O Papel da Estatística em áreas do conhecimento

Marcos Oliveira Prates

2019/01

Site: http://www.est.ufmg.br/˜marcosop
e-mail: marcosop@est.ufmg.br

1. O Pensamento Estatístico
2. Coleta de dados
3. Modelos Mecanicistas e Empíricos
4. Probabilidade e Modelos de Probabilidade

Distribuição de Pontos

- 3 provas - 33, 33 e 34 pontos respectivamente;

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Identificar o papel da estatística na resolução de problemas.
- Discutir como a variabilidade afeta os dados coletados e usados para tomada de decisão.
- Diferenciar estudos enumerativos e analíticos.
- Discutir métodos de coletas de dados.

- Identificar as vantagens de experimentos planejados.
- Diferenciar modelos mecanicistas de modelos empíricos.
- Discutir como probabilidade e modelos de probabilidade são usados em administração e em ciência.

O Pensamento Estatístico

- Objetivo de um pesquisador:
  - resolver problemas de interesse da sociedade através da aplicação do método científico.

- Método científico $\Rightarrow$ abordagem para formular e resolver problemas.

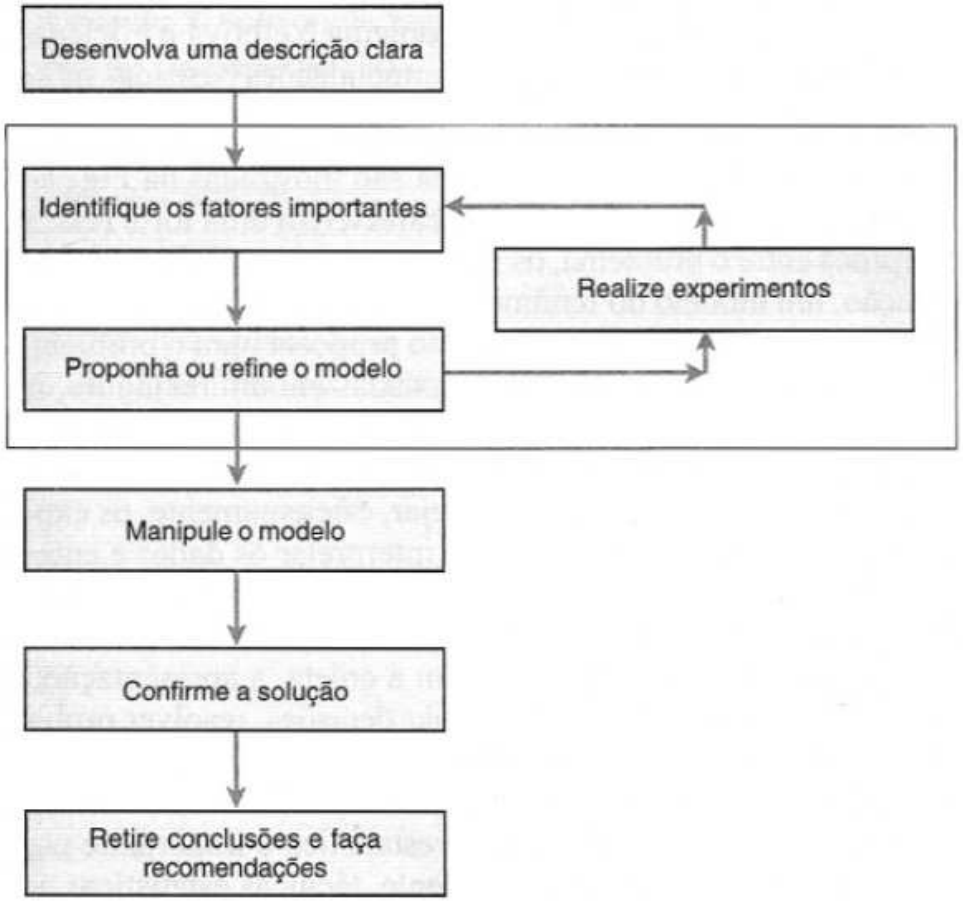

Fig. 1.1 O método de solução de um problema.

O que pesquisadores devem saber:

- planejar eficientemente experimentos;
- coletar dados;
- analisá-los e interpretá-los;
- entender como os dados estão relacionados com o modelo proposto em estudo.

Estatística lida com:

- coleta, apresentação e análise de dados;
- uso desses dados para tomar decisões e resolver problemas.

- O conhecimento de estatística é importante para qualquer profissional.
- Técnicas estatísticas podem ser uma ajuda poderosa no planejamento de novos produtos e sistemas.
- Podem ser usadas melhorando projetos existentes e planejando, desenvolvendo e melhorando sistemas existentes.

- Métodos estatísticos nos ajudam a entender a variabilidade.
- Variabilidade $\Rightarrow$ sucessivas observações de um sistema não produzem o mesmo resultado.
- Pensamento estatístico $\Rightarrow$ nos ensina a incorporar essa variabilidade na tomada de decisão.

Exemplo: Desempenho do consumo de gasolina de um carro

- O desempenho não é o mesmo para cada tanque de combustível.
- Depende de fatores como:
  - tipo de estrada usada;
  - mudanças nas condições do veículo;
  - marca da gasolina;
  - condições climáticas.

- São fatores potenciais de variabilidade.
- Estatística nos fornece uma estrutura para
  - descrever a variabilidade;
  - aprender quais fontes são mais importantes.

Exemplo: projeto de um conector de náilon para aplicação automotiva

- Especificação da espessura da parede: 3/32 polegadas.
- Está inseguro quanto a força de remoção.
- Produz oito unidades do protótipo e mede suas forças de remoção.
- Valores observados: 12,9; 13,4; 12,3; 13,6; 13,5; 12,6; 13,1.
- Nem todos protótipos tem a mesma força de remoção.

- Consideramos a força de remoção como uma variável aleatória X.
- Podemos pensar em X como

$$X = \mu + \epsilon$$

onde μ é uma constante e ϵ é uma perturbação aleatória.

- A constante permanece a mesma em cada medida.
- O valor de ϵ pode mudar, devido a
  - mudanças no ambiente;
  - diferenças na peças individuais;

- O diagrama de pontos permite visualizar localização e dispersão dos dados.

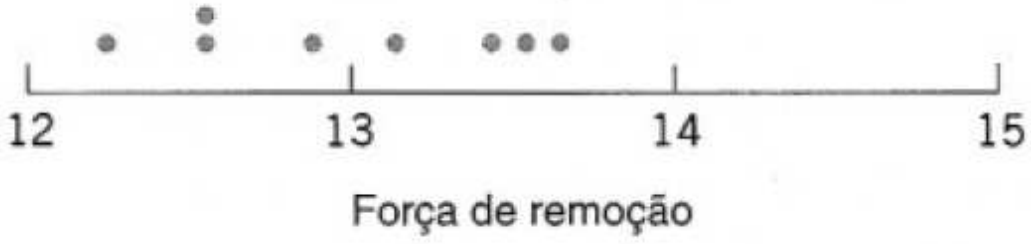

Fig. 1.2 Diagrama de pontos dos dados da força de remoção, quando a espessura da parede for 3/32 polegada.

Inferência Estatística

- raciocinar a partir de um conjunto de medidas para casos mais gerais.

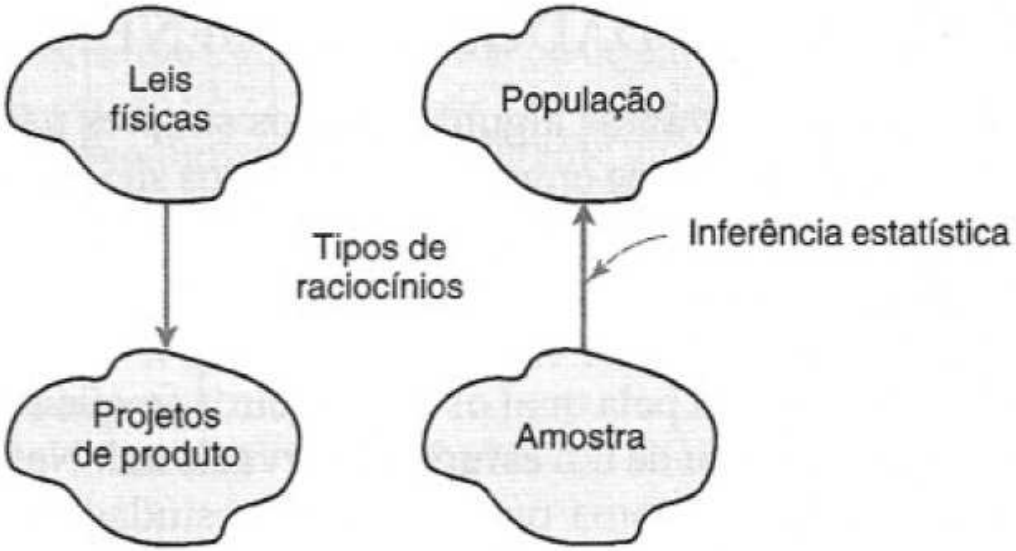

Fig. 1.6 Inferência estatística é um tipo de raciocínio.

- O raciocínio baseado em alguns objetos pode resultar em erros:
  - erros amostrais.
- Se amostra é selecionada adequadamente,
  - esses riscos podem ser quantificados e um tamanho apropriado de amostra é determinado.

Coleta de dados

Três métodos básicos de coletar dados são:

- estudo **retrospectivo** usando dados históricos;
- estudo de **observação**;
- um **experimento planejado**.

Estudo Retrospectivo

- Utiliza dados arquivados.
- Esses dados podem conter informação relativamente de pouca utilidade para o problema.
- Alguns dados relevantes podem ser omitidos.
- Podem haver erros de transcrição ou registro, resultando em *outliers* (valores não-usuais).

Exemplo:

- Deseja-se medir a concentração de acetona em um destilado.
- Fatores que afetam:
  - temperatura do refervedor,
  - temperatura do condensado,
  - taxa de refluxo.
- Uma amostra de dados arquivados é utilizada.
- Pode-se estar interessado em descobrir relações entre as duas temperaturas e a taxa de refluxo sobre a concentração de acetona.

Estudo de observação

- Observa-se o processo perturbando-o tão pouco quanto possível.
- Ocorrem por um curto período de tempo.
- Assim, variáveis que não são rotineiramente medidas podem ser incluídas.

No exemplo de destilação:

- O cientista poderia planejar uma forma de medir as temperaturas e a taxa de refluxo quando as medidas de concentração são feitas.
- Poderia ser possível também medir concentrações da corrente de alimentação e estudar seu impacto.

Experimentos planejados

- O cientista faz variações deliberadas em variáveis controláveis do sistema.
- Observa os dados de saída resultante.
- Faz uma inferência sobre quais variáveis são responsáveis pelas mudanças observadas.

Exemplo:

- No exemplo do conector de náilon.
- O cientista escolhe duas espessuras da parede e mede a força de remoção para cada uma delas.
- Nesse tipo de **experimento comparativo** o interesse é em determinar se existe diferença entre os dois projetos.

Observando Processos ao Longo do Tempo

- Dados coletados ao longo do tempo.
- É útil plotar os dados versus tempo em um gráfico de **série temporal**.
- Nesse tipo de gráfico, fenômenos que possam afetar o sistema se tornam mais visíveis.
- O conceito de estabilidade pode ser mais bem julgado.

Exemplo:

- No exemplo da destilação.
- Poderíamos fazer um gráfico de série temporal da concentração de acetona como esse mostrado abaixo.

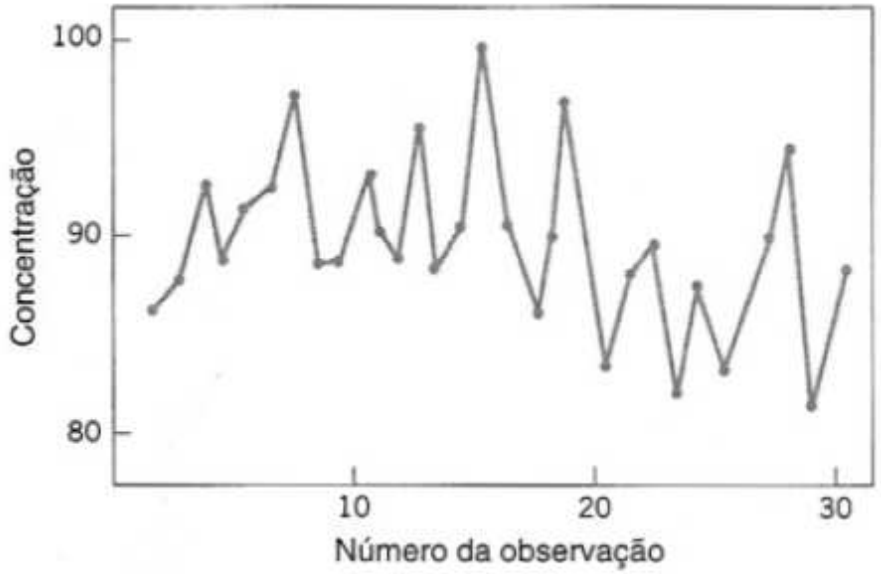

Fig. 1.15 Um gráfico temporal de concentração provê mais informações do que o diagrama de pontos.

Modelos Mecanicistas e Empíricos

- Modelos desempenham um papel importante em quase todos problemas.
- Suponha que estejamos medindo a corrente em um fio de cobre.
- O modelo para esse fenômeno pode ser a lei de Ohm:

$$\text{Corrente=voltagem/resistência}$$

ou

$$I = E/R$$

- Esse tipo de modelo é chamado **mecanicista**.
- É construído a partir do nosso conhecimento do mecanismo físico que relaciona essas variáveis.

- Se fizermos o processo de medição mais de uma vez, a corrente observada poderá diferir.
- Isso pode ser consequência de: mudanças de temperatura, flutuações no desempenho de medidor, etc.
- O modelo mais realista seria

$$I = E/R + \epsilon$$

  sendo ϵ uma medida de erro.
- ϵ inclui efeitos de todas as fontes de variabilidade que não fazem parte do modelo.
- Esse tipo de modelo é chamado **modelo empírico**.

- Em muitas situações não temos modelos mecanicistas simples ou bem entendidos.
- Suponha que estamos interessados no peso molecular médio (M_n) de um polímero.
- Sabemos que está relacionado com:
  - viscosidade do material (V);
  - quantidade de catalisador (C);
  - temperatura (T).
- A relação entre M_n e essas variáveis é

$$M_n = f(V, C, T)$$

onde f é uma função desconhecida.

- Talvez um modelo razoável seja

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T$$

sendo os β's parâmetros desconhecidos.

- Esse modelo não descreve exatamente o fenômeno.
- Desse modo devemos adicionar o termo ϵ ao modelo

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \epsilon$$

- Veremos que esse tipo modelo é chamado **modelo de regressão**.
- Os termos β's podem ser encontrados através de métodos de **mínimos quadrados**

Probabilidade e Modelos de Probabilidade

- Vimos que as decisões são tomadas tendo como base os objetos de uma amostra.
- Precisamos saber o quão bem essa amostra representa a população.
- Devemos quantificar os riscos de decisões tomadas tendo como base a amostra.
- Os **modelos de probabilidade** ajudam a quantificar esses riscos.

- Suponha que temos um lote com 25 pastilhas.
- Se todas forem defeituosas ou todas boas:
  - as pastilhas da amostra serão todas defeituosas ou todas boas.
- Suponha que só uma pastilha é defeituosa.
- Uma amostra retirada pode ou não conter essa pastilha defeituosa.
- Para quantificarmos os riscos de uma pastilha defeituosa ser detectada devemos usar:
  - um modelo de probabilidade
  - um método para selecionar a amostra.