Estatística

1 – Medidas de Tendência Central

2 – Medidas de Posição

3 – Medidas de Dispersão

Renata Souza

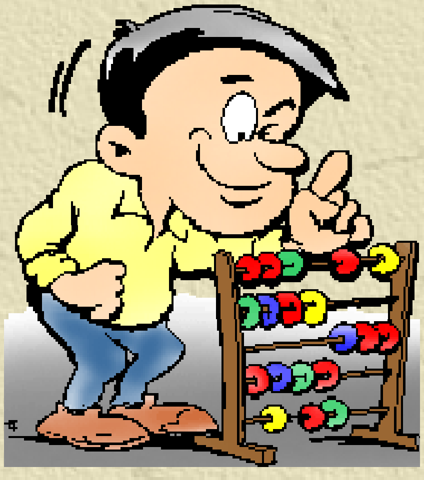
Medidas

- Depois que você conheceu os conceitos de coleta de dados, variação, causas comuns e causas especiais, chegou a hora de estudarmos algumas formas de medir os resultados.

- Para melhor interpretar os resultados obtidos com uma amostra, são definidas algumas medidas:
  - medidas de posição central
  - medidas de posição
  - medidas de dispersão.

Medidas de Tendência Central

- Mostram a tendência dos pontos se concentrarem em torno de um determinado valor



Medidas de Tendência Central

- Há várias medidas de tendência central. Entre elas citamos a média aritmética, a mediana, a média harmônica, geométrica, etc.

- Cada uma dessas medidas apresenta vantagens e desvantagens, e a escolha depende dos objetivos desejados.

Média Aritmética

- A média aritmética, ou simplesmente média, de um conjunto de n valores $x_1, ..., x_n$ é definida como:

$$\overline{X} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- As letras gregas são usadas para representar parâmetros populacionais e as letras comuns parâmetros amostrais.
- A média de uma amostra é representada por $\overline{X}$ e média de uma população é representada pela letra grega µ.

- Exemplo: A média aritmética de 7,5 7,9 8,1 8,2 8,7 é

$$\overline{X} = \frac{7{,}5 + 7{,}9 + 8{,}1 + 8{,}2 + 8{,}7}{5} = 8{,}08$$

Média Aritmética Ponderada

- Algumas vezes associa-se a cada observação um peso $W_{i.}$ onde esse peso representa a importância atribuída a cada observação. Nesse caso a média ponderada é calculada como:

$$\bar{X} = \frac{w_1 x_1 + \ldots . + w_n x_n}{w_1 + \ldots . + w_n} = \frac{\sum_{i=1}^{n} w_i \, x_i}{\sum_{i=1}^{n} w_i}$$

- Exemplo: O exame de seleção pode ser composto de três provas onde as duas primeiras tem peso 1 e a terceira tem peso 2. Um candidato com notas 70 75 e 90 terá média final:

$$\bar{X} = \frac{1(70) + 1(75) + 2(90)}{4} = 81{,}25$$

Mediana

- Dado um conjunto de valores em ordem crescente, a mediana é definida como:
  - Se *n* é impar, o valor central;
  - Se *n* é par, a média simples dos dois valores centrais.
- Exemplos
- Exemplo 1: Na amostra 25 26 26 28 30 a mediana é

$$\tilde{x} = 26$$

- Exemplo 2: Na amostra 71 73 74 75 77 79 a mediana é

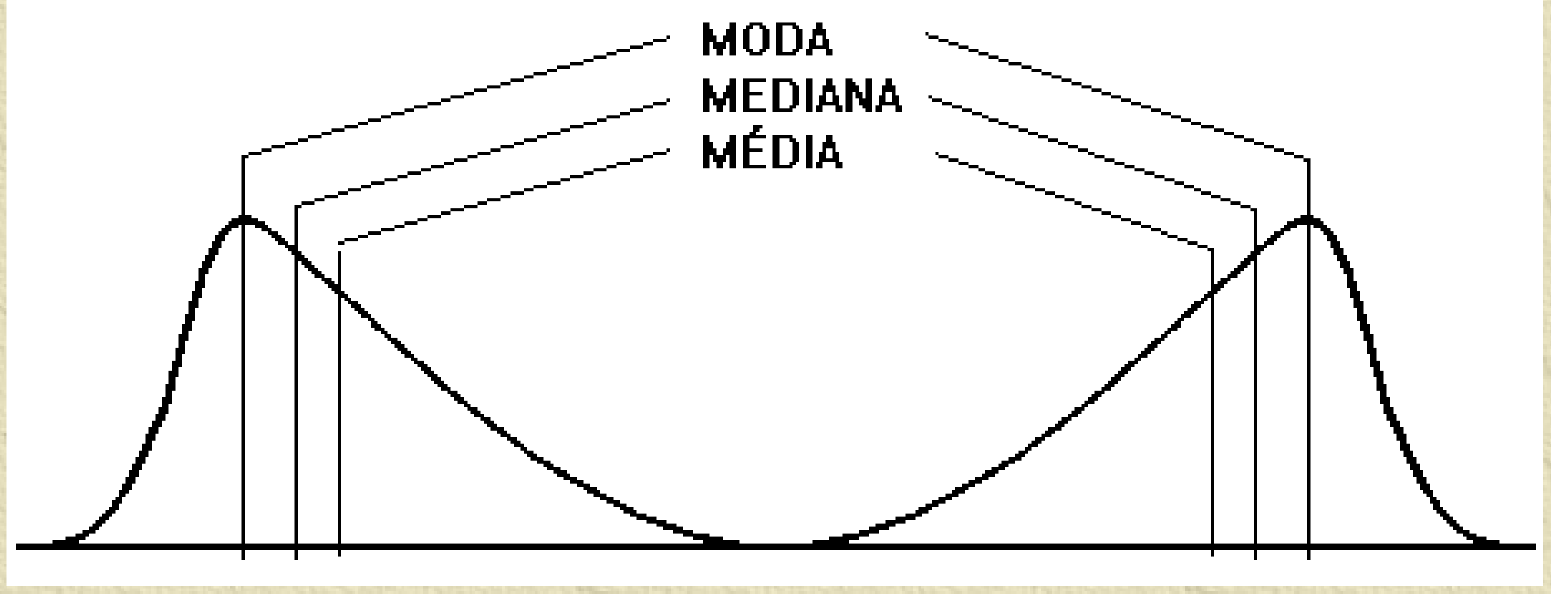
$$\tilde{x} = \frac{(74 + 75)}{2} = 74{,}5$$

Moda

- A moda é o valor que ocorre com maior freqüência, ou seja, é o valor mais comum.
- Exemplos
- Exemplo 1: A amostra 23 25 25 26 26 26 27 29 tem moda 26.
- Exemplo 2: A amostra 71 73 73 75 76 77 77 79 81 tem moda 73 e 77.
- A moda pode ser múltipla ou pode não existir.

Relações Empíricas entre Média, Moda e Mediana

- Para distribuições simétricas a média, a mediana e a moda coincidem aproximadamente.
- Para distribuições assimétricas observa-se o seguinte:



Relações Empíricas entre Média, Moda e Mediana

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

A	Distribuição simétrica	10 12 14 16 18	$\bar{x} = 14 = \tilde{x} = 14$
B	Distribuição assimétrica à direita	10 12 14 16 23	$\bar{x} = 15 > \tilde{x} = 14$
C	Distribuição assimétrica à esquerda	05 12 14 16 18	$\bar{x} = 13 < \tilde{x} = 14$

Comparação entre Média, Moda e Mediana

- Quão freqüente?
  - Média: mais familiar
  - Mediana: usada comumente
  - Moda: usada às vezes
- Existência
  - Média: existe sempre.
  - Mediana: existe sempre.
  - Moda: pode não existir; pode haver mais de uma moda
- Afetada pelos extremos?
  - Média: sim
  - Mediana: não
  - Moda: não

Comparação entre Média, Moda e Mediana

- Vantagens e desvantagens:
  - Média: funciona bem com muitos métodos estatísticos
  - Mediana: costuma ser uma boa escolha se há alguns valores extremos.
  - Moda: apropriada para dados ao nível nominal

Média Geométrica (G)

- É a raiz de ordem *n* do produto dos valores da amostra:

$$G = \sqrt[n]{X_1 X_2 \ldots . X_n}$$

- Exemplo
- A média geométrica de 12 14 16 é:

$$G = \sqrt[3]{12 \times 14 \times 16} = 13{,}90$$

- É usada em administração e economia para achar taxas médias de variação, de crescimento, ou razões médias

Média Harmônica (H)

- É o inverso da média aritmética dos inversos das observações.

$$H = \frac{1}{\frac{1}{n}\sum \frac{1}{X_i}} = \frac{n}{\sum \frac{1}{X_i}}$$

$$H = \frac{3}{\frac{1}{12} + \frac{1}{14} + \frac{1}{16}} = 13{,}81$$

- Exemplo
- A média harmônica de 12 14 16 é:

Relação entre Média Aritmética, Geométrica e Harmônica:

A média geométrica e a média harmônica são menores, ou no máximo igual, à média aritmética.

A igualdade só ocorre no caso em que todos os valores da amostra são idênticos.

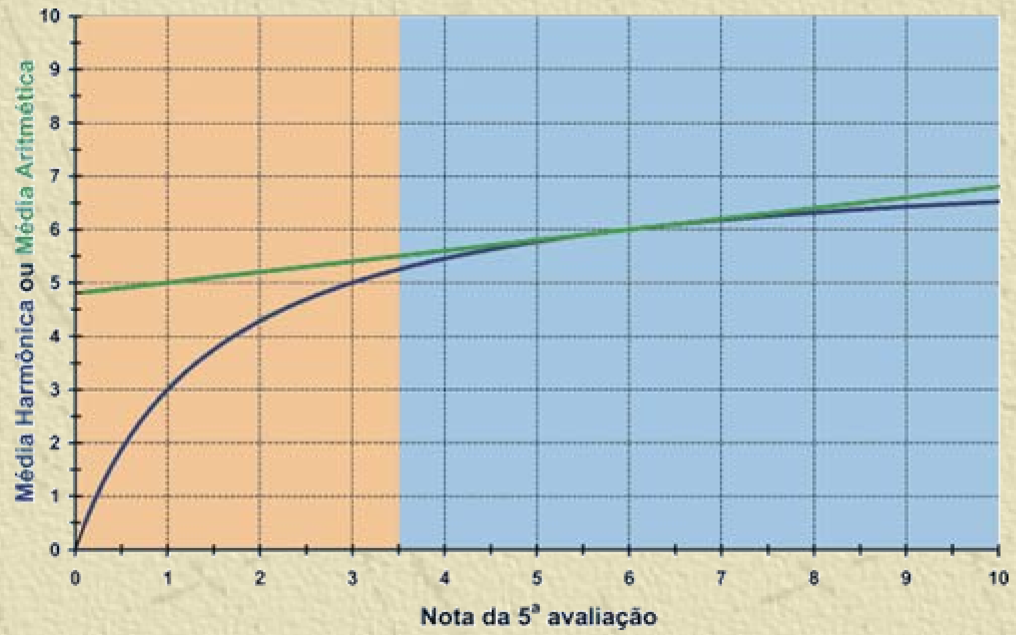
Quanto maior a variabilidade, maior será a diferença entre as médias harmônica e geométrica e a média aritmética.

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

- Exemplo: Para a amostra *12 14 16* tem-se
- $\boldsymbol{H = 13{,}81 < G = 13{,}90 <}$ $\bar{X} = 14{,}00$

Comparação Média Aritmética e Média Harmônica

O gráfico abaixo mostra uma simulação comparativa entre a **Média Harmônica** e a **Média Aritmética**, calculadas para cinco avaliações, onde as notas de quatro avaliações correspondem a 6,0 e a nota da 5ª avaliação varia de 0 a 10.



Medidas de Dispersão

- Invariavelmente as observações individuais irão apresentar alguma dispersão em torno do valor médio. Isso é chamado de variabilidade ou dispersão dos dados.
- Há muitas medidas de variabilidade, como por exemplo, a amplitude total, o desvio padrão, a amplitude inter-quartílica ou o coeficiente de variação.
- Os valores mínimos e máximos também podem ser usados como medidas de variabilidade

Amplitude total

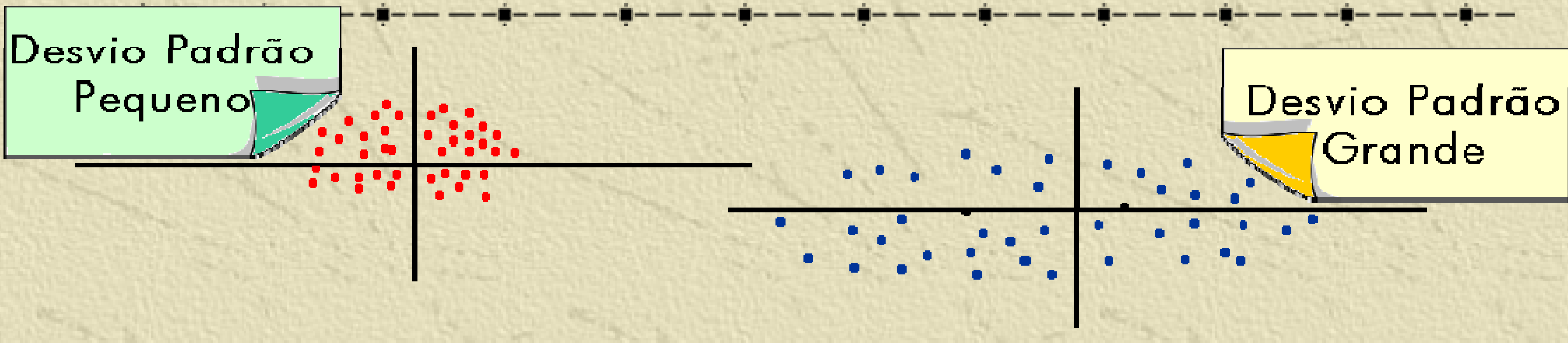
- É definida como a diferença entre o maior e o menor valor das observações.
- Exemplo : 8,5 8,7 8,9 10,1 10,5 10,7 11,5 11,9
- A amplitude é total: $R = 11,9 - 8,5 = 3,4$

- A amplitude é fácil de calcular e fornece uma idéia da magnitude da faixa de variação dos dados.
- Não informa a respeito da dispersão dos valores que caem entre os dois extremos.

Desvio Padrão

- Para uma amostra de n observações, $x_1, \ldots, x_n$, o desvio padrão S é definido como:
- $$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$
- A vantagem do desvio padrão é que trata-se de uma medida de variabilidade que leva em conta toda a informação contida na amostra.
- O desvio-padrão de uma população é representado por σ e o desvio padrão de uma amostra por S.

Medidas de Dispersão



- As medidas mais utilizadas para representar a dispersão é a VARIÂNCIA e o DESVIO PADRÃO.
- Uma dificuldade é que a variância não é expressa nas mesmas unidades dos dados originais.

Desvio Padrão

- Exemplo: para a amostra 10 12 14 16 18
- A média é $\bar{x} = 14$ e o desvio-padrão é calculado:
- Os desvios de cada valor em relação à média totalizam zero pois a média é o valor central:

$$10 - 14 = -4$$
$$12 - 14 = -2$$
$$14 - 14 = 0$$
$$16 - 14 = +2$$
$$18 - 14 = +4$$

$$S = \sqrt{\frac{(10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2}{n-1}} = 3{,}16$$

Variância

- A variância S^2 é definida como o quadrado do desvio padrão.

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

- A variância de uma população é representada pela letra grega σ^2.

- A variância é o quadrado do desvio padrão, ou seja,

$$\sigma^2 = 3{,}16^2 = 9{,}98$$

Amplitude Inter-quartílica

- É definida como a amplitude do intervalo entre o primeiro e o terceiro quartis, ou seja:

$$Q = Q_3 - Q_1$$

- Às vezes também é usada a semi-amplitude inter-quartílica, que é a metade da anterior.
- Trata-se de uma medida de variabilidade bastante robusta, que é pouco afetada pela presença de dados atípicos.
- A amplitude inter-quartílica guarda a seguinte relação aproximada com o desvio padrão:

$$Q = (4/3) \times \textit{desvio padrão}$$

Coeficiente de Variação

- É definido como o quociente entre o desvio padrão e a média e, em geral, é expresso em percentual.

$$CV = 100 \times \frac{S}{\bar{X}}$$

- O coeficiente de variação 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.

Medidas de Posição: Quartis

- Tanto a média como o desvio padrão podem não ser medidas adequadas para representar dados, pois:
  - São afetadas por valores extremos
  - Apenas com estes dois valores não temos idéia da simetria ou assimetria da distribuição dos dados
- Se um conjunto de dados é organizado em ordem crescente, o valor central é a mediana.
- 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.
- Resumo dos cinco números: Q_1, Q_2, Q_3 e os valores mínimo e máximo.

Relações

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

Cálculo do k ésimo percentil

- Ordenar os dados do menor para o maior
- Calcular:
  - $L=(k/100)\times n$
  - n=número de valores
  - k=percentil desejado
- Se L não é inteiro: arredonde L para o próximo inteiro acima dele. P_k é L-ésimo valor da lista ordenada.

Quartis: Exemplo

- Exemplo: Para a amostra a seguir calcular o primeiro e terceiro quartis:

13,3 13,5 17,2 13,8 12,3 12,7 13,0

14,5 14,9 15,8 13,1 13,3 14,1

x(i)	i
12,3	1
12,7	2
13,0	3
13,1	4
13,3	5
13,3	6
13,5	7
13,8	8
14,1	9
14,5	10
14,9	11
17,2	13

Exemplo: Quartis

- 1º quartil = 25º percentil
  - $L=(25/100)\times 13=3{,}25$
  - $L=4$
  - $P_{25}=Q_1=13{,}1$
- 3º quartil = 75º percentil
  - $L=(75/100)\times 13=9{,}25$
  - $L=10$
  - $P_{25}=Q_3=14{,}5$

Percentis: Dados agrupados

- P_i

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

$i \in \{1,2,3,4,5,6,...,99,100\}$

l_{Pi} - limite inferior da classe de P_i

Σf - soma das freqüências anteriores a classe de P_i

h – amplitude da classe de P_i

f_{Pi} – freqüência da classe P_i

Percentis: Exemplo com dados agrupados

Intervalos de classe	Freqüência absoluta	Freqüência absoluta
12,51 a 13,50	3	0,06
13,51 a 14,50	8	0,22
14,51 a 15,50	15	0,52
15,51 a 16,50	13	0,78
16,51 a 17,50	9	0,96
17,51 a 18,50	2	100,00

- 1º quartil = 25º percentil

$$P_{25} = 14{,}51 + \frac{\left(\frac{25 \times 50}{100} - 11\right) \times 1{,}01}{15} = 14{,}51 + 0{,}1 = 14{,}52$$

Variável Reduzida ou Padronizada

$$Z = \frac{X - \bar{X}}{S}$$

- Ela mede a magnitude do desvio em relação à média, em unidades do desvio padrão.
- *Z = 1,5* significa uma observação desviada 1,5 desvios padrão para cima da média.
- A variável reduzida é muito útil para comparar distribuições e detectar dados atípicos.
- Dados são considerados atípicos quando *Z > 3*.

Exemplo

- O engenheiro está analisando as espessuras de peças fabricadas em duas máquinas de corte.
- O operador mediu uma peça da máq. A com espessura de 90 mm e outra peça da máq. B com espessura de 100 mm.

engenheiro deve considerar esses dados reais ou atípicos?

- A máq. A possui média 51mm e desvio-padrão de 12mm.

Máq. A $$Z = \frac{X - \bar{X}}{S} = \frac{90 - 51}{12} = 3{,}25$$

Como $Z > 3$
é dado atípico

- A máq. B possui média 72mm e desvio-padrão de 16mm.

Máq. B $$Z = \frac{X - \bar{X}}{S} = \frac{100 - 72}{16} = 1{,}75$$

Como $Z < 3$
não é dado atípico

Exemplo

- Supondo que 51 fosse a média em uma prova de inglês, onde o desvio padrão é 12, para um candidato que obtivesse 90 acertos tem-se:

$$Z = \frac{X - \bar{X}}{S} = \frac{90 - 51}{12} = 3{,}25$$

- Conclui-se que na prova de inglês este candidato está 3,25 desvios-padrão acima da média.

Medidas de assimetria e curtose

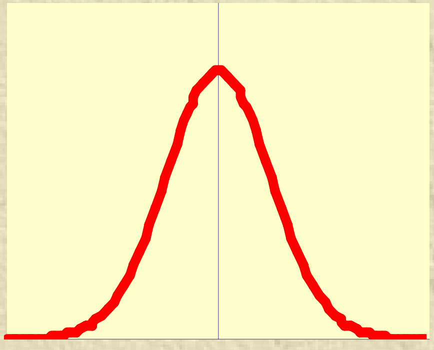
- As características mais importantes são o grau de deformação ou assimetria e o grau de achatamento ou afilamento da curva de freqüências ou do histograma, chamado curtose.

Assimetria: ***skewness***

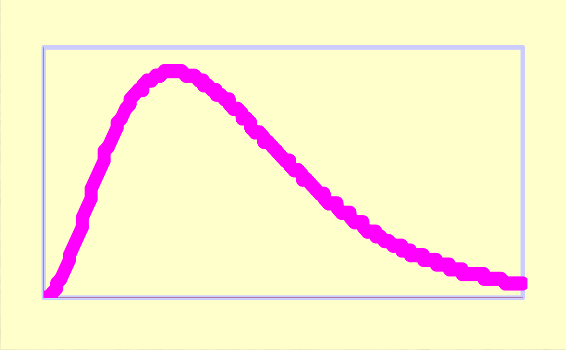
Coeficiente = 0 (Simétrica)

Coeficiente > 0 (Assimetria positiva)

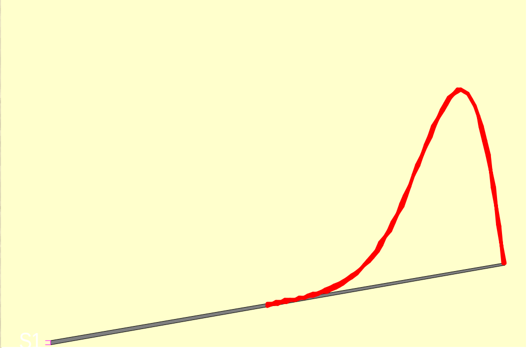
Coeficiente < 0 (Assimetria negativa)



Média=Mediana=Moda



Moda < Mediana < Média



Moda > Mediana> Média

Cálculo da assimetria

- Conhecido como primeiro coeficiente de assimetria de Pearson
- S: desvio padrão amostral
- Mo: moda
- $\overline{X}$: média
- Assimetria assume valores entre -1 e +1

$$Sk = \frac{\overline{X} - Mo}{S}$$

Curtose: ***kurtosis***

- Coeficiente de curtose de Pearson

$$a_4 = m_4/s^4, \text{ onde } m_4 = \Sigma(X - \bar{X})^4/n$$

$a_4 = 3$ (Mesocúrtica)

$a_4 > 3$ (Leptocúrtica)

$a_4 < 3$ (Platocúrtica)

- A distribuição normal tem curtose igual a 3