



Estatística

2020.1

— Aula 17 - Unidade III —

Danilo Borges da Silva
danilo@frn.uespi.br

Análise Exploratória de Dados (Parte 2)

Cronograma

- Motivação
- Medidas de Centro
- Medidas de Dispersão
- Medidas de Posição
- Medidas de Assimetria e Curtose

Motivação

Motivação

As medidas são ferramentas básicas importantes para a medição e descrição de diferentes características de um conjunto de dados.

Medida de Posição Central

Medidas de Posição Central

Definição

Representam os fenômenos pelos seus valores médios, em torno dos quais tendem a concentrar-se os dados.

Dentre todas as medidas de tendência central, veremos:

- Média;
- Mediana;
- Moda

Média

Definição. É o valor médio de uma distribuição, determinado segundo uma regra estabelecida *a priori* e que se utiliza para representar todos os valores da distribuição. Representada por \overline{x}

Podem ser:

- Aritmética
- Ponderada
- Harmônica
- Geométrica

Média Aritmética

É a mais utilizada, dada pela fórmula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \cdots + x_n)$$

Onde:

- n é o número de valores de uma amostra
- x_i é cada variável que representa os valores individuais dos dados

Média Aritmética para Dados Agrupados

É calculada quando a informação disponível é o valor médio do intervalo $i (X_i)$ e a frequência de intervalo $i (f_i)$:

$$\overline{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i}$$

Exemplo

Considere os seguintes dados:

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,30	17,48	17,80	18,47

...continuando

Intervalos de classes	Frequência absoluta
12,51 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

...continuando

Intervalos de classes	Frequência absoluta
12,51 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

$$\overline{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i} = \frac{3 \cdot 13 + 8 \cdot 14 + 15 \cdot 15 + 13 \cdot 16 + 9 \cdot 17 + 2 \cdot 18}{50} ;$$

...continuando

Intervalos de classes	Frequência absoluta
12,51 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

$$\overline{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i} = \frac{3 \cdot 13 + 8 \cdot 14 + 15 \cdot 15 + 13 \cdot 16 + 9 \cdot 17 + 2 \cdot 18}{50} = 15,46$$

Média Ponderada

Nos cálculos envolvendo média aritmética simples, todas as ocorrências têm exatamente a mesma importância ou o mesmo peso. No entanto, existem casos onde as ocorrências têm importância relativa ou pesos relativos diferentes. Nestes casos, o cálculo da média deve levar em conta esta importância relativa ou peso relativo. Este tipo de média chama-se *média aritmética ponderada*.

Média Ponderada

Nos cálculos envolvendo média aritmética simples, todas as ocorrências têm exatamente a mesma importância ou o mesmo peso. No entanto, existem casos onde as ocorrências têm importância relativa ou pesos relativos diferentes. Nestes casos, o cálculo da média deve levar em conta esta importância relativa ou peso relativo. Este tipo de média chama-se *média aritmética ponderada*.

É dada por:

$$\overline{X} = \frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i}$$

onde w_i é o peso de cada x_i

Exemplo

O exame de seleção pode ser composto de três provas onde a primeira tem peso 3 a segunda tem peso 2 e a terceira tem peso 1. O candidato 1 com notas 70, 85 e 100 terá média final de quando? E o Candidato 2 com 95, 85 e 60 terá que média?

Exemplo

O exame de seleção pode ser composto de três provas onde a primeira tem peso 3 a segunda tem peso 2 e a terceira tem peso 1. O candidato 1 com notas 70, 85 e 100 terá média final de quando? E o Candidato 2 com 95, 85 e 60 terá que média?

Candidato 1: 80,00

Candidato 2: 85,83

Média Harmônica

A média harmônica está relacionada ao cálculo matemático das situações envolvendo as grandezas inversamente proporcionais. Como exemplo, temos a relação entre velocidade e tempo ([site](#)). A média harmônica equivale ao inverso da média aritmética dos inversos de n valores.

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Média Geométrica

Como o próprio nome indica, a média geométrica sugere interpretações geométricas ([site](#)). Essa média é a raiz de ordem n do produto dos valores da amostra:

$$G = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Relação entre Médias

- A média geométrica e a média harmônica são menores, ou no máximo, iguais à aritmética

Relação entre Médias

- A média geométrica e a média harmônica são menores, ou no máximo, iguais à aritmética
- A igualdade só ocorre no caso em que os valores são idênticos

Relação entre Médias

- A média geométrica e a média harmônica são menores, ou no máximo, iguais à aritmética
- A igualdade só ocorre no caso em que os valores são idênticos
- Quanto maior a variabilidade, maior será a diferença entre as médias harmônicas e geométrica e a média aritmética

$$H \leq G \leq \overline{X}$$

Relação entre Médias

- A média geométrica e a média harmônica são menores, ou no máximo, iguais à aritmética
- A igualdade só ocorre no caso em que os valores são idênticos
- Quanto maior a variabilidade, maior será a diferença entre as médias harmônicas e geométrica e a média aritmética

$$H \leq G \leq \overline{X}$$

Exemplo: Para a amostra 12, 14 e 16 temos:

$$13,81 < 13,90 < 14,00$$

Mediana

Definição. É um número que caracteriza as observações de uma determinada variável de tal forma que este número de um grupo de dados ordenados separa a metade inferior da amostra, população ou distribuição de probabilidade, da metade superior. Representada por \tilde{x} ou Md

Mediana

Definição. É um número que caracteriza as observações de uma determinada variável de tal forma que este número de um grupo de dados ordenados separa a metade inferior da amostra, população ou distribuição de probabilidade, da metade superior. Representada por \tilde{x} ou Md

Para calcular ordena-se a amostra de forma crescente:

- Se n é ímpar, a mediana é o valor central:
 - Na amostra 30, 32, 35, 48, 76 a mediana é 35
- Se n é par, mediana é a média simples entre os dois valores centrais:
 - Na amostra 30, 32, 35, 48, 76, 81 a mediana é $(35+48)/2 = 41$

Mediana para Dados Agrupados

1. Calcula-se $n/2$
2. Achar qual das classes esse valor se encontra a partir das frequências absolutas
3. Usar a fórmula

$$Md = l_{Md} + \frac{\left(\frac{n}{2} - \sum f\right) \cdot h}{f_{Md}}$$

onde:

- l_{Md} é o limite inferior da classe
- f_{Md} é a frequência da classe da mediana
- $\sum f$ é a soma das frequências anteriores a classe da mediana
- h é a amplitude da classe da mediana

Exemplo

Intervalos de classe	Frequência absoluta	Frequência acumulada
12,51 a 13,50	3	3
13,51 a 14,50	8	11
14,51 a 15,50	15	26
15,51 a 16,50	13	39
16,51 a 17,50	9	48
17,51 a 18,50	2	50

Exemplo

Intervalos de classe	Frequência absoluta	Frequência acumulada
12,51 a 13,50	3	3
13,51 a 14,50	8	11
14,51 a 15,50	15	26
15,51 a 16,50	13	39
16,51 a 17,50	9	48
17,51 a 18,50	2	50

- n
- l_{Md}
- f_{Md}
- Σf
- h

Exemplo

Intervalos de classe	Frequência absoluta	Frequência acumulada
12,51 a 13,50	3	3
13,51 a 14,50	8	11
14,51 a 15,50	15	26
15,51 a 16,50	13	39
16,51 a 17,50	9	48
17,51 a 18,50	2	50

- $n = 50$
- $l_{Md} = 14,51$
- $f_{Md} = 15$
- $\Sigma f = 11$
- $h = 0,99$ (15,50-14,51)

Exemplo

Intervalos de classe	Frequência absoluta	Frequência acumulada
12,51 a 13,50	3	3
13,51 a 14,50	8	11
14,51 a 15,50	15	26
15,51 a 16,50	13	39
16,51 a 17,50	9	48
17,51 a 18,50	2	50

- $n = 50$
- $l_{Md} = 14,51$
- $f_{Md} = 15$
- $\Sigma f = 11$
- $h = 0,99$ (15,50-14,51)

$$Md = l_{Md} + \frac{\left(\frac{n}{2} - \Sigma f\right) \cdot h}{f_{Md}}$$

Exemplo

Intervalos de classe	Frequência absoluta	Frequência acumulada
12,51 a 13,50	3	3
13,51 a 14,50	8	11
14,51 a 15,50	15	26
15,51 a 16,50	13	39
16,51 a 17,50	9	48
17,51 a 18,50	2	50

- $n = 50$
- $l_{Md} = 14,51$
- $f_{Md} = 15$
- $\Sigma f = 11$
- $h = 0,99$ (15,50-14,51)

$$Md = 14,51 + \frac{(25 - 11) \cdot 0,99}{15} = 15,434$$

Moda

Definição. É o valor que ocorre com mais frequência. Representada por Mo . Numa amostra, Mo pode não existir ou ser múltipla (amostra multimodal).

Exemplos:

- Na amostra 21 24 27 27 28 28 31 31 31 $Mo = 31$
- Na amostra 45 46 49 52 52 60 60 76 79 tem moda 52 e 60

Moda para Dados Agrupados

Utiliza-se a fórmula de King:

$$Mo = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

onde:

- l é o limite inferior da classe modal
- Δ_1 é a diferença entre a frequência da classe e a anterior
- Δ_2 é a diferença entre a frequência da classe e a posterior
- h é a amplitude da classe modal

Moda para Dados Agrupados

Notas	Número de Alunos
0 - 20	2
20 - 40	7
40 - 60	23
60 - 80	16
80 - 100	3
Total	51

- l é o limite inferior da classe modal
- Δ_1 é a diferença entre a frequência da classe e a anterior
- Δ_2 é a diferença entre a frequência da classe e a posterior
- h é a amplitude da classe modal

Moda para Dados Agrupados

Notas	Número de Alunos
0 - 20	2
20 - 40	7
40 - 60	23
60 - 80	16
80 - 100	3
Total	51

- $l = 40$
- $\Delta_1 = 16$
- $\Delta_2 = 7$
- $h = 20$

Moda para Dados Agrupados

Notas	Número de Alunos
0 - 20	2
20 - 40	7
40 - 60	23
60 - 80	16
80 - 100	3
Total	51

- $l = 40$
- $\Delta_1 = 16$
- $\Delta_2 = 7$
- $h = 20$

$$Mo = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

Moda para Dados Agrupados

Notas	Número de Alunos
0 - 20	2
20 - 40	7
40 - 60	23
60 - 80	16
80 - 100	3
Total	51

- $l = 40$
- $\Delta_1 = 16$
- $\Delta_2 = 7$
- $h = 20$

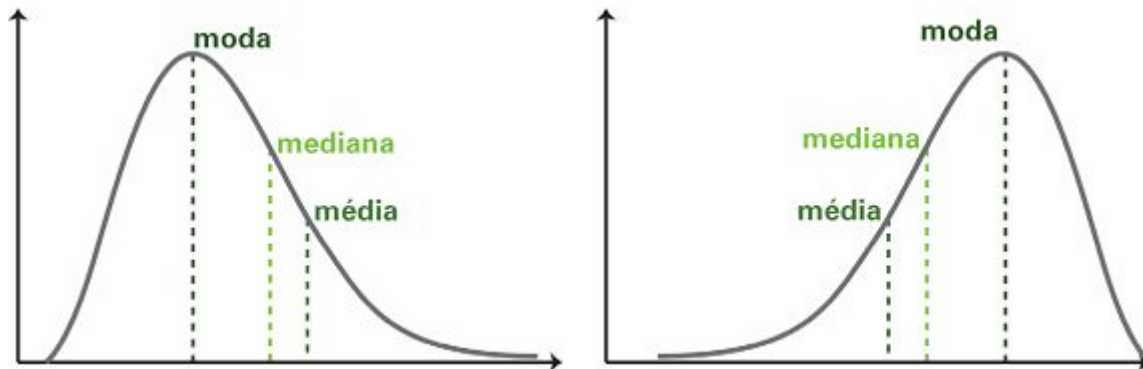
$$Mo = 40 + \frac{16}{16 + 7} \cdot 20 = 53,91$$

Comparação

- Para distribuições *simétricas*, a média, mediana e moda são aproximadamente iguais;

Comparação

- Para distribuições *simétricas*, a média, mediana e moda são aproximadamente iguais;
- Para *assimétricas*, observa-se o seguinte:



Relações Empíricas entre Medidas de Posição

Exemplo

A relação entre média e mediana para as amostras a seguir é:

A	Distribuição Simétrica	10 12 14 16 18	
B	Distribuição Assimétrica à direita	10 12 14 16 23	
C	Distribuição Assimétrica à esquerda	05 12 14 16 18	

Medidas de Dispersão

Medidas de Dispersão

Definição. É um valor que busca quantificar o quanto os valores da amostra estão afastados ou dispersos relativos à média amostral;

As medidas utilizadas para representar dispersão são:

- 2.1 Amplitude Total
- 2.2 Desvio Padrão;
- 2.3 Variância;
- 2.4 Amplitude Interquartílica.

Amplitude Total

Definição. Também chamado somente por Amplitude, é a diferença entre o maior e o menor valor de um conjunto de dados.

A amplitude é muito fácil de ser calculada, mas como depende apenas dos valores maior e menor, não é tão útil quanto as outras medidas de variação que usam todos os valores

Desvio Padrão

Definição. É uma medida da variação dos valores em torno da média em um conjunto de valores amostrais. Representado por s (para amostral) e σ (para populacional).

Desvio Padrão

Definição. É uma medida da variação dos valores em torno da média em um conjunto de valores amostrais. Representado por s (para amostral) e σ (para populacional).

- Para uma população de N indivíduos:
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Desvio Padrão

Definição. É uma medida da variação dos valores em torno da média em um conjunto de valores amostrais. Representado por s (para amostral) e σ (para populacional).

- Para uma população de N indivíduos:
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$
- Para uma amostra com n observações, x_1, \dots, x_n :
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Desvio Padrão

Definição. É uma medida da variação dos valores em torno da média em um conjunto de valores amostrais. Representado por s (para amostral) e σ (para populacional).

- Para uma população de N indivíduos:
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

- Para uma amostra com n observações, x_1, \dots, x_n :
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

onde:

- x_i é o valor de cada variável
- \bar{x} é a média amostral e μ é a média populacional

Exemplo

Para a amostra 10, 12, 14, 16 e 18. Qual o valor do desvio padrão?

Exemplo

Para a amostra 10, 12, 14, 16 e 18. Qual o valor do desvio padrão?

R=3,16

Desvio padrão para Dados Agrupados

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

Desvio padrão para Dados Agrupados

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

Intervalos de Classe	Frequência Absoluta
12,50 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

→ Ponto médio do intervalo

Desvio padrão para Dados Agrupados

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

Intervalos de Classe	Frequência Absoluta
12,50 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

→ Ponto médio do intervalo

$$\sigma = \sqrt{\frac{3 \cdot (13 - 15,46)^2 + 8 \cdot (14 - 15,46)^2 + 15 \cdot (15 - 15,46)^2 + 13 \cdot (16 - 15,46)^2 + 9 \cdot (17 - 15,46)^2 + 2 \cdot (18 - 15,46)^2}{50}} = 1,308$$

Coeficiente de Variação

Definição. Para um conjunto de dados amostrais ou populacionais, expresso como um percentual, descreve o desvio padrão relativo à média, e é dado pelo seguinte:

- Para população: $cv = \frac{\sigma}{\mu}$
- Para amostra: $cv = \frac{s}{\bar{x}}$

Coeficiente de Variação

Definição. Para um conjunto de dados amostrais ou populacionais, expresso como um percentual, descreve o desvio padrão relativo à média, e é dado pelo seguinte:

- Para população: $cv = \frac{\sigma}{\mu}$
- Para amostra: $cv = \frac{s}{\bar{x}}$
 - É uma medida *dimensional*, útil para comparar resultados de amostras ou populações cujas unidades podem ser diferentes;
 - Uma desvantagem do coeficiente de variação é que ele *deixa de ser útil* quando a média é próxima de zero.

Variância

Definição. É uma medida de variação igual ao quadrado do desvio padrão.

Representada por s^2 ou σ^2 :

- Para população: $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$
- Para amostra: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Variância

Definição. É uma medida de variação igual ao quadrado do desvio padrão.

Representada por s^2 ou σ^2 :

- Para população: $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$
- Para amostra: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Uma dificuldade é que a variância não é expressa nas mesmas unidades dos dados originais.

Amplitude Interquartílica

Definição. É utilizada na estatística para medir a dispersão de um conjunto de dados pois exclui valores discrepantes. É a amplitude do intervalo entre o primeiro e o terceiro quartil. Representada por Q ;

$$Q = Q_3 - Q_1$$

Amplitude Interquartílica

Definição. É utilizada na estatística para medir a dispersão de um conjunto de dados pois exclui valores discrepantes. É a amplitude do intervalo entre o primeiro e o terceiro quartil. Representada por Q ;

$$Q = Q_3 - Q_1$$

Às vezes também é usada a semi amplitude interquartílica. Essa é utilizada quando a média não pode ou não deve ser utilizada, nessa situação utiliza-se a mediana. Sendo calculada como metade de Q

Amplitude Interquartílica

Definição. É utilizada na estatística para medir a dispersão de um conjunto de dados pois exclui valores discrepantes. É a amplitude do intervalo entre o primeiro e o terceiro quartil. Representada por Q ;

$$Q = Q_3 - Q_1$$

Às vezes também é usada a semi amplitude interquartílica. Essa é utilizada quando a média não pode ou não deve ser utilizada, nessa situação utiliza-se a mediana. Sendo calculada como metade de Q

Relações aproximadas com o desvio padrão:

$$Q = \frac{4}{3}\sigma \qquad Q = \frac{4}{3}s$$

Medida de Posição

Medida de Posição

Definição. São medidas que dividem a área de uma distribuição de frequências em regiões de áreas iguais.

As principais medidas de posição são:

- Quartil;
- Percentil.

Quartil

Definição. É qualquer um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais, e assim cada parte representa $\frac{1}{4}$ da amostra ou população.

Valores que dividem o conjunto em quatro partes iguais são representados por Q_1 , Q_2 , Q_3 e denominam-se primeiro, segundo e terceiro quartis, respectivamente:

- Q_1 separa os 25% inferiores dos 75% dos superiores;
- Q_2 é a mediana;
- Q_3 separa os 75% inferiores dos 25% dos superiores.



Quartil para Dados Agrupados

Determinação de Q_1 :

- 1) Calcula-se $n/4$
- 2) Identifica-se a classe Q_1 pela frequência acumulada
- 3) Aplica-se a fórmula:

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

Quartil para Dados Agrupados

Determinação de Q_2 :

- 1) Calcula-se $3n/4$
- 2) Identifica-se a classe Q_3 pela frequência acumulada
- 3) Aplica-se a fórmula:

$$Q_3 = l_{Q_3} + \frac{\left(\frac{3n}{4} - \sum f\right)}{F_{Q_3}} \cdot h$$

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo


Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$


$$816/4 = 204$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

$$816/4 = 204$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

$$816/4 = 204$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

$$816/4 = 204$$

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} h$$

Diagram illustrating the calculation of Q_1 using the formula and the data table:

- $\left(\frac{n}{4}\right)$ corresponds to $816/4 = 204$.
- $\sum f$ corresponds to the cumulative frequency 206.
- F_{Q_1} corresponds to the cumulative frequency 206.
- h corresponds to the class width 16.

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

$$816/4 = 204$$

Exemplo

Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

$$Q_1 = l_{Q_1} + \frac{\left(\frac{n}{4} - \sum f\right)}{F_{Q_1}} \cdot h$$

$$816/4 = 204$$

$$Q_1 = 12 + \frac{(204 - 99)}{206} \cdot 4 = 14,03$$

Exemplo

Qual é o mínimo de pontos necessários para um aluno se classificar entre os 25% mais fortes?

$$Q_3 = l_{Q_3} + \frac{\left(\frac{3n}{4} - \sum f\right)}{F_{Q_3}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o mínimo de pontos necessários para um aluno se classificar entre os 25% mais fortes?

$$Q_3 = l_{Q_3} + \frac{\left(\frac{3n}{4} - \sum f\right)}{F_{Q_3}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

$$Q_3 = 20 + \frac{(612 - 524)}{155} \cdot 4 = 22,27$$

Percentil

Definição. É um valor que divide o conjunto ordenado de dados em cem partes iguais, e assim cada parte representa 1/100 da amostra ou população.

O k-ésimo percentil corresponde a frequência cumulativa de $Nk/100$, onde N é o tamanho do espaço amostral

$$P_k = \frac{Nk}{100}$$

Percentil

Definição. É um valor que divide o conjunto ordenado de dados em cem partes iguais, e assim cada parte representa 1/100 da amostra ou população.

O k-ésimo percentil corresponde a frequência cumulativa de $Nk/100$, onde N é o tamanho do espaço amostral (arredondar para inteiro mais próximo)

$$P_k = \frac{Nk}{100}$$

Percentil

Definição. É um valor que divide o conjunto ordenado de dados em cem partes iguais, e assim cada parte representa 1/100 da amostra ou população.

O k-ésimo percentil corresponde a frequência cumulativa de $Nk/100$, onde N é o tamanho do espaço amostral (arredondar para inteiro mais próximo)

$$P_k = \frac{Nk}{100}$$

$$P_i = l_{P_i} + \frac{\left(\frac{i \cdot n}{100} - \sum f\right)}{F_{P_i}} \cdot h$$

Exemplo

Qual é o mínimo de pontos para que um aluno esteja entre os 5% mais fortes?

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o mínimo de pontos para que um aluno esteja entre os 5% mais fortes?

$$P_i = l_{P_i} + \frac{\left(\frac{i \cdot n}{100} - \sum f\right)}{F_{P_i}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

Exemplo

Qual é o mínimo de pontos para que um aluno esteja entre os 5% mais fortes?

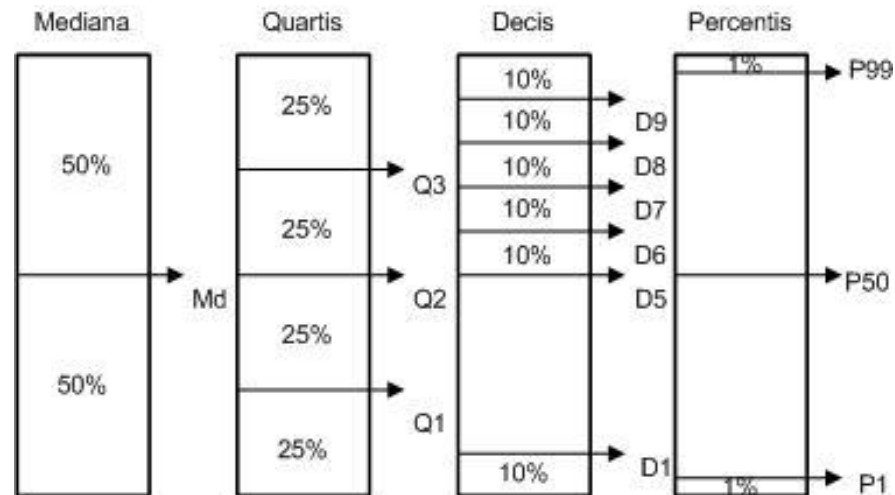
$$P_{95} = l_{P_{95}} + \frac{\left(\frac{95 \cdot n}{100} - \sum f\right)}{F_{P_{95}}} \cdot h$$

Pontos Alcançados no Teste	Nº de Alunos	Frequência Acumulada
4 8	10	10
8 12	89	99
12 16	206	305
16 20	219	524
20 24	155	679
24 28	78	757
28 32	30	787
32 36	18	805
36 40	11	816
Total	816	816

$$P_{95} = 28 + \frac{(775 - 757)}{30} \cdot 4 = 30,4$$

Relações

- 1º quartil = 25º percentil;
- Mediana = 5º decil = 50º percentil;
- 3º quartil = 75º percentil.



Medida de Assimetria e Curtose

Medida de Assimetria

As medidas de assimetria possibilitam analisar uma distribuição de acordo com as relações entre suas medidas de moda, média e mediana, quando observadas graficamente ou analisando apenas os valores;

- Uma distribuição é dita *simétrica* quando apresenta o mesmo valor para a moda, a média e a mediana;
- É dita *assimétrica* quando essa igualdade não ocorre.

Medida de Assimetria

Para o cálculo de assimetria, usa-se o coeficiente de assimetria de Pearson:

$$Sk = \frac{\bar{x} - Mo}{s}$$

Medida de Assimetria

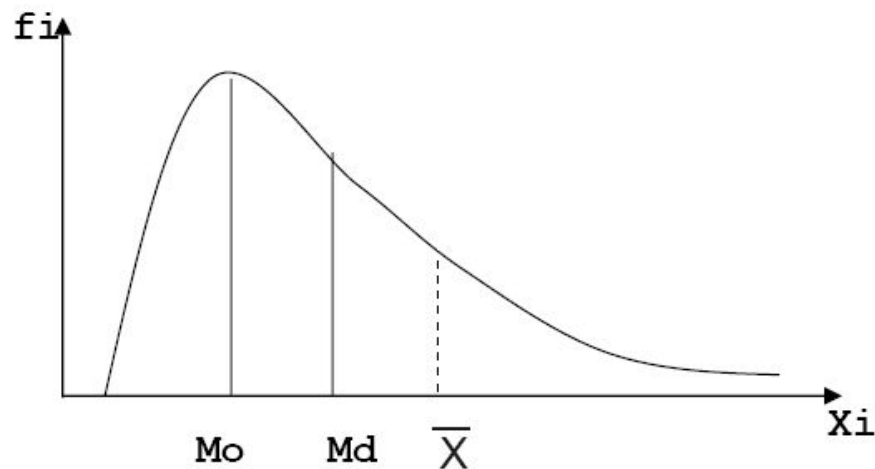
Para o cálculo de assimetria, usa-se o coeficiente de assimetria de Pearson:

$$Sk = \frac{\bar{x} - Mo}{s}$$

Os valores estão dentro do intervalo $[-1, +1]$

Medida de Assimetria

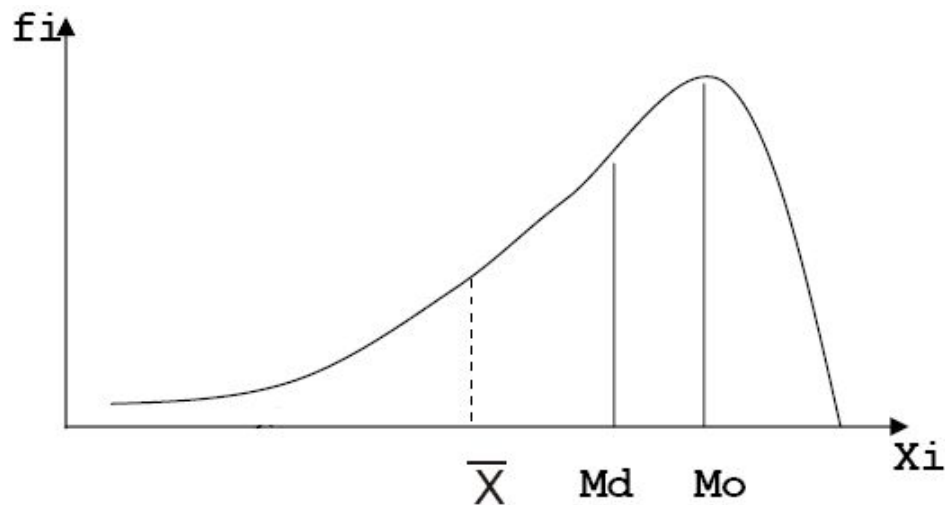
- Quando a cauda da curva da distribuição declina para *direita*, temos uma distribuição com curva assimétrica *positiva*;
- Coeficiente > 0 .



Medida de Assimetria

Quando a cauda da curva da distribuição declina para *esquerda*, temos uma distribuição com curva assimétrica *negativa*;

- Coeficiente < 0 ;



Medida de Curtose

Definição. Curtose é o grau de achatamento da distribuição. Ou o quanto uma curva de frequência será achatada em relação a uma curva normal de referência.

Medida de Curtose

Definição. Curtose é o grau de achatamento da distribuição. Ou o quanto uma curva de frequência será achatada em relação a uma curva normal de referência.

Para o cálculo da curtose, usa-se o coeficiente de curtose de Pearson:

$$\alpha_4 = \frac{m_4}{s_4^4}, \text{ onde } m_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}$$

Medida de Curtose

Definição. Curtose é o grau de achatamento da distribuição. Ou o quanto uma curva de frequência será achatada em relação a uma curva normal de referência.

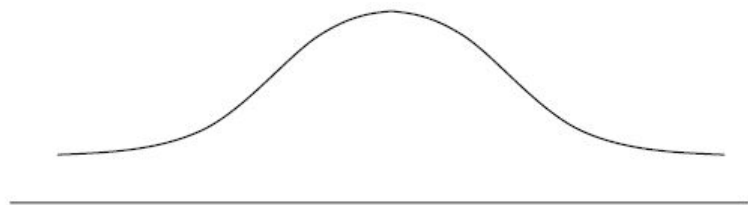
Para o cálculo da curtose, usa-se o coeficiente de curtose de Pearson:

$$\alpha_4 = \frac{m_4}{s_4}, \text{ onde } m_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}$$

Pode ser:

- Mesocúrtica: $\alpha_4 = 3$
- Leptocúrtica: $\alpha_4 > 3$
- Platicúrtica: $\alpha_4 < 3$

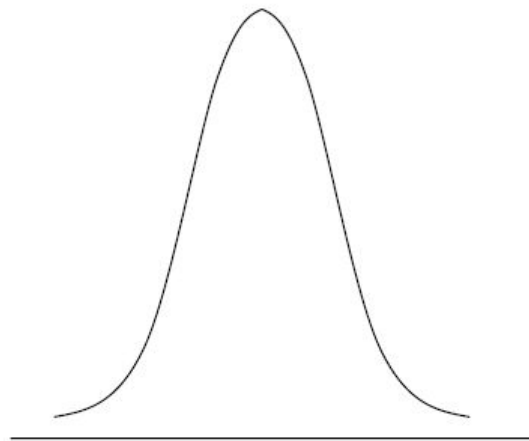
4. Medida de Assimetria e Curtose



Mesocúrtica



Platicúrtica



Leptocúrtica

Exercícios

- 1) Foram feitas coletas do tempo (ms) de acesso de uma página na internet e obteve-se os valores:

85,3 84,3 79,5 82,5 80,2 84,6 79,2 70,9 78,6 86,2 74,0 83,7

Calcule:

- a) Média
- b) Mediana
- c) Desvio Padrão

Exercícios

2) Dada a amostra:

28 33 27 30 31 30 33 30 33 29 27 33 31 27 31 28 27 29 31 24 31 33 30
32 30 33 27 33 31 33 23 29 30 24 28 34 30 30 18 17 18 15 16 17 17 18
19 19 20 29

- a) Construir a tabela com a distribuição de frequência;
- b) Calcular a média;
- c) Moda;
- d) Mediana;
- e) O coeficiente de variação;
- f) Determinar a curtose.

Exercícios

- 3) O Sr. Malaquias, cujas habilitações literárias não vão além da 4ª ano de escolaridade, respondeu a 2 anúncios de oferta de emprego. As empresas trabalham no mesmo ramo, pelo que o serviço que o Sr. Malaquias iria fazer seria semelhante em qualquer das empresas. Resolveu saber alguma coisa sobre os ordenados processados nos dois sítios, tendo obtido a seguinte informação:

	Empresa A	Empresa B
Média	R\$ 445	R\$ 475
Mediana	R\$ 400	R\$ 350
Desvio padrão	R\$ 160	R\$ 190

Qual das empresas aconselharia o Sr. Malaquias a escolher? Explique porquê.