CURSO DE SPSS
AULA 2

MEDIDAS DESCRITIVAS

UFBA/FACED

José Albertino Lordello

Sheila Regina Pereira

MEDIDAS RESUMO

Uma maneira conveniente de descrever um grupo como um todo é achar um número único que represente o que é médio, ou típico, daquele conjunto de dados, ou seja, **é a tendência de dados quantitativos de se agruparem ao redor de um valor central.**

As medidas de tendência central geralmente são localizadas mais para o meio, ou o centro de uma distribuição, onde a maior parte dos dados tende a se concentrar-se.

As medidas de tendência central mais comuns são a **média aritmética**, a **mediana** e **moda**.

Moda

A moda é o valor mais frequente, mais típico ou mais comum em uma distribuição.

A moda é a menos empregada. No entanto, é adequada para caracterizar situações onde estejam em causa os casos ou valores mais usuais.

É a única medida de tendência central de que dispomos para variável qualitativa nominal.

Bimodal: possui dois valores modais.
Amodal: não possui moda.
Multimodal: possui mais do que dois valores modais.

EXEMPLO:
A moda de {maçã, banana, laranja, laranja, laranja, pêssego} é laranja.

A série {1, 3, 5, 5, 6, 6} apresenta duas modas (BIMODAL): 5 e 6.

A série {1, 3, 2, 5, 8, 7, 9} não apresenta moda (AMODAL).

A série {1, 3, 5, 5, 6, 6, 7, 7} apresenta mais do que duas modas (MULTIMODAL): 5, 6 e 7

Média Aritmética

É a mais usada para descrever resumidamente uma distribuição de frequência. Há vários tipos de médias a mais utilizada é a média aritmética. Ela pode ser **simples** ou **ponderada.**

Definição: É um valor que representa um ponto de equilíbrio de uma distribuição. É o somatório do conjunto de dados dividido pelo tamanho da amostra.

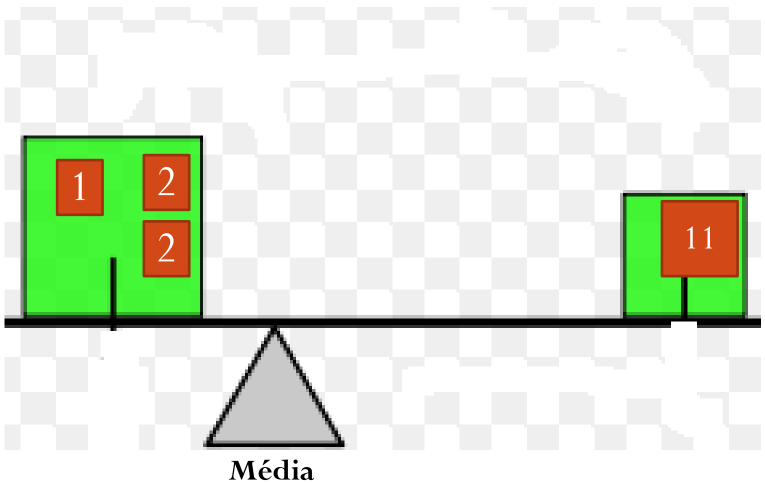
$$\mu = \frac{x_1 + \ldots + x_N}{N}$$

População

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n}$$

Amostra

È o ponto de uma distribuição em torno do qual os valores acima dele se equilibram com os que estão abaixo.



Exemplo: Rendimento Acadêmico dos estudantes da turma A do curso de engenharia civil

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

Média 6,8

- média é um valor típico, característico, do conjunto de dados;
- É a principal medida de tendência central;
- Leva em consideração todas as observações efetuadas;
- Porém é sensível a valores excessivamente pequenos ou grandes, em relação às demais observações do conjunto de dados.

Exemplo: Rendimento Acadêmico dos estudantes da turma A do curso de engenharia civil

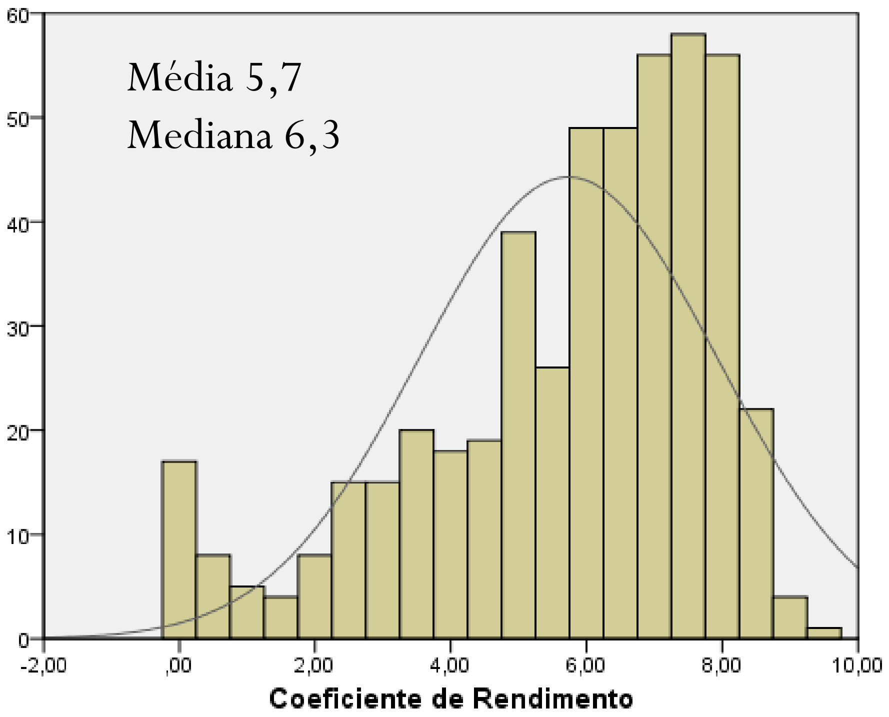
3,3	3,4	2,9	3,8	3,5	3,1	10	10

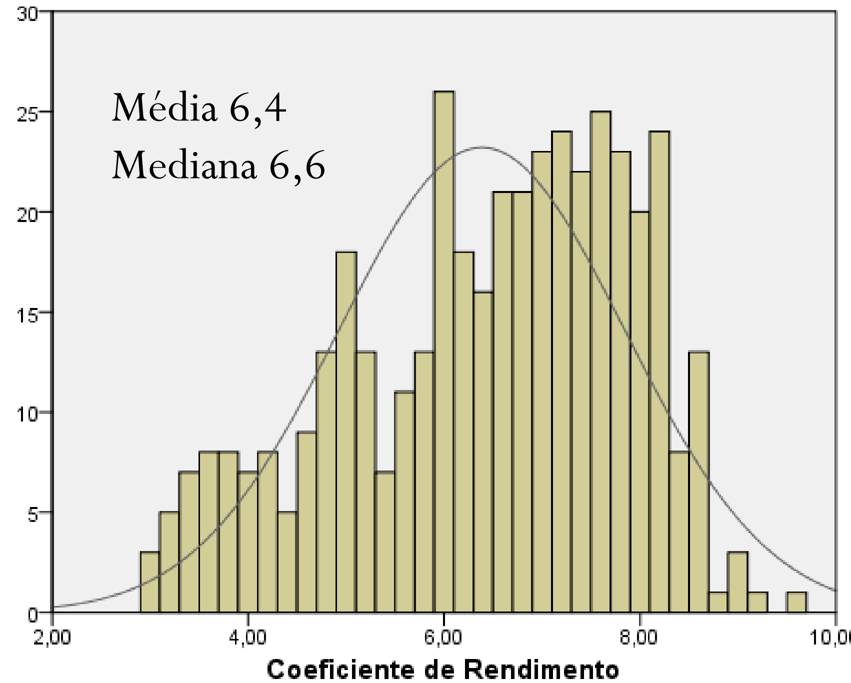
Média 5,0

Exemplo: salário médio mensal de cinco empregados de uma certa empresa.
Dados em reais: **123 145 210 225 2.500**

Média 640,6 reais

CR dos estudantes do curso de engenharia da UFBA em 2005





Mediana

- É o valor central em uma distribuição, quando o conjunto de dados está **ordenado**.
- Divide a distribuição em duas partes iguais, de modo que 50% dos valores observados são iguais ou inferiores ao valor mediano e 50% iguais ou superiores a esse valor;
- Se o total de observações for ÍMPAR, a mediana, é o valor que está localizado exatamente ao meio dos dados ordenados;
- Se o total de observações for PAR, a mediana é a média dos dois valores centrais.

Exemplo: Rendimento Acadêmico dos estudantes da turma A do curso de engenharia civil.

2,9	3,1	3,3	3,4	3,5	3,8	10	10

Média5,0   Mediana 3,45

Se *n* é ímpar:

$$Md = X_{\left(\frac{n+1}{2}\right)}$$

Se *n* é par:

$$Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$$

A mediana é uma medida de posição resistente, pois é pouco afetada por mudanças de pequena porção dos dados, ao contrário da média que é sensível a valores atípicos (discrepantes). O cálculo da mediana torna-se trabalhoso quando o número de observações é grande, devido a necessidade de ordenar os dados.

Exemplo Calcule a média e a mediana.

Conjunto 1 = 200, 250, 250, 300, 450, 460,510

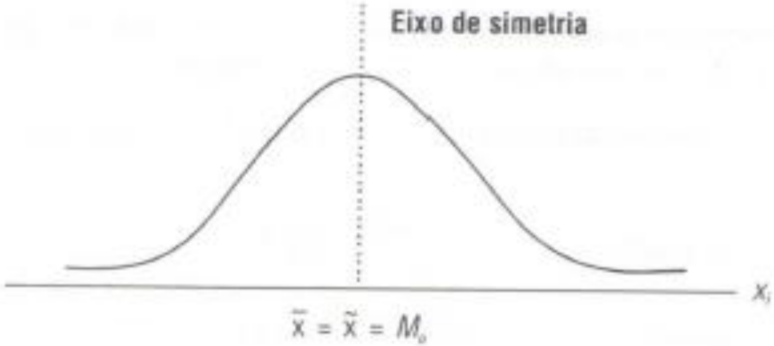
Média 345,71 mediana 300

Conjunto 2 = 200, 250, 250, 300, 450, 460 2.300

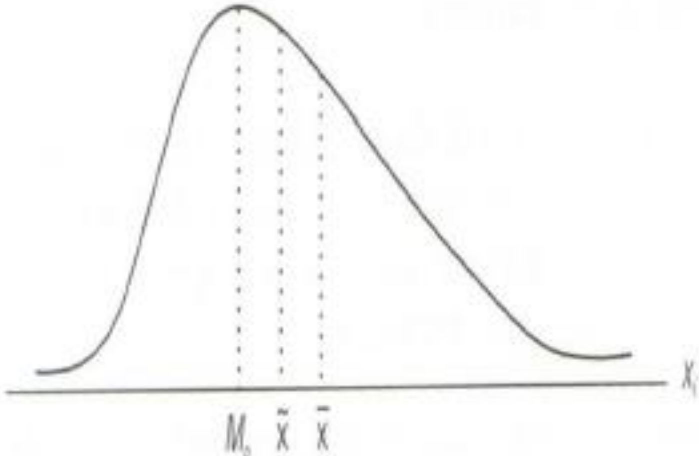
Média 601,43 mediana 250

Relação entre a média, moda e mediana

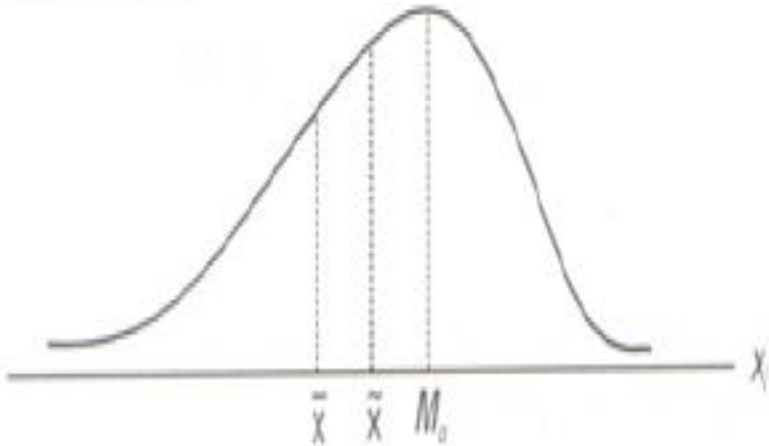
a) $\overline{X} = \widetilde{X} = Mo,$ distribuição é simétrica



b) $\overline{X} \geq \widetilde{X} \geq Mo,$ distribuição é assimétrica positiva ou à direita



c) $\overline{X} \leq \widetilde{X} \leq Mo,$ distribuição é assimétrica negativa ou à esquerda



- O calculo da mediana exige a ordenação das categorias, assim, a mediana só pode ser obtida para dados **ordinai**s ou **intervalares**.

- O uso da média está restrito exclusivamente a dados **intervalares**. Sua aplicação a dados ordinais ou nominais dá um resultado desprovido de sentido, que em geral não indica nenhuma tendência central.

- Para descrever uma distribuição assimétrica, o pesquisador em geral escolhe a mediana, porque ela tende a proporcionar uma representação equilibrada dos escores extremos.

- Para uma medida precisa de distribuições que são ao menos aproximadamente simétrica, a média tende a ser preferida em relação à mediana, porque pode ser usada facilmente em análises estatísticas mais avançadas.

-

Medidas de variabilidade

- Vimos anteriormente que um conjunto de dados podem ser sintetizado, por meio de procedimentos matemáticos, em poucos valores representativos - **Média aritmética, mediana e moda.**

- A análise de um conjunto de informação com base em uma única medida de tendência central não nos fornece informações suficientes para o conjunto de valores

Exemplo de motivação: A produção diária da peça Z de uma certa indústria foi observada em três empregados no período de 15 à 19 de abril de 2000.

Empregado	Dia					Média Diária
	1º	2º	3º	4º	5º	
Carlos	82	70	65	60	73	70
Daniel	60	78	68	62	82	70
Eduardo	53	72	75	75	75	70

Suponha que o interesse do administrador da empresa é que os empregados apresentem produção elevada e a mais homogênea possível. Qual dos três empregados apresentou melhor desempenho no trabalho no período observado?

Exemplo : Os dados abaixo referem-se as sentenças proferidas em caso de condenação por assalto. pelo juiz A e o juiz.

							Média
Juiz A	34	30	31	33	36	34	33,0
Juiz B	26	43	22	35	20	34	30,0

Para um advogado qual a melhor opção?

- Chamamos de dispersão ou variabilidade a maior ou menor diversificação dos valores de uma variável aleatória em torno de um valor de tendência central;

- As medidas de dispersão servem para avaliar o grau de variabilidade ou dispersão dos valores de um conjunto de dados;

- As medidas de dispersão permitem estabelecer comparações entre fenômenos de mesma natureza ou de natureza distintas e, em geral, essa variabilidade

Variância

É uma medida de variabilidade que utiliza todos os dados, baseada na diferença entre o valor de cada observação , xi ; e a média.

Variância Populacional

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Variância Amostral

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

Desvio Padrão

É a medida de dispersão mais usada e mais importante. Assim como a variância, mede a concentração dos dados em torno da média. Porém, tem a unidade de medida igual a unidade de medida original da variável.

Desvio-Padrão Populacional

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Desvio-Padrão Amostral

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

Interpretação do desvio padrão (análoga à da variância)

O desvio padrão mede a variação entre valores. Assim:

- Se os valores estiverem próximos uns dos outros, então o desvio padrão será pequeno, e consequentemente os dados serão mais homogêneos.

- Se os valores estiverem distantes uns dos outros, então o desvio padrão será grande, e consequentemente os dados serão heterogêneos.

Exemplo: Suponhamos que o juiz A e o juiz B acusem uma média **24 meses** em suas sentenças de prisão proferidas em caso de condenação por assalto.

Poderíamos facilmente ser levados a pensar que os juízes concordam em suas filosofias quanto a uma sentença adequada.

							Média	Var	SD	CV (%)
Juiz A	34	30	31	33	36	34	33,0	4,0	2,0	6,1
Juiz B	26	43	22	35	20	34	30,0	65,0	8,1	27,0

Com base nas evidências, poderíamos dizer que o juiz A é mais rigoroso

O juiz B é menos rigoroso, porém inconsistente.

Para um advogado, a melhor opção seria o juiz A. Mesmo correndo o risco de uma sentença mais longa (em virtude da média mais alta), ele por certo não ariscaria ver seu cliente submetido às severas sentenças que o juiz B costuma proferir.

Coeficiente de Variação de Pearson

- Trata-se de uma medida relativa de dispersão, útil para comparar a variabilidade de duas ou mais distribuições, mesmo quando essas se referem a diferentes fenômenos e sejam expressas em unidades de medida distintas.

- É empregada para estimar a precisão de experimentos e representa o desvio padrão expresso como porcentagem da média. **Sua principal qualidade é a capacidade de comparação de distribuições diferentes.**

Definição: O Coeficiente de Variação de Pearson para um conjunto de n observações é definido como sendo o quociente entre o desvio padrão e a média aritmética da distribuição.

$$CV = \frac{\sigma}{\overline{X}} \qquad\qquad CV = \frac{S}{\overline{X}}$$

em que, σ é o desvio padrão populacional e S é o desvio padrão amostral. geralmente o CV é dado em %.

- Se CV < 15% : tem-se baixa dispersão
- Se 15% CV < 30%: tem-se média dispersão
- Se CV = 30% : tem-se elevada dispersão

Exemplo de motivação: A produção diária da peça Z de uma certa indústria foi observada em três empregados no período de 15 à 19 de abril de 2000.

						média	VAR	SD	CV (%)
Carlos	82	70	65	60	73	70	69,5	8,3	11,9
Daniel	60	78	68	62	82	70	94	9,7	13,9
Eduardo	53	72	75	75	75	70	92	9,6	13,7

Exercício 1: Calcular as medidas descritivas do banco de Dados1.

Próxima aula

Introdução Ao SPSS

1. Identificar o SPSS

2. Montar arquivos de dados

3. Operar comandos básicos dos menus do SPSS

4. Gerar e interpretar tabelas de cruzamento de dados.

5. Calcular medidas de dispersão de dados.