

Apresentação da Estatística

A Estatística estuda fenômenos que envolvem conjuntos de dados, formados por elementos que apresentam semelhanças entre si com relação a certos atributos e diferenças com relação a outros.

A Estatística e a Teoria das Probabilidades procuram fornecer ferramentas e técnicas quantitativas para o tratamento desses conjuntos de dados, permitindo sua descrição e a obtenção e a interpretação das suas propriedades.

Praticamente todas as áreas do conhecimento requerem algum uso da Estatística e da Teoria das Probabilidades, desde as ciências naturais até as ciências sociais. Portanto, é importante que os profissionais atuantes nessas disciplinas tenham pelo menos os conceitos básicos e alguma habilidade operacional com a Estatística, para poder desenvolver suas funções com um mínimo de autonomia.

Brevíssimo Resumo Histórico

- Levantamentos estatísticos vêm sendo feitos desde a Antigüidade, no Egito, Mesopotâmia, China, Grécia, Roma etc, tendo por objetivo a elaboração de censos, cadastros etc. Na Idade Média, a instituição que mais se preocupou em levantar dados estatísticos foi a Igreja.
- Origem do termo *estatística*: Achenwall (Göttingen, Alemanha, Séc. XVIII). "Ciência das coisas que pertencem ao Estado, chamando Estado a tudo que constitui uma sociedade civil e ao país em que ela habita, com tudo quanto se encontra de ativo e de efetivo; a estatística ocupa-se dos fenômenos que podem favorecer ou defender a prosperidade do

Estado. A política ensina como devem ser os estados, a estatística explica como o são realmente."

- Graunt (Londres, Inglaterra, Séc. XVIII). Utilizando dados demográficos retirados das paróquias de Londres sobre o número de mortes pela peste, estimou a população da cidade. Precursor do atual campo da *inferência estatística*.
- *Teoria das probabilidades*. Desenvolvida por matemáticos dos Sécs. XVII-XIX interessados inicialmente em jogos de azar: Cardano, Fermat, Pascal, de Moivre, Bernoulli, Poisson, Laplace, etc. Este último estabeleceu uma síntese da teoria das probabilidades e da estatística, lançando as bases da moderna *estatística matemática*.
- *Teoria dos erros*. Desenvolvida para o tratamento de dados experimentais em ciências naturais (física, química etc): Gauss, Bessel, Laplace, Quetelet (Séc. XIX).
- Estatística moderna (Sécs. XIX-XX). Testes de significância, análise da variância, regressão e correlação, teoria da amostragem, processos estocásticos, inferência bayesiana, etc: Quetelet, Gauss, Bayes, Galton, Pearson, Gosset, Charlier, Fisher, Neyman e muitos outros.

Atividade para casa (não precisa entregar, é apenas para praticar)

Procure na Web (use o Google: <http://www.google.com> ou outra ferramenta) por páginas sobre “estatística” e/ou “teoria das probabilidades” e algum tema de seu interesse (tente em português e inglês).

- a) Quantas páginas foram encontradas em português?
- b) E em inglês?
- c) Passeie livremente pelas páginas encontradas (em português e inglês). Observe o escopo e a abrangência das aplicações da estatística e da teoria das probabilidades no tema escolhido. Tente levantar em quantas

áreas do seu tema existem aplicações e quais as naturezas delas (isto é, descrição de dados, testes estatísticos, modelagem etc).

Divisões da Estatística

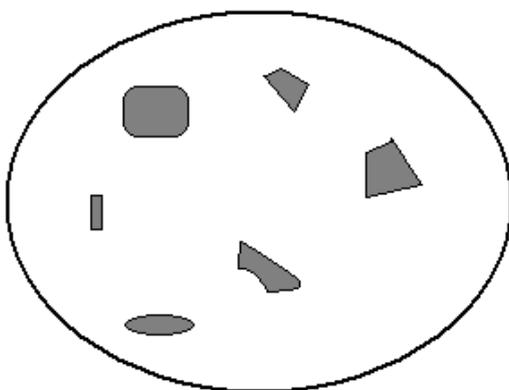
A metodologia estatística pode ser separada em duas componentes:

- Estatística descritiva, ou dedutiva: trata da apresentação de dados em gráficos e tabelas e do cálculo de parâmetros numéricos para descrição de dados, tais como frequências, médias, medianas, porcentagens e faixas de variação.
- Inferência estatística, ou estatística indutiva: trata de técnicas que permitam tirar conclusões ou tomar decisões sobre uma população a partir de evidências apresentadas pelos dados numéricos relativos à população, ou a uma amostra dela.

Estatística Descritiva

População e Amostra

A figura abaixo ilustra o que se quer dizer em estatística por população e amostra.



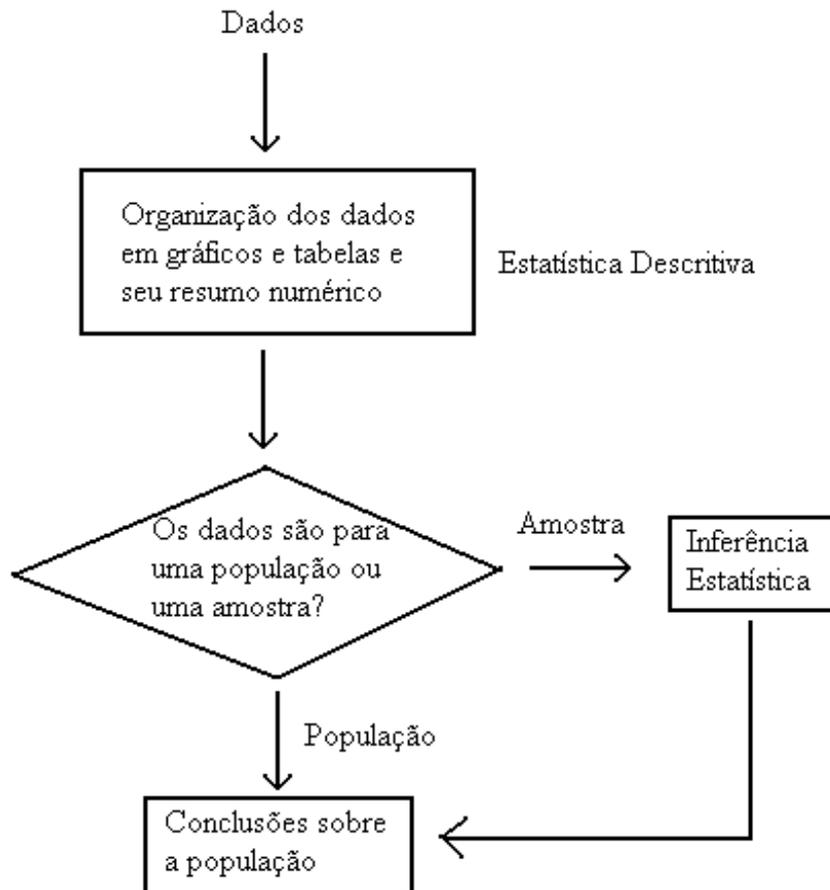
Área total: população = conjunto total dos elementos do grupo.

Área pintada: amostra = dados do grupo que são realmente observados ou medidos.

Exemplos:

- Números de horas semanais dedicadas ao estudo de um grupo de 32 estudantes de graduação de uma universidade (amostra) escolhidos aleatoriamente do conjunto total de estudantes de graduação da universidade (população).
- Tempos de resposta a um estímulo luminoso de um grupo de 40 estudantes com idades entre 18 e 20 anos (amostra) que, por hipótese, representam todos os estudantes com idades entre 18 e 20 anos que já existiram, que existem e que vão existir (população).
- Níveis de glicose no sangue de um grupo de 15 pacientes (amostra) selecionados aleatoriamente de uma lista de pacientes diabéticos de um hospital (população).
- Preços das ações de um grupo de 20 empresas cujas ações são negociadas na BOVESPA (amostra) selecionadas aleatoriamente do conjunto total de empresas cujas ações são negociadas na BOVESPA (população).

O diagrama a seguir ilustra a metodologia da análise estatística:



Este diagrama ilustra também a importância da amostragem. Se a amostra não for representativa da população, toda a análise estatística, por mais sofisticados que sejam os métodos matemáticos adotados, pode ser invalidada.

Dados: registros de medidas ou observações feitas.

Tipos de Dados ou Variáveis:

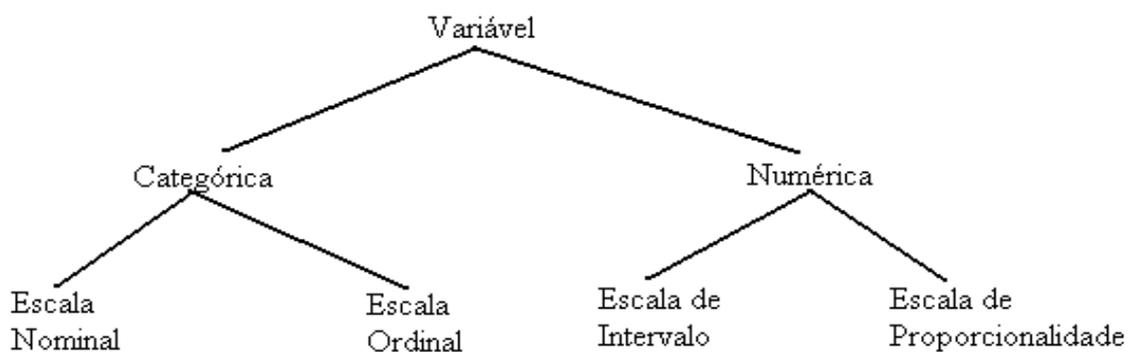
- **Numéricos:** resultam de medidas numéricas, como peso, altura, tempo e distância; ou de contagens, como número de filhos, número de televisores em casa e número de bactérias em uma colônia.

- **Catagóricos:** resultam de observações sobre características, como sexo, raça, religião, tipo de ocupação econômica, tipo de dieta ao qual foi submetido etc.
- Os dados catagóricos são, em geral, *qualitativos*, enquanto que os dados numéricos são *quantitativos*.

Tipos de dados numéricos

- **Discretos:** resultam de um processo de contagem e podem assumir os valores inteiros: 0, 1, 2, 3, ... Exemplos: quantos filhos um casal tem; quantos exemplares de uma certa espécie vegetal existem em uma floresta.
- **Contínuos:** resultam de um processo de medição e podem assumir qualquer valor dentro de um intervalo contínuo, dependendo da precisão do instrumento de medida. Exemplos: temperatura; altura de uma pessoa; tempo de duração de uma reação química.

Escalas para dados ou variáveis catagóricas e numéricas



Variáveis Categóricas

- Escala Nominal: as categorias às quais a variável pertence não podem ser ordenadas. Exemplos: sexo (masculino ou feminino); possui seguro de vida (sim ou não); cor preferida (azul, vermelho, verde etc).
- Escala Ordinal: as categorias às quais a variável pertence podem ser ordenadas. Exemplos: classe social (A, B, C, D ou E; ou simplesmente alta, média ou baixa); dureza de uma madeira (muito dura, dura, dureza média, pouco dura); tipo sanguíneo (A, B, AB ou O).

Variáveis Numéricas

- Escala de Intervalo: é uma escala ordenada em que as *diferenças* entre os valores das variáveis são significativas. Exemplos: temperatura (se a temperatura for de 30°, ela é dois graus mais alta do que 28°; da mesma forma, uma temperatura de 8° é dois graus mais alta que 6°); calendário (a diferença em anos entre 2001 e 1999 é igual à diferença entre 1997 e 1995, dois anos).
- Escala de Proporcionalidade: é uma escala de intervalo em que existe um zero absoluto. Exemplos: altura (uma pessoa com 2 m de altura é duas vezes mais alta do que uma com 1 m); valor depositado em conta bancária (uma pessoa com R\$ 10.000,00 na conta tem cinco vezes mais dinheiro do que uma com R\$ 2.000,00).

Dados Primários: são dados coletados por nós mesmos.

Exemplos:

- Dados obtidos por nossa observação. Ex.: quantas células foram contadas em uma observação ao microscópio; quanto tempo um animal gasta explorando uma dada região; quantas pessoas entraram em uma agência bancária em um dado dia.

- Dados que coletamos por experimentos. Ex.: medição das alturas e pesos de um grupo de pessoas; comportamento de uma amostra de ratos submetidos a uma dieta especial.
- Dados obtidos por questionários ou entrevistas. Ex.: dados censitários; questionário de avaliação da disciplina.
- Dados de informação pessoal. Ex.: informações sobre a situação interna de um departamento de uma universidade fornecidas por alguém que trabalha nele.

Dados Secundários: são dados coletados por outra(s) pessoa(s).

Exemplos:

- Estatísticas oficiais. Ex.: dados sobre o Brasil fornecidos pelo Instituto Brasileiro de Geografia e Estatística - IBGE (<http://www.ibge.gov.br>) ou sobre o Estado de São Paulo fornecidos pela Fundação Sistema Estadual de Análise de Dados – SEADE (<http://www.seade.gov.br/>);
- Estatísticas não-oficiais. Ex.: dados fornecidos por agências não-governamentais e institutos de pesquisa independentes: DataFolha (<http://www1.folha.uol.com.br/folha/datafolha/>), Instituto Gallup (<http://www.gallup.com/>), Instituto Vox Populi (<http://www.voxpopuli.com.br/>), SOS Mata Atlântica (<http://www.sosmatatlantica.org.br/>) etc;
- Estatísticas sobre instituições, organizações, empresas etc. Ex.: dados sobre um banco, uma universidade, uma indústria, um hospital etc.

Atividade para casa (também não é necessário entregar):

Explore os *sites* do IBGE (<http://www.ibge.gov.br>) e da Fundação SEADE (<http://www.seade.gov.br/>). Verifique o que os dois sites trazem sobre Ribeirão Preto. As informações são iguais? Obtenha o valor mais recente da população do município de Ribeirão Preto-SP pelas duas organizações.

Explore outros *sites* também, veja o que eles contêm e o que mais lhe interessa. Em particular, de uma olhada nos seguintes *sites*:

- Biblioteca Nacional:
<http://www.bn.br/Script/index.asp>
- DATASUS: <http://www.datasus.gov.br/>
- Base de dados tropical:
<http://www.bdt.fat.org.br/index>
- IPEA (Instituto de Pesquisa Econômica Aplicada):
<http://www.ipea.gov.br/>
- Banco Central do Brasil: <http://www.bc.gov.br/>
- SENAI: <http://www.senai.br/>
- FIESP: <http://www.fiesp.com.br/>
- Fundação Getúlio Vargas: <http://www.fgv.br/>
- CNI (Confederação Nacional da Indústria): <http://www.cni.org.br/>
- SIBI (Sistema Integrado de Bibliotecas da USP): <http://www.usp.br/sibi/>
- Biblioteca da UNICAMP:
<http://www.unicamp.br/bc/>
- IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia):
<http://www.ibict.br/>
- UNESCO: <http://www.unesco.org/>
- ONU: <http://www.un.org/>
- FAO: <http://www.fao.org/>
- WHO (World Health Organization):
<http://www.who.int/en/>
- CID-10 (Classificação Internacional das Doenças):
<http://www.datasus.gov.br/cid10/cid10.htm>
- Biblioteca Virtual de Direitos Humanos:
<http://www.direitoshumanos.usp.br/principal.html>

- BLAST (Basic Local Alignment Search Tool):
<http://www.ncbi.nlm.nih.gov/BLAST>
- ISO (International Organization for Standardization): <http://www.iso.ch>
- ABNT (Associação Brasileira de Normas Técnicas):
<http://www.abnt.org.br>
- ANSI (American National Standards Institute): <http://www.ansi.org/>
- CEN (European Committee for Standardization): <http://www.cenorm.be>
- W3C (World Wide Web Consortium): <http://www.w3c.org>
- FAPESP: <http://www.fapesp.br>
- CNPq: <http://www.cnpq.br>
- CAPES: <http://www.capes.gov.br/>
- ISSI (International Society for Scientometrics and Informetrics):
<http://www.issi-society.info/>
- NSF (National Science Foundation):
<http://www.nsf.gov/>
- NIH (National Institutes of Health):
<http://www.nih.gov/>
- PubMed (Serviço público da Biblioteca Nacional de Medicina dos Estados Unidos):
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>
- CiteSeer (Biblioteca digital de literatura científica mantida pela NEC):
<http://citeseer.nj.nec.com/cs>
- SCIRUS (Ferramenta de busca de artigos e informação científica na Internet mantida pela editora Elsevier): <http://www.scirus.com/>
- BioMedNet (Fornece acesso a uma grande quantidade de informação e artigos na área biomédica): <http://journals.bmn.com/journals>
- ISI (Web of Knowledge):
<http://isi2.isiknowledge.com/portal.cgi>

- Nature: <http://www.nature.com/>
- Science: <http://www.sciencemag.org/>
- Scientific American: <http://www.sciam.com>
- Scientific American Brasil:
<http://www2.uol.com.br/sciam/>
- HIWW (Health Informatics World Wide):
<http://www.hiww.org/>;
- PloS Biology (Public Library of Science Biology):
<http://biology.plosjournals.org/perlserv/?request=index-html&issn=1545-7885>;
- Neurosciences on the Internet:
(<http://www.neuroguide.com/>);
- Neuroscience for Kids:
(<http://faculty.washington.edu/chudler/interr.html>)
- International Society for Computational Biology:
(<http://www.iscb.org/>).