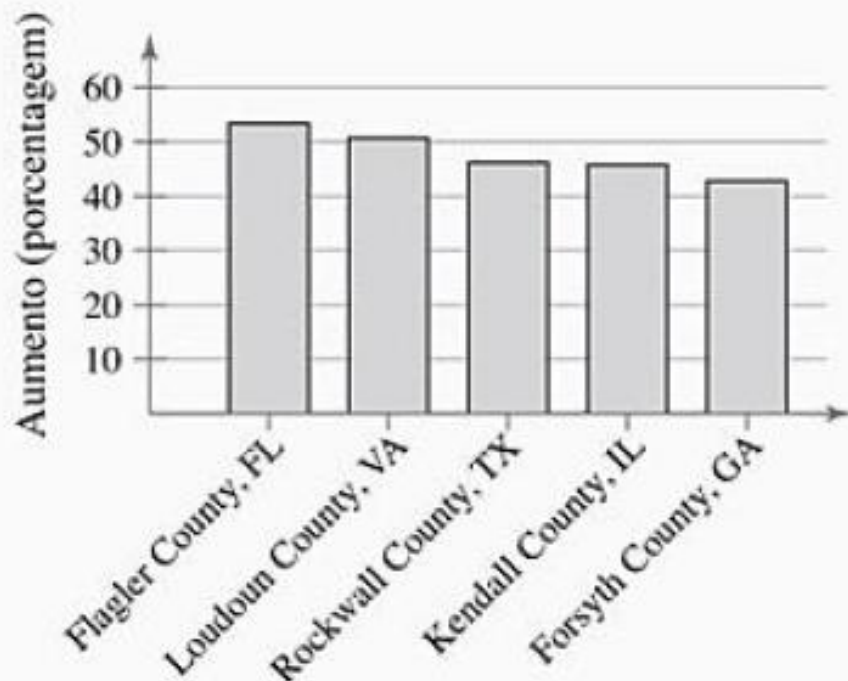
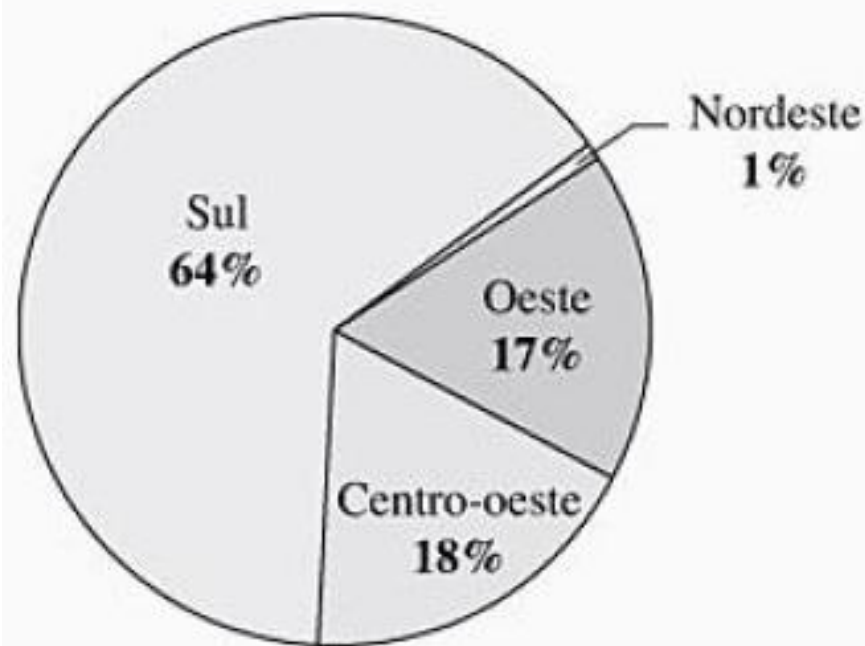


# Introdução à estatística

Regiões americanas com crescimento mais rápido



Localização das 100 regiões de crescimento mais rápido



o rápido crescimento de regiões dos Estados Unidos com 10.000 ou mais habitantes, de 2000 a 2005, e onde estão as regiões com os 100 crescimentos mais rápidos, de 2000 a 2005.

## Definição de estatística

---

Conforme começamos o curso, você pode se perguntar: *o que é estatística? Por que eu devo estudar estatística? Como o estudo da estatística pode me ajudar profissionalmente?* Quase todos os dias somos expostos à estatística. Por exemplo, considere os trechos a seguir, retirados de jornais e publicações recentes.

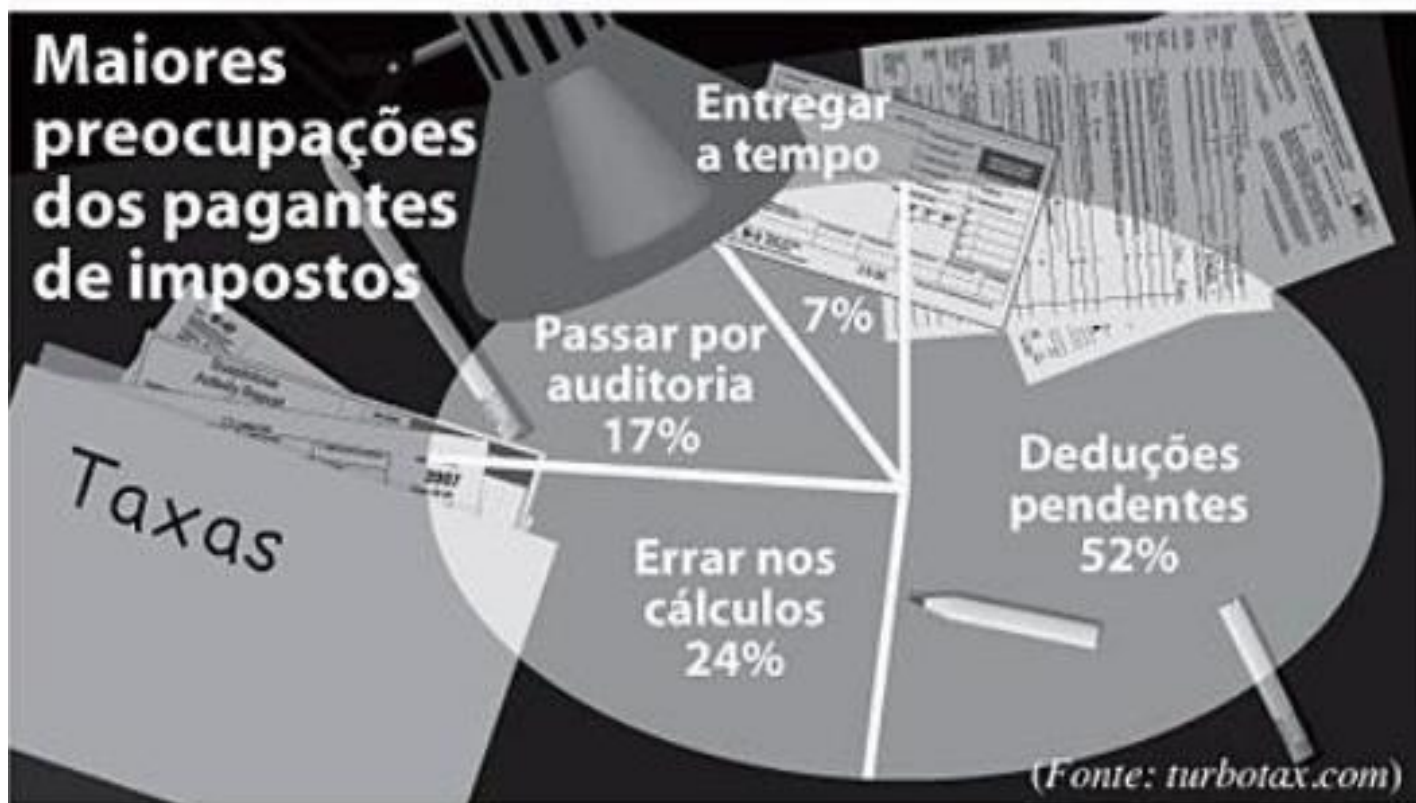
- “As pessoas que comem três porções diárias de grãos integrais têm risco de sofrer problemas cardíacos reduzido em 37%.” (Fonte: Whole Grains Council.)
- “Setenta por cento dos 1.500 danos à espinha dorsal em menores de idade resultam de acidentes de carro e 68 dos feridos não estavam usando o cinto de segurança.” (Fonte: UPI.)

afirmações que você acabou de ler são baseadas na coleta de **dados**.

## Definição

Dados consistem em informações que vêm de observações, contagens, medições ou respostas.

Às vezes, os dados são apresentados graficamente.



## Definição

---

Estatística é a ciência que coleta, organiza, analisa e interpreta dados para a tomada de decisões.

---

## Conjuntos de dados

---

Há dois tipos de conjuntos de dados usados em estatística. Esses conjuntos são chamados de população e amostra.

## Definição

---

Uma população é uma coleção de todos os resultados, respostas, medições ou contagens que são de interesse.

Uma amostra é um subgrupo de uma população.

---

Dados amostrais podem ser usados para formar conclusões sobre populações. Os dados amostrais devem ser coletados usando o método apropriado, tal como a seleção aleatória. Se os dados não forem coletados usando-se o método apropriado, eles não terão valor.

# Exemplo

1

## Identificando conjunto de dados

Em uma pesquisa recente, foi perguntado a 1.708 adultos americanos se eles consideram o aquecimento global um problema que exige uma ação imediata do governo. Novecentos e trinta e nove deles responderam que sim. Identifique a população e a amostra. Descreva o conjunto de dados. *(Adaptado de: Pew Research Center.)*

### Solução

A população consiste das respostas de todos os adultos nos Estados Unidos, e a amostra consiste das respostas de 1.708 adultos nos Estados Unidos na pesquisa. A amostra é um subgrupo das respostas de todos os adultos nos Estados Unidos. O conjunto de dados consiste de 939 sim e 769 não.

Respostas de todos os adultos nos Estados Unidos (população)

Respostas dos adultos na pesquisa (amostra)

**Tente  
você**

**1**

O departamento de energia dos Estados Unidos conduz pesquisas semanais em aproximadamente 800 postos de gasolina para determinar o preço médio por galão de gasolina comum. Em 12 de fevereiro de 2007, o preço médio era \$ 2,24 por galão. Identifique a população e a amostra. (Fonte: *Energy Information Administration*.)

■ Identifique a *população*.

a. A população consiste dos preços por galão de gasolina comum em todos os postos de gasolina dos Estados Unidos.

■ Identifique a *amostra*.

b. A amostra consiste dos preços por galão de gasolina comum em 800 postos pesquisados.

■ Do que consiste o conjunto de dados?

c. O conjunto de dados consiste em 800 preços.

Se o conjunto de dados é uma população ou uma amostra normalmente depende do contexto da vida real. No caso do Exemplo 1, a população era o conjunto de respostas de todos os adultos dos Estados Unidos. Dependendo da proposta da pesquisa, a população poderia ter sido o grupo de respostas de todos os adultos que moram na Califórnia ou daqueles que têm telefones celulares ou que leem determinado jornal.

## Definição

---

Um **parâmetro** é a descrição numérica de uma característica *populacional*.

Uma **estatística** é a descrição numérica de uma característica *amostral*.

---

### Dica de estudo

Os termos *parâmetro* e *estatística* são fáceis de lembrar se você usar o recurso mnemônico de se relacionar as **primeiras** letras em *população parâmetro* e as **últimas** letras em *amostra e estatística*.



## Exemplo 2

### Distinguindo entre parâmetro e estatística

Decida se o valor numérico descreve um parâmetro populacional ou uma estatística amostral. Explique seu raciocínio.

1. Uma pesquisa recente de uma amostra de MBAs reportou que o salário médio para um MBA é mais do que \$ 82.000. (*Fonte: The Wall Street Journal.*)

Em razão de a média de \$ 82.000 ser baseada em um subgrupo de uma população, ela é uma estatística amostral.

2. Os salários iniciais para 667 MBAs graduados na Escola de Negócios da Universidade de Chicago aumentaram 8,5% em comparação ao ano anterior.

Devido ao fato de o aumento porcentual de 8,5% ser baseado nos salários iniciais de todos os 667 graduandos, ele é um parâmetro populacional.

**Tente  
você  
2**

Em 2006, a liga dos times de beisebol gastou um total de \$ 2.326.706.685 nos salários dos jogadores. Esse valor numérico descreve um parâmetro populacional ou uma estatística amostral? (*Fonte: USA Today.*)

- a. Decida se o valor numérico é de uma *população* ou uma *amostra*.
- b. Especifique se o valor numérico é um *parâmetro* ou uma *estatística*.

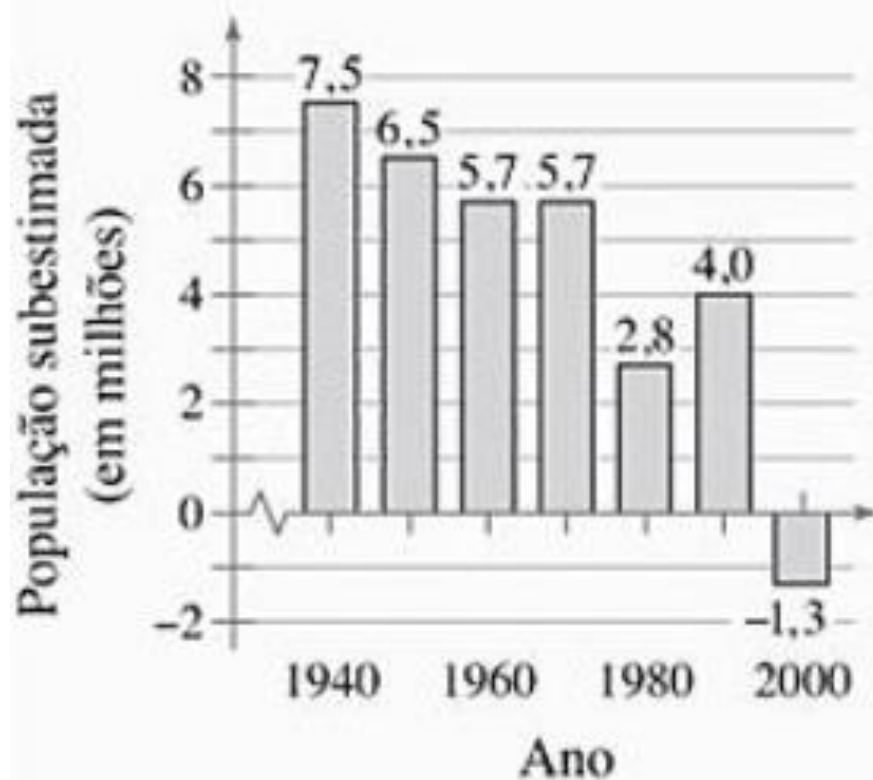
a. População.

b. Parâmetro.

## Retratando o mundo

Quão preciso é o censo americano? De acordo com uma avaliação pós-censo conduzida pelo *Bureau*, o censo de 1990 contou de forma equivocada a população dos Estados Unidos: aproximadamente 4 milhões de pessoas a menos do que, de fato, havia no país. O censo de 1990 foi o primeiro desde 1940 a ser menos preciso do que seu anterior. Note que a contagem errada para o censo de 2000 foi de -1,3 milhões de pessoas. Isso significa que o censo de 2000 contou a população dos EUA para mais 1,3 milhões.

## Contagem do censo dos EUA para menos



*Quais são algumas das dificuldades ao se coletar dados de uma população?*

## Ramos da estatística

---

O estudo de estatística tem duas ramificações consideráveis: **estatística descritiva** e **estatística inferencial**.

## Definição

---

**Estatística descritiva** é o ramo da estatística que envolve a organização, o resumo e a representação dos dados.

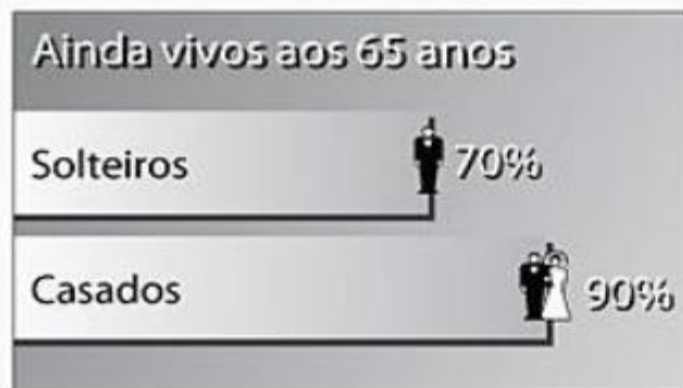
**Estatística inferencial** é o ramo da estatística que envolve o uso de uma amostra para chegar a conclusões sobre uma população. Uma ferramenta básica no estudo da estatística inferencial é a probabilidade.

# Exemplo 3

## Estatística descritiva e inferencial

Decida qual parte do estudo representa o ramo descritivo da estatística. Que conclusões podem ser tomadas do estudo usando estatística inferencial?

1. Uma grande amostra de homens, com 48 anos de idade, foi estudada durante 18 anos. Para os que são solteiros, 70% ainda estavam vivos aos 65 anos. Para os casados, 90% estavam vivos aos 65 anos. (Fonte: *The Journal of Family Issues*.)



## Solução

1. A estatística descritiva envolve afirmações tais como "Para os que são solteiros, 70% ainda estavam vivos aos 65 anos" e "Para os casados, 90% ainda estavam vivos aos 65 anos." Uma inferência possível tirada do estudo é que o fato de ser casado está associado a uma vida mais longa.

2. Em uma amostra dos analistas de Wall Street, a porcentagem dos que previram incorretamente os lucros de empresas de alta tecnologia em um ano recente foi de 44%. (Fonte: Bloomberg News.)

2. A parte do estudo que representa o ramo descritivo da estatística envolve afirmações como “a porcentagem dos que previram incorretamente os lucros de empresas de alta tecnologia em um ano recente era de 44%.” Uma inferência possível com base no estudo é que o mercado de ações é difícil de ser previsto, até mesmo para os profissionais.

**Tente  
você**  
**3**

Uma pesquisa conduzida entre 1.017 homens e mulheres pela Corporação Internacional de Pesquisa de Opinião descobriu que 76% das mulheres e 60% dos homens haviam passado por exames físicos no ano anterior. (*Fonte: Men's Health.*)

- a. Identifique o aspecto descritivo da pesquisa.
- b. Quais inferências podem ser retiradas com base nessa pesquisa?
  - a. Estatísticas descritivas envolvem a afirmação "76% das mulheres e 60% dos homens fizeram exames físicos durante o ano anterior."
  - b. Uma inferência a partir do estudo é que uma porcentagem maior de mulheres passou por exames físicos no ano anterior.

Durante esse curso, você verá aplicações em ambas as ramificações. Um tema principal desse curso será como usar os dados estatísticos amostrais para fazer inferências sobre parâmetros populacionais desconhecidos.



# Classificação dos dados

---

## O que você deve aprender

- Como distinguir entre dados qualitativos e quantitativos.
- Como classificar dados em relação aos 4 níveis de mensuração: nominal, ordinal, intervalar ou racional.

## Definição

---

**Dados qualitativos** consistem de atributos, rótulos ou entradas não numéricas.

**Dados quantitativos** consistem de medidas numéricas ou contagens.

# Exemplo 1

## Classificando dados por tipo

Os preços-base para diversos veículos são apresentados na tabela a seguir. Quais dados são qualitativos e quais são quantitativos? Explique seu raciocínio. (Fonte: Ford Motor Company.)

Modelo	Preço-base
Fusion 14 S	\$ 17.795
F-150 XL	\$ 18.710
Five Hundred SEL	\$ 23.785
Escape XLT Sport	\$ 24.575
2007 Explorer Sport Trac Limited	\$ 26.775
Freestar SEL	\$ 27.500
Crown Victoria LX	\$ 28.830
Expedition XLT	\$ 35.480

## Solução

A informação mostrada na tabela pode ser separada em dois conjuntos de dados. Um grupo contém os nomes dos modelos de veículos e o outro contém os preços-base para os modelos. Os nomes são entradas não numéricas, portanto, são dados qualitativos. Os preços são entradas numéricas, portanto, são dados quantitativos.

**Tente  
você  
1**

As populações de diversas cidades norte-americanas são apresentadas na tabela. Quais dados são qualitativos e quais são quantitativos? (Fonte: US Census Bureau.)

- Identifique os conteúdos de cada conjuntos de dados.
- Decida se cada conjunto de dados consiste em entradas numéricas ou não numéricas.
- Especifique os dados qualitativos e os quantitativos.

Cidade	População
Cleveland, OH	452.208
Detroit, MI	886.671
Houston, TX	2.016.582
Las Vegas, NV	545.147
Portland, OR	533.427
Topeka, KS	121.946

- Nomes das cidades e população das cidades.
- Nome da cidade: não numérico.  
População da cidade: numérico.
- Nome da cidade: qualitativo.  
População da cidade: quantitativo.

## Níveis de mensuração

---

Outra característica de dados é o nível de mensuração. O nível de mensuração determina quais cálculos estatísticos são significantes. Os quatro níveis de medida, em ordem do mais baixo para o mais alto, são nominal, ordinal, intervalar e racional.

### Definição

---

Dados no nível nominal de mensuração são apenas qualitativos. Dados neste nível podem ser categorizados usando-se nomes, rótulos ou qualidades. Não são realizados cálculos matemáticos neste nível.

Dados no nível ordinal de mensuração são qualitativos ou quantitativos. Dados neste nível podem ser organizados em ordem ou posição, mas as diferenças entre as entradas de dados não são significantes.

---

## Importante

Quando números estão no nível nominal de mensuração, eles simplesmente representam um rótulo. Exemplos de números usados como rótulos incluem o número do seguro social e os números nos uniformes esportivos. Por exemplo, não faria sentido somar os números dos uniformes dos Chicago Bears (time de futebol americano).

## Exemplo 2

### Classificando dados por nível

A seguir temos dois conjuntos de dados. Que conjunto de dados consiste em dados no nível nominal? Que conjunto de dados consiste em dados no nível ordinal? Explique. (Fonte: Nielsen Media Research.)

#### Os cinco programas de TV mais assistidos (de 12/02/2007 a 18/02/2007)

1. American Idol — terça-feira
2. American Idol — quarta-feira
3. Grey's anatomy
4. House
5. CSI

#### Afiliações das redes em Pittsburgh, PA

- |            |
|------------|
| WTAE (ABC) |
| WPXI (NBC) |
| KDKA (CBS) |
| WPGH (FOX) |

## *Solução*

O primeiro conjunto de dados lista a posição de cinco programas de TV. Os dados consistem das posições 1, 2, 3, 4 e 5. Em razão de as posições poderem ser listadas em ordem, esses dados estão no nível ordinal. Note que a diferença entre uma posição 1 e 5 não tem significado matemático. O segundo grupo consiste das siglas de cada afiliada de rede de Pittsburg. As siglas simplesmente nomeiam as afiliadas, então esses dados estão no nível nominal.

## Definição

Dados no nível de mensuração intervalar podem ser ordenados e você pode calcular diferenças significativas entre as entradas de dados. No nível intervalar, um registro nulo simplesmente representa uma posição em uma escala; a entrada não é um zero inerente.

Dados no nível de mensuração racional são similares aos dados no nível intervalar, com uma propriedade adicionada: neste nível, um registro nulo é um zero inerente. Uma razão de dois valores de dados pode ser formada de modo que um valor de dado possa ser significativamente expresso como o múltiplo de outro.

Um zero inerente é um zero que significa “nada”. Por exemplo, a quantia de dinheiro que você tem em sua poupança pode ser zero dólares. Neste caso, o zero representa nenhum dinheiro; é um zero inerente. Por outro lado, a temperatura de  $0^{\circ}\text{C}$  não representa uma condição no qual o aquecimento não está presente. A temperatura de  $0^{\circ}\text{C}$  simplesmente representa uma posição na escala Celsius; não é um zero inerente.

Para distinguir entre dados no nível intervalar e no nível racional, determine se a expressão “duas vezes mais” tem algum sentido no contexto dos dados. Por exemplo, \$ 2 é duas vezes mais que \$ 1, então estes dados estão no nível racional. Por outro lado,  $2^{\circ}\text{C}$  não é duas vezes mais quente que  $1^{\circ}\text{C}$ , então esses dados estão no nível intervalar.



# Exemplo 3

## Classificando dados por nível

Temos dois conjuntos de dados. Qual conjunto de dados está no nível intervalar? Qual conjunto de dados está no nível racional? Explique. (Fonte: Major League Baseball.)

### Vitórias do New York Yankees na série mundial (anos)

1923, 1927, 1928, 1932, 1936, 1937, 1938, 1939, 1941, 1943, 1947, 1949, 1950, 1951, 1952, 1953, 1956, 1958, 1961, 1962, 1977, 1978, 1996, 1998, 1999, 2000.

### Totais de *home runs*\* da Liga Americana em 2006 (por time)

Baltimore	164
Boston	192
Chicago	236
Cleveland	196
Detroit	203
Kansas City	124
Los Angeles	159
Minnesota	143
Nova Iorque	210
Oakland	175
Seattle	172

## *Solução*

Ambos os conjuntos de dados contêm dados quantitativos. Considere as datas das vitórias dos Yankees na série mundial. Faz sentido encontrar diferenças entre dados específicos. Por exemplo, o tempo entre a primeira e a última vitória dos Yankees é

$$2000 - 1923 = 77 \text{ anos.}$$

Mas não faz sentido dizer que um ano é múltiplo do outro. Então, esses dados estão no nível intervalar. Usando o total de *home runs*\*, podemos encontrar diferenças e escrever razões. Com base nos dados, podemos ver que o Detroit atingiu 31 *home runs* a mais do que o Seattle, e que o Chicago atingiu duas vezes mais do que o Kansas City. Então, esses dados estão no nível racional.

As tabelas a seguir resumem quais operações são significativas em cada um dos quatro níveis de mensuração. Quando identificar o nível de mensuração do conjunto de dados, use o nível mais alto que for adequado.

<b>Nível de mensuração</b>	<b>Categorizar os dados</b>	<b>Ordenar os dados</b>	<b>Subtrair os valores dos dados</b>	<b>Determinar se um valor de dado é múltiplo de outro</b>
Nominal	Sim	Não	Não	Não
Ordinal	Sim	Sim	Não	Não
Intervalar	Sim	Sim	Sim	Não
Racional	Sim	Sim	Sim	Sim