar a comparação e a interpretação das diferentes categorias.

! Importante

Os gráficos de barras podem apresentar uma variação na qual as categorias são apresentadas no eixo vertical e as barras são horizontais.



Vieira (2019) destaca que essa variação do gráfico de barras pode ser utilizada quando os nomes das categorias são muito extensos. Outra possibilidade para o uso desse tipo de gráfico é quando existem muitas categorias distintas e deseja-se representar todas elas em um mesmo gráfico.



O **gráfico de setores** também é muito utilizado para variáveis categóricas em que se deseja fazer uma comparação da parte com o todo. Neste tipo de gráfico, as categorias são expressas em diferentes porções de um todo circular, destacando a parcela de cada categoria que formam a totalidade dos dados levantados. Observe a tabela e as diferentes variações de um mesmo gráfico de setores apresentados em Vieira (2019) sobre o estado civil da população brasileira acima de 10 anos retirada do censo demográfico de 2000.

Estado civil	Frequência	Percentual
Casado	50.703.610	37,0%
Desquitado, separado	2.661.741	1,9%
Divorciado	2.319.575	1,7%
Viúvo	6.231.273	4,6%
Solteiro	74.994.159	54,8%
Total	136.910.358	100,0%

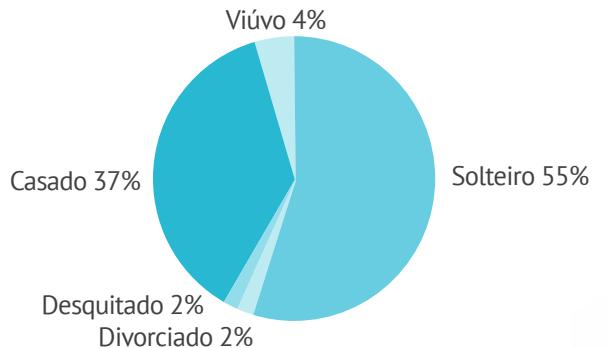


Figura 5: Gráfico de setores. Fonte: Adaptado de Vieira (2019).

Na figura, vemos representadas as proporções de cada uma das categorias que são possíveis para o estado civil da população brasileira. Neste caso, temos um gráfico de setores bidimensional, em que cada proporção é calculada de acordo com a proporcionalidade dos indivíduos em cada categoria.

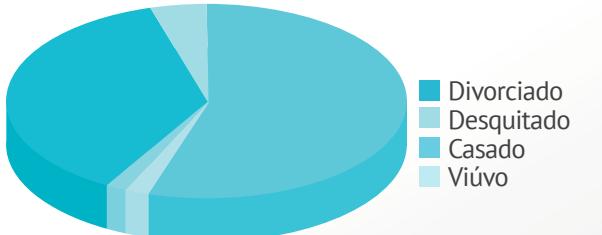
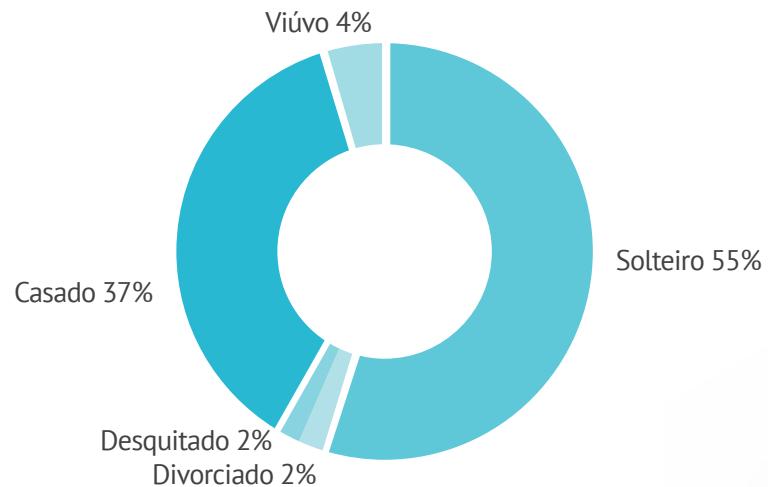
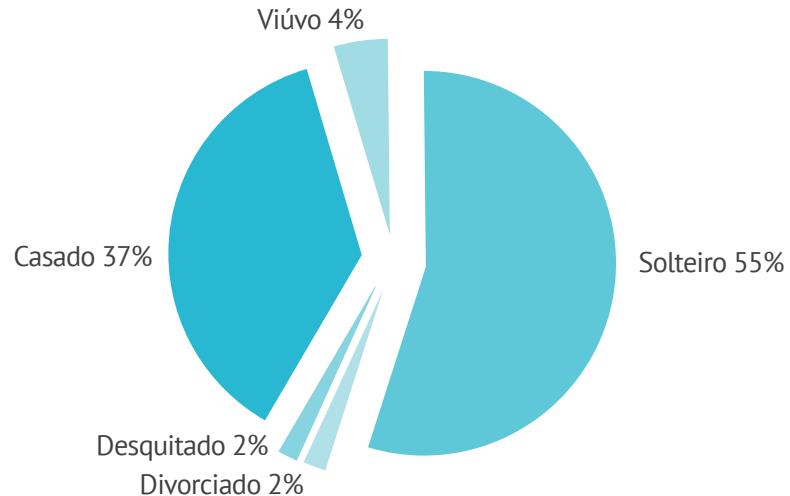


Figura 6: Gráfico de setores tridimensional. Fonte: Adaptado de Vieira (2019).

Assim como no gráfico de barras, também temos a variação tridimensional do gráfico de setores apresentada na figura.

! Importante

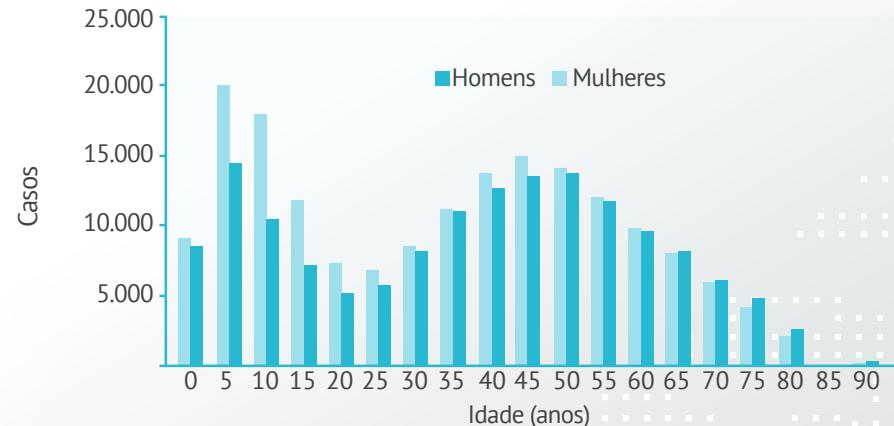
Visando a facilitar e a proporcionar maior independência a todos os leitores, os gráficos de setores podem apresentar variações com cores ou até mesmo representar as diferentes parcelas ligeiramente separadas.



Isso possibilita, por exemplo, que pessoas com baixa visão ou daltonismo consigam interpretar os gráficos de setores. Tratam-se de variações simples e que não comprometem as informações, mas que podem fazer a diferença para diversas pessoas.

Os **histogramas** são bem parecidos com os gráficos de barras, mas diferem quanto à natureza da variável que está sendo representada. Nesse caso, serão variáveis quantitativas. Observe o seguinte exemplo de histograma apresentado em Moore, Notz e Flingner (2017) sobre os casos da doença de Lyme, de acordo com a idade e o sexo.

Casos confirmados da doença de Lyme por idade e sexo - Estados Unidos, 2001-2010



Casos relatados da doença de Lyme são mais comuns entre meninos de idade 5-9 anos.

Figura 7: Histograma - Casos da doença de Lyme de acordo com a idade e o sexo.

Fonte: Adaptado de MOORE, NOTZ e FLINGNER (2017).

Repare que, neste caso, temos, além da divisão por idade, a divisão por sexo em cada um dos intervalos de idade. Essa é uma forma de representar as informações quando deseja-se destacar a diferença entre uma mesma categoria. Note que o número de casos entre os meninos com menos de 15 anos é significativamente maior do que os casos da doença de Lyme entre meninas na mesma faixa etária.

O último tipo de gráfico que apresentamos aqui são os **gráficos de linha**, que também trabalham com uma variável quantitativa, mas, geralmente, no caso específico, tal variável é o tempo. No exemplo a seguir, temos representada a variação da população da cidade de Elmsted na Inglaterra entre os anos de 1881 e 2011.

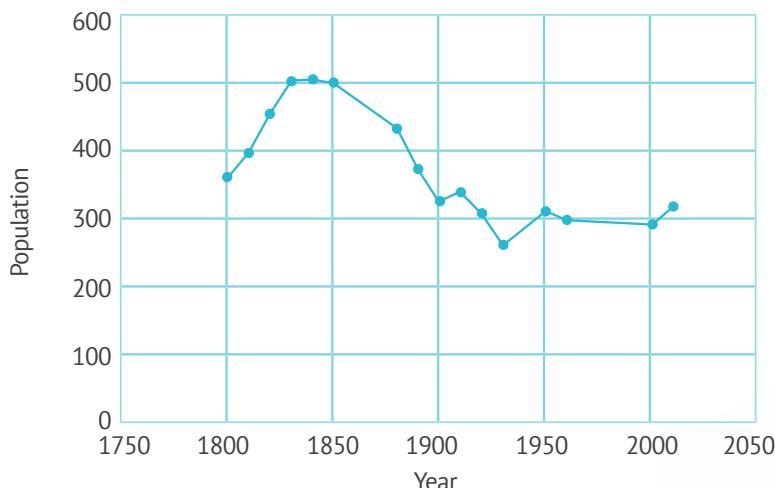


Figura 8: Gráfico de linha. Fonte: Wikimedia.

Mais de uma linha pode ser representada no mesmo gráfico para que haja uma comparação entre duas variáveis que podem (ou não) estar relacionadas, como mostra o exemplo a seguir, representando as taxas de natalidade e mortalidade da Suécia entre 1735 e 2016.

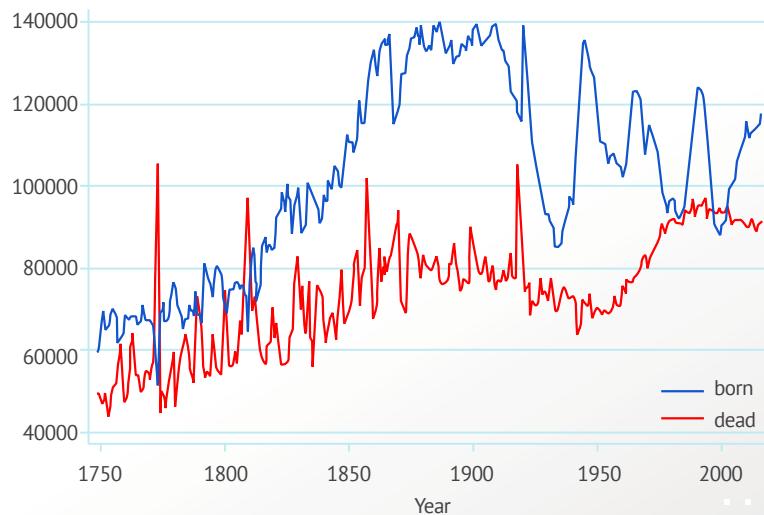


Figura 9: Gráfico de linha. Fonte: Wikimedia.

2. Tipos de tabelas

Outro tipo de representação muito utilizada para a disponibilização de dados são as **tabelas**. De acordo com o IBGE de 1993 (VIEIRA, 2019):

“Tabela é uma forma não discursiva de apresentar informações, das quais o dado numérico se destaca como informação central”. Vamos analisar detalhadamente cada elemento que compõe uma tabela e entender como podemos construí-las para transmitir as informações de maneira mais adequada.

As tabelas sempre apresentam suas informações dispostas em **linhas** e **colunas**. Geralmente, na primeira linha da tabela, temos um **cabeçalho**, em que são encontrados os rótulos de cada coluna, enquanto que na primeira coluna estão dispostos os **indicadores de linha** que classificam a quem pertence os dados dispostos em cada coluna daquela respectiva linha. Uma **célula** é o espaço da tabela determinado por uma linha e uma coluna.

Município	População (2000)	População (2010)	Variação (+)
Marataízes	30.603	34.140	11,55%
Castelo	32.756	34.747	6,08%
Presidente Kennedy	9.555	10.314	7,94%
Cachoeiro	174.879	189.889	8,58%
Itapemirim	28.121	30.988	10,19%
Vitória	292.304	327.801	12,14%
Campo dos Goytacazes - RJ	406.989	463.731	13,94%
Guarapari	88.400	105.286	19,10%
Vila Velha	345.965	414.586	19,83%

Piúma	14.987	18.123	20,92%
Anchieta	19.176	23.902	24,64%
Serra	321.181	409.267	27,42%
Macaé - RJ	132.461	206.728	56,07%
Espírito Santo	3.097.232	3.514.952	13,49%
Rio de Janeiro (estado)	14.391.282	15.989.929	11,11%
Brasil	169.799.170	190.755.799	12,34%

Figura 10: Exemplo de tabela. Fonte: Wikimedia.

Na tabela acima, temos, na primeira linha, a descrição de qual informação será apresentada em cada coluna. Na primeira célula de cada linha, temos um identificador referente ao município, estado ou país dos quais as informações serão dispostas.

! Importante

Outra informação fundamental para a inserção de tabelas e gráficos em qualquer trabalho acadêmico, artigo científico ou qualquer outro tipo de publicação é a **fonte**. A relevância dos dados apresentados depende da forma como tais dados foram coletados e tratados, e isso só pode ser verificado caso a fonte seja devidamente apresentada. Caso o levantamento de dados e a elaboração das tabelas ou gráficos sejam feitos pelo próprio autor ou autores, então não é necessário inserir a fonte após cada um desses recursos.

Os dados de uma tabela trazem de maneira objetiva os resultados de uma pesquisa ou estudo e, portanto, são fundamentais para a credibilidade de um trabalho. Pode-se dizer que as informações contidas em uma tabela são mais imparciais do que uma representação gráfica, isso porque em um gráfico podemos selecionar informações e apresentá-las de acordo com o viés da publicação. É importante ter essa visão crítica sobre as intenções por trás de cada informação retratada em um gráfico ou tabela.

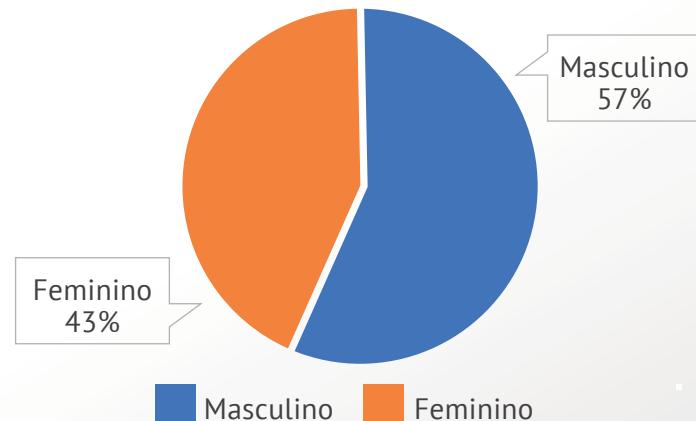
3. Distribuição de frequências

Uma distribuição de frequências faz referência ao número de vezes que uma determinada categoria ou grupo foi verificada dentre os indivíduos da amostra. Por exemplo, uma das divisões mais comuns de ocorrer em estudos com seres humanos é a divisão de frequência entre os sexos dos entrevistados que participaram do estudo. Dessa forma, a distribuição de frequências é um modo de apresentar as categorias dos dados avaliados com suas respectivas frequências.

Podemos subdividir categorias de acordo com o tipo de dado que ela faz referência. A variável sexo, por exemplo, é um **dado nominal** e pode ser apresentado tanto em forma de tabela como em forma de gráfico. Veja os exemplos fictícios abaixo.

Sexo	Frequência
Masculino	17
Feminino	13
Total	30

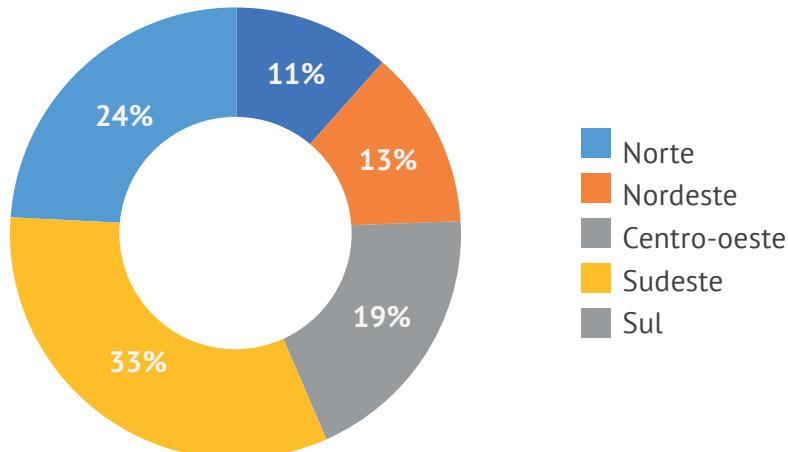
DIVISÃO DOS ENTREVISTADOS POR SEXO



Região	Número de entrevistados	Percentual
Norte	125	11%
Nordeste	146	13%
Centro-oeste	214	19%

Sudeste	362	33%
Sul	266	24%
Total	1113	100%

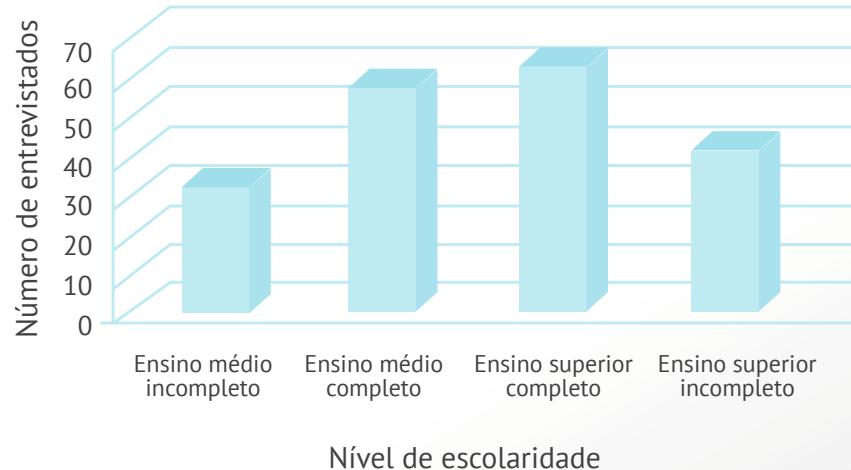
ENTREVISTADOS DIVIDIDOS POR REGIÕES



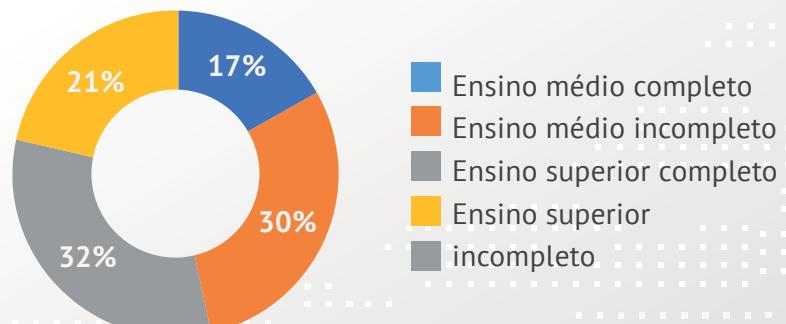
Podemos usar a distribuição de frequências para apresentar **dados ordinais**, em que podemos definir uma ordem crescente dentre os valores possíveis.

A seguir, apresentamos uma distribuição de frequências fictícias, em que as categorias fazem referência ao nível de escolaridade dos entrevistados.

ENTREVISTADOS DIVIDIDOS POR NÍVEL DE ESCOLARIDADE



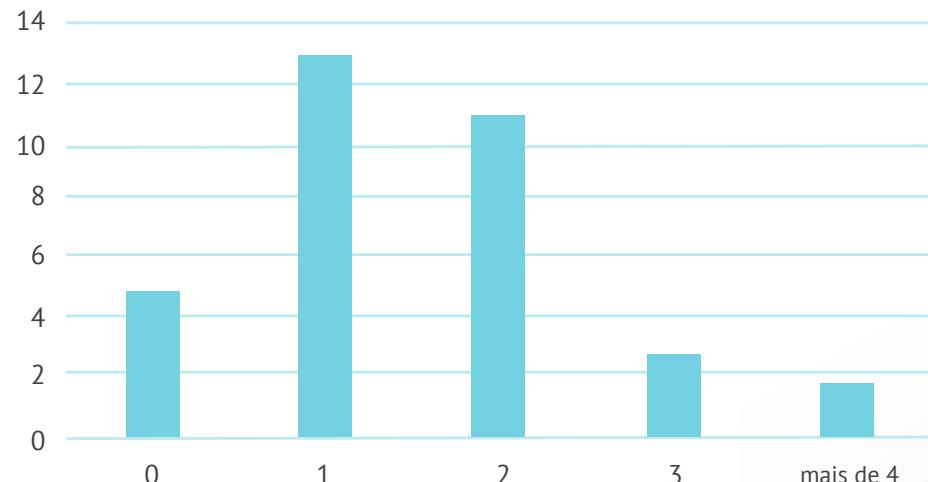
NÍVEL DE ESCOLARIDADE DOS ENTREVISTADOS



Outro tipo de dado que pode ser descrito por uma distribuição de frequências são os **dados discretos**, ou seja, variáveis numéricas que assumem apenas valores inteiros como número de filhos.

Número de filhos	Frequência
0	5
1	13
2	11
3	3
Mais de 4	2

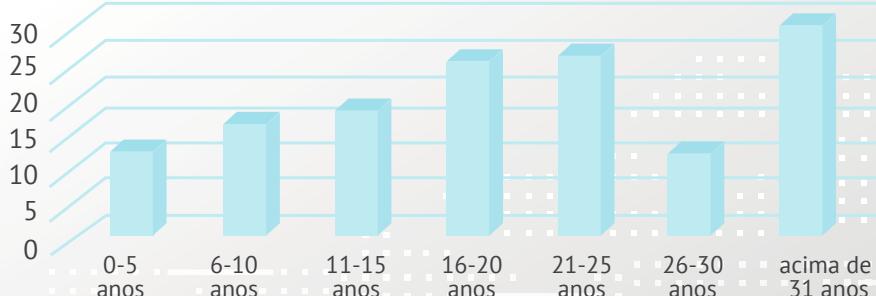
NÚMERO DE FILHO POR CASAL ENTREVISTADO



De maneira geral, sempre podemos representar as diferentes variáveis do nosso estudo por meio de distribuição de frequências, mesmo quando trabalhamos com variáveis quantitativas contínuas como a idade dos participantes. Neste caso, devemos estabelecer **intervalos de classe** para a representação das frequências, como mostra o exemplo a seguir.

Idade	Frequência
0 a 5 anos	12
6 a 10 anos	16
11 a 15 anos	18
16 a 20 anos	25
21 a 25 anos	26
26 a 30 anos	12
Acima de 31 anos	30

IDADE DOS ENTREVISTADOS



Os intervalos de classe são definidos de modo a apresentar de maneira agrupada os dados de um determinado estudo estatístico. Esse tipo de representação é geralmente utilizada para descrever dados numéricos contínuos ou variáveis que possuam uma grande variabilidade de valores. No exemplo anterior, temos intervalos para representar a frequência das idades dentro da amostra. Uma notação para os intervalos de classe é a seguinte: $a \text{ } \text{---} \text{ } b$ onde a é o limite inferior da classe e b o limite superior. Neste caso, o valor a pertence a esse intervalo de classe, mas o valor b não pertence, isso é, somente os valores estritamente menores que b pertence a esse intervalo. Utilizando essa notação, a tabela apresentada anteriormente ficaria:

Idade (anos)	Frequência
0 --- 6	12
6 --- 11	16
11 --- 16	18
16 --- 21	25
21 --- 26	26
26 --- 31	12
Acima de 31 anos	30

Os intervalos de classe devem ser determinados de modo a otimizar a leitura dos dados e, geralmente, os intervalos de classe têm sempre o mesmo comprimento ou amplitude; isso é, o espectro que cada intervalo cobre é geralmente de mesmo tamanho. Para organizar o procedimento de determinar intervalos de classe, os seguintes passos podem ser tomados:

1. Organizar todos os dados em ordem crescente.
2. Observar o maior e o menor valor dentro das observáveis e calcular a diferença entre eles.
3. Determinar o número de classes que serão utilizadas na representação tabular ou gráfica.
4. Dividir a variação encontrada no item 2 pelo número de classes e assim obter o comprimento de cada um dos intervalos de classe.
5. Determinar os limites inferior e superior de cada classe e, finalmente, agrupar os dados em suas respectivas classes e, por consequência, determinar a frequência de cada uma das classes.



Note que o número de intervalos ou classes que serão construídas pode variar de acordo com a necessidade da representação. Esse tipo de escolha também pode influenciar a leitura dos dados e o viés que se deseja imprimir em uma representação tabular ou gráfica.

Neste capítulo, trabalhamos com as representações de dados por meio de gráficos e tabelas. Esse tipo de representação pode sistematizar uma pesquisa ou, ainda, indicar uma relação entre variáveis de maneira visualmente mais atraente e intuitiva. É importante destacar e compreender a intenção dos gráficos presentes em cada publicação e os recursos utilizados para transmitir certas informações. Podemos adaptar tais representações para dar destaque a determinadas relações e comparações, dependendo do viés que a nossa publicação tem. Analisamos também os diferentes tipos de distribuição de frequência para diversos tipos de variáveis e como as representações gráficas podem auxiliar a transmissão de informações concisas. Apresentamos ainda diversos exemplos reais e fictícios para ilustrar cada uma das definições trabalhadas.



Referências

FONSECA, J. S. da; MARTINS, G. de A. **Curso de estatística**. 6. ed. São Paulo: Atlas, 2012.

MOORE, D.; NOTZ, W.; FLIGNER, M. **A estatística básica e sua prática**. 7. ed. Rio de Janeiro: LTC, 2017.

MORETTIN, P.; BUSSAB, W. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.

SPIEGEL, M.; SCHILLIER, J.; SRINIVASAN, A. **Probabilidade e estatística**. 3. ed. Porto Alegre: Bookman, 2013.

TRIOLA, M. **Introdução à estatística**. 12. ed. Rio de Janeiro: LTC, 2017.

VIEIRA, S. **Fundamentos de Estatística**. 6. ed. São Paulo: Atlas, 2019.