Aula 2

Distribuições de freqüência e seus gráficos

Professor: Phelipe Fabres

Aula 2

Exercício 1.

A direção de um parque contratou uma equipe de pesquisadores para coletar algumas informações sobre seus frequentadores. Os cem entrevistados responderam a questões sobre **sexo, idade, número de vezes por semana que vão ao parque, período de visita** (manhã, tarde, começo da noite), **tempo de permanência e quantia gasta nas dependências do parque**. Cada um desses objetos de estudo corresponde a uma variável. Classifique as variáveis quanto ao tipo (qualitativa ou quantitativa).

Aula 2

Exercício 1 - Solução.

A direção de um parque contratou uma equipe de pesquisadores para coletar algumas informações sobre seus frequentadores. Os cem entrevistados responderam a questões sobre sexo, idade, número de vezes por semana que vão ao parque, período de visita (manhã, tarde, começo da noite), tempo de permanência e quantia gasta nas dependências do parque. Cada um desses objetos de estudo corresponde a uma variável. Classifique as variáveis quanto ao tipo (qualitativa ou quantitativa).

São qualitativas: sexo e período de visita. As demais são quantitativas.

Aula 2

Exercício 2.

Uma concessionária de automóveis tem cadastrados 3 500 clientes e fez uma pesquisa sobre a preferência de compra em relação a "cor" (branco, vermelho ou azul), "preço", "número de portas" (duas ou quatro) e "estado de conservação" (novo ou usado). Foram consultados 210 clientes. Diante dessas informações, responda:

a) Qual é a população e qual é a amostra dessa pesquisa?
b) Quais são as variáveis e qual é o tipo de cada uma?
c) Quais os possíveis valores da variável "cor" nessa pesquisa?

Aula 2

Exercício 2 - Solução.

a) População: conjunto formado pela totalidade dos clientes (3 500); Amostra: conjunto formado pelos clientes consultados (250);
b) Cor: qualitativa nominal; preço: quantitativa contínua; número de portas: quantitativa discreta; estado de conservação: qualitativa ordinal;
c) Branca, vermelha e azul.

Aula 2

Exercício 3.

Indique se as populações a seguir devem ser consideradas finitas ou infinitas:

- Todos os eleitores inscritos do estado da Califórnia.

- Todos os aparelhos de televisão que poderiam ser produzidos pelo parque industrial da TV-M Company, em Allentown, Pensilvânia.

- Todos os pedidos que poderiam ser processados por uma empresa de encomenda postal.

- Todas as chamadas telefônicas de emergência que poderiam ser feitas a uma delegacia de polícia local.

- Todos os componentes que a Fibercon, Inc., produziu no segundo turno de trabalho no dia 17 de maio.

Amostra sistemática

Selecione aleatoriamente um valor inicial. Depois, escolha os membros da amostra a intervalos regulares.

Digamos que vamos selecionar cada *k*° membro. Nesse caso, $k = 5$. Logo, cada 5º membro da população será selecionado.

Amostra por agrupamento

Divida a população em unidades individuais ou grupos. Em seguida, selecione aleatoriamente uma ou mais unidades. A amostra consistirá em todos os membros da(s) unidade(s) selecionada(s).

Amostra aleatória estratificada

Divida a população em grupos (estratos) e selecione uma amostra aleatória de cada grupo. Os estratos podem ser faixas etárias, gêneros ou graus de escolaridade, por exemplo.

Amostra Aleatória

Amostra aleatória: Cada membro da população tem a mesma chance de ser selecionado.

Amostra aleatória simples: Todas as amostras de mesmo tamanho são igualmente prováveis.

Atribua um número a cada membro da população.
Números aleatórios podem ser gerados por uma tabela apropriada, por um software ou ainda por uma calculadora.
Os dados dos membros da população que correspondam a tais números passarão a ser os membros da amostra.

Amostra - Exemplo

Exemplo: Você está realizando um estudo para determinar a opinião dos estudantes em sua faculdade sobre a pesquisa de células-tronco. Identifique a técnica de amostragem que você usaria se selecionasse as amostras listadas.

1. Você seleciona uma classe aleatoriamente e questiona cada aluno da classe.
2. Você divide a população de estudantes com relação às graduações, seleciona aleatóriamente e questiona alguns de cada curso de graduação.
3. Você designa um número para cada aluno e gera números aleatóriamente. Então, questiona cada estudante cujo número é selecionado aleatóriamente.

Amostra – Exemplo - Resposta

1. Amostra por agrupamento: Você questiona cada aluno da classe, que é um subgrupo da população.
2. Amostra estratificada: Os estudantes são divididos em estratos(graduações).
3. Amostra aleatória simples: Todo mundo tem a mesma chance de ser selecionado.

- Exercícios do PLT, páginas 26 e 27.

Distribuições de freqüência

• **Distribuição de frequência** é uma tabela que mostra classes ou intervalos de entradas de dados com uma contagem do número de entrada de cada classe.

• A frequência *f* de uma classe é o número de entrada de dados em uma classe.

Distribuições de freqüência

Minutos gastos ao telefone

102	124	108	86	103	82
71	104	112	118	87	95
103	116	85	122	87	100
105	97	107	67	78	125
109	99	105	99	101	92

Faça uma tabela de distribuição de freqüência com cinco classes.

Valores-chave:

Valor mínimo = 67
Valor máximo = 125

Passos para construir uma distribuição de freqüência

1. Decida o número de classes, que deve ficar entre 5 e 15.

   (Para este problema use 5.)

2. Calcule a amplitude das classes.

   Primeiro calcule: amplitude total = valor máximo – mínimo. Em seguida, divida o resultado pelo número de classes. Por fim, arredonde até o próximo número conveniente. (125 – 67)/5 = 11,6 (arredondado para 12)

3. Calcule os limites das classes.

   O limite inferior da classe é o valor mais baixo que pertence a ela e o limite superior é o mais alto. Use o valor mínimo (67) como limite inferior da primeira classe.

4. Marque um risco | em cada entrada de dado na classe apropriada.

   Quando todos os valores estiverem marcados, conte os riscos em cada classe para determinar a freqüência dessa classe.

Construa uma distribuição de freqüência

Mínimo = 67, Máximo = 125
Número de classes = 5
Amplitude de classe = 12

<table>
<tr><th>Classe</th><th>Limites</th><th>Riscos</th><th>f</th></tr>
<tr><td>67</td><td>78</td><td>|||</td><td>3</td></tr>
<tr><td>79</td><td>90</td><td>~~||||~~</td><td>5</td></tr>
<tr><td>91</td><td>102</td><td>~~||||~~ |||</td><td>8</td></tr>
<tr><td>103</td><td>114</td><td>~~||||~~ ||||</td><td>9</td></tr>
<tr><td>115</td><td>126</td><td>~~||||~~</td><td>5</td></tr>
<tr><td></td><td></td><td></td><td>$\Sigma f = 30$</td></tr>
</table>

Faça primeiro todos os limites inferiores.

Outras informações

Ponto médio: (limite inferior + limite superior)/2

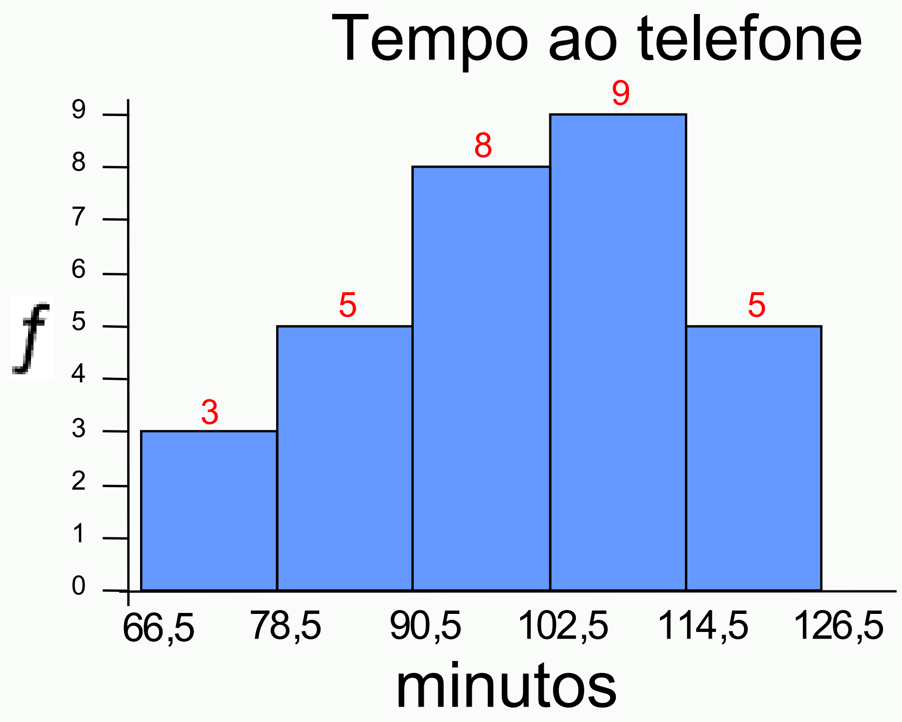
Freqüência relativa: freqüência da classe/freqüência total

Freqüência cumulativa: número de valores em determinada classe ou abaixo dela

Classe	f	Ponto médio	Freqüência relativa	Freqüência cumulativa
67–78	3	72,5	0,10	3
79–90	5	84,5	0,17	8
91–102	8	96,5	0,27	16
103–114	9	108,5	0,30	25
115–126	5	120,5	0,17	30

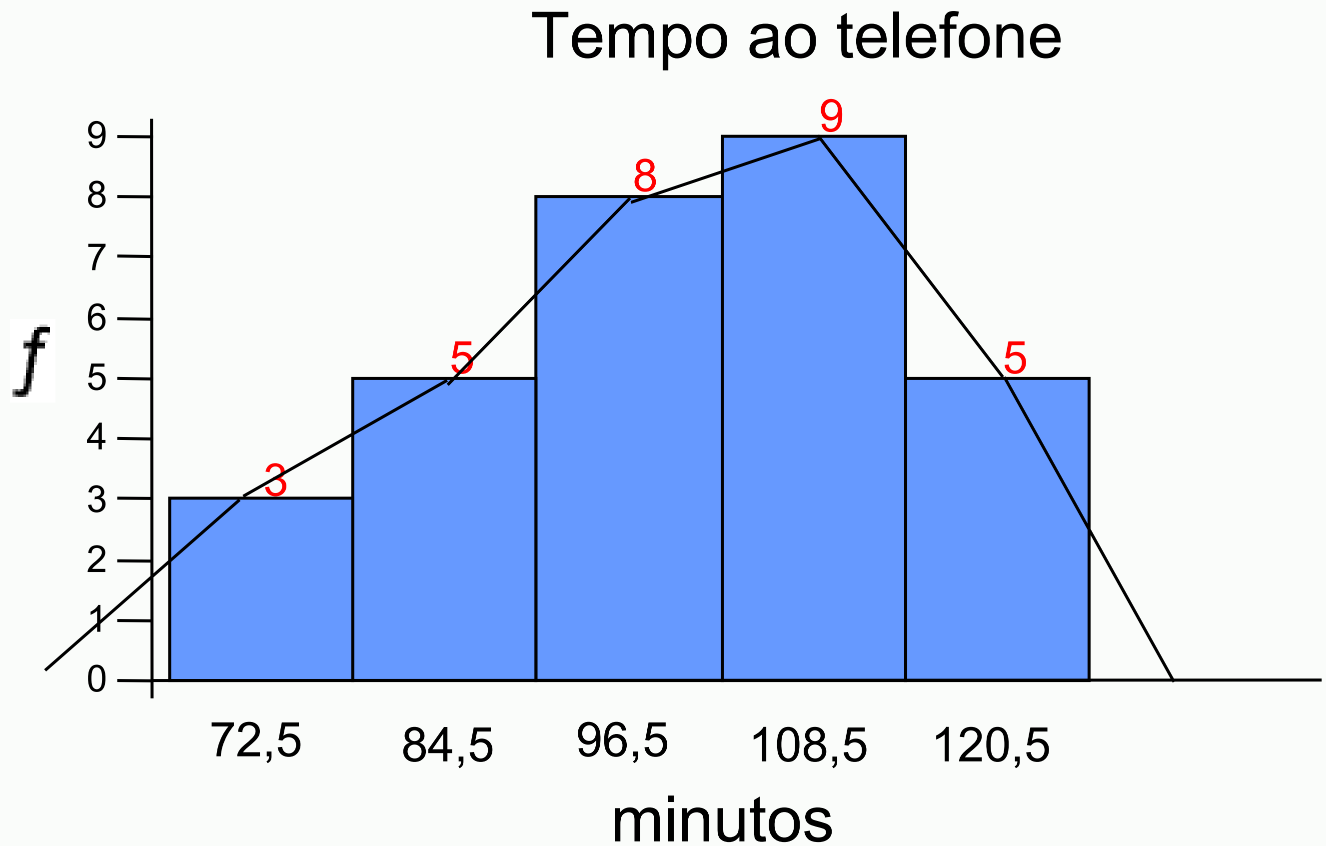
Histograma de freqüência

Classe	*f*	Fronteiras
67–78	3	66,5–78,5
79–90	5	78,5–90,5
91–102	8	90,5–102,5
103–114	9	102,5–114,5
115–126	5	114,5–126,5



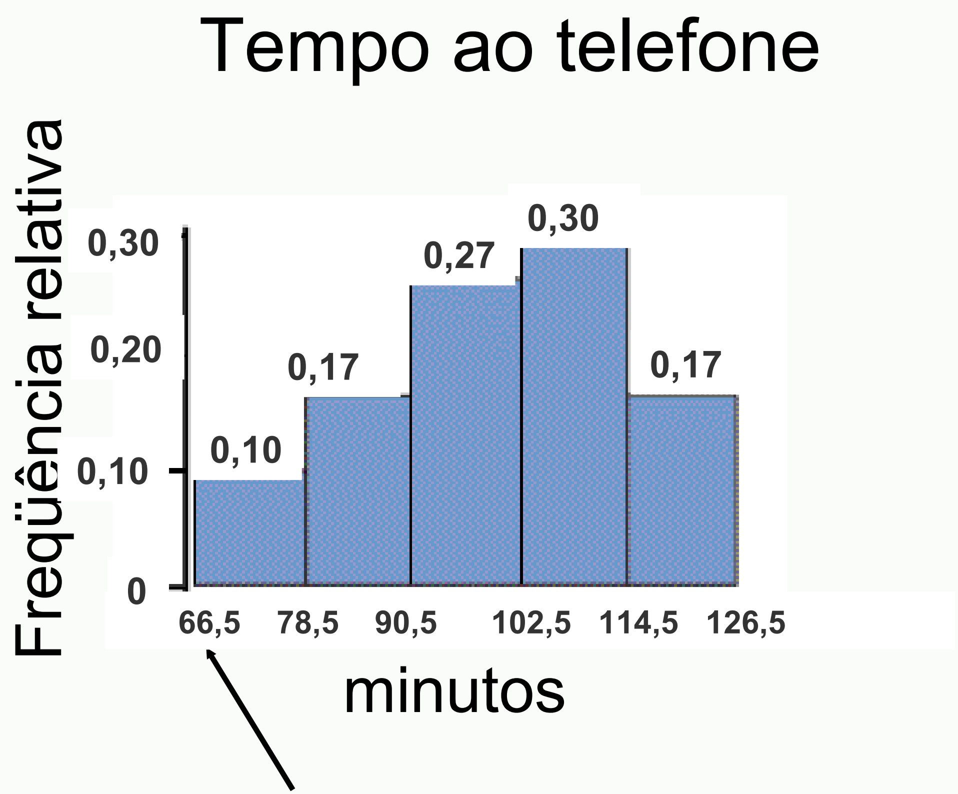
Polígono de freqüência

Classe	f
67–78	3
79–90	5
91–102	8
103–114	9
115–126	5



Marque o ponto médio no topo de cada barra. Conecte os pontos médios consecutivos. Estenda o polígono até os eixos.

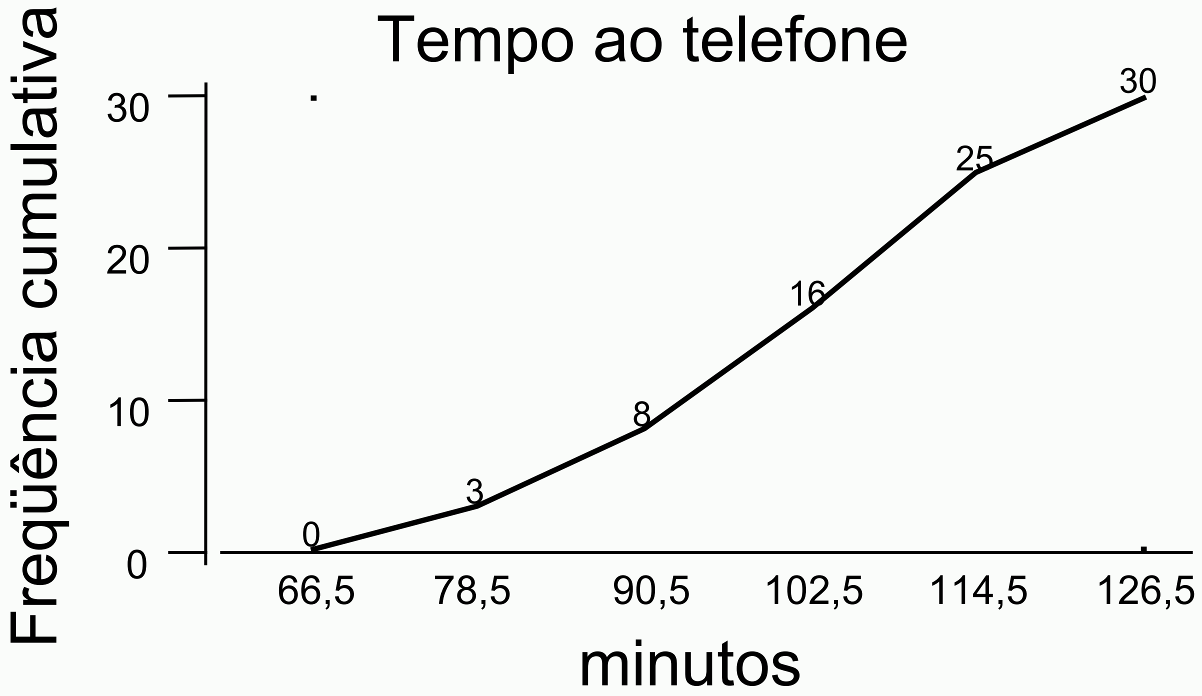
Histograma de freqüência relativa



A escala vertical mede as freqüências relativas.

Gráfico de freqüência cumulativa (ogiva)

Um gráfico de freqüência cumulativa (ou ogiva) mostra o número de valores, em um conjunto de dados, que são iguais ou inferiores a um dado valor *x*.



Seção 2.2

Mais gráficos e representações

Plote tronco-e-folhas

Se o valor mais baixo é 67 e o mais alto é 125, o tronco vai de 6 a 12.

102 124 108 86 103 82

Tronco	Folhas		
6 |			
7 |			
8 |	6	2	
9 |			
10 |	2	8	3
11 |			
12 |	4		

Para ver a representação completa, avance até o próximo slide.

Plote tronco-e-folhas

Chave: 6|7 significa 67

6	7
7	1 8
8	2 5 6 7 7
9	2 5 7 9 9
10	0 1 2 3 3 4 5 5 7 8 9
11	2 6 8
12	2 4 5

Plote tronco-e-folhas com duas linhas por tronco

Chave: 6|7 significa 67

Dígitos da 1ª linha 0 1 2 3 4 → 8 | 2

Dígitos da 2ª linha 5 6 7 8 9 → 8 | 5 6 7 7

Dígitos da 1ª linha 0 1 2 3 4 → 11 | 2

Dígitos da 2ª linha 5 6 7 8 9 → 11 | 6 8

6	7
7	1
7	8
8	2
8	5 6 7 7
9	2
9	5 7 9 9
10	0 1 2 3 3 4
10	5 5 7 8 9
11	2
11	6 8
12	2 4
12	5

Plote de pontos

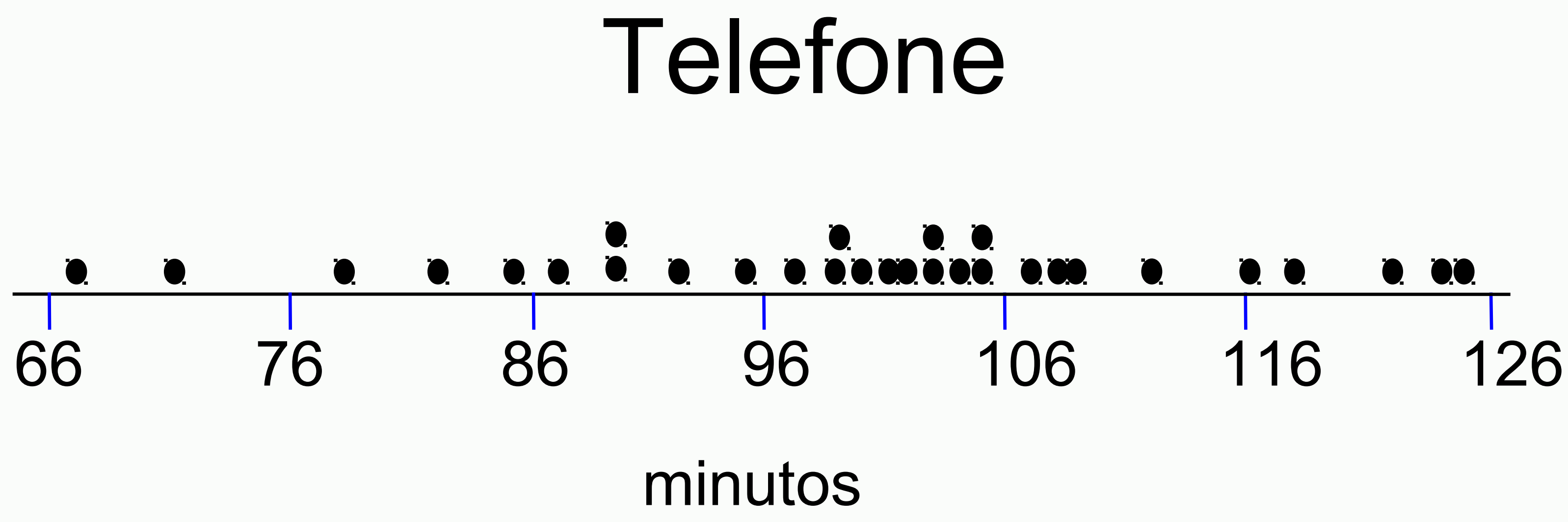


Diagrama de pizza

- É usado para descrever as partes de um todo.
- O ângulo central para cada segmento é:

$$\frac{\text{Número na categoria}}{\text{Número total}} \; 360^\circ$$

O orçamento da Nasa (em bilhões de dólares) dividido em três categorias

	Bilhões de US$
Vôo espacial humano	5,7
Tecnologia	5,9
Apoio às missões	2,7

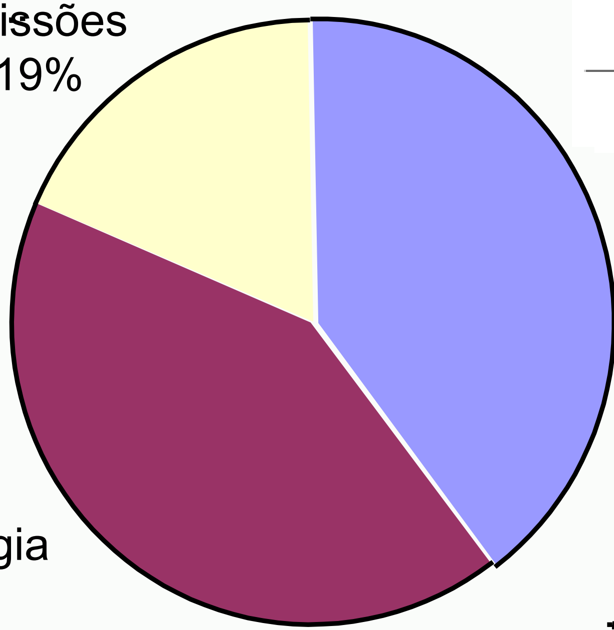
Construa um diagrama de pizza para esses dados.

Diagrama de pizza

	Bilhões de US$	Graus
Vôo espacial humano	5,7	143
Tecnologia	5,9	149
Apoio às missões	2,7	68
Total	14,3	360

$$\frac{5{,}7}{14{,}3} \times 360^\circ = 143^\circ$$

$$\frac{5{,}9}{14{,}3} \times 360^\circ = 149^\circ$$

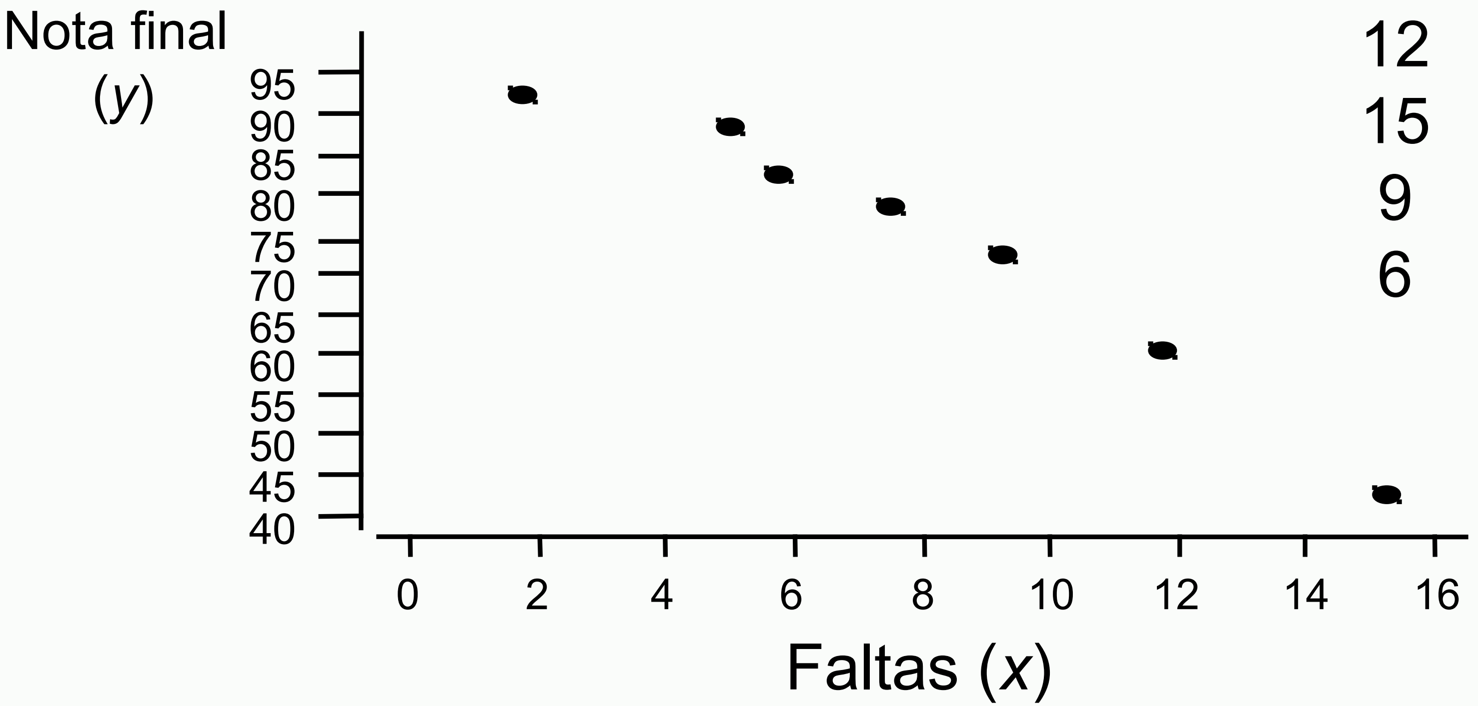


Orçamento da Nasa

(em bilhões de dólares)

Mapa de dispersão

Faltas x	Nota final y
8	78
2	92
5	90
12	58
15	43
9	74
6	81



Seção 2.3

Medidas de tendência central

Medidas de tendência central

Média: A soma de todos os valores dividida pelo número de valores.

Em uma população: Em uma amostra:

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

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

Mediana: Ponto que tem um número igual de valores acima e abaixo de si.

Moda: O valor com a maior freqüência.

Um instrutor registra a média de faltas de seus alunos em determinado semestre. Em uma amostra aleatória, os dados são:

2 4 2 0 40 2 4 3 6

Calcule a média, a mediana e a moda.

Média: $\bar{x} = \frac{\Sigma x}{n}$ $\quad \Sigma x = 63 \quad n = 9 \quad \bar{x} = \frac{63}{9} = 7$

Mediana: Ordene os dados.

0 2 2 2 $\boxed{3}$ 4 4 6 40

O valor que fica no meio é 3, logo a mediana é 3.

Moda: A moda é 2, pois esse é o valor que ocorre mais vezes.

Suponha que o aluno com 40 faltas abandone o curso. Calcule a média, a mediana e a moda dos valores restantes. Compare o efeito da mudança para cada tipo de média.

2 4 2 0 2 4 3 6

Calcule a média, a mediana e a moda.

Média: $\bar{x} = \frac{\Sigma x}{n}$ $\Sigma x = 23$ $n = 8$ $\bar{x} = \frac{23}{8} = \mathbf{2{,}875}$

Mediana: Coloque os dados em ordem.

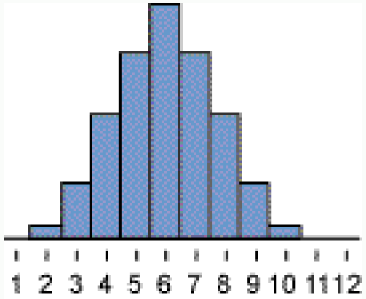
0 2 2 $\boxed{2 \quad 3}$ 4 4 6

Os valores que ficaram no meio são 2 e 3, logo a mediana é 2,5.

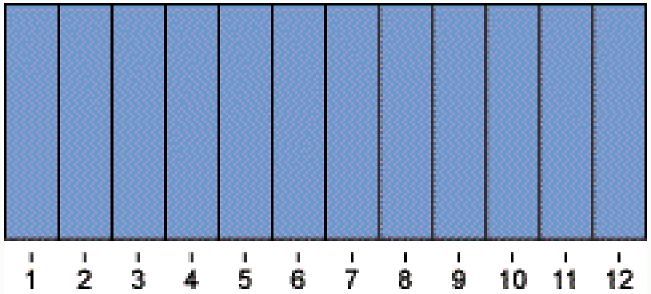
Moda: A moda é 2, pois esse é o valor que ocorre mais vezes.

Aspecto das distribuições

Simétrica

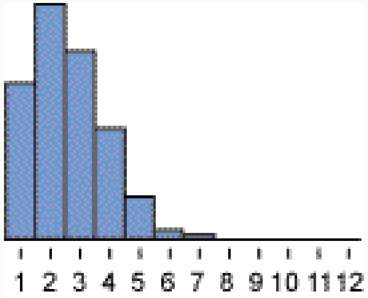


Uniforme



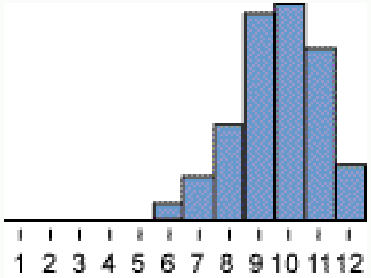
Média = Mediana

Anti-simétrica à direita



Média > Mediana

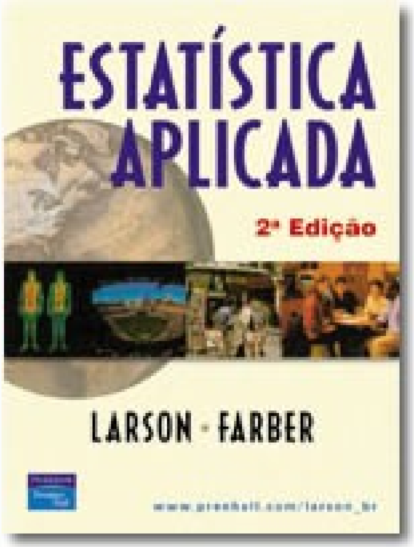
Anti-simétrica à esquerda



Média < Mediana

Seção 2.4

Medidas de variação



Dois conjuntos de dados

O preço de fechamento atingido por dois pacotes de ações foi registrado em dez sextas-feiras consecutivas. Calcule a média, a mediana e a moda de cada pacote.

Ações A		**Ações B**
56	33	
56	42	
57	48	
58	52	
61	57	
63	67	
63	67	
67	77	
67	82	
67	90	

Ações A:

Média = 61,5
Mediana = 62
Moda = 67

Ações B:

Média = 61,5
Mediana = 62
Moda = 67

Medidas de variação

Amplitude total = valor máximo – valor mínimo

Amplitude total de A = 67 – 56 = US$ 11

Amplitude total de B = 90 – 33 = US$ 57

A amplitude total é fácil de calcular porque só usa dois números do conjunto de dados.

Medidas de variação

Para aprender a calcular medidas de variação que usem todo e qualquer valor do conjunto de dados, primeiro você precisa saber o que é um desvio.

O **desvio** de cada valor *x* é a diferença entre o valor de *x* e a média do conjunto de dados.

Em uma **população**, o desvio de cada valor *x* é: $x - \mu$

Em uma **amostra**, o desvio de cada valor *x* é: $x - \bar{x}$

Desvios

Ações A	Desvio		
56	– 5,5	←	56 – 61,5
56	– 5,5	←	56 – 61,5
57	– 4,5	←	57 – 61,5
58	– 3,5	←	58 – 61,5
61	– 0,5		
63	1,5		
63	1,5		
67	5,5		
67	5,5		
67	5,5		

$\mu = 61{,}5$

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

A soma dos desvios é sempre zero.

Variância populacional

Variância populacional: a soma dos quadrados dos desvios, dividida por *N*.

x		
56	– 5,5	30,25
56	– 5,5	30,25
57	– 4,5	20,25
58	– 3,5	12,25
61	– 0,5	0,25
63	1,5	2,25
63	1,5	2,25
67	5,5	30,25
67	5,5	30,25
67	5,5	30,25
		188,50

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

$$\sigma^2 = \frac{188{,}50}{10} = 18{,}85$$

Soma dos quadrados

Desvio padrão populacional

Desvio padrão populacional: a raiz quadrada da variância populacional.

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{18{,}85} = 4{,}34$$

O desvio padrão populacional é US$ 4,34.

Variância e desvio padrão amostrais

Para calcular uma variância amostral, divida a soma do quadrados por n – 1.

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad s^2 = \frac{188{,}50}{9} = 20{,}94$$

Para calcular o desvio padrão amostral, *s*, tire a raiz quadrada da variância amostral.

$$s = \sqrt{s^2}$$

$$s = \sqrt{20{,}94} = 4{,}58$$

Resumo

Amplitude total = valor máximo – valor mínimo

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

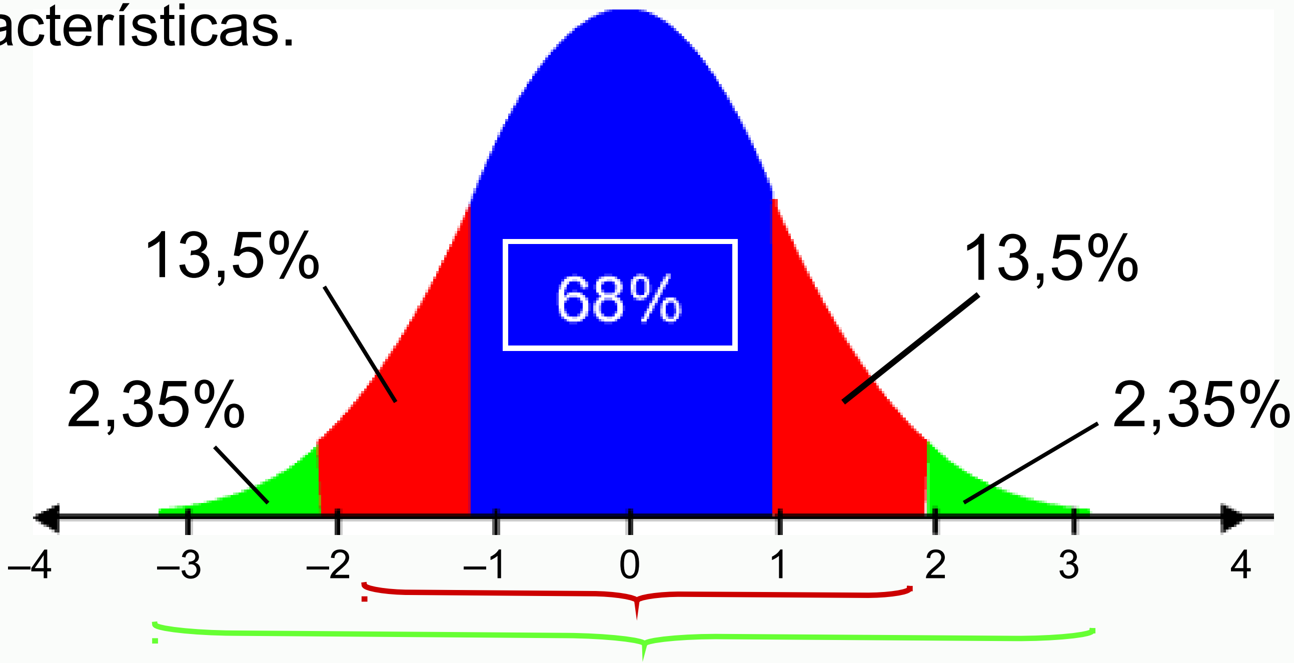
Desvio padrão populacional $$\sigma = \sqrt{\sigma^2}$$

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

Desvio padrão amostral $$s = \sqrt{s^2}$$

Regra Empírica (68-95-99,7%)

Dados com distribuição **simétrica na forma de sino** têm as seguintes características.



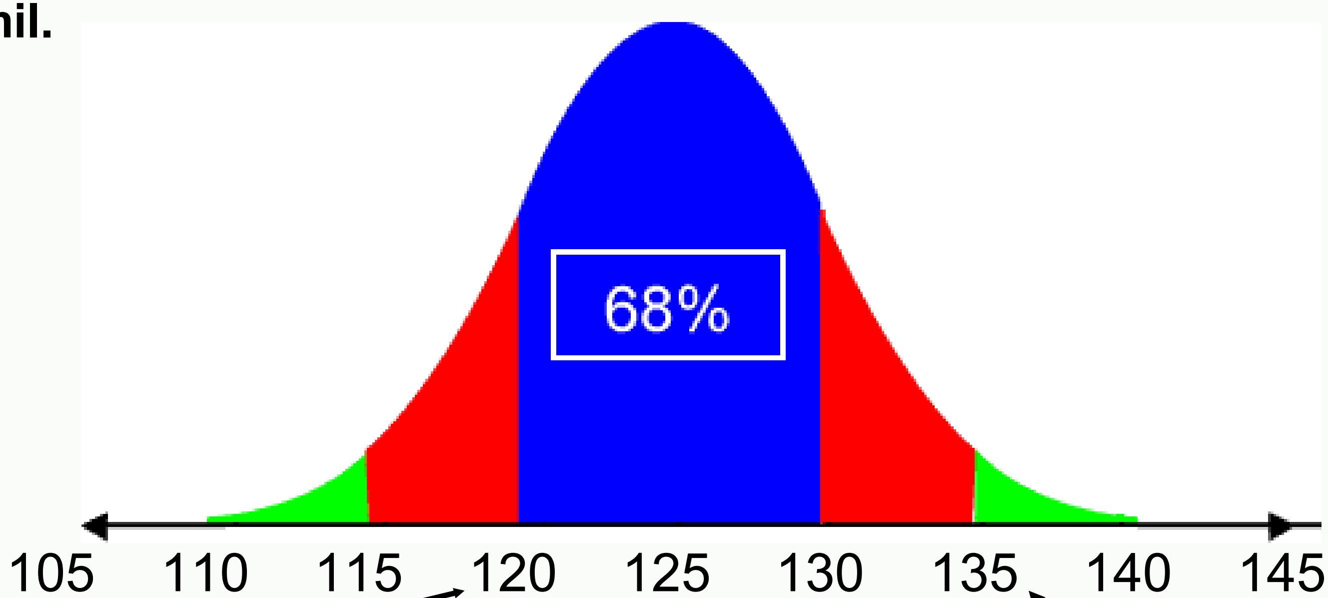
Cerca de 68% dos dados estão a até 1 desvio padrão da média.

Cerca de 95% dos dados estão a até 2 desvios padrão da média.

Cerca de 99,7% dos dados estão a até 3 desvios padrão da média.

Como usar a Regra Empírica

O valor médio das casas de determinada rua é de US$ 125 mil,com um desvio padrão de US$5 mil. O conjunto de dados tem uma distribuição na forma de sino. Estime o porcentual de casas que custam entre US$ 120 e US$ 135 mil.



US$ 120 mil fica 1 desvio padrão abaixo da média e US$ 135 mil fica 2 desvios padrão acima da média.

68% + 13,5% = 81,5%

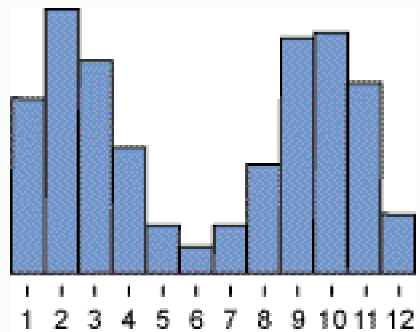
Logo, 81,5% das casas custam entre US$ 120 e US$ 135 mil.

Teorema de Chebychev

Em *qualquer* distribuição, independentemente de sua forma, a porção de dados que está dentro de *k* desvios padrão (*k* > 1) da média é *pelo menos* 1 – 1/*k*2.

$\mu = 6$

$\sigma = 3{,}84$



Se *k* = 2, ***pelo menos*** 1 – 1/4 = ¾ = 75% dos dados estão dentro de dois desvios padrão da média.

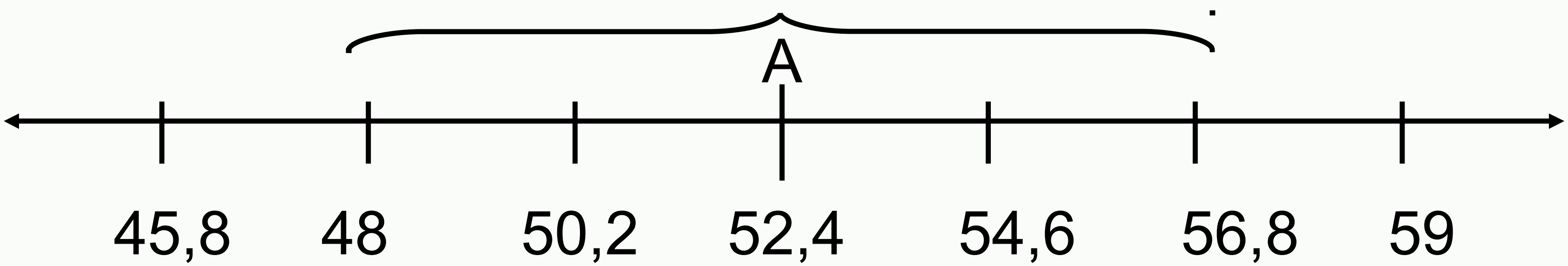
Se *k* = 3, ***pelo menos*** 1 – 1/9 = 8/9 = 88,9% dos dados estão dentro de três desvios padrão da média.

Teorema de Chebychev

A média feminina nos 400 metros rasos é de 52,4 segundos,com um desvio padrão de 2,2 segundos. Aplique o Teorema de Chebychev com *k* = 2.

Trace uma linha de números nas unidades de desvio padrão.

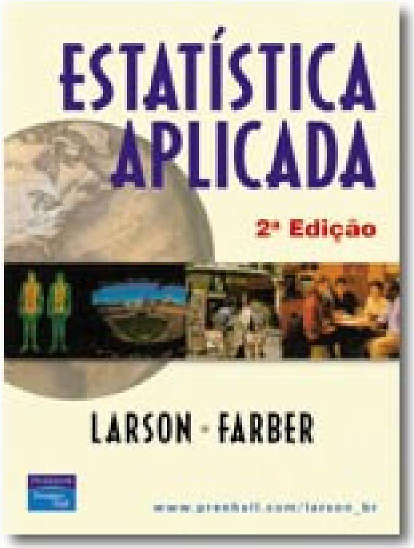
2 desvios padrão



Pelo menos 75% dos tempos femininos nos 400 metros rasos estarão entre 48 e 56,8 segundos.

Seção 2.5

Medidas de posição



Quartis

Os três quartis Q_1, Q_2 e Q_3 dividem os dados em quatro partes iguais.

Q_2 é igual à mediana.

Q_1 é a mediana dos dados que ficaram abaixo de Q_2.

Q_3 é a mediana dos dados que ficaram acima de Q_2.

Você é gerente de uma loja. A média de vendas em 27 dias do ano passado, selecionados aleatoriamente, é dada abaixo. Determine Q_1, Q_2 e Q_3.

28 43 48 51 43 30 55 44 48 33 45 37 37 42 27
47 42 23 46 39 20 45 38 19 17 35 45

Como determinar quartis

Coloque os dados (n = 27) em ordem:

17 19 20 23 27 28 30 33 35 37 37 38 39 42 42 43 43 44 45 45 45 46 47 48 48 51 55.

A posição da mediana é: (27 + 1)/2 = 14. A mediana = Q_2 = 42.

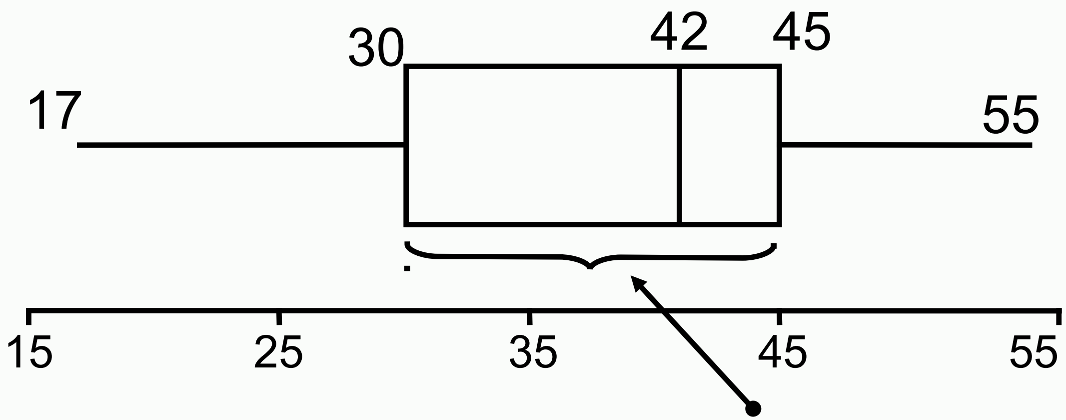
Há 13 valores abaixo da mediana.
A posição de Q_1 = 7. Q_1 é 30.
Q_3 é a posição 7 a partir do último valor. Q_3 é 45.

A amplitude interquartil é $Q_3 - Q_1$ = 45 – 30 = 15.

Plote maria-chiquinha

Um plote maria-chiquinha usa cinco valores-chave para descrever um conjunto de dados: Q_1, Q_2 e Q_3, o valor mínimo e o valor máximo.

Q_1	30
Q_2 = a mediana	42
Q_3	45
Valor mínimo	17
Valor máximo	55



Amplitude interquartil = 45 – 30 = 15

Percentis

Os percentis dividem os dados em cem partes. Existem, portanto, 99 percentis: P_1, P_2, $P_3 \ldots P_{99}$.

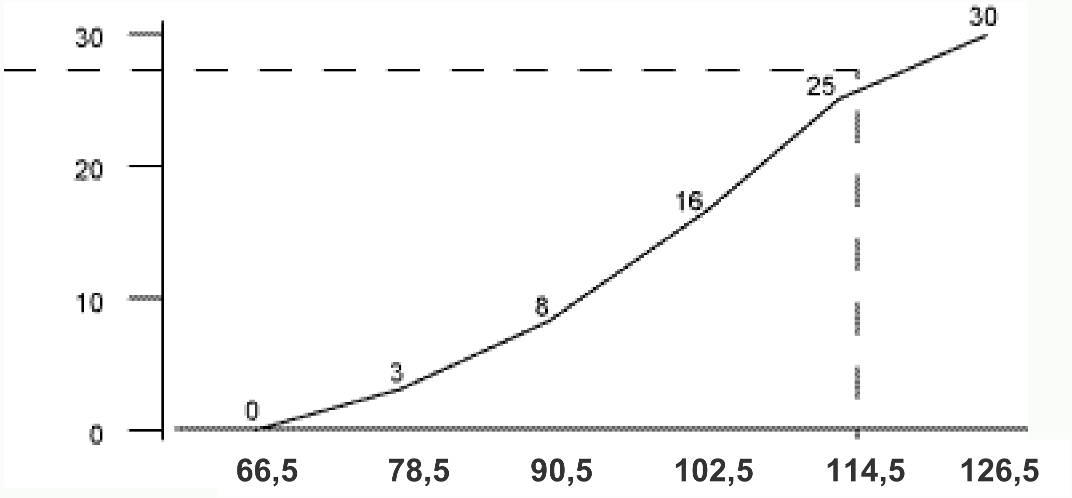
$$P_{50} = Q_2 = \text{a mediana}$$

$P_{25} = Q_1$

$P_{75} = Q_3$

Uma pontuação no 63º percentil indica que ela é igual ou superior a 63% das pontuações e igual ou inferior a 37% delas.

Percentis



Distribuições cumulativas podem ser usadas para calcular percentis.

114,5 é igual ou inferior a 25 dos 30 valores.
25/30 = 83,33.
Logo, você pode aproximar que 114 = P_{83}.

Escore padrão

O escore padrão, ou escore *z*, representa o número de desvios padrão que separa um valor *x* da média.

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

Em um concurso público, as pontuações tiveram uma média de 152 e desvião padrão de 7. Calcule o escore *z* para um candidato com uma pontuação de:

(a) 161 **(b) 148**

(c) 152

Cálculos de escore *z*

(a)

$$z = \frac{161 - 152}{7}$$

$$z = \mathbf{1{,}29}$$

Um valor de *x* = 161 está 1,29 desvio padrão acima da média.

(b)

$$z = \frac{148 - 152}{7}$$

$$z = -\mathbf{0{,}57}$$

Um valor de *x* = 148 está 0,57 desvio padrão abaixo da média.

(c)

$$z = \frac{152 - 152}{7}$$

$$z = 0$$

Um valor de *x* = 152 é igual à média.