

Palestra - Matemática Aplicada

Marinho G. Andrade

SME/ICMC/USP

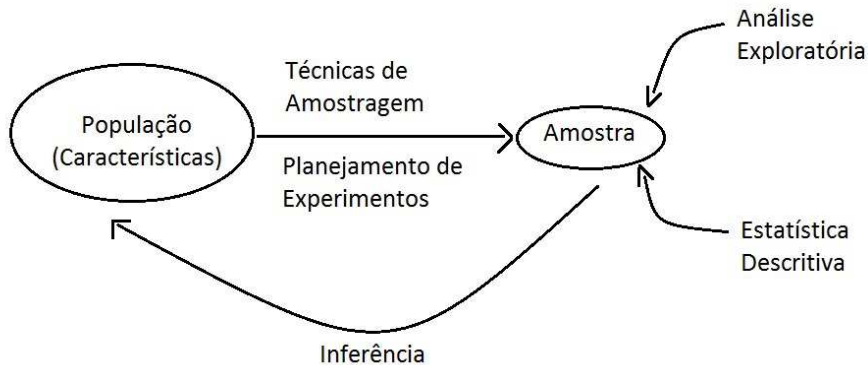
Maio de 2018

Primeiras Ideias

- Visão geral da abordagem estatística dos problemas;
- O papel da incerteza nos problemas científicos;
- A importância da coleta de dados (amostras);
- Os tipos de dados.

Análise Descritiva e Exploratória de Dados

● Ideia:



Algumas Definições

- **Dados:** São observações coletadas (por exemplo, medidas, sexo, respostas de um questionário, etc).
- **Estatística:** É uma ciência baseada na **Teoria da Probabilidade**, cujo o objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir dos dados.

Observação: A coleta de dados representa somente um dos aspectos da Estatística.

Resumo: Estatística \Leftrightarrow Incerteza

Algumas Definições

Informalmente, podemos definir “**Estatística**” da seguinte maneira:

- **Estatística:** É um conjunto de métodos para o planejamento de estudos e experimentos:
 - Obtenção de dados;
 - Organização dos dados;
 - Resumo ou redução dos dados;
 - Apresentação, análise e interpretação dos dados;
 - Elaboração de conclusão baseado nos dados;
 - Conclusão: de forma clara, amigável e objetiva.

Algumas Definições

- **População:** É uma coleção completa de todos os elementos a serem estudados e que possuem certa característica comum.

A coleção é completa no sentido de que inclui todos os sujeitos a serem estudados.

Exemplos de população:

- Brasileiros entre 16 e 25 anos;
- Crianças com menos de 1 ano no Brasil;
- Peças produzidas em uma linha de produção de uma fábrica;
- Cor de carros.

Classificação da População:

- **População finita:** indústrias situadas no Estado de São Paulo.
- **População infinita:** pressão atmosféricas ocorridas em diversos pontos do Continente em determinado momento.

Qual a característica que mais interessa em uma população?

Algumas Definições

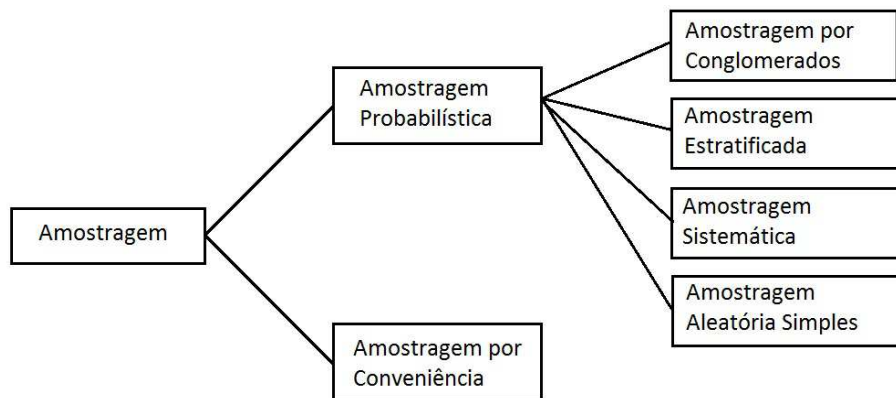
- **Variável:** Qualquer característica de interesse associada aos elementos de uma população.
- **Censo:** É um conjunto de dados obtidos de **todos** os membros da população.
- **Amostra:** É um subconjunto finito de membros selecionados de uma população.

Observação: Existem diferentes formas de se coletar dados, nem todos são adequados para uma análise estatística.

Coleta de dados

- Os dados amostrais devem ser coletados de modo apropriado, por exemplo: usando um processo de seleção aleatória.
- Se os dados não são coletados de modo apropriado podem ser inúteis para análise estatística.
- A coleta apropriada dos dados está relacionada diretamente aos objetivos do experimento e as conclusões que se deseja chegar.
- Existem diferentes técnicas de amostragem para coleta de dados. A ética é um ponto importante na elaboração da coleta de dados.

Técnicas de Amostragem



Técnicas de Amostragem

- **Amostragem por conveniência:** Coleta-se elementos de fácil acesso ou de interesse para o estudo.
- **Amostragem probabilística:** Cada elemento da população tem uma chance conhecida de ser selecionado.
 - **Amostragem por conglomerado:** Divide-se a população em blocos (homogêneos) e amostra-se os blocos aleatoriamente e todos os indivíduos dentro do bloco são entrevistados.
 - **Amostragem estratificada:** Divide-se a população em grupo segundo algum estrato (pelo menos dois), amostra-se aleatoriamente indivíduos dentro de cada grupo.
 - **Amostragem sistemática:** Selecionar um elemento da população a cada k .
 - **Amostragem aleatória simples:** Cada indivíduo tem mesma chance de ser amostrado.

Objetivo Principal

- Usar os dados amostrais para fazer inferência (ou generalizações) sobre uma população inteira.

Devemos compreender os conceitos de:

- Tipos de dados: qualitativos e quantitativos.
- Amostras representativas de uma população;
- Parâmetro de uma população;
- Estatística ou estimador dos parâmetros da população (uma função da amostra que reduz os dados);
- Inferência estatística;

Conceito × Notação

Alguns nomes:

Parâmetro	Estimador	Estimativa
Média populacional (μ)	Média amostral ($\hat{\mu}$ ou \bar{X})	$\hat{\mu}(\mathbf{x})$ ou \bar{x}
Variância populacional (σ^2)	Variância amostral ($\hat{\sigma}^2$ ou S^2)	$\hat{\sigma}^2(\mathbf{x})$ ou s^2
Proporção populacional (p)	Proporção amostral (\hat{p})	$\hat{p}(\mathbf{x})$

Classificação de variáveis

Classificação segundo à natureza

Qualitativa {
Nominal
Ordinal

Quantitativa {
Discreta
Contínua

Classificação de variáveis

Observações:

- Variável CEP?
- Variável Idade?
- Variáveis codificadas (rotuladas).

Tipos de Experimentos

- **Estudo transversal:** Dados coletados em um ponto do tempo.
- **Estudo retrospectivo ou controle de caso:** Dados são coletados do passado.
- **Estudo prospectivo, longitudinal ou coorte:** Dados são coletados no presente e futuro.

Tipos de Experimentos

- **Outras Definições:**

- Experimento cego (o paciente não sabe que recebe placebo, evita o efeito placebo);
- Experimento duplo-cego (nem pacientes nem médicos sabem quem recebe placebo).

Organização dos dados

- **Objetivo principal:**

- Facilitar a leitura dos dados.
- Compreensão dos dados.
- Obter informações sobre a população.
- Fazer comparações.

- **Série temporal, cronológica ou histórica:**

Tabela: Variação mensal do Índice de Preços ao Consumidor da categoria alimentação - São Paulo, Jan/2015- Dez/2015.

Mês	Variação (%)
Jan	1,57
Fev	1,00
Mar	0,72
Abr	0,83
Mai	0,88
Jun	0,58
Jul	0,77
Ago	-0,52
Set	-0,04
Out	1,31
Nov	2,26
Dez	1,55

Fonte: <http://www.fipe.org.br/pt-br/indices/ipc/#indice-mensal>

- **Série geográfica ou espacial:**

Tabela: População Residente segundo os municípios da Região Metropolitana de Salvador - 2010.

Municípios	População
Camaçari	242.970
Candeias	83.158
Dias D'Avila	66.440
Itaparica	20.725
Lauro de Freitas	163.449
Madre de Deus	17.376
Salvador	2.675.656
São Francisco do Conde	33.183
Simões Filho	118.047
Vera Cruz	37.567
Mata de São João	40.183
Pojuca	33.066
São Sebastião do Passé	42.153
Total	3.573.973

Fonte: IBGE, Censo Demográfico 2010, Bahia.

Características dos Dados

- **Medida central:** Indica onde se localiza o meio dos dados.
- **Variação:** Indica quanto os valores dos dados variam entre si.
- **Distribuição:** Como os dados se distribuem de forma simétrica, assimétrica, na forma de sino uniforme.
- **Valores discrepantes (ou outliers):** Valores atípicos presentes em um conjunto de dados.
- **Tempo:** Como os dados mudam com o tempo (tendência, ciclos, sazonalidade).

Características dos Dados

- **Ferramentas usadas:**

- Ordenação de dados;
- Representação na forma de tabelas;
- Representação na forma de gráficos;
- Fórmulas matemáticas para calcular grandezas.

- **Tecnologia ou cálculo manual?**

Fazer cálculos manuais com alguns exemplos pode ajudar a compreensão dos resultados obtidos pela tecnologia.

Áreas de Atuação

Algumas áreas de atuação do Estatístico

- Mercado Financeiro
- Indústria: Gestão de Qualidade);
- Indústria e Serviços: Confiabilidade;
- Saúde: Análise de Sobrevivência;
- Setor Público: Modelagem Estatística
- Planejamento de Produção: Modelagem Estatística
- Pesquisa de Mercado: Amostragem.