

MAE116 – Farmácia 2017

Estatística Descritiva (I)

O que é Estatística

- A estatística desempenha o papel importante em muitos processos de tomada de decisão. Um pesquisador, em muitas situações, necessita trabalhar os dados relevantes ao seu particular objeto de estudo, transformá-los em informações com finalidade de compará-los com outros resultados, ou ainda para julgar sua adequação a alguma teoria.

O que é Estatística

- **Exemplo 1:** antes que um novo remédio genérico possa ser colocado no mercado, a ANVISA exige que seja submetido a um estudo de bioequivalência – estudo experimental envolvendo indivíduos humanos saudáveis ou pacientes controlados. Os dados desse estudo precisam ser compilados e analisados para que seja determinado se o remédio é efetivo e seguro.

O que é Estatística

- **Exemplo 2:** a questão relacionada aos programas de Previdência Social, as decisões do governo devem ser tomadas baseadas nas previsões sobre a longevidade da população da nação, ou seja, deveríamos ser capazes de prever o número de anos que cada indivíduo viverá.

O que é Estatística

- **Exemplo 3:** muitas outras questões também necessitam ser respondidas, por exemplo, onde o governo deve investir seus recursos para reduzir a mortalidade infantil? O fato de ter sido amamentado pela mãe é um fator de proteção para o câncer de mama? O consumo de grande quantidade de álcool nas mulheres grávidas prejudica a saúde de seus bebês? Que fatores aumentam o risco de um indivíduo desenvolver uma doença cardíaca coronariana?

O que é Estatística

- A análise da estatística explora a coleta, a organização, a análise e a interpretação dos dados numéricos.
- Portanto, a estatística é uma ciência (ou método) baseada na teoria das probabilidades, cujo objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir de informações numéricas.

Procedimentos estatísticos

1. **Análise descritiva**
2. **Probabilidade**
3. **Inferência estatística**

Estatística Descritiva

Etapa inicial da análise utilizada para descrever, organizar e resumir os dados coletados.

A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou esta área da Estatística.

Probabilidade

A teoria das probabilidades nos permite modelar fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza. É uma ferramenta fundamental para inferência estatística.

Inferência Estatística

procura os argumentos estatísticos para fazer afirmações sobre as características de uma população, com base em informações dadas por amostras.

Exemplo 1: observe como uma cozinheira verifica se o prato que ela está preparando está pronto ou não.

Exemplo 2: como você concluiria se um bolo está delicioso ou não.

Essas são decisões baseadas em procedimentos amostrais.

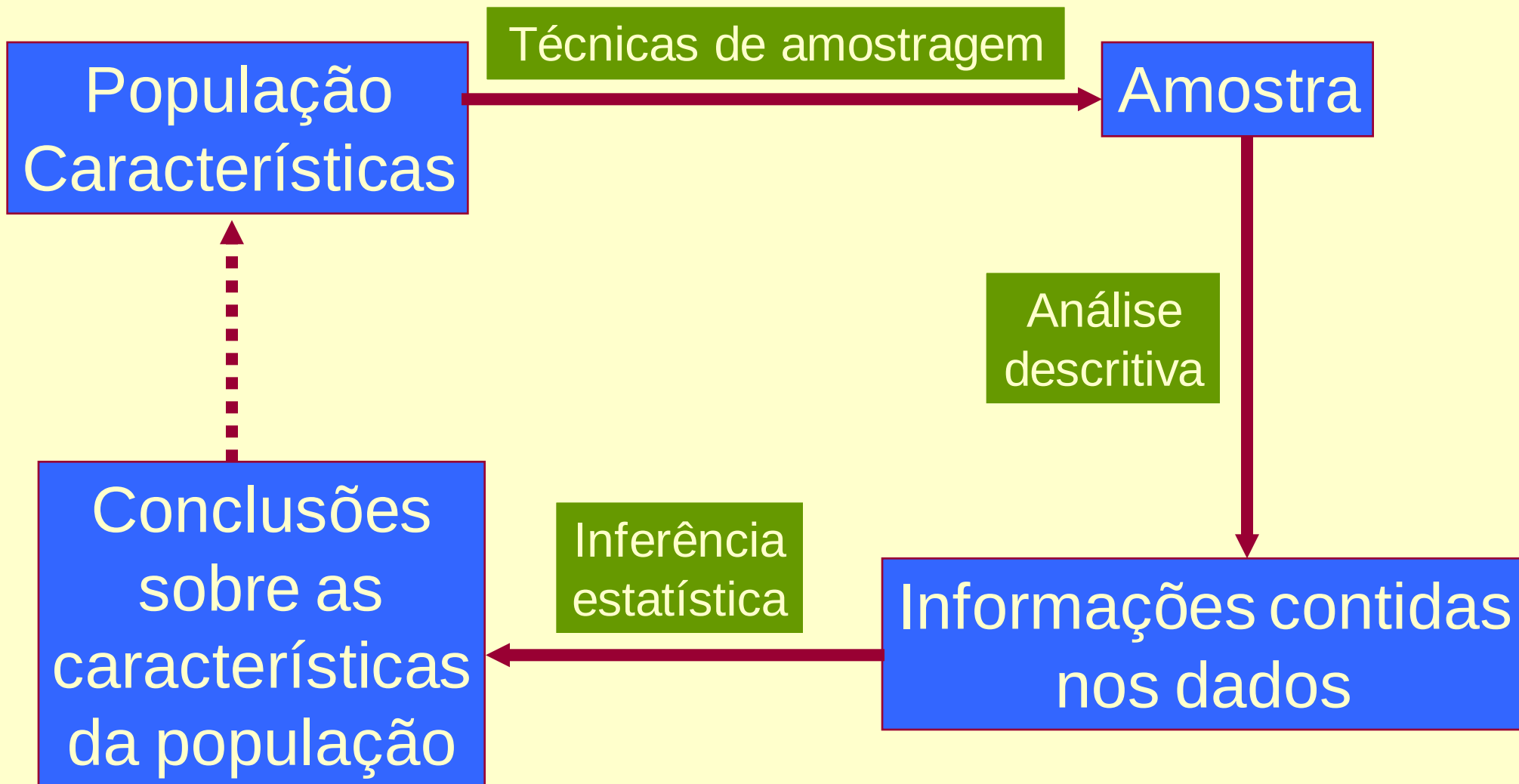
Amostragem

Associada a coleta de dados, a tecnologia da amostragem desenvolveu um conjunto de técnicas para obtenção de amostras representativas da população de interesse

Exemplos de utilização:

Pesquisa de mercado, pesquisa de opinião pública, ensaios de medicamentos e em praticamente todo experimento.

Estatística



Conceitos Básicos

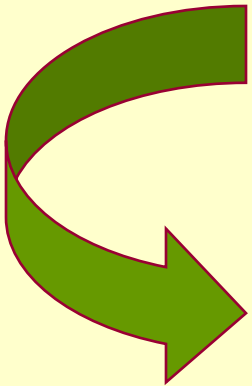
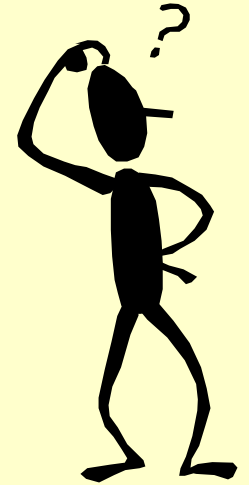
- **População:** é o conjunto de todos os elementos ou resultados sob investigação.
- **Amostra:** é qualquer subconjunto da população.
- **Parâmetro:** é uma medida (constante) usada para descrever uma característica da população. Por exemplo, média e variância populacional da variável de interesse X , $\mu(X)$ e $\sigma^2(X)$, respectivamente.

Conceitos Básicos

- **Estatística:** uma característica da amostra, ou seja, uma estatística T é uma função da amostra $(X_1, X_2, \dots, X_n)$ da variável de interesse X . Por exemplo, média amostral de X , que também é uma variável aleatória.
- **Variável:** qualquer característica associada a uma população, por exemplo, peso, altura da população brasileira.

Estatística Descritiva

O que fazer com as observações
que coletamos?



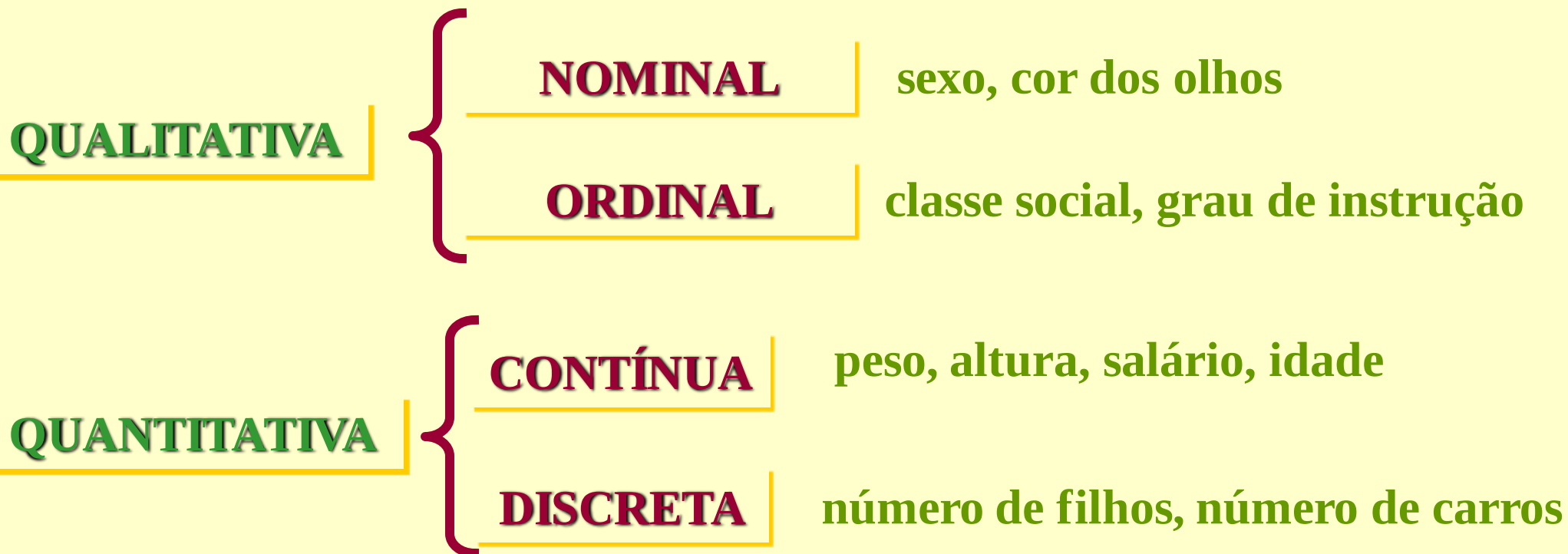
Primeira Etapa:

Resumo dos dados = Estatística descritiva

Variável:

Qualquer característica associada a uma população.

Classificação das variáveis:



Variável:

Classificação das variáveis:



Variáveis Quantitativas

MEDIDAS DE POSIÇÃO:

Mínimo, Máximo, Moda, Média, Mediana, Percentis

MEDIDAS DE DISPERSÃO:

Amplitude, Intervalo-Interquartil, Variância, Desvio Padrão, Coeficiente de Variação.

Medidas de Posição

- **Máximo (max)**: a maior observação
- **Mínimo (min)**: a menor observação
- **Moda (mo)**: é o valor (ou atributo) que ocorre com maior frequência.

Dados: 4, 5, 4, 6, 5, 8, 4

$$\text{max} = 8$$

$$\text{min} = 4$$

$$\text{mo} = 4$$

- Média:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Dados: 2, 5, 3, 7, 8

$$\bar{X} = \frac{2 + 5 + 3 + 7 + 8}{5} = 5$$

- **Mediana:**

A mediana é o valor da variável que ocupa a **posição central** de um conjunto de n **dados ordenados**.

Posição da mediana: $\frac{n+1}{2}$

Exemplos:

Dados: 2, 6, 3, 7, 8 $\Rightarrow n = 5$ (ímpar)

Dados ordenados: 2 3 6 7 8 $\Rightarrow \frac{5+1}{2} = 3 \Rightarrow Md=6$

Posição da Mediana $\uparrow$

Dados: 4, 8, 2, 1, 9, 6 $\Rightarrow n = 6$ (par)

Dados ordenados: 1 2 4 6 8 9 $\Rightarrow \frac{6+1}{2} = 3,5$

$\uparrow$
 Md

$$Md = (4 + 6) / 2 = 5$$

Percentis:

O percentil de ordem $p \times 100$ ($0 < p < 1$), em um conjunto de dados de tamanho n , é o valor da variável que ocupa a posição $p \times (n + 1)$ do conjunto de dados ordenados.

Casos particulares:

percentil 50 = mediana ou segundo quartil (Md)

percentil 25 = primeiro quartil (Q_1)

percentil 75 = terceiro quartil (Q_3)

percentil 10 = primeiro decil

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7 $\Rightarrow n=10$

Posição de Md : $0,5(n+1) = 0,5 \times 11 = 5,5 \Rightarrow Md = (3 + 3,1)/2 = 3,05$

Posição de $Q1$: $0,25 (11) = 2,75 \Rightarrow Q_1 = (2 + 2,1)/2 = 2,05$

Posição de $Q3$: $0,75 (11) = 8,25 \Rightarrow Q_3 = (3,7 + 6,1)/2 = 4,9$

$$Md = 3,05$$

$$Q_1 = 2,05$$

$$Q_3 = 4,9$$

Dados: 0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0 33,6

$$\Rightarrow n=11$$

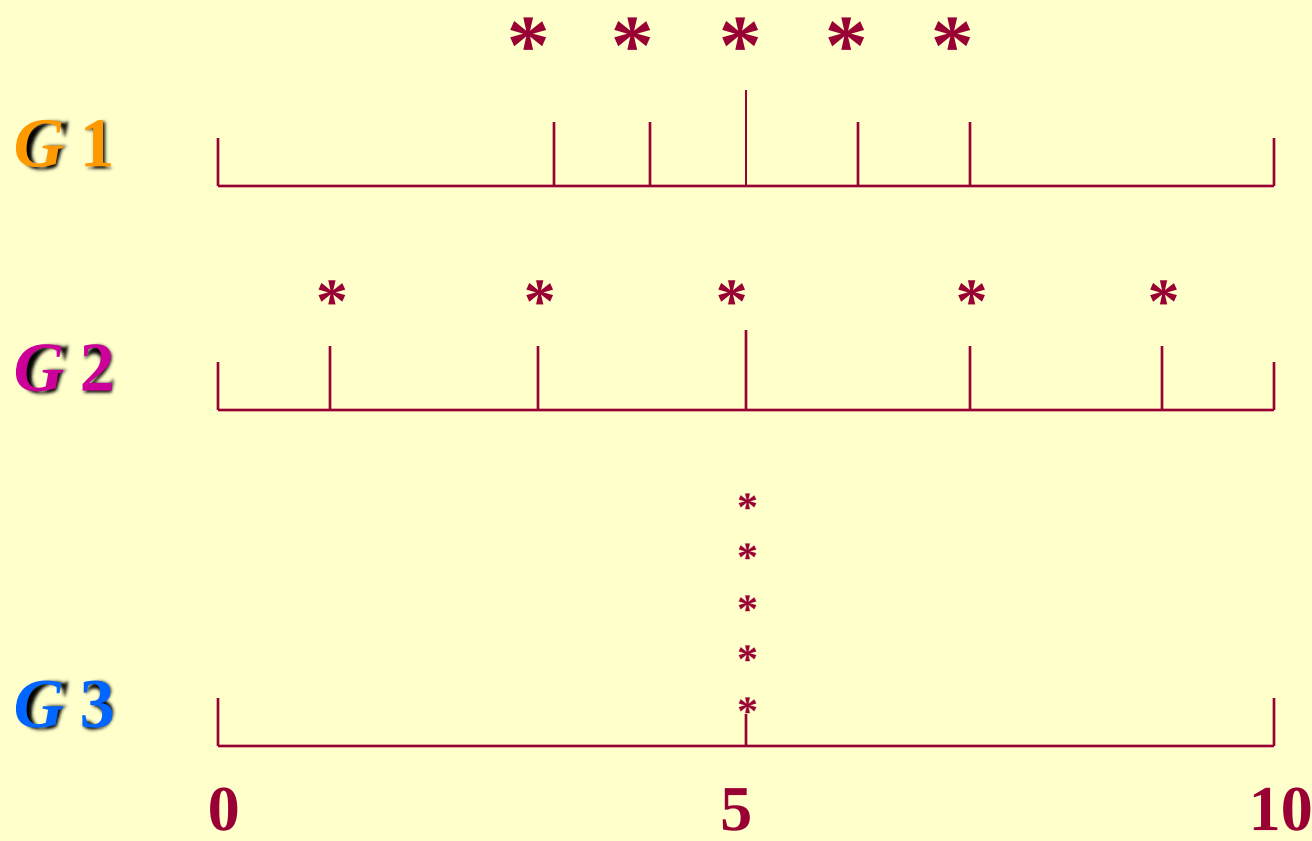
$$Md = 5,3$$

$$Q1 = 1,7$$

$$Q3 = 12,9$$

Exemplo 2: Considere as notas de um teste de 3 grupos de alunos

Grupo 1: 3,4,5,6,7 Grupo 2: 1, 3, 5, 7, 9 Grupo 3: 5,5,5,5,5



Temos: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$ e $md_1 = md_2 = md_3 = 5$

Medidas de Dispersão

Finalidade: encontrar um valor que resuma a variabilidade de um conjunto de dados.

O resumo de um conjunto de dados por uma única medida representativa de posição central esconde toda a informação sobre a variabilidade do conjunto de observações.

- **Amplitude (A):**

$$A = \textit{máx} - \textit{min}$$

Para os grupos anteriores, temos:

Grupo 1, $A = 4$

Grupo 2, $A = 8$

Grupo 3, $A = 0$

•Intervalo-Interquartil:

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja, $Q3 - Q1$.

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7

$$Q1 = 2,05 \quad e \quad Q3 = 4,9$$

$$Q3 - Q1 = 4,9 - 2,05 = 2,85$$

- Variância:

$$\text{Variância} = s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

- Desvio padrão:

$$\text{Desvio Padrão} = s = \sqrt{\text{Variância}}$$

Cálculo para os grupos:

$$\mathbf{G1: } s^2 = \frac{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{4}$$

$$\Rightarrow s^2 = 10/4 = 2,5 \Rightarrow s = 1,58$$

$$\mathbf{G2: } s^2 = 10 \Rightarrow s = 3,16$$

$$\mathbf{G3: } s^2 = 0 \Rightarrow s = 0$$

Fórmula alternativa:

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

Em **G1**: $\sum X_i^2 = 9 + 16 + 25 + 36 + 49 = 135$

$$\Rightarrow S^2 = \frac{135 - 5 \times (5)^2}{4} = 2,5$$

• Coeficiente de Variação (CV)

- é uma medida de dispersão relativa
- elimina o efeito da magnitude dos dados
- exprime a variabilidade em relação à média

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Exemplo 3:

Altura e peso de alunos

	Média	Desvio Padrão	Coef. de Variação
Altura	1,143m	0,063m	5,5%
Peso	50 kg	6kg	12%

Conclusão: Os alunos são, aproximadamente, duas vezes mais dispersos quanto ao peso do que quanto à altura.

Exemplo 4:

Altura (em *cm*) de uma amostra de recém-nascidos e de uma amostra de adolescentes

	Média	Desvio padrão	Coef. de variação
Recém-nascidos	50	6	12%
Adolescentes	160	16	10%

Conclusão: Em relação às médias, as alturas dos adolescentes e dos recém-nascidos apresentam variabilidade quase iguais.

Exemplo 5:

5 voluntários submeteram-se a um estudo de bioequivalência, obtendo-se as seguintes valores de $T_{\max}$ (h):

Para medicamento referência (X): 1,5; 1,25; 1,5; 1,25 e 1,75

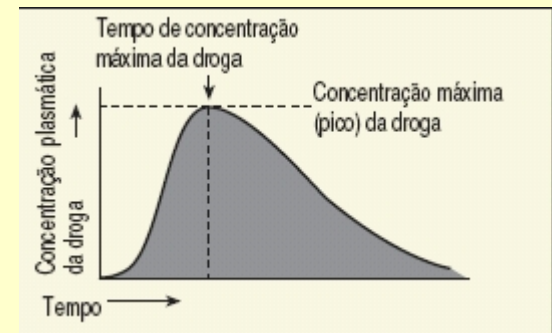
Para medicamento teste (Y): 1,25; 0,75; 1,5; 1,25; 2,5

$$\bar{X} = \bar{Y} = 1,45$$

Amplitude (A): $A = \text{máx} - \text{min}$

➤ grupo referência, $A = 1,75 - 1,25 = 0,5$

➤ grupo teste: $A = 2,5 - 0,75 = 1,75$.



Referência: $s^2 = \frac{(1,5 - 1,45)^2 + (1,25 - 1,45)^2 + (1,5 - 1,45)^2 + (1,25 - 1,45)^2 + (1,75 - 1,45)^2}{5 - 1} = 0,04375$

$$S = 0,2092, \text{ CV} = 0,2092 / 1,45 = 14,43\%$$

Teste: $s^2 = \frac{(1,25 - 1,45)^2 + (0,75 - 1,45)^2 + (1,5 - 1,45)^2 + (1,25 - 1,45)^2 + (2,5 - 1,45)^2}{5 - 1} = 0,41875$

$$S = 0,6471, \text{ CV} = 0,6471 / 1,45 = 44,63\%.$$