

POLI-PMI  
ENGENHARIA DE PETRÓLEO E GÁS  
ESTATÍSTICA

Aulas 2, 3 e 4 – Estatística Descritiva

# Estatística Descritiva

- Como organizar e descrever conjunto de dados
  - Observar
    - Centro
    - Variabilidade (ou amplitude)
    - Forma
  - Agrupar os dados em intervalos (classes) e formar distribuição de frequência

# Objetivos

- Como construir uma distribuição de frequência
  - ▣ Limites
  - ▣ Pontos médios
  - ▣ Frequências relativas
  - ▣ Frequências acumuladas
- Como construir
  - ▣ Histogramas de frequências
  - ▣ Polígonos de frequências
  - ▣ Histogramas de frequências relativas
  - ▣ Ogivas

# Distribuição de frequência

## Distribuição de frequência

- Tabela que mostra **classes** ou **intervalos** de dados com uma contagem do número de entradas em cada classe
- A **frequência,  $f$** , de uma classe é o número de entradas de dados na classe

Classe	Frequência, $f$
1 – 5	5
6 – 10	8
11 – 15	6
16 – 20	8
21 – 25	5
26 – 30	4

Tamanho da classe  $6 - 1 = 5$

Limite inferior da classe

Limite superior da classe

# Construir Distribuição de Frequência

1. Definir o número de classes (entre 5 e 20)
2. Encontrar a largura da classe
  - ▣ Determine a amplitude dos dados
  - ▣ Dividir a amplitude pelo número de classes (aproximar para cima)
3. Encontrar os limites inferiores e superiores das classes
4. Registrar os dados nas respectivas classes
5. Calcular a frequência para cada classe

# Construir distribuição de frequência

A amostra seguinte lista o número de minutos que 50 usuários da internet passaram conectados durante a sessão mais recente. Construa uma distribuição de frequência para as sete classes.

50 40 41 17 11 7 22 44 28 21 19 23 37 51  
54 42 86 41 78 56 72 56 17 7 69 30 80 56  
29 33 46 31 39 20 18 29 34 59 73 77 36 39  
30 62 54 67 39 31 53 44

# Construir Distribuição de Frequência

50 40 41 17 11 7 22 44 28 21 19 23 37  
51 54 42 **86** 41 78 56 72 56 17 **7** 69 30  
80 56 29 33 46 31 39 20 18 29 34 59 73  
77 36 39 30 62 54 67 39 31 53 44

1. Número de classes = 7 (dados)
2. Encontrar o tamanho da classe

$$\frac{\text{max} - \text{min}}{\text{\#classes}} = \frac{86 - 7}{7} \approx 11.29$$

Arredondando  
para  
cima: 12

3. Usar 7 (valor mínimo) como o primeiro limite mínimo. Adicionar 12 (tamanho da classe) para definir o limite mínimo da próxima classe

$$7 + 12 = 19$$

Encontre os limites mínimos restantes

Tamanho da classe = 12

Limite mínimo	Limite máximo
7	
19	
31	
43	
55	
67	
79	

O limite máximo da primeira classe é 18 (um a menos que o limite mínimo da segunda classe).

Some o tamanho da classe, 12, para definir o limite máximo da próxima classe.

$$18 + 12 = 30$$

Encontre os limites máximos restantes.

Limite mínimo	Limite máximo
7	18
19	30
31	42
43	54
55	66
67	78
79	90

Tamanho da classe = 12

4. Faça um registro para cada entrada de dado na fileira da classe apropriada.
5. Conte os registros para encontrar a frequência total  $f$  para cada classe.

Classe	Registro	Frequência, $f$
7 – 18	IIII I	6
19 – 30	IIII IIII	10
31 – 42	IIII IIII III	13
43 – 54	IIII III	8
55 – 66	IIII	5
67 – 78	IIII I	6
79 – 90	II	2

$$\Sigma f = 50$$

# Determinar o Ponto Médio

## Ponto médio de uma classe

$$\frac{(\text{Limite M\u00ednimo da Classe}) + (\text{Limite M\u00e1ximo da Classe})}{2}$$

Classe	Ponto m\u00e9dio	Frequ\u00eancia, $f$
7 - 18	$\frac{7+18}{2} = 12.5$	6
19 - 30	$\frac{19+30}{2} = 24.5$	10
31 - 42	$\frac{31+42}{2} = 36.5$	13

Tamanho da classe = 12

# Determinar a Frequência Relativa

## Frequência relativa de uma classe

- Porcentagem dos dados que se encaixa em um classe em particular

- Frequência relativa = 
$$\frac{\text{Frequência da classe}}{\text{Tamanho da amostragem}} = \frac{f}{n}$$

Classe	Frequência, $f$	Frequência relativa
7 – 18	6	$\frac{6}{50} = 0.12$
19 – 30	10	$\frac{10}{50} = 0.20$
31 – 42	13	$\frac{13}{50} = 0.26$

## Frequência acumulada de uma classe

A soma das frequências daquela classe e de todas as classes anteriores.

Classe	Frequência, $f$	Frequência acumulada
7 – 18	6	6
19 – 30	10	16
31 – 42	13	29

# Distribuição de Frequência Expandida

Classe	Frequência, $f$	Ponto médio	Frequência relativa	Frequência acumulada
7 – 18	6	12.5	0.12	6
19 – 30	10	24.5	0.20	16
31 – 42	13	36.5	0.26	29
43 – 54	8	48.5	0.16	37
55 – 66	5	60.5	0.10	42
67 – 78	6	72.5	0.12	48
79 – 90	2	84.5	0.04	50

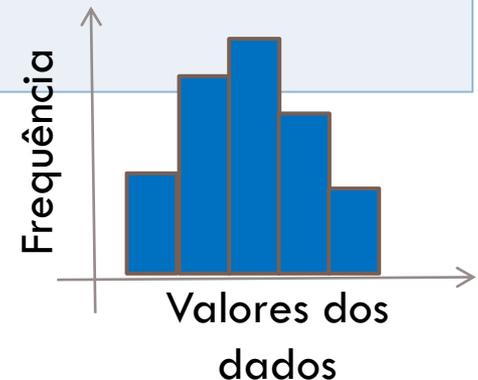
$$\Sigma f = 50$$

$$\Sigma \frac{f}{n} = 1$$

# Gráficos de Distribuição de Frequência

## Histograma de frequência

- Um gráfico de barras que representa a distribuição da frequência
- O eixo horizontal é quantitativo e mede os valores dos dados
- O eixo vertical mede as frequências das classes
- Barras consecutivas precisam se tocar



# Fronteiras de classes

- Os números separam as classes sem formar espaços entre elas

- A distância do limite superior da primeira classe para o limite inferior da segunda é  $19 - 18 = 1$
- A metade dessa distância é  $0,5$

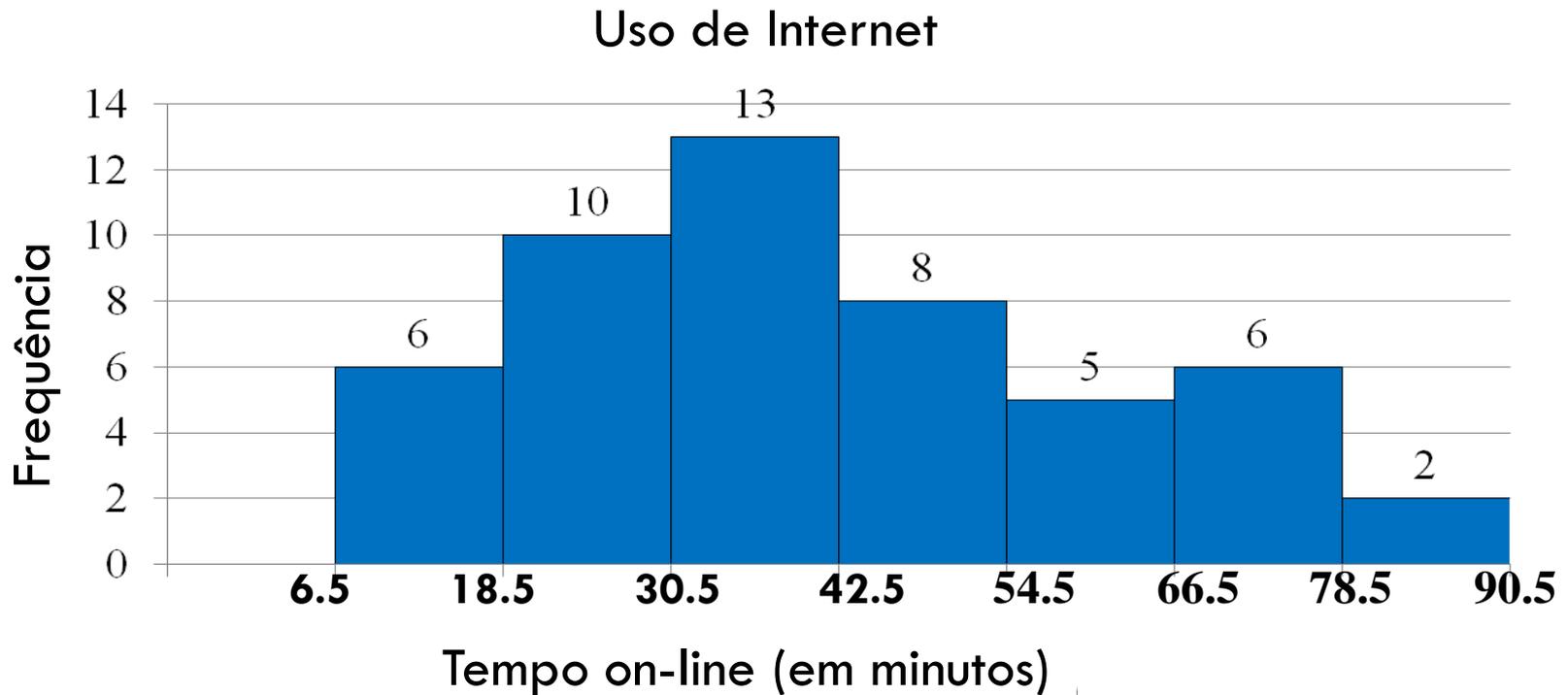
- Fronteira inferior da primeira classe =  $7 - 0.5 = 6.5$
- Fronteira superior da primeira classe =  $18 + 0.5 = 18.5$

Classe	Fronteiras de classes	Frequência, $f$
7 – 18	6.5 – 18.5	6
19 – 30		10
31 – 42		13

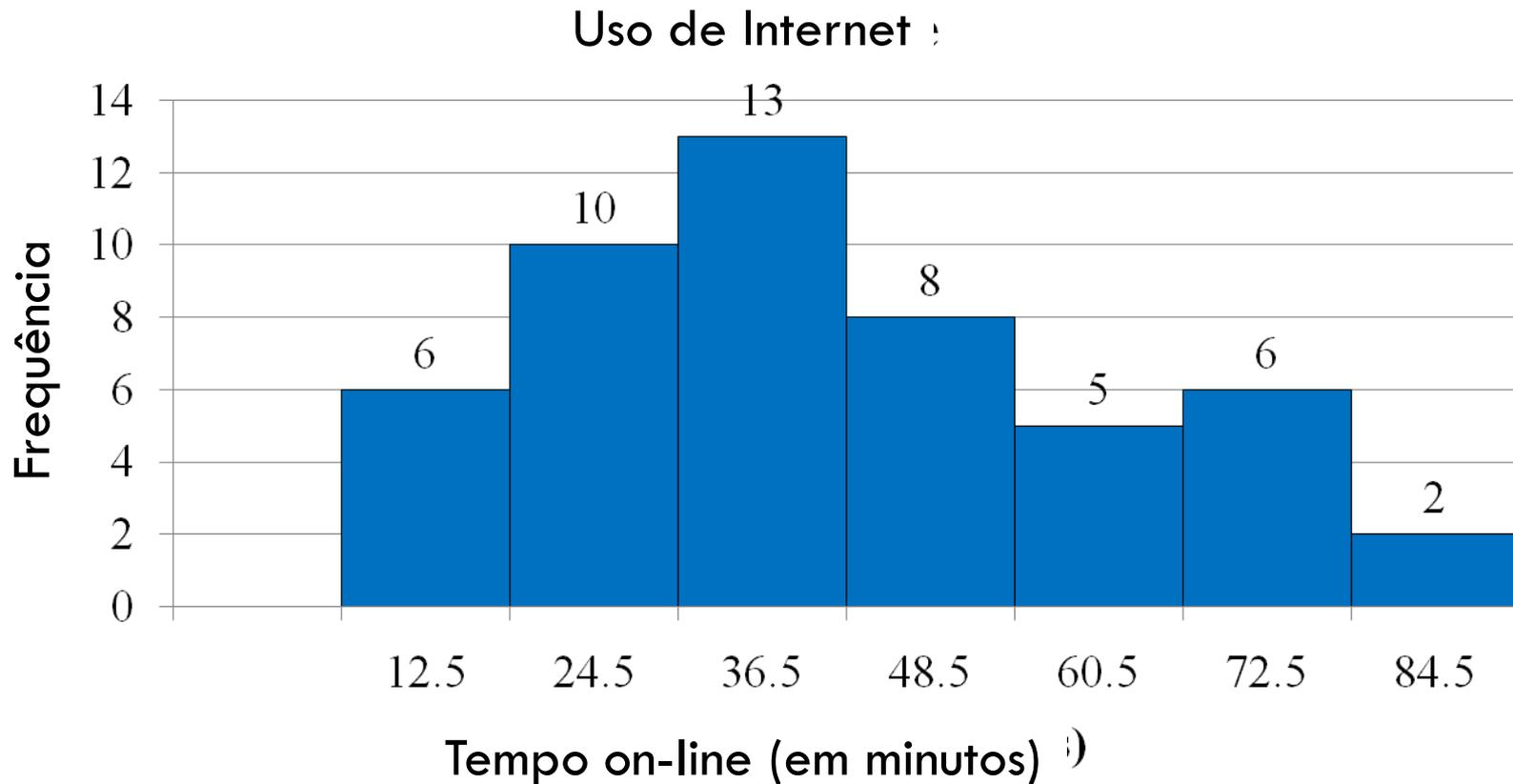
- Construir um histograma para a distribuição de frequência do uso da internet

<b>Classe</b>	<b>Fronteiras de classes</b>	<b>Frequência, <math>f</math></b>
7 – 18	6.5 – 18.5	6
19 – 30	18.5 – 30.5	10
31 – 42	30.5 – 42.5	13
43 – 54	42.5 – 54.5	8
55 – 66	54.5 – 66.5	5
67 – 78	66.5 – 78.5	6
79 – 90	78.5 – 90.5	2

# Histograma de frequência (usando fronteiras de classes)



# Histograma de frequência (usando pontos médios)



# Exercício 2.1

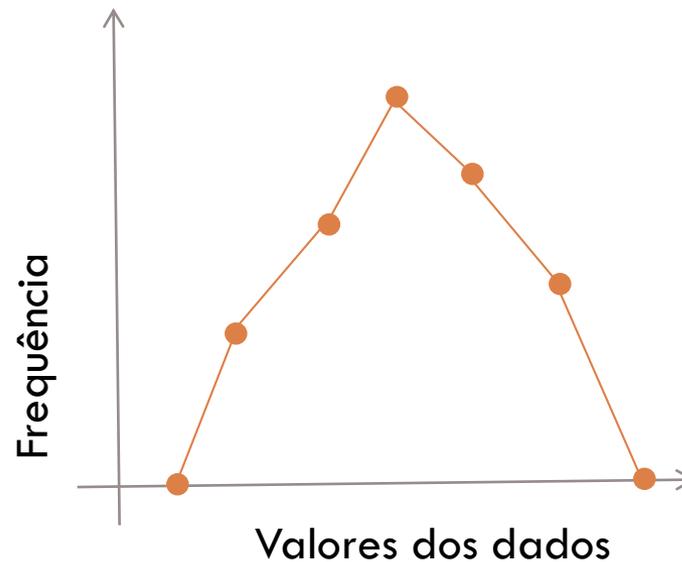
Construa a distribuição de frequência e o histograma de frequência relativa para o conjunto de dados usando 5 classes. Qual classe tem maior frequência relativa e qual tem a menor?

Conjunto de dados: uma amostra dos saques realizados em caixa eletrônicos (em real)

35	10	30	25	75	10	30	20	20	10	40
50	40	30	60	70	25	40	10	60	20	80
40	25	20	10	20	25	30	50	80	20	

# Polígono de frequência

Gráfico em linha que enfatiza a mudança contínua nas frequências.



# Construir um polígono de frequência

- ❑ Usar a mesma escala (horizontal e vertical) do histograma dos pontos médios
- ❑ Represente o ponto médio e a frequência
- ❑ Conecte os pontos
- ❑ O gráfico deve começar e terminar no eixo horizontal. Sendo assim, prolongue uma classe do ponto médio

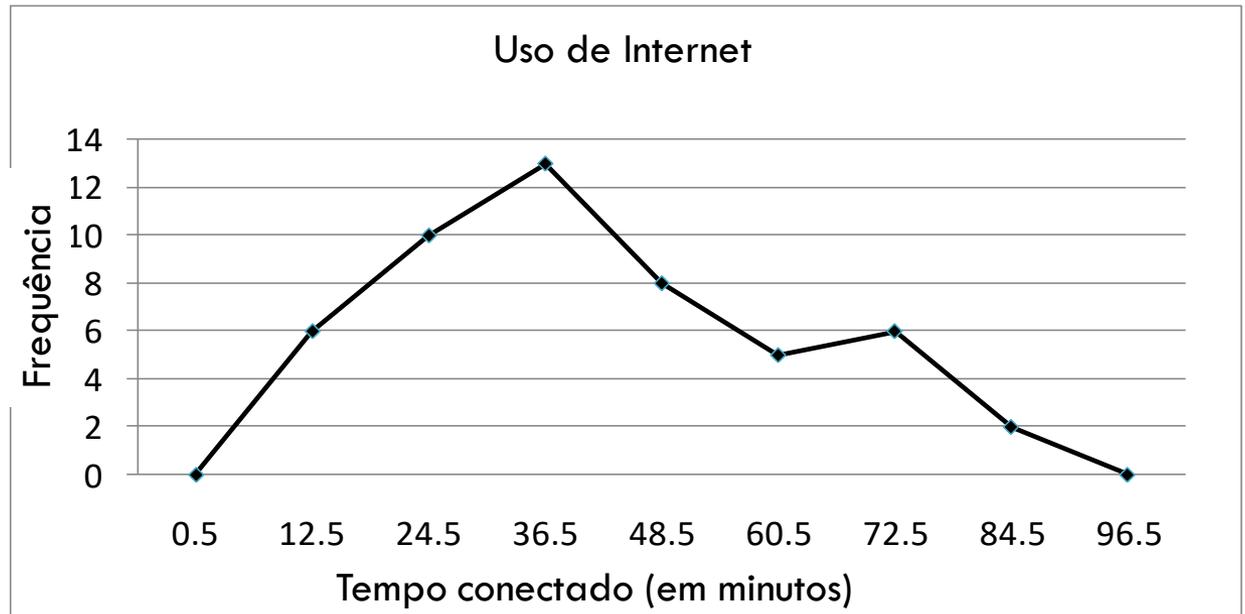
# Polígono de frequência

Construa um polígono de frequência para a distribuição da frequência do uso de Internet.

Classe	Ponto médio	Frequência, $f$
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2

# Polígono de Frequência

Gráfico deve começar e terminar no eixo horizontal, então estenda o lado esquerdo até o tamanho de uma classe antes do ponto médio da primeira classe e estenda o lado direito até o tamanho de uma classe depois do ponto médio da última classe.



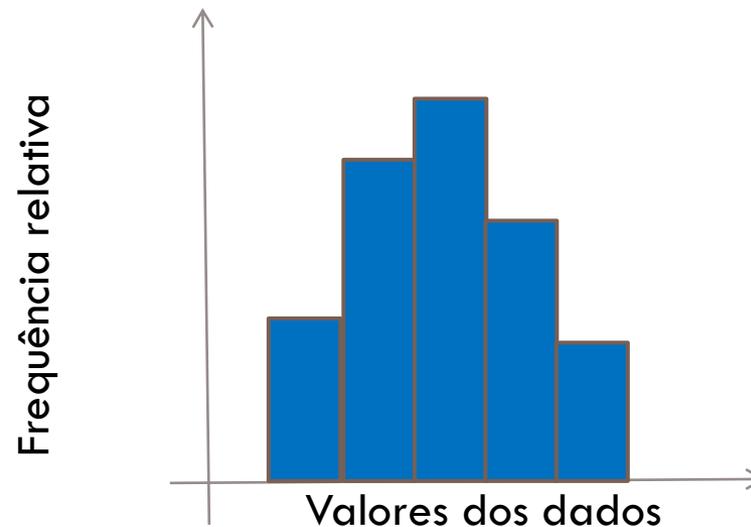
Pode-se perceber que a frequência dos assinantes aumenta até 36,5 minutos e então diminui.

# Exercício 2.2

- Construa polígono de frequência e descreva padrões que puder encontrar

# Histograma de Frequência Relativa

- Tem o mesmo formato e eixo horizontal que o histograma de frequência correspondente
- O eixo vertical mede as **frequências relativas**

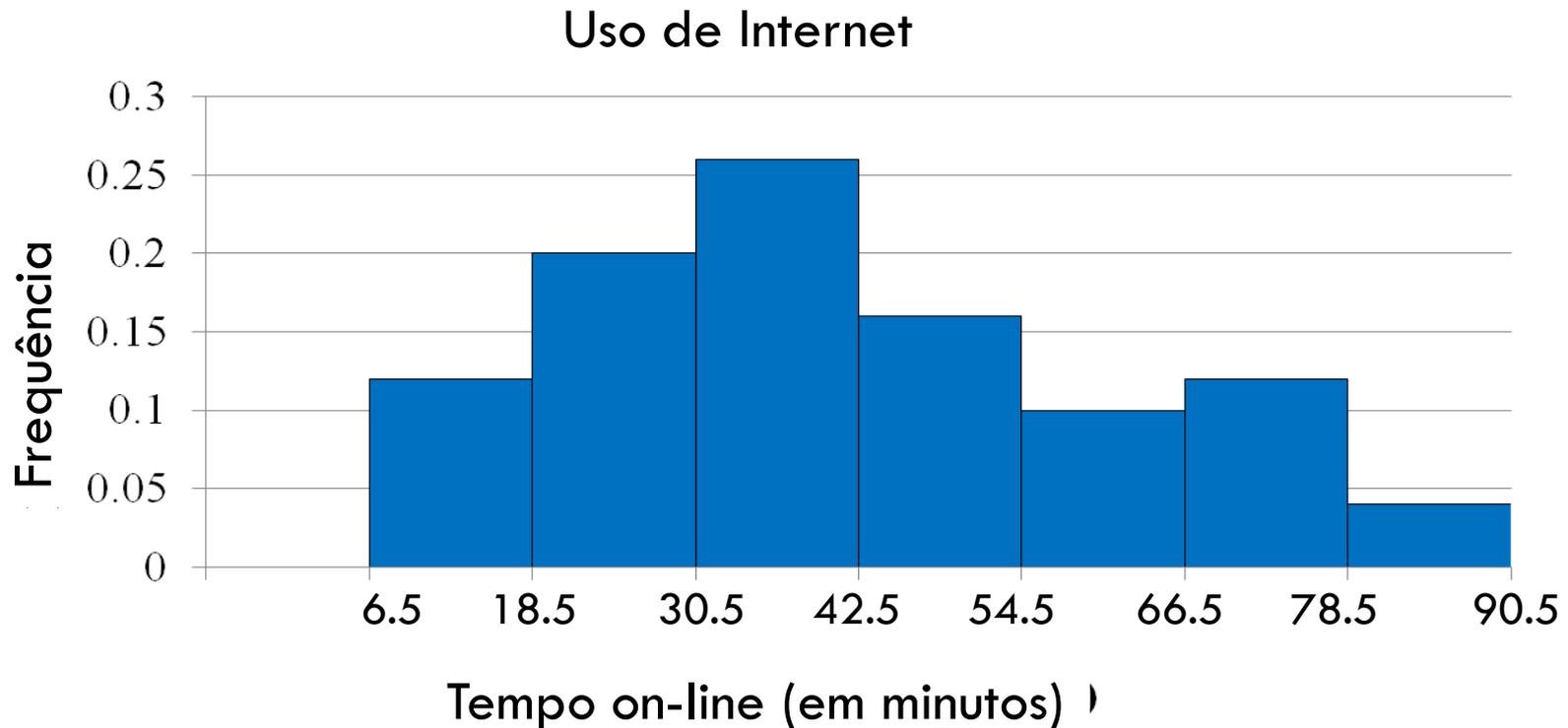


# Histograma de Frequência Relativa

Construir um histograma de frequência relativa para a distribuição da frequência do uso de Internet.

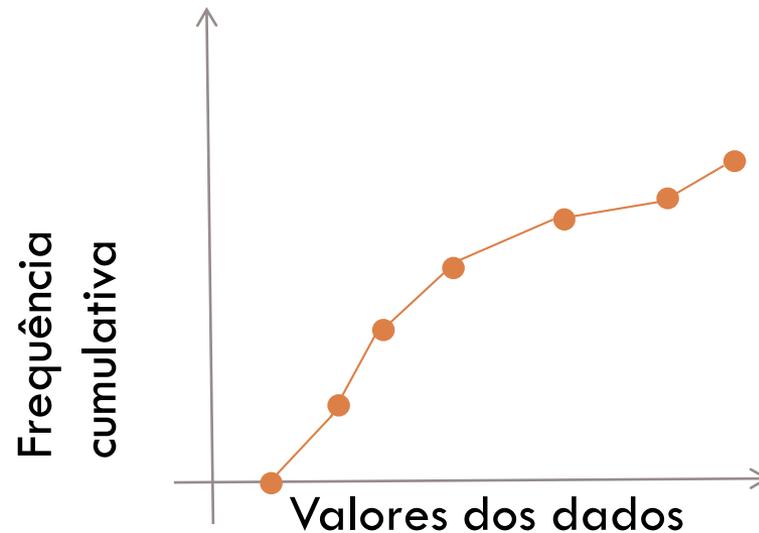
<b>Classe</b>	<b>Fronteiras de classes</b>	<b>Frequência, <math>f</math></b>	<b>Frequência relativa</b>
7 – 18	6.5 – 18.5	6	0.12
19 – 30	18.5 – 30.5	10	0.20
31 – 42	30.5 – 42.5	13	0.26
43 – 54	42.5 – 54.5	8	0.16
55 – 66	54.5 – 66.5	5	0.10
67 – 78	66.5 – 78.5	6	0.12
79 – 90	78.5 – 90.5	2	0.04

# Histograma de Frequência Relativa



# Gráficos de Frequências Cumulativas ou Ogiva

- ❑ Gráfico de linhas que demonstra a frequência cumulativa de cada classe em sua fronteira superior
- ❑ Fronteiras superiores demarcadas no eixo horizontal
- ❑ Frequências cumulativas demarcadas no eixo vertical



# Construir uma ogiva

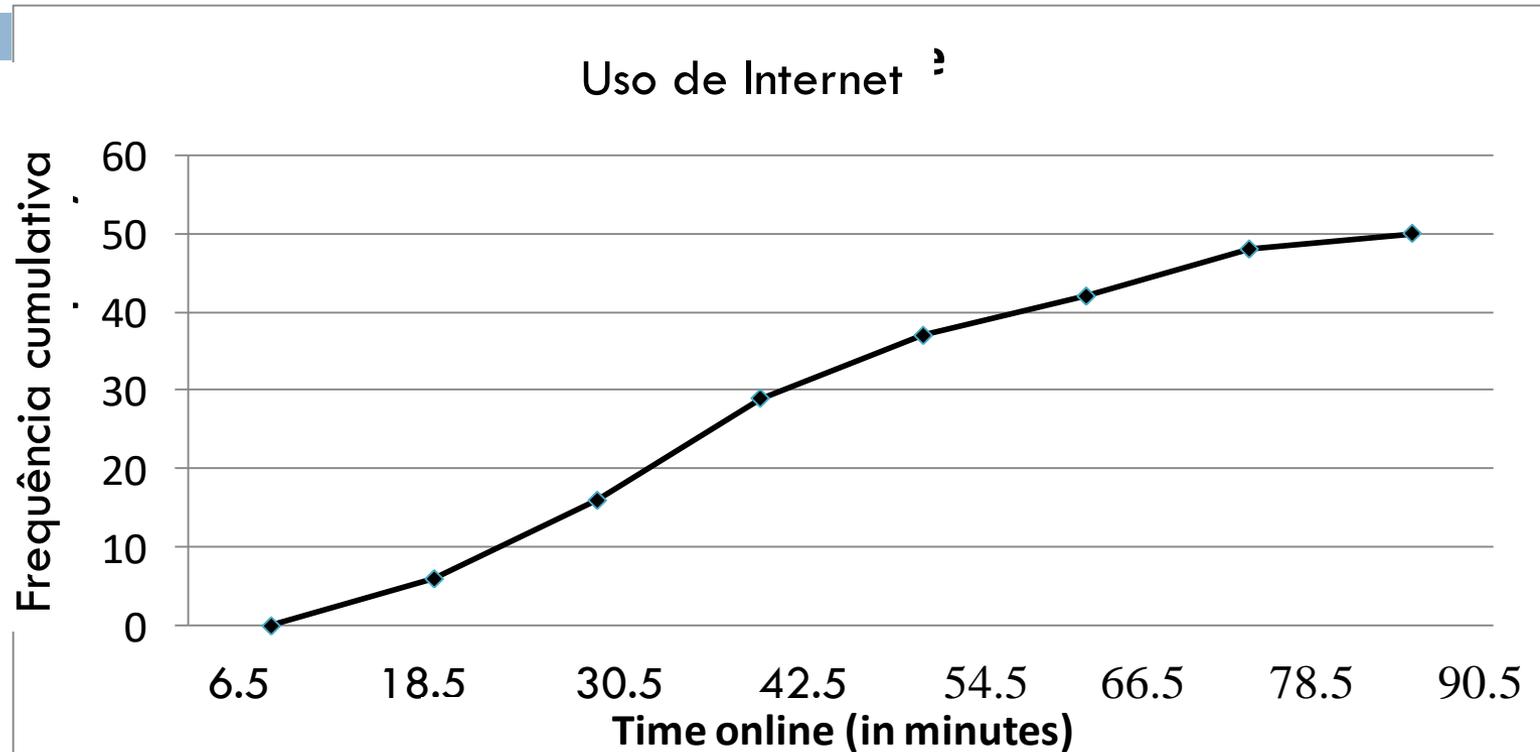
1. Construa uma distribuição de frequência considerando frequências cumulativas
2. Especifique os eixos horizontal (Fronteiras Superiores das Classes) e vertical (Frequências Cumulativas)
3. Represente os pontos das fronteiras superiores das classes e suas frequências cumulativas correspondentes
4. Gráfico começa na fronteira inferior da primeira classe (frequência cumulativa = 0) e termina na fronteira superior da última classe (frequência cumulativa = tamanho da amostra).

# Construir uma ogiva

Construir uma ogiva para a distribuição de frequência do uso da Internet.

Classe	Fronteiras de classes	Frequência, $f$	Frequência cumulativa
7 – 18	6.5 – 18.5	6	6
19 – 30	18.5 – 30.5	10	16
31 – 42	30.5 – 42.5	13	29
43 – 54	42.5 – 54.5	8	37
55 – 66	54.5 – 66.5	5	42
67 – 78	66.5 – 78.5	6	48
79 – 90	78.5 – 90.5	2	50

# Construir uma ogiva



Aproximadamente 40 assinantes passaram menos de 60 minutos conectados em sua última sessão.

O maior aumento no uso ocorre entre 30,5 minutos e 42,5 minutos.

## Exercício 2.3

- Construa uma ogiva e descreva a localização do maior aumento de frequência.

# Exercício 2.4

- Você trabalha em um banco e deve decidir a quantia de dinheiro que será colocada em um caixa eletrônico a cada dia. Você não quer colocar dinheiro em excesso (por razão de segurança) ou pouco dinheiro (que pode causar problemas com os clientes). As quantias retiradas diariamente (em centenas de reais) em um período de 30 dias são mostradas abaixo.

72	84	61	76	104	76	86	92	80	88
98	76	97	82	84	67	70	81	82	89
74	73	86	81	85	78	82	80	91	83

- A. construa um histograma de frequência relativa para os dados usando 8 classes
- B. se você colocar \$ 9.000 no caixa eletrônico a cada dia, qual a porcentagem de dias em um mês na qual você deve esperar ficar sem dinheiro no caixa? Explique o raciocínio
- C. se você deseja que o caixa fique sem dinheiro em 10% dos dias, quanto dinheiro, em centenas de reais, você deveria colocar no caixa eletrônico a cada dia? Explique o raciocínio

# Exercício 2.5

- Construa uma distribuição de frequência e um polígono de frequência para o conjunto de dados. Descreva quaisquer padrões (número de classes = 5)

Notas de exame para uma classe de estatística

83	92	94	82	73	98	78	85	72	90
89	92	96	89	75	85	63	47	75	82

# Exercício 2.6

- Construa uma distribuição de frequência acumulada e uma ogiva para os conjuntos de dados usando 6 classes. Descreva a localização do maior aumento de frequência.

38	32	34	39	40	54	32	17	29	33
57	40	25	36	33	24	42	16	31	33

# Mais gráficos e representações

Aula 3

# Objetivos

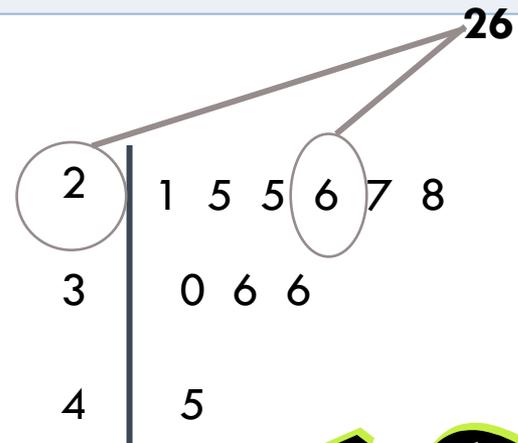
- Fazer gráficos de dados quantitativos
  - Diagrama de ramos e folhas
  - Diagrama de Pontos
- Fazer gráficos de dados qualitativos
  - Gráficos de Pizza
  - Gráfico de Pareto
- Fazer gráficos de dados emparalhedos
  - Gráficos de Dispersão
  - Gráficos da Série Temporal

# Gráficos de Dados Quantitativos

## Diagrama de ramos-e-folhas

- Cada número é separado em um **ramo** e uma **folha**
- Similar a um histograma
- Ainda contém os valores dos dados originais

Dados: 21, 25, 25, **26**, 27, 28, 30,  
36, 36, 45



# Diagrama de Ramos-e-Folhas

- ❑ Liste todos os ramos possíveis
- ❑ Liste as folhas de cada entrada de dados à direita de seu ramo
- ❑ Reescreva o diagrama ramo-e-folhas de modo que as folhas estejam ordenadas
- ❑ Use o gráfico para tirar conclusões

# Diagrama de Ramos-e-Folhas

Os números abaixo representam a quantidade de mensagens de texto enviadas no mês passado pelos usuários de celulares em um andar do dormitório da faculdade. Mostre os dados em um diagrama de ramos-e-folhas.

155	159	144	129	105	145	126	116	130	114
122	112	112	142	126	118	108	122	121	109
140	126	119	113	117	118	109	109	119	139
122	78	133	126	123	145	121	134	124	119
132	133	124	129	112	126	148	147		

# Diagrama de Ramos-e-Folhas

```
155 159 144 129 105 145 126 116 130 114 122 112 112 142
126 118 108 122 121 109 140 126 119 113 117 118 109 109
119 139 122 78 133 126 123 145 121 134 124 119 132 133
124 129 112 126 148 147
```

- Entradas de dados variam do 78 até 159
- Usar o dígito mais à direita como a folha
  - Por exemplo,  
 $78 = 7 \mid 8$  e  $159 = 15 \mid 9$
- Listar os ramos – 7 até 15 – à esquerda de uma linha vertical
- Para cada entrada de dados, listar uma folha à direita de seu ramo

# Diagrama de Ramos-e-Folhas

**Número de mensagens de texto enviadas**      **Número de mensagens de texto enviadas**

7 | 8      Key: 15|5 = 155  
8 |  
9 |  
10 | 5 8 9 9 9  
11 | 6 4 2 2 8 8 9 3 7 8 9 9 2  
12 | 9 6 2 6 2 1 6 2 6 3 1 4 4 9 6  
13 | 0 9 9 3 4 2 3  
14 | 4 5 2 0 5 8 7  
15 | 5 9

7 | 8      Key: 15|5 = 155  
8 |  
9 |  
10 | 5 8 9 9 9  
11 | 2 2 2 3 4 6 7 8 8 8 9 9 9  
12 | 1 1 2 2 2 3 4 4 6 6 6 6 6 9 9  
13 | 0 2 3 3 4 9 9  
14 | 0 2 4 5 5 7 8  
15 | 5 9

**Diagrama de ramos-e-folhas não ordenado**      **Diagrama de ramos-e-folhas ordenado**

Mais de 50% dos usuários de celular enviaram entre 110 e 130 mensagens de texto.

# Diagrama Ramo-e-Folhas

- Organize os dados usando um diagrama de ramo-e-folhas que tenha duas fileiras para cada ramo. O que você pode concluir?

# Exercício 3.1

- Use um diagrama ramo-e-folhas para representar os dados. Os dados representam a espessura do gelo medido em 20 localidades diferentes de um lago congelado.

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

# Diagrama de Pontos

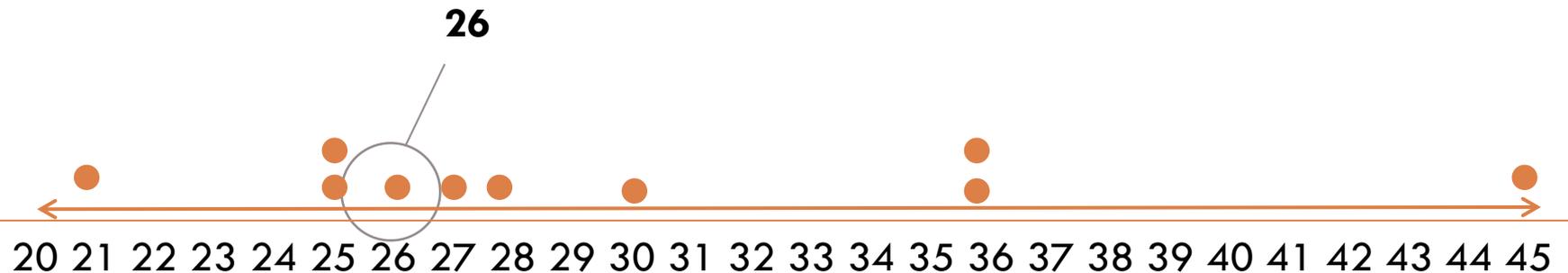
- Entrada de dados é indicada usando um ponto acima de um eixo horizontal

Dados: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45

# Diagrama de Pontos

- Entrada de dados é indicada usando um ponto acima de um eixo horizontal

Dados: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



# Diagrama de Pontos

- Escolha uma escala adequada para o eixo horizontal
- Represente cada entrada de dados por um diagrama de pontos
- Descreva quaisquer padrões

# Exercício 3.2

Use um diagrama de pontos para organizar os dados das mensagens de texto.

```
155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147
```

- Para que cada entrada seja incluída no diagrama de pontos, o eixo horizontal deve incluir números entre 70 e 160
- Para representar uma entrada, insira um ponto acima da posição da entrada no eixo
- Se uma entrada for repetida, faça outro ponto acima do ponto anterior

## Exercício 3.3

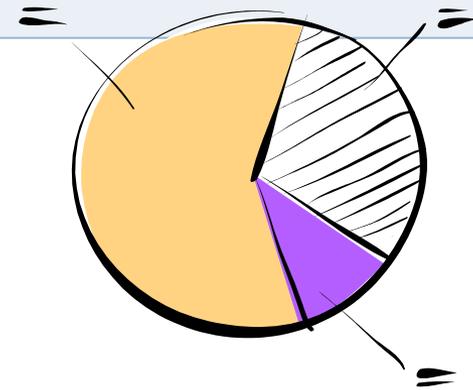
- Use um diagrama de pontos para representar os dados. Os dados representam a espessura do gelo medido em 20 localidades diferentes de um lago congelado.

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

# Gráficos de Dados Qualitativos

## Gráfico de pizza

- Círculo é dividido em vários setores que representam categorias
- A área de cada setor é proporcional à frequência de cada categoria



# Construir um Gráfico de Pizza

- Encontre a frequência relativa de cada categoria
- Use o ângulo central pra encontrar a porção que corresponde a cada categoria
- Analise o gráfico

# Construir um Gráfico de Pizza

O número de ocupantes de veículos motorizados mortos em acidentes em 2005 é exibido na tabela. Use um gráfico de pizza para organizar os dados (*Fonte: U.S. Department of Transportation, National Highway Traffic Safety Administration*).

Tipo de veículo	Mortes
Carros	18.440
Caminhões	13.778
Motocicletas	4.553
Outros	823

# Construir um Gráfico de Pizza

- Encontre a frequência relativa de cada categoria

Tipo de Veículo	Frequência, $f$	Frequência relativa
Carros	18.440	$\frac{18440}{37594} \approx 0.49$
Caminhões	13.778	$\frac{13778}{37594} \approx 0.37$
Motocicletas	4.553	$\frac{4553}{37594} \approx 0.12$
Outros	823	$\frac{823}{37594} \approx 0.02$

# Construir um Gráfico de Pizza

- Construa o gráfico de pizza usando o ângulo central que corresponda à cada categoria.
  - ▣ Para encontrar o ângulo central, multiplique  $360^\circ$  pela frequência relativa da categoria.
  - ▣ Por exemplo, o ângulo central para carros é

$$360(0,49) \approx 176^\circ$$

# Construir um Gráfico de Pizza

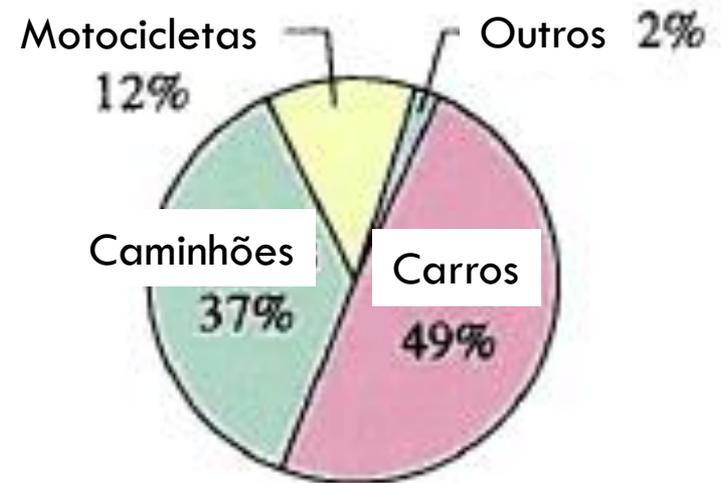
Tipo de veículo	Frequência, f	Frequência relativa	Ângulo Central
Carros	18.440	0,49	$360^\circ(0,49) \approx 176^\circ$
Caminhões	13.778	0,37	$360^\circ(0,37) \approx 133^\circ$
Motocicletas	4.553	0,12	$360^\circ(0,12) \approx 43^\circ$
Outros	823	0,02	$360^\circ(0,02) \approx 7^\circ$

# Construir um Gráfico de Pizza

Tipo de veículo	Frequência relativa	Ângulo central
Carros	0,49	176°
Caminhões	0,37	133°
Motocicletas	0,12	43°
Outros	0,02	7°

## Ocupantes de veículos

### motorizados mortos em 2005



A partir do gráfico, pode-se concluir que a maioria dos acidentes fatais em veículos automotivos envolve carros.

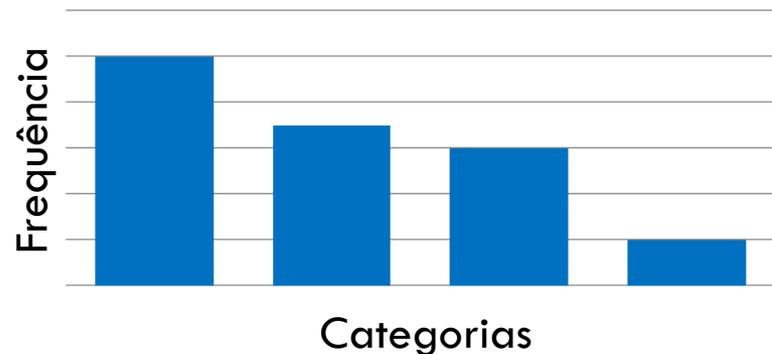
# Exercício 3.4

O número de ocupantes de veículos motorizados mortos em acidentes em 1995 é mostrado na tabela. Use um gráfico de pizza para organizar os dados e compare com os dados de 2005.

<b>Tipo de veículo</b>	<b>Mortes</b>
Carros	22.423
Caminhões	10.216
Motocicletas	2.227
Outros	425

# Gráficos de Pareto

- Um gráfico de barras verticais no qual a altura de cada barra representa uma frequência ou uma frequência relativa
- As barras são posicionadas por ordem decrescente de altura; a barra mais alta é posicionada à esquerda



# Gráficos de Pareto

- Encontre a frequência ou a frequência relativa para cada entrada de dados
- Posicione as barras em ordem decrescente de acordo com a frequência ou frequência relativa
- Interprete os resultados no contexto dos dados

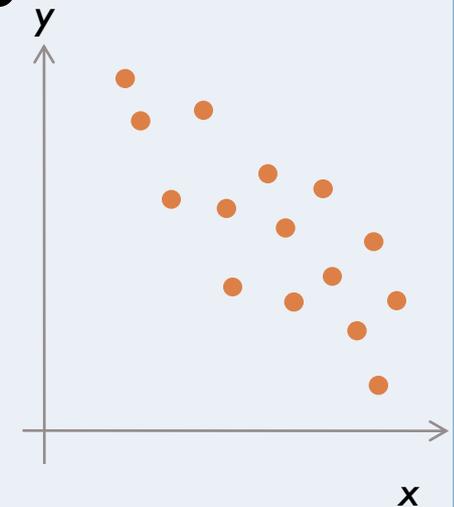
## Exercício 3.5

Recentemente, a indústria de varejo perdeu 41 milhões com redução nos estoques. A redução de estoque é uma perda de estoque por meio de quebra, roubo de carga, roubo em lojas etc. As causas da redução de estoque são erro administrativo (7,8 milhões), roubo por funcionários (15,6 milhões), furto das lojas (14,7 milhões) e fraude dos vendedores (2,9 milhões). Se você fosse um varejista, para qual causa de redução de estoque você olharia primeiro? *(Fonte: National Retail Federation and Center for Retailing Education, University of Florida.)*

# Gráficos de Conjunto de Dados Emparelhados

## Conjunto de dados emparelhados

- Cada entrada em um conjunto de dados corresponde à outra entrada em um segundo conjunto de dados
- Gráficos usam **um gráfico de dispersão**
  - ▣ Os pares ordenados são representados como pontos em um plano coordenado
  - ▣ Usado para representar a relação entre duas variáveis quantitativas



# Gráfico de dispersão

- Classifique os eixos horizontal e vertical
- Represente graficamente os dados emparelhados
- Descreva as tendências

# Exercício 3.6

O estatístico britânico Ronald Fisher apresentou um famoso conjunto de dados chamado de conjunto de dados de Íris de Fisher. Esse conjunto de dados descreve várias características físicas, como o comprimento de pétalas e a sua largura (em milímetros) para três espécies de íris (flor). No gráfico de dispersão mostrado, os comprimentos de pétalas formam o primeiro conjunto de dados e as larguras formam o segundo conjunto de dados. Conforme o comprimento da pétala aumenta, o que tende a acontecer com a largura? (*Fonte: Fisher, R. A., 1936.*)



# Gráficos de Conjunto de Dados Emparelhados

O período de emprego e os salários de 10 funcionários estão listados na tabela. Faça o gráfico para os dados usando um gráfico de dispersão. O que podemos concluir?

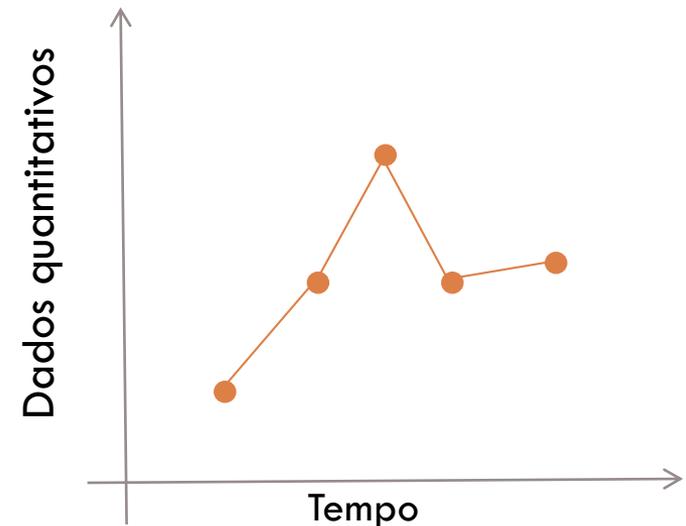
- Classifique os eixos horizontal e vertical
- Represente graficamente os dados emparelhados
- Descreva tendências

Tempo de emprego	Salário (\$)
5	32.000
4	32.500
8	40.000
4	27.350
2	25.000
10	43.000
7	41.650
6	39.225
9	45.100
3	28.000

# Gráficos de Série Temporal

## Série temporal

- Conjuntos de dados são compostos de entradas quantitativas tomadas em intervalos regulares em um período de tempo
  - Por exemplo, a quantidade de precipitações medidas a cada dia por um mês
- Gráfico de **períodos de tempo**



# Construir Gráfico de Série Temporal

- Classifique os eixos horizontais e verticais
- Represente graficamente os dados emparelhados
- Conecte os dados com segmentos de linha
- Descreva padrões

## Exercício 3.7

A tabela lista o número de telefones celulares (em milhões) para os anos de 1995 até 2005. Construa um gráfico de série temporal do número de telefones celulares. (Fonte: *Cellular Telecommunication & Internet Association.*)

Ano	Usuários (em milhões)
1995	33.8
1996	44.0
1997	55.3
1998	69.2
1999	86.0
2000	109.5
2001	128.4
2002	140.8
2003	158.7
2004	182.1
2005	207.9

# Objetivos

- Fazer gráficos de dados quantitativos
  - Diagrama de ramos e folhas
  - Diagrama de Pontos
- Fazer gráficos de dados qualitativos
  - Gráficos de Pizza
  - Gráfico de Pareto
- Fazer gráficos de dados emparelhados
  - Gráficos de Dispersão
  - Gráficos da Série Temporal

# Medidas de Tendência Central

Aula 3

# Objetivos

- Determinar
  - ▣ Média, Moda e Mediana da população e da amostra
  - ▣ Média Ponderada de um conjunto de dados
  - ▣ Média da Distribuição de Frequência
- Descrever formas da distribuição

# Medida de Tendência Central

- Valor que representa uma entrada típica ou central de um conjunto de dados
- Medidas de tendência central mais comuns:
  - Média
  - Mediana
  - Moda

# Medidas de Tendência Central

□ **Média:** soma de todas as entradas de dados dividida pelo número de entradas

□ **Média populacional:**

$$\mu = \frac{\Sigma x}{N}$$

□ **Média amostral:**

$$\bar{x} = \frac{\Sigma x}{n}$$

## Exercício 3.8

Os preços (em dólares) para uma amostra de viagens feitas de Chicago em Illinois para Cancun no México são listados. Qual o preço médio dos voos? R. \$527,90

872 432 397 427 388 782 397



# Medidas de Tendência Central

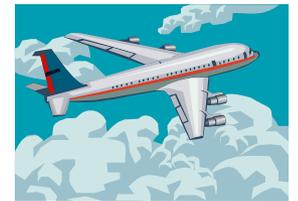
**Mediana:** valor que está no meio dos dados quando o conjunto dos dados é **ordenado**

- Divide os dados em duas partes iguais
- Se o conjunto de dados possui um número de entradas:
  - ▣ **ímpar:** o mediano é o elemento do meio
  - ▣ **par:** o mediano será a média dos dois elementos centrais

## Exercício 3.9

Os preços (em dólares) para uma amostra de viagens feitas de Chicago, Illinois, para Cancun, México, são listados. Encontre a mediana.

872 432 397 427 388 782 397



# Exercício 3.10

O preço dos voos em \$ 432 não está mais disponível.  
Qual o preço mediano dos voos restantes?

872 397 427 388 782 397



# Medidas de Tendência Central

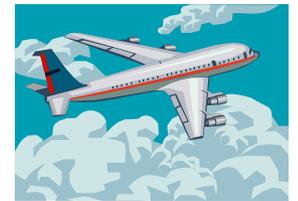
## Moda

- Entrada de dados que ocorre com a maior frequência
- Se não houver entradas repetidas, o conjunto de dados não tem moda
- Se duas entradas ocorrem com a mesma e mais alta frequência, cada entrada é um moda (**bimodal**)

# Exercício 3.11

Os preços (em dólares) para uma amostra de viagens feitas de Chicago, Illinois, para Cancun, México, são listados. Encontre a moda dos preços dos voos.

872 432 397 427 388 782 397



# Exercício 3.12

Em um debate político, uma amostra de membros da audiência foi questionada à respeito de seus partidos políticos. Suas respostas estão na tabela. Qual a moda de suas respostas?

Partido político	Frequência, $f$
Democrata	34
Republicano	56
Outros	21
Não responderam	9



# Comparar Média, Mediana e Moda

- As três medidas descrevem uma entrada típica de um conjunto de dados
- Média
  - ▣ Vantagem: medida confiável porque leva em conta cada entrada do conjunto de dados
  - ▣ Desvantagem: afetada por valores discrepantes (uma entrada que é muito distante das outras entradas no conjunto de dados)

# Exercício 3.13

Encontre a média, a mediana e a moda da amostra de idades de uma classe. Qual medida de tendência central descreve melhor uma entrada típica desse conjunto de dados? Existe algum valor discrepante?

Idades em uma classe						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

# Exercício 3.14

Remova a entrada de dados 65 do conjunto de dados anterior. Como a ausência do valor discrepante muda cada uma destas medidas?

- a. Encontre a média, a mediana e a moda
- b. Compare essas medidas da tendência central com o exemplo anterior

# Média Ponderada

- Média de um conjunto de dados cujas entradas possuem pesos variantes

- $$\bar{x} = \frac{\sum(x \cdot w)}{\sum w}$$

- $w$  é o peso de cada entrada  $x$

# Encontrar a Média Ponderada

Você está frequentando uma aula onde sua nota é determinada com base em 5 fontes: 50% da média de seu exame, 15% do seu exame bimestral, 20% de seu exame final, 10% de seu trabalho no laboratório de informática e 5% de seus deveres de casa. Suas notas são: 86 (média do exame), 96 (exame bimestral), 82 (exame final), 98 (laboratório) e 100 (dever de casa). Qual é a média ponderada de suas notas? A média mínima para um A é 90, você obteve uma nota A?

# Encontrar a Média Ponderada

Fonte	Notas, $x$	Peso, $w$	$x \cdot w$
Média do exame			
Exame bimestral			
Exame final			
Laboratório			
Dever de casa			
		$\Sigma w =$	$\Sigma(x \cdot w) =$

Sua média ponderada foi .....e, portanto, não tirou A

# Encontrar a Média Ponderada

Fonte	Notas, $x$	Peso, $w$	$x \cdot w$
Média do exame	86	0,50	$86(0,50) = 43,0$
Exame bimestral	96	0,15	$96(0,15) = 14,4$
Exame final	82	0,20	$82(0,20) = 16,4$
Laboratório	98	0,10	$98(0,10) = 9,8$
Dever de casa	100	0,05	$100(0,05) = 5,0$
		$\Sigma w = 1$	$\Sigma(x \cdot w) = 88,6$

$$\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w} = \frac{88.6}{1} = 88.6$$

Sua média ponderada foi 88,6 e, portanto, não tirou A

# Exercício 3.15

Houve um erro no cálculo da nota de seu exame final. Ao invés de 82 você obteve 98. Qual sua nova média ponderada?

# Média de Dados Agrupados

## Média de uma distribuição de frequência

- Aproximada por

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} \quad n = \Sigma f$$

Onde  $x$  = ponto médio da classe

$f$  = frequência da classe

# Encontrar Média da Distribuição de Frequência

## *Em palavras*

1. Encontre o ponto médio de cada classe.
2. Encontre a soma dos produtos dos pontos médios e das frequências.
3. Encontre a soma das frequências.
4. Encontre a média da distribuição das frequências.

## *Em símbolos*

$$x = \frac{(\text{Limite inferior}) + (\text{Limite superior})}{2}$$

$$\Sigma(x \cdot f)$$

$$n = \Sigma f$$

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n}$$

# Encontrar a Média da Distribuição da Frequência

Usar a distribuição de frequência para aproximar a média do número de minutos que uma amostra de internautas passou conectada em sua última sessão.

Classe	Ponto médio	Frequência, $f$
7 – 18		6
19 – 30		10
31 – 42		13
43 – 54		8
55 – 66		5
67 – 78		6
79 – 90		2

# Encontrar a Média da Distribuição da Frequência

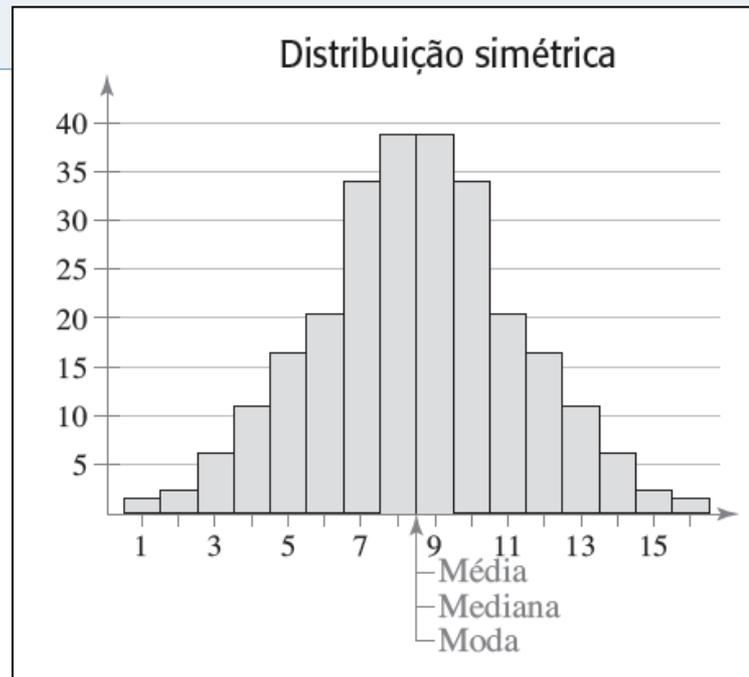
Classe	Ponto médio, $x$	Frequência, $f$	$(x \cdot f)$
7 – 18	12,5	6	$12,5 \cdot 6 = 75,0$
19 – 30	24,5	10	$24,5 \cdot 10 = 245,0$
31 – 42	36,5	13	$36,5 \cdot 13 = 474,5$
43 – 54	48,5	8	$48,5 \cdot 8 = 388,0$
55 – 66	60,5	5	$60,5 \cdot 5 = 302,5$
67 – 78	72,5	6	$72,5 \cdot 6 = 435,0$
79 – 90	84,5	2	$84,5 \cdot 2 = 169,0$
		$n = 50$	$\Sigma(x \cdot f) = 2.089,0$

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} = \frac{2089}{50} \approx 41.8 \text{ minutes}$$

# Forma das distribuições

## Distribuição simétrica

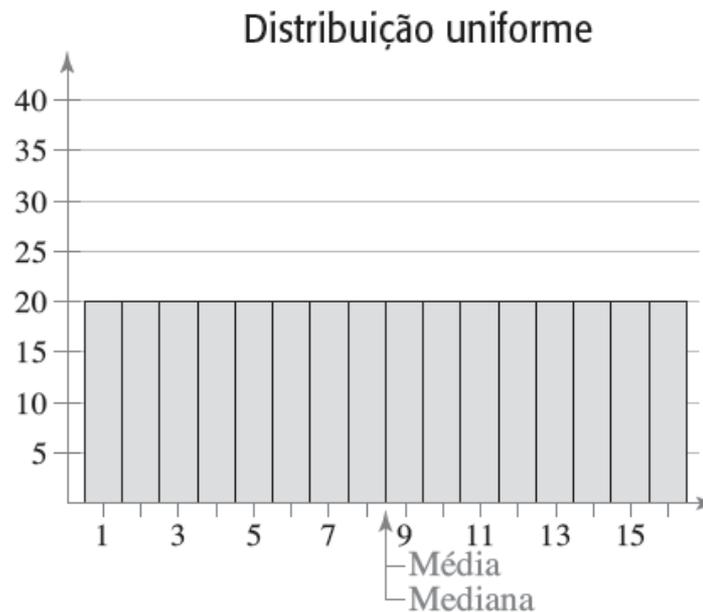
Uma linha vertical pode ser traçada do meio do gráfico de distribuição e as metades resultantes são quase idênticas.



# Forma das Distribuições

## Distribuição uniforme (retangular)

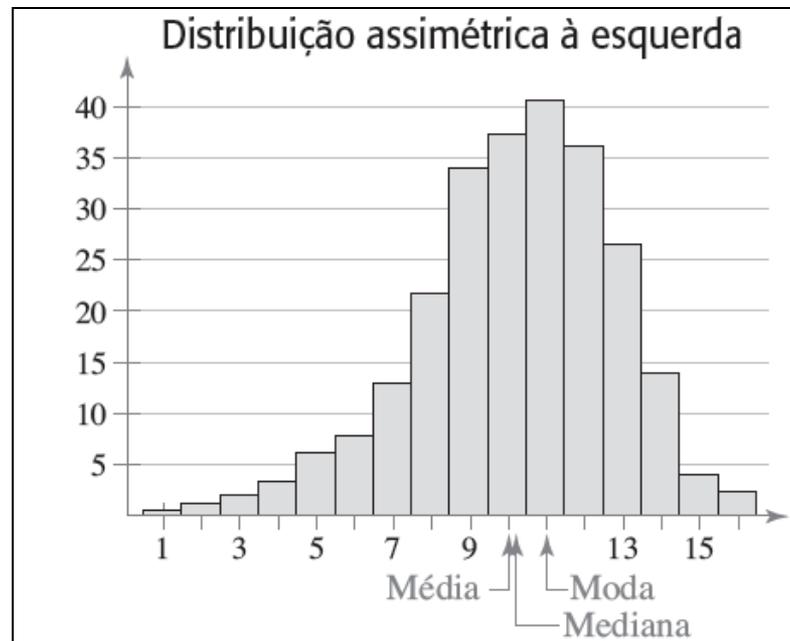
- Todas as entradas têm frequências iguais ou quase iguais
- Simétrica



# Forma das Distribuições

## Distribuição assimétrica à esquerda (negativamente)

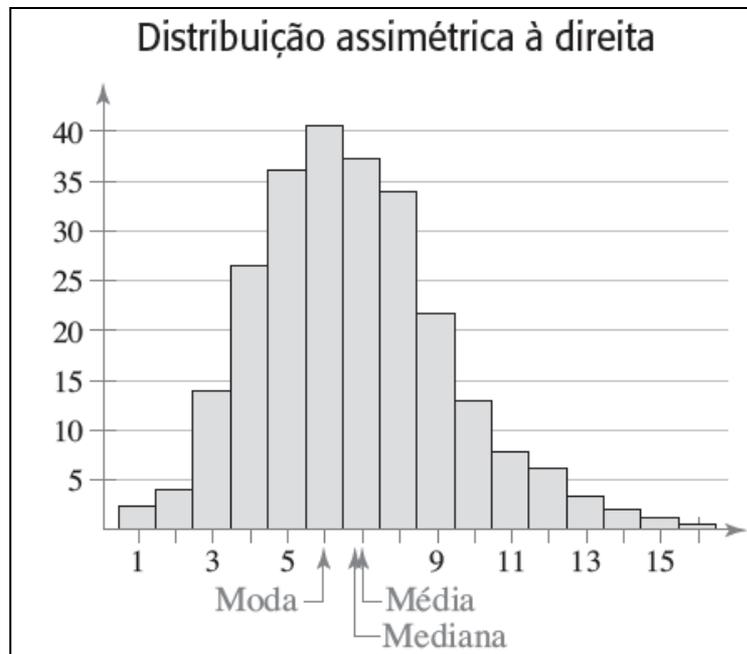
- A “cauda” do gráfico se alonga mais à esquerda
- A média fica à esquerda da mediana



# Forma das Distribuições

## Distribuição assimétrica à direita (positivamente)

- A “cauda” do gráfico se alonga mais à direita
- A média fica à direita da mediana



# Exercício 3.16

Encontre a média, a mediana e a moda dos dados  
“Pesos em kg dos cães em um canil” Chave 1/0 = 10

1	0 2
2	1 4 7
3	7 8
4	1 5 5
5	0 7
6	5
7	
8	
9	
10	6

# Objetivos

- Determinar
  - Média, Moda e Mediana da população e da amostra
  - Média Ponderada de um conjunto de dados
  - Média da Distribuição de Frequência
- Descrever formas da distribuição

# Medidas de Variação

Aula 4

# Objetivos

- Determinar
  - ▣ Amplitude de um conjunto de dados
  - ▣ Variância e Desvio Padrão da População e da Amostra
- Usar a Regra Empírica e o Teorema de Chebychev para interpretar o desvio padrão
- Aproximar o desvio padrão da amostra para dados agrupados

# Amplitude

- Diferença entre entradas máxima e mínima em um conjunto de dados
- Os dados devem ser quantitativos
- $\text{Amplitude} = \text{Entrada Máxima} - \text{Entrada Mínima}$



# Exercício 4.1

As empresas A e B contrataram 10 graduados. Os salários iniciais de cada um são demonstrados abaixo. Encontre a amplitude dos salários iniciais.

Salários iniciais para empresa A (milhares de \$)

41	38	39	45	47	41	44	41	37	42
----	----	----	----	----	----	----	----	----	----

Salários iniciais para empresa B (milhares de \$)

40	23	41	50	49	32	41	29	52	58
----	----	----	----	----	----	----	----	----	----

Encontre a média, mediana e a moda para as duas empresas

# Desvio, Variância e Desvio Padrão

## Desvio

- Entrada de Dados – Média
- Conjunto de dados da população:
  - ▣ Desvio de  $x = x - \mu$
- Conjunto de dados da amostra:
  - ▣ Desvio de  $x = x - \bar{x}$

# Encontrar o Desvio

Uma corporação contratou 10 graduados. Os salários iniciais de cada um são demonstrados abaixo. Encontre o desvio dos salários iniciais.

Salários iniciais (milhares de \$)

41 38 39 45 47 41 44 41 37 42

# Encontrar o Desvio

Uma corporação contratou 10 graduados. Os salários iniciais de cada um são demonstrados abaixo. Encontre o desvio dos salários iniciais.

Salários iniciais (milhares de \$)

41 38 39 45 47 41 44 41 37 42

## Solução:

- Primeiro, determine a média dos salários iniciais

$$\mu = \frac{\Sigma x}{N} = \frac{415}{10} = 41.5$$

# Encontrar o Desvio

Determine o desvio para cada entrada

Salário (\$ 1.000s), $x$	Desvio: $x - \mu$
41	$41 - 41,5 = -0,5$
38	$38 - 41,5 = -3,5$
39	$39 - 41,5 = -2,5$
45	$45 - 41,5 = 3,5$
47	$47 - 41,5 = 5,5$
41	$41 - 41,5 = -0,5$
44	$44 - 41,5 = 2,5$
41	$41 - 41,5 = -0,5$
37	$37 - 41,5 = -4,5$
42	$42 - 41,5 = 0,5$

$$\Sigma x = 415$$

$$\Sigma(x - \mu) = 0$$

# Desvio, Variância e Desvio Padrão

## Variância da população

□ 
$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N}$$

← Soma dos quadrados,  $SQ_x$

## Desvio padrão da população

□ 
$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x - \mu)^2}{N}}$$

# Encontrar Variância Populacional e o Desvio Padrão

## *Em palavras*

1. Encontre a média do conjunto de dados da população.
2. Encontre o desvio de cada entrada.
3. Eleve os desvios ao quadrado.
4. Some para obter a soma dos quadrados.

## *Em símbolos*

$$\mu = \frac{\sum x}{N}$$

$$x - \mu$$

$$(x - \mu)^2$$

$$SS_x = \sum (x - \mu)^2$$

# Encontrar Variância Populacional e o Desvio Padrão

## *Em palavras*

5. Divida por  $N$  para obter a **variância populacional**.
6. Encontre a raiz quadrada para obter o **desvio padrão populacional**.

## *Em símbolos*

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N}$$

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{N}}$$

## Exercício 4.2

Uma corporação contratou 10 graduados. Os salários iniciais de cada um são demonstrados abaixo. Encontre a variância e o desvio padrão dos salários iniciais.

Salários iniciais (milhares de \$)

41 38 39 45 47 41 44 41 37 42

Lembrar  $\mu = 41,5$

# Desvio, Variância e Desvio Padrão

Variância da amostra

$$\square s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

Desvio padrão da amostra

$$\square s = \sqrt{s^2} = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

# Encontrar Variância e Desvio Padrão da Amostra

## *Em palavras*

1. Encontre a média do conjunto de dados da amostra.
2. Encontre o desvio de cada entrada.
3. Eleve cada desvio ao quadrado.
4. Some-os para obter a soma dos quadrados.

## *Em símbolos*

$$\bar{x} = \frac{\sum x}{n}$$

$$x - \bar{x}$$

$$(x - \bar{x})^2$$

$$SS_x = \sum (x - \bar{x})^2$$

# Encontrar Variância e Desvio Padrão da Amostra

## *Em palavras*

5. Divida por  $n - 1$  para obter a **variância da amostra**.
6. Encontre a raiz quadrada para obter o **desvio padrão da amostra**.

## *Em símbolos*

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

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

## Exercício 4.3

A empresa tem várias outras filiais e você planeja usar os salários iniciais da filial de Chicago para estimar os salários iniciais das populações maiores. Encontre o desvio padrão dos salários iniciais da amostra.

Salários iniciais (milhares de \$)

41 38 39 45 47 41 44 41 37 42

# Usar tecnologia para encontrar o desvio padrão

A amostra dos aluguéis de escritórios (em dólares por metro quadrado ao ano) no distrito comercial central de Miami é exibida na tabela. Use uma calculadora ou um computador para encontrar a média dos aluguéis e o desvio padrão da amostra (*Adaptado de: Cushman & Wakefield Inc.*)

Preço dos aluguéis		
35,00	33,50	37,00
23,75	26,50	31,25
36,50	40,00	32,00
39,25	37,50	34,75
37,75	37,25	36,75
27,00	35,75	26,00
37,00	29,00	40,50
24,50	33,00	38,00

# Usar tecnologia para encontrar o desvio padrão

## MINITAB

### Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev
Rental Rates	24	33.73	35.38	33.88	5.09
Variable	SE Mean	Minimum	Maximum	Q1	Q3
Rental Rates	1.04	23.75	40.50	29.56	37.44

## TI-83/84

### 1-Var Stats

$$\bar{x} = 33.72916667$$

$$\Sigma x = 809.5$$

$$\Sigma x^2 = 27899.5$$

$$S_x = 5.089373342$$

$$\sigma_x = 4.982216639$$

$$n = 24$$

: média

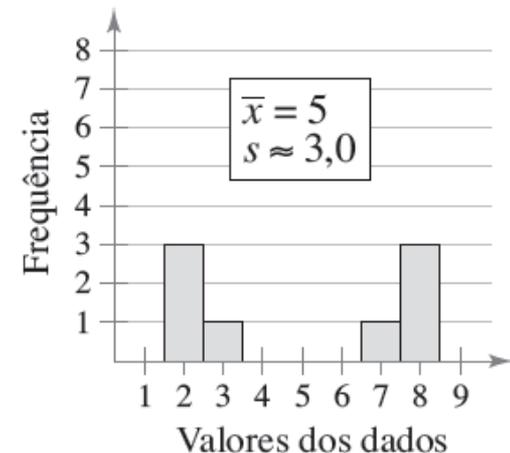
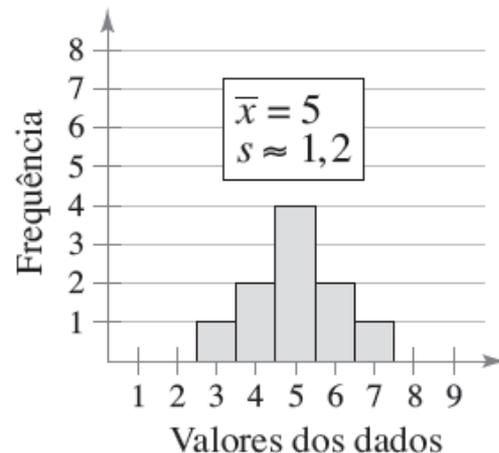
: desvio padrão

## EXCEL

	A	B
1	Média	33,72917
2	Erro padrão	1,038864
3	Média	35,375
4	Moda	37
5	Desvio padrão	5,089373
6	Variância da amostra	25,90172
7	Kurtosis	-0,74282
8	Skewness	-0,70345
9	Extensão	16,75
10	Mínimo	23,75
11	Máximo	40,5
12	Soma	809,5
13	Conta	24

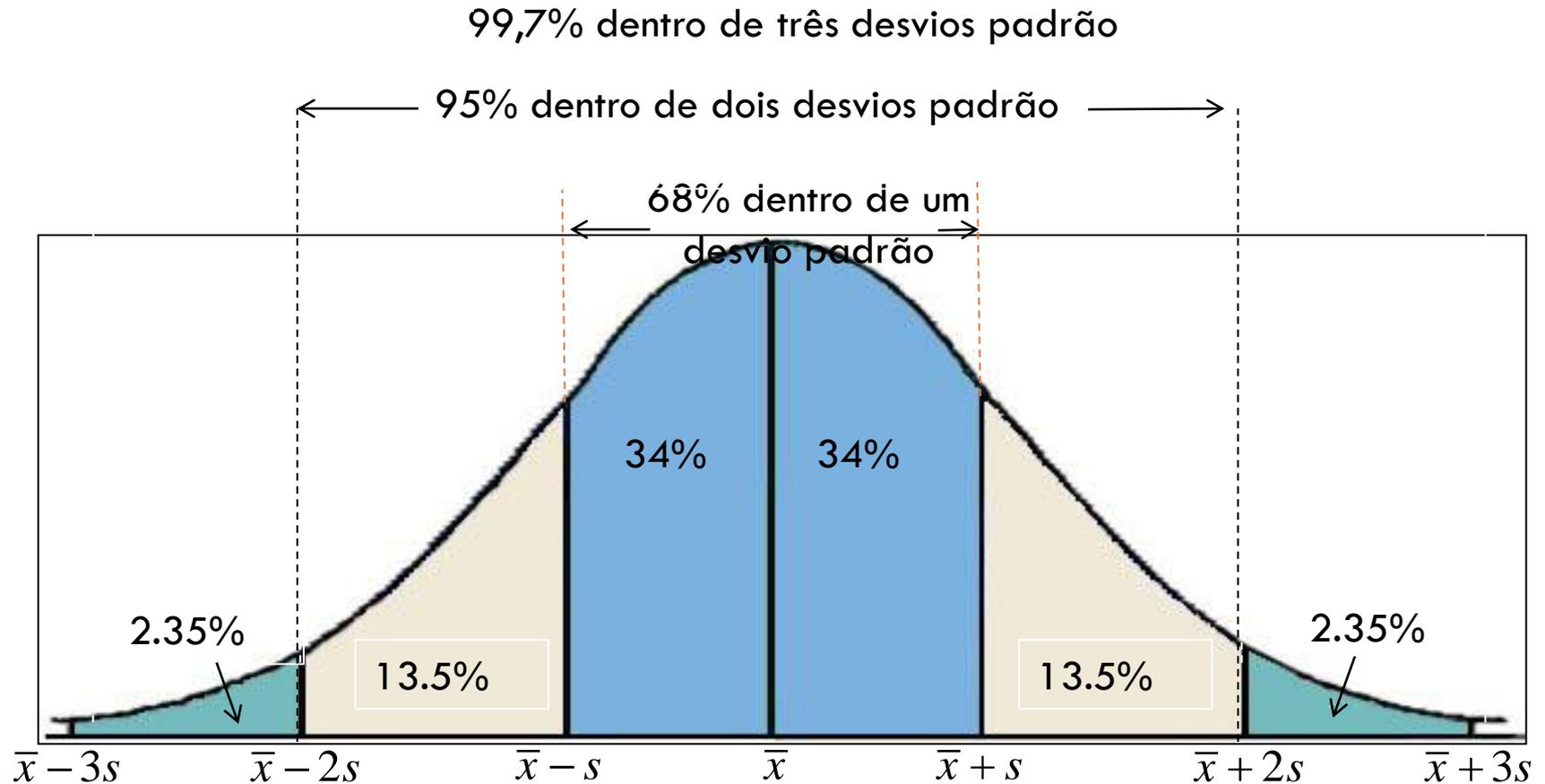
# Interpretar Desvio Padrão

- Medida de quanto uma entrada típica se desvia da média
- Quanto mais as entradas estão espalhadas, maior o desvio padrão



# Interpretar desvio Padrão

## Regra Empírica (Regra 68 – 95 – 99.7)



# Interpretar desvio Padrão

## Regra Empírica (Regra 68 – 95 – 99.7)

Para dados com uma distribuição em formato de sino (simétrica), o desvio padrão tem as seguintes características:

- Cerca de **68%** dos dados estão dentro de um desvio padrão da média
- Cerca de **95%** dos dados estão dentro de dois desvios padrão da média
- Cerca de **99,7%** dos dados estão dentro de três desvios padrão da média

## Exercício 4.4

Em uma pesquisa conduzida pelo National Center for Health Statistics, a amostragem média da altura de mulheres nos Estados Unidos (20-29 anos) era de 64 polegadas, com um desvio padrão da amostragem de 2,71 polegadas. Estime a porcentagem de mulheres que estão entre 64 e 69,42 polegadas. E entre 61,29 e 64 polegadas?

# Teorema de Chebychev

A porção de qualquer conjunto de dados postos dentro de  $k$  desvios padrão ( $k > 1$ ) da média é no mínimo

$$1 - \frac{1}{k^2}$$

# Teorema de Chebychev

A porção de qualquer conjunto de dados postos dentro de  $k$  desvios padrão ( $k > 1$ ) da média é no mínimo

$$1 - \frac{1}{k^2}$$

Se  $k = 2$ :

$$1 - \frac{1}{2^2} = \frac{3}{4} \text{ or } 75\%$$

Pelo menos 75% dos dados estão dentro de dois desvios padrão da média

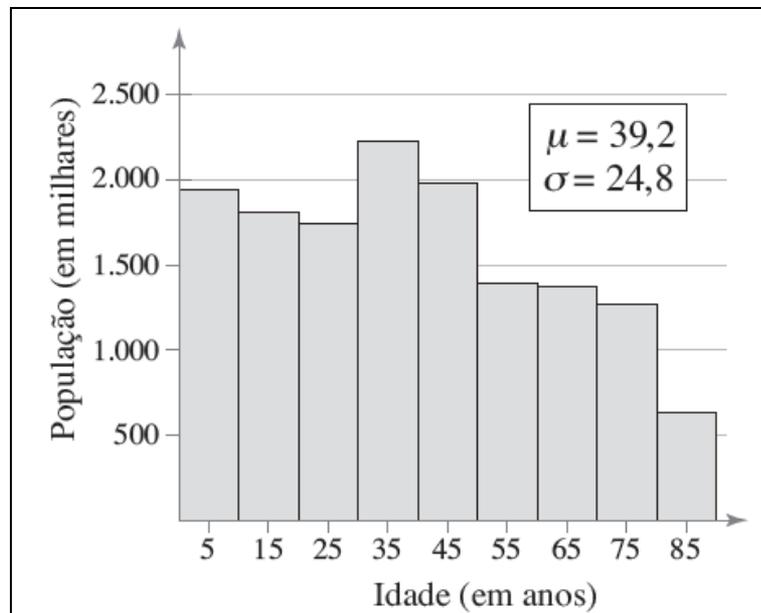
Se  $k = 3$ :

$$1 - \frac{1}{3^2} = \frac{8}{9} \text{ or } 88.9\%$$

Pelo menos 89% dos dados estão dentro de três desvios padrão da média

## Exercício 4.5

A distribuição de idade na Flórida é mostrada no histograma. Aplique o Teorema de Chebychev aos dados usando  $k = 2$ . O que se pode concluir?



## Exercício 4.6

- O Old Faithful é um famoso gêiser no Yellowstone National Park. A partir de uma amostra com  $n=32$ , a duração média das erupções do Old é de 3,32 minutos e o desvio padrão é de 1,09 minutos. Usando o teorema de Chebychev, determine quantas erupções (pelo menos) duraram entre 1,14 e 5,5 minutos. R. 24

# Desvio Padrão para Dados Agrupados

Desvio padrão de uma amostra para uma distribuição de frequência

$$s = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n - 1}}$$

Onde  $n = \sum f$   
(número de entradas no conjunto de dados)

Quando uma distribuição de frequência tem classes, estimar a média da amostra e o desvio padrão usando o ponto médio de cada classe

# Exercício 4.7

Você coleta uma amostragem aleatória do número de crianças por casa em uma região. Encontre a média e o desvio padrão da amostra do conjunto de dados

**Número de crianças em  
50 casas**

1	3	1	1	1
1	2	2	1	0
1	1	0	0	0
1	5	0	3	6
3	0	3	1	1
1	1	6	0	1
3	6	6	1	2
2	3	0	1	1
4	1	1	2	2
0	3	0	2	4

# Exercício 4.8

- 1000 adultos foram questionados sobre o quanto gastam na preparação de viagens pessoais no ano. Os resultados estão apresentados na tabela. Estime o desvio padrão amostral do conjunto de dados. R. 160,3

viajantes	Gastos antes de viagens
380	Menos de \$100
230	Entre \$100 e \$199
210	Entre \$200 e \$299
50	Entre \$300 e \$399
60	Entre \$400 e \$499
70	\$500 ou mais

# Exercício 4.9

- Listamos uma amostra de salários anuais (em milhares de dólares) para os funcionários municipais das cidades A e B.
  - A: 20,2; 26,1; 20,9; 32,1; 35,9; 23; 28,2; 31,6; 18,3
  - B: 20,9; 18,2; 20,8; 21,1; 26,5; 26,9; 24,2; 25,1; 22,2
  - Encontre a amplitude, a variância e o desvio padrão de cada conjunto de dados
  - Interprete os resultados no contexto de um cenário real

# Exercício 4.10

- O valor médio de terras e construções por acre de uma amostra de fazendas é \$1500 com desvio padrão \$200. O conjunto de dados tem distribuição em forma de sino. Estime a porcentagem de fazendas cujos valores das construções e terras por acre estejam entre \$1300 e \$1700. R. 68%

# Exercício 4.11

- O número de vitórias para diferentes times de futebol em 2014 está listado a seguir. Faça a distribuição de frequência (usando cinco classes) para o conjunto de dados. Então aproxime a média populacional e o desvio padrão populacional para o conjunto de dados.

12	10	7	6	13	8	8	4	12	8	8	6
14	9	9	2	10	9	8	5	13	8	6	3
10	8	7	4	9	8	7	5				

# Objetivos

- Determinar
  - ▣ Amplitude de um conjunto de dados
  - ▣ Variância e Desvio Padrão da População e da Amostra
- Usar a Regra Empírica e o Teorema de Chebychev para interpretar o desvio padrão
- Aproximar o desvio padrão da amostra para dados agrupados

# MEDIDAS DE POSIÇÃO

Aula 4

# Objetivos

- Determinar Medidas de Posição
  - Quartis
  - Amplitude interquartil
  - Pontuação Padrão (z-escore)
- Criar um gráfico de caixa-e-bigodes

# Quartis

- **Quartis:** números que dividem um conjunto de dados ordenados em quatro partes iguais
  - **Primeiro quartil,  $Q_1$** 
    - 25% dos dados possuem valores menores ou iguais a  $Q_1$
  - **Segundo quartil,  $Q_2$** 
    - 50% dos dados possuem valores menores ou iguais a  $Q_2$  (mediana)
  - **Terceiro quartil,  $Q_3$** 
    - 75% dos dados dados possuem valores menores ou iguais a  $Q_3$

# Encontrar quartis

As pontuações dos testes de 15 empregados matriculados em um curso de primeiros socorros são listadas. Encontre o primeiro, o segundo e o terceiro quartil das pontuações dos testes.

13 9 18 15 14 21 7 10 11 20 5 18 37 16 17

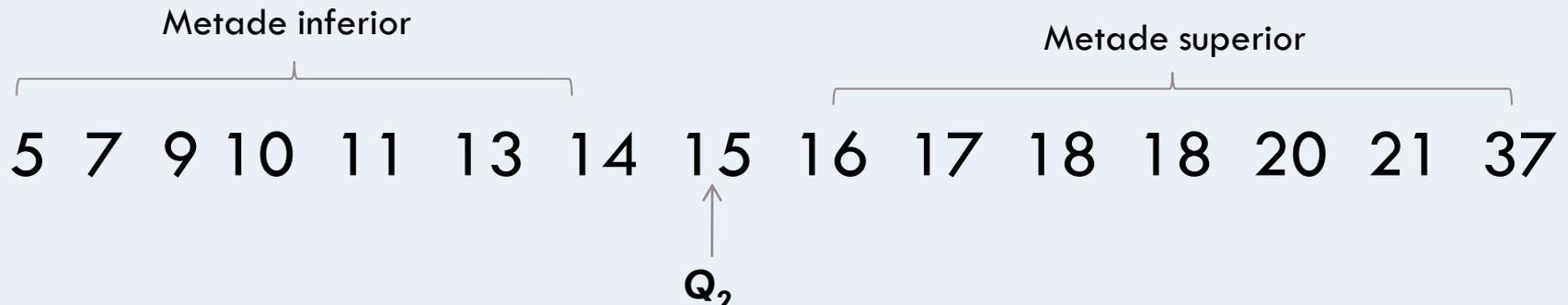
# Encontrar quartis

As pontuações dos testes de 15 empregados matriculados em um curso de primeiros socorros são listadas. Encontre o primeiro, o segundo e o terceiro quartil das pontuações dos testes.

13 9 18 15 14 21 7 10 11 20 5 18 37 16 17

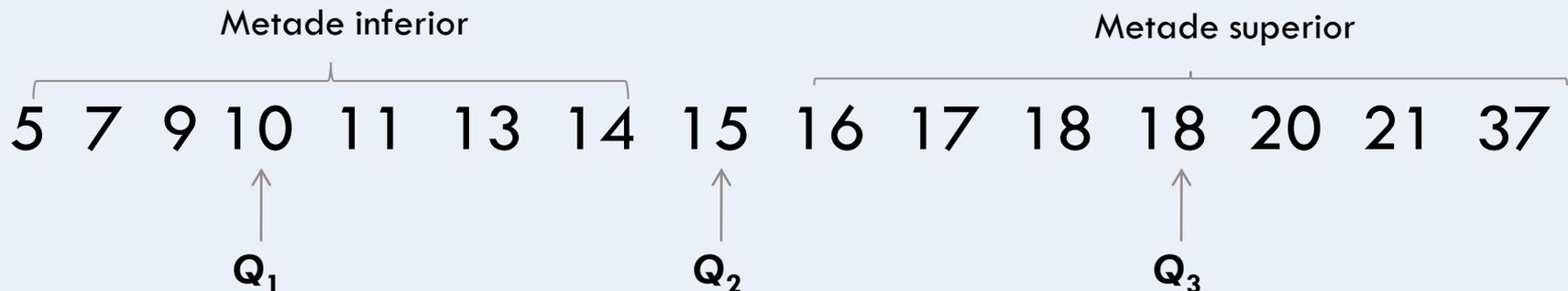
## Solução:

- $Q_2$  divide o conjunto de dados em duas metades



# Encontrar Quartis

O primeiro e o terceiro quartis são as medianas das metades inferior e superior do conjunto de dados



Cerca 25% dos funcionários obteve nota 10 ou menos; cerca 50% obteve 15 ou menos; e cerca 75% obteve 18 ou menos.

## Exercício 4.12

As pontuações dos testes de 14 empregados matriculados em um curso de primeiros socorros são listadas. Encontre o primeiro, o segundo e o terceiro quartil das pontuações dos testes.

9 18 15 14 21 7 10 11 20 5 18 37 16 17

# Exercício 4.13

As pontuações dos testes de 12 empregados matriculados em um curso de primeiros socorros são listadas. Encontre o primeiro, o segundo e o terceiro quartil das pontuações dos testes.

9 18 15 14 7 10 11 20 5 18 16 17

# Encontrar Amplitude Interquartil

## Amplitude Interquartil (IQR)

- Diferença entre o terceiro e o primeiro quartis
- $IQR = Q_3 - Q_1$

# Encontrar a Amplitude Interquartil

Encontre a amplitude interquartil das notas dos testes.

Lembre-se:  $Q_1 = 10$ ,  $Q_2 = 15$  e  $Q_3 = 18$

**Solução:**

- $IQR = Q_3 - Q_1 = 18 - 10 = 8$

As notas do teste na porção da metade do conjunto de dados variam no máximo em 8 pontos.

# Que valores numéricos são importantes para descrever o conjunto de dados?

5 7 9 10 11 13 14 15 16 17 18 18 20 21 37

Metade inferior

Metade superior



$Q_1$



$Q_2$



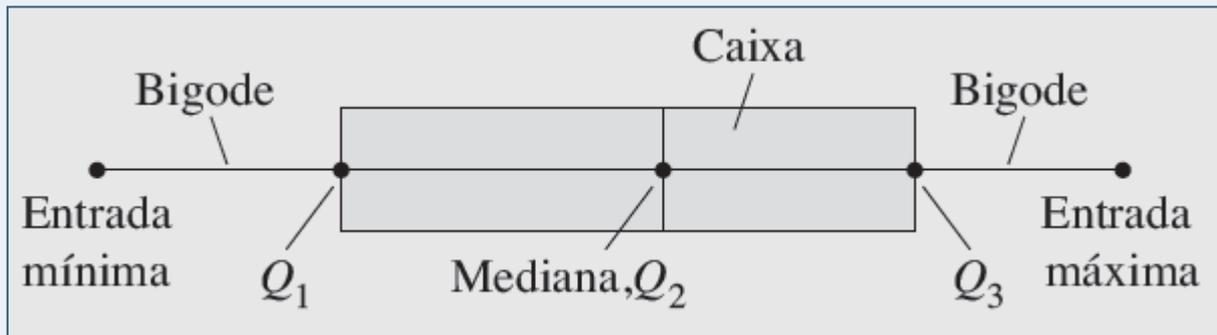
$Q_3$

# Gráfico de caixa-e-bigodes

- Ferramenta exploratória de análise de dados
- Destaca qualidades importantes do conjunto de dados
- Requer Sumário de Cinco Números:
  - Entrada mínima
  - Primeiro quartil  $Q_1$
  - Mediana  $Q_2$
  - Terceiro quartil  $Q_3$
  - Entrada máxima

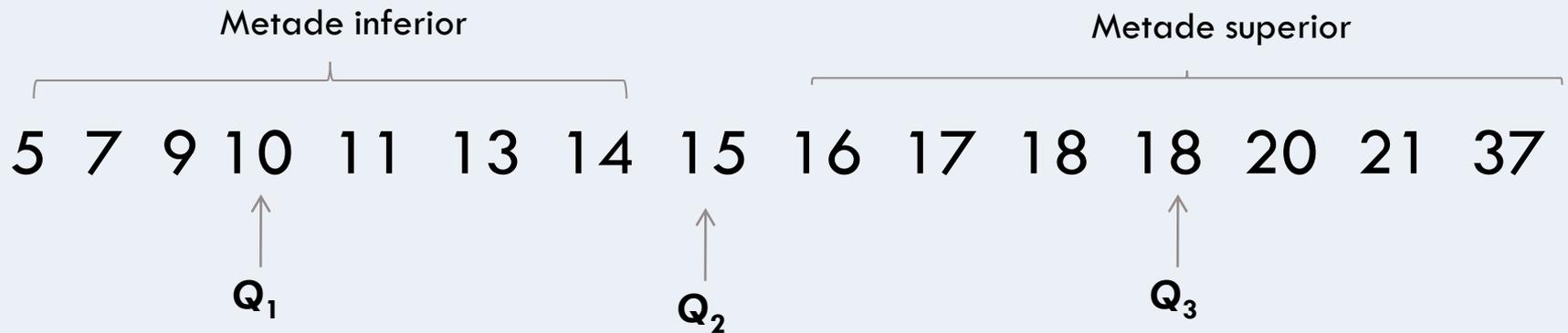
# Desenhar Gráfico de Caixa-e-Bigodes

- ❖ Encontre o sumário de cinco números do conjunto de dados.
- ❖ Construa uma escala horizontal que cubra a variância dos dados.
- ❖ Ponha os cinco números acima da escala horizontal.
- ❖ Desenhe uma caixa acima da escala horizontal de  $Q_1$  até  $Q_3$  e desenhe uma linha vertical na caixa em  $Q_2$ .
- ❖ Desenhe bigodes saindo da caixa para as entradas mínima e máxima.



# Desenhar Gráfico de caixa-e-bigodes

Desenhe um gráfico de caixa-e-bigodes que represente as 15 pontuações dos testes.

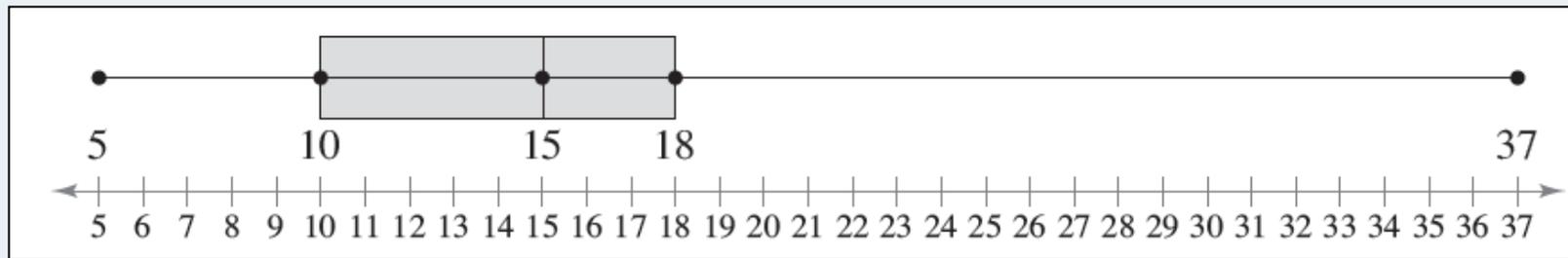


# Desenhar Gráfico de caixa-e-bigodes

Desenhe um gráfico de caixa-e-bigodes que represente as 15 pontuações dos testes.

Lembre-se: Mín. = 5  $Q_1 = 10$   $Q_2 = 15$   $Q_3 = 18$

Máx. = 37



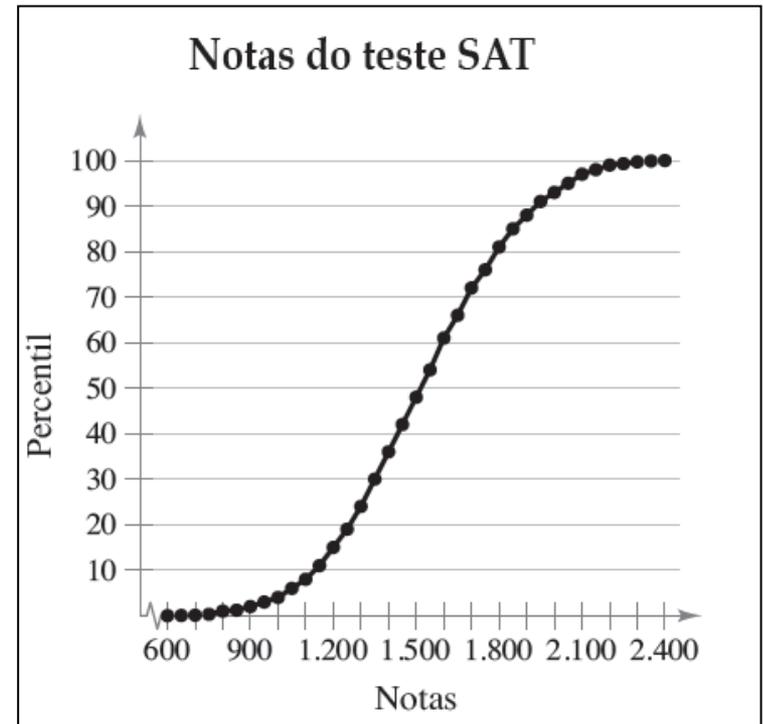
Cerca de metade das notas estão entre 10 e 18. Olhando para o comprimento do bigode direito, pode-se concluir que 37 é um possível valor discrepante.

# Percentis e outros fractis

Fractis	Sumário	Símbolos
Quartis	Divide os dados em 4 partes iguais	$Q_1, Q_2, Q_3$
Decis	Divide os dados em 10 partes iguais	$D_1, D_2, D_3, \dots, D_9$
Percentis	Divide os dados em 100 partes iguais	$P_1, P_2, P_3, \dots, P_{99}$

# Exercício 4.14

A ogiva representa a distribuição de frequência cumulativa para as provas do SAT (vestibular dos EUA) de estudantes em um ano recente. Qual nota representa o 72º percentil? Como você deve interpretar isso? (*Fonte: College Board Online.*)



# Pontuação padrão

## Pontuação padrão (escore z)

- Representa o número de desvios padrão que um dado valor  $x$  está a partir da média  $\mu$

$$z = \frac{\text{valor} - \text{média}}{\text{desvio padrão}} = \frac{x - \mu}{\sigma}$$

## Exercício 4.15

- A velocidade média de veículos em uma reta de uma rodovia é de 56 milhas por hora, com desvio padrão de 4 milhas por hora. Você mede a velocidade de três carros que estão passando pela reta da rodovia como 62 milhas por hora, 47 milhas por hora e 56 milhas por hora. Encontre o z-escore que corresponde a cada velocidade. O que podemos concluir?

# Exercício 4.16

Em 2007, o ator Forest Whitaker ganhou o Oscar de melhor ator, aos 45 anos de idade, por sua atuação no filme *O Último Rei da Escócia*. A atriz Helen Mirren ganhou o prêmio de melhor atriz aos 61 anos por seu papel em *A Rainha*. A idade média para todos os vencedores do prêmio de melhor ator é 43,7, com desvio padrão de 8,8. A idade média para as vencedoras do prêmio de melhor atriz é 36, com desvio padrão de 11,5. Encontre o z-escore que corresponda à idade de cada ator ou atriz. Depois, compare os resultados.



# Objetivos

- Determinar
  - ▣ Quartis de um conjunto de dados
  - ▣ Amplitude interquartil de um conjunto de dados
  - ▣ Escore Padrão (z-escore)
- Criar um gráfico de caixa-e-bigodes
- Interpretar outros fractis

# Exercício 4.17

## Notas de Estatística

60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	68	97	70	64	70	70
58	78	89	44	55	85	82	83	72	77
72	86	50	94	92	80	91	75	76	78

Média = 74,92  
Desvio Padrão = 14,20

Posso afirmar que as notas de Estatística seguem uma distribuição é normal?

# Exercício 4.18

Resultado do teste de aptidão física por ordem de classificação

1	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6
8	9	9	9	9	9	9	10	10	11	12	12	12	13	14	14
14	15	15	15	15	16	16	16	17	17	17	17	18	18	18	18
19	19	19	19	20	20	20	21	21	21	22	22	22	23	24	24

Teste aplicado nos alunos do terceiro ano do ensino fundamental da Escola X

Posso afirmar que os resultados seguem uma distribuição normal?