Professor: Adriano Sales
Matéria: Teoria Estatística

O que é Estatística

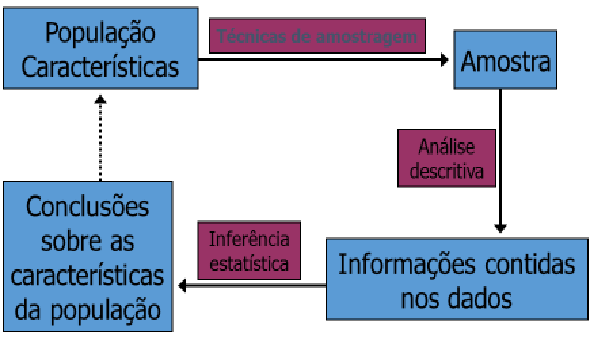
Para muitos, a Estatística não passa de conjuntos de tabelas de dados numéricos. Os estatísticos são as pessoas que coletam esses dados.

A Estatística originou-se com a coleta e construção de tabelas de dados para o governo.
A situação evoluiu e esta coleta de dados representa somente um dos aspectos da Estatística.

No século XIX, o desenvolvimento do cálculo de probabilidade e outras metodologias matemáticas, tais como a técnica de Mínimos Quarados, foram fundamentais para o desenvolvimento da Estatística.

Somente no século XX a Estatística desenvolve-se como uma área específica do conhecimento a partir do desenvolvimento da Inferência Estatística; uma metodologia baseada em probabilidade que tem ampla aplicação nas ciências experimentais.

A Estatística hoje consiste numa metodologia cientifica para obtenção, organização e análise de dados, oriundos das mais variadas áreas das ciências experimentais, cujo objetivo principal é auxiliar a tomada de decisões em situações de incerteza.



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 Descritiva

Etapa inicial da análise utilizada para descrever e resumir os dados.
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 os fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza.
É uma ferramenta fundamental para inferência estatística.

Inferência Estatística

Um conjunto de técnicas baseadas em probabilidade, que a partir de dados **amostrais** nos permite tirar conclusões sobre a **população** de interesse.

Exemplo 1:

Numa pesquisa eleitoral, um instituto de pesquisa tem como objetivo prever o resultado da eleição, utilizando uma amostra da população.

Considere o Candidato "A":

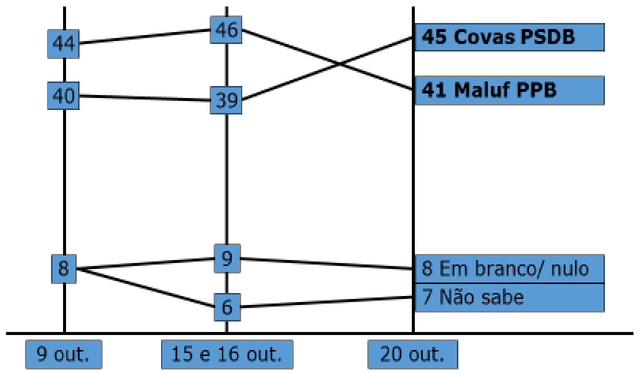
Denomine por p a proporção de pessoas (na população) que votarão em "A" na eleição.
Denomine por p a proporção de pessoas no levantamento de opinião que expressam intenção de voto em "A".
Estimação: Podemos usar o valor de p para estimar a proporção p da população.

Na eleição presidencial, para governadores e prefeitos, os institutos de pesquisa de opinião colhem periodicamente amostras de eleitores para obter as estimativas de intenção de voto da população. As estimativas são fornecidas com um valor e uma margem de erro.

IBOPE / Opinião

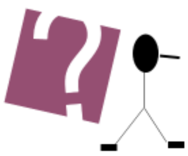
Os quadros do IBOPE/Opinião a seguir referem-se à intenção de voto para prefeito de São Paulo para o primeiro e segundo turno das eleições de 2004. Estimar a proporção p da população.
A resposta foi estimula e única.
Pergunta realizada: Se a eleição para prefeito fosse hoje e os candidatos fossem estes, em quem o (a) Sr. (Sra) votaria?

Governador de São Paulo – 1998
Resposta estimulada e única, em % do total de votos.
A última pesquisa ouviu 3.389 eleitores – Margem de erro de 2%.



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:

- QUALITATIVA
  - NOMINAL: sexo, cor dos olhos
  - ORDINAL: classe social, grau de instrução
- QUANTITATIVA
  - CONTÍNUA: peso, altura, salário, idade
  - DISCRETA: número de filhos, número de carros

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 (máx): a maior observação
Mínimo (min): a menor observação
Moda (Mo): é o valor (ou atributo) que ocorre com maior freqüência.

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

Xmáx = 8 Xmin = 4 Mo = 4

Média:

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

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 → n = 5 (ímpar)

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

Posição da Mediana

Dados: 3, 5, 2, 1, 8, 6 → n = 6 (par)

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

Md = $\frac{3+5}{2} = 4$ Md

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 (D_1)

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7 → n=10

Posição de Md: 0,5(n+1) = 0,5×11= 5,5 → Md = (3+3,1)/2 = 3,05
Posição de Q1: 0,25 (11) = 2,75 → Q_1 = (2+2,1)/2 = 2,05
Posição de Q3: 0,75 (11) = 8,25 → Q_3 = (3,7+6,1)/2 = 4,9

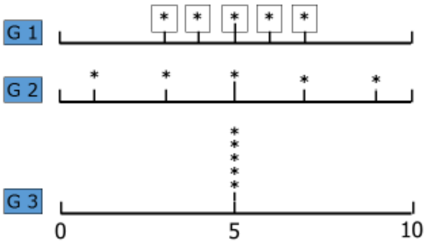
Md = 3,05 Q1 = 2,05 Q3 = 4,9

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

Amplitude (A):

$$A = X_{\max} - X_{\min}$$

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 e Q3 = 4,9

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

Variância:

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

Desvio padrão:

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

Cálculo para os grupos:

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

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

$$G2: s^2 = 10 \quad s = 3,16$$

$$G3: s^2 = 0 \quad s = 0$$

Fórmula alternativa:

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

Em G1: SX_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	6 kg	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%