



Introdução à estatística

Texto baseado no livro:

Estatística Aplicada - Larson / Farber – Editora Pearson – 2010



Descrição do capítulo

1.1 Uma visão geral sobre estatística

1.2 Classificação dos dados

1.3 Planejamento experimental



Seção 1.1

Uma visão geral sobre estatística



Objetivos da seção 1.1

Definir estatística

Diferenciar população e amostra

Diferenciar parâmetros e dados estatísticos

Diferenciar estatística descritiva de estatística inferencial



O que são dados?

Dados

Consistem em informações advindas de observações, contagens, medidas ou respostas.

- “Pessoas que comem grãos integrais em três de suas refeições diárias têm risco reduzido de (...) derrames em 37%.”
(*Fonte: WGC.*)
- “Setenta por cento das 1.500 lesões na coluna vertebral em menores nos EUA são resultado de acidentes em veículos e 68% deles não estavam usando cinto de segurança.”
(*Fonte: UPI.*)



O que é estatística?

Estatística

É a ciência de coletar, organizar, analisar e interpretar dados para a tomada de decisões.

Conjunto de dados



População

Coleção de *todos* os resultados, respostas, medições ou contagens que são de interesse.

Amostra

Um subgrupo da população.



Exemplo: identificando conjunto de dados

Em uma pesquisa recente, 1.708 adultos foram questionados se achavam que o aquecimento global era um problema que requeria ação governamental imediata. Novecentos e trinta e nove dos adultos disseram que sim. Identifique a população e a amostra. Descreva o conjunto de dados.



A população consiste nas respostas de todos os adultos daquela região

A amostra consiste nas respostas dos 1.708 adultos da pesquisa

A amostra é uma parcela das respostas de todos os adultos

O conjunto de dados consiste de 939 sim e 769 não

Respostas dos adultos
(população)

Respostas dos
adultos na pesquisa
(amostragem)

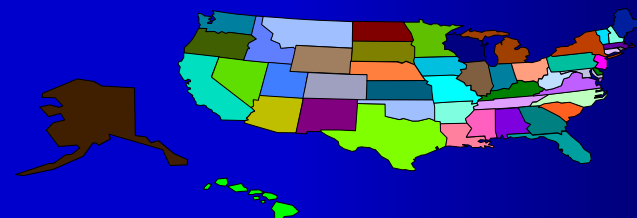


Parâmetro e estatística

Parâmetro

Um número que descreve uma característica populacional.

Ex.: Idade média de todas as pessoas no Brasil



Estatística

Um número que descreve uma característica amostral.

Idade média das pessoas de uma amostra de três estados





Exemplo: distinguindo entre parâmetro e estatística

Decida se o valor numérico descreve um parâmetro populacional ou uma estatística amostral.

1. Uma pesquisa recente de uma amostragem de MBAs revelou que o salário médio de um MBA é superior a 82.000. (*Fonte: The Wall Street Journal.*)

Solução:

Estatística amostral (a média de 82.000 é baseado em uma parte da população).



Decida se o valor numérico descreve um parâmetro populacional ou uma estatística amostral.

2. O salário inicial dos 667 MBAs da *University of Chicago Graduate School of Business* aumentou 8,5% desde o ano passado.

Solução:

Parâmetro populacional (a porcentagem de 8,5% é baseada nos salários iniciais de todos os 667 graduados).



Ramos da estatística

Estatística descritiva

Envolve organização, o resumo e a representação dos dados.

Ex.: tabelas, gráficos, médias.

Estatística inferencial

Envolve o uso de *amostras* para chegar a conclusões sobre uma *população*.



Exemplo: estatísticas descritiva e inferencial

Decida qual parte desse estudo representa o ramo descritivo da estatística. Quais conclusões podem ser tiradas do estudo usando estatística inferencial?

Uma amostra grande de homens de 48 anos foi estudada por 18 anos. Para homens não casados, aproximadamente 70% estavam vivos aos 65 anos. Para os casados, 90% estavam vivos aos 65 anos.
(*Fonte: The Journal of Family Issues.*)

Ainda vivos aos 65 anos

Homens solteiros – 70%

Homens casados – 90%



Solução: estatísticas descritiva e inferencial

Estatística descritiva envolve afirmações como “Para homens não casados, aproximadamente 70% estavam vivos aos 65 anos” e “Para homens casados, 90% estavam vivos aos 65 anos”.

Uma inferência possível tirada desse estudo é que ser casado está associado com uma vida mais longa para homens.

Ainda vivos aos 65 anos

Homens solteiros – 70%

Homens casados – 90%



Sumário da Seção 1.1

Definimos estatística

Diferenciamos população e amostra

Diferenciamos parâmetro e dado estatístico

Diferenciamos estatística descritiva de estatística inferencial



Seção 1.2

Classificação dos dados



Objetivos da Seção 1.2

Distinguir dados qualitativos e dados quantitativos

Classificar dados em relação a quatro níveis de mensuração

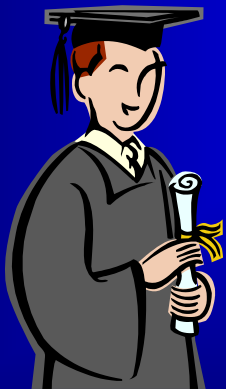


Tipos de dados

Dados qualitativos

Consistem em atributos, qualidades ou entradas não numéricas.

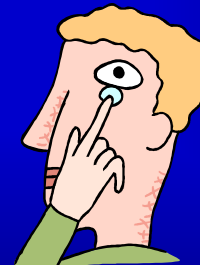
Graduação



Local de nascimento



Cor dos olhos





Dados quantitativos

Medidas ou contagens numéricas.

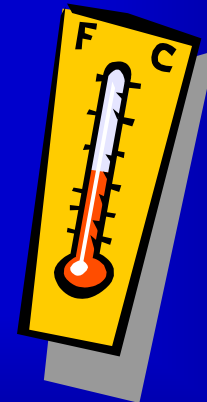
Idade



Peso de uma carta



Temperatura





Exemplo: classificando dados por tipo

Os preços-base para diversos veículos são apresentados na tabela a seguir. Quais dados são qualitativos e quais são quantitativos? (*Fonte: Ford Motor Company.*)

Modelo	Preço-base
Fusion 14 S	\$ 17.795
F-150 XL	\$ 18.710
Five Hundred SEL	\$ 23.785
Escape XLT Sport	\$ 24.575
2007 Explorer Sport	
Trac Limited	\$ 26.775
Freestar SEL	\$ 27.500
Crown Victoria LX	\$ 28.830
Expedition XLT	\$ 35.480

**Modelo**

Fusion 14 S

F-150 XL

Five Hundred SEL

Escape XLT Sport

2007 Explorer Sport

Trac Limited

Freestar SEL

Crown Victoria LX

Expedition XLT

Preço-base

\$ 17.795

\$ 18.710

\$ 23.785

\$ 24.575

\$ 26.775

\$ 27.500

\$ 28.830

\$ 35.480

Dados qualitativos (nomes dos modelos automotivos são entradas não numéricas).

Dados quantitativos (preços-base dos modelos automotivos são entradas numéricas).



Níveis de mensuração

Nível nominal de mensuração

Somente dados qualitativos

São classificados com nomes, etiquetas ou qualidades

Não podem ser feitos cálculos matemáticos

Nível ordinal de mensuração

- Dados qualitativos ou quantitativos
- Dados podem ser organizados em ordem ou posição
- Diferenças entre as entradas de dados não são significativas



Exemplo: classificando dados por nível

A seguir temos dois conjuntos de dados. Qual conjunto de dados está no nível nominal? E qual está no nível ordinal? Explique. (*Fonte: Nielsen Media Research.*)

Os cinco programas de TV mais assistidos (de 12/2/2007 a 18/2/2007)

1. American Idol – terça-feira
2. American Idol – quarta-feira
3. Grey's anatomy
4. House
5. CSI

Afiladas das redes em Pittsburg, PA

- WTAE (ABC)
- WPXI (NBC)
- KDKA (CBS)
- WPGH (FOX)



Solução: classificando dados por nível

**Os cinco programas de TV
mais assistidos (de 12/2/2007
a 18/2/2007)**

1. American Idol – terça-feira
2. American Idol – quarta-feira
3. Grey's anatomy
4. House
5. CSI



**Nível ordinal (lista das colocações
de cinco programas de TV; os
dados podem ser postos em
ordem;).**

**Afiladas das
redes em
Pittsburg, PA
WTAE (ABC)
WPXI (NBC)
KDKA (CBS)
WPGH (FOX)**



**Nível nominal (lista das afiliadas
da rede de TV; as siglas nomeiam
as afiliadas).**



Níveis de medida

Nível de mensuração intervalar

Dados quantitativos

Dados podem ser organizados

Diferenças entre as entradas de dados são significativas

Zero representa uma posição em uma escala (e não um zero inerente – zero não implica em ‘nulo’)

Nível de mensuração racional



Similar ao nível intervalar

Uma entrada zero é um zero inerente (implica em 'nulo')

Pode ser feita uma proporção de dois valores de dados

Um valor de dado pode ser expresso como o múltiplo de outro



Exemplo: classificando dados por nível

A seguir temos dois conjuntos de dados. Qual conjunto de dados está no nível intervalar? E qual está no nível racional? Explique. (*Fonte: Major League Baseball.*)

Vitórias do New York Yankees na série mundial (anos)

1923, 1927, 1928,
1932, 1936, 1937,
1938, 1939, 1941,
1943, 1947, 1949,
1950, 1951, 1952,
1953, 1956, 1958,
1961, 1962, 1977,
1978, 1996, 1998,
1999, 2000

Totais de *home runs* da Liga Americana em 2003 (por time)

Baltimore	164
Boston	192
Chicago	236
Cleveland	196
Detroit	203
Kansas City	124
Los Angeles	159
Minnesota	143
Nova Iorque	210
Oakland	175
Seattle	172
Tampa Bay	190
Texas	183
Toronto	199

Solução: classificando dados por nível



Vitórias do New York Yankees na série mundial (anos)

•1923, 1927, 1928, 1932, 1936, 1937, 1938, 1939, 1941, 1943, 1947, 1949, 1950, 1951, 1952, 1953, 1956, 1958, 1961, 1962, 1977, 1978, 1996, 1998, 1999, 2000



Nível intervalar (dados quantitativos; pode-se encontrar uma diferença entre dois dados, mas uma proporção não faz sentido).

Totais de *home runs* da Liga Americana em 2003 (por time)

Baltimore	164
Boston	192
Chicago	236
Cleveland	196
Detroit	203
Kansas City	124
Los Angeles	159
Minnesota	143
Nova Iorque	210
Oakland	175
Seattle	172
Tampa Bay	190
Texas	183
Toronto	199



Nível racional (podem-se encontrar diferenças e formar proporções).



Sumário dos quatro níveis de mensuração

Nível de mensuração	Categorizar os dados	Ordenar os dados	Subtrair os valores dos dados	Determinar se um valor de dado é múltiplo de outro
Nominal	Sim	Não	Não	Não
Ordinal	Sim	Sim	Não	Não
Intervalar	Sim	Sim	Sim	Não
Racional	Sim	Sim	Sim	Sim



Sumário da Seção 1.2

Distinguimos dados qualitativos de dados quantitativos

Classificamos dados de acordo com os quatro níveis de mensuração



Seção 1.3

Planejamento experimental



Objetivos da Seção 1.3

Discutir como planejar um estudo estatístico

Discutir técnicas de coleta de dados

Discutir como planejar um experimento

Discutir técnicas de amostragem



Planejando um estudo estatístico

1. Identifique a(s) variável(is) de interesse (foco) e a população do estudo.
2. Desenvolva um plano detalhado para coleta de dados. Se você usar uma amostra, certifique-se de que ela é representativa da população.
3. Colete os dados.
4. Descreva os dados usando as técnicas de estatística descritiva.
5. Interprete os dados e tome decisões sobre a população usando estatística inferencial.
6. Identifique quaisquer erros possíveis.



Coleta de dados

Estudo observacional

Um pesquisador observa e mede as características interessantes de uma parte da população

Pesquisadores observaram e registraram o comportamento oral de crianças acima de três anos de idade com objetos não comestíveis (*Fonte: Pediatric Magazine.*)

Experimento



Um tratamento é aplicado em uma parte da população e suas respostas são observadas

Foi realizado um experimento no qual diabéticos tomaram extrato de canela diariamente, enquanto um grupo de controle, não. Depois de 40 dias, os diabéticos que tomaram canela reduziram o risco de doenças do coração, enquanto o grupo de controle não apresentou mudanças (*Fonte: Diabetes Care.*)

Simulação



Usa um modelo matemático ou físico para reproduzir as condições de uma situação ou processo

Frequentemente envolve o uso de computadores

Empresas automobilísticas usam simulações com bonecos para estudar os efeitos de acidentes em pessoas

Pesquisa



Investigação de uma ou mais características de uma população

Comumente feita por entrevistas, correspondências ou telefone

Um estudo é conduzido em uma amostra de médicos para determinar se a razão primária na escolha de suas carreiras é a estabilidade financeira



Exemplo: métodos de coleta de dados

Considere os seguintes estudos estatísticos. Qual método de coleta de dados você usaria em cada um destes casos?

1. Um estudo do efeito da mudança dos padrões de voo no número de acidentes com aviões.

Solução:

Simulação (não é prático criar essa situação).



2. Um estudo sobre o efeito da aveia para abaixar a pressão sanguínea.

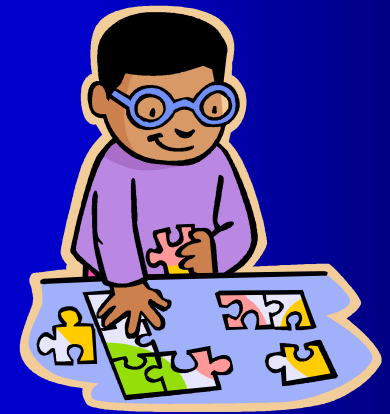


Solução:

Experimento (mede o efeito do tratamento – comer aveia).



3. Um estudo sobre como alunos da quarta série montam um quebra-cabeça.



Solução:

Estudo observacional (observa e mede uma certa característica da população).



4. Um estudo dizendo se os residentes de um país aprovam seu presidente ou não.



Solução:

Pesquisa (perguntar “Você aprova o modo como o presidente está lidando com o cargo?”).



Elementos-chave do planejamento experimental

Controle

Aleatorização

Replicação

controle



Controle para efeitos alheios ao que está sendo medido

Variáveis confusas

Ocorre quando um experimento não consegue mostrar a diferença entre os efeitos de fatores diversos em uma variável

O dono de uma cafeteria reforma sua loja ao mesmo tempo em que um shopping próximo inaugura. Se as vendas da cafeteria aumentarem, não há como saber se a causa foi a reforma da cafeteria ou a inauguração do shopping



Efeito placebo

Um sujeito reage favoravelmente a um placebo quando na verdade ele não está recebendo nenhum tratamento

Teste cego é uma técnica na qual os sujeitos não sabem se estão recebendo um remédio ou um placebo

Teste duplo-cego é quando nem o sujeito nem o experimentador sabem se o sujeito está recebendo tratamento ou um placebo



Aleatorização é o processo de distribuir aleatoriamente sujeitos a diferentes grupos

Desenho totalmente aleatorizado

Os sujeitos são distribuídos a diferentes grupos por seleção aleatória

Desenho em blocos aleatórios

Divide os sujeitos com características semelhantes em **blocos** e, então, dentro de cada bloco, distribui aleatoriamente os sujeitos em diferentes grupos



Desenho em blocos aleatórios

Um experimento testando os efeitos de uma nova bebida emagrecedora pode primeiramente dividir os sujeitos em categorias por idade. Então, dentro de cada um desses grupos, distribuí-los aleatoriamente para grupos de tratamento ou de controle.

Todos os sujeitos	De 30 a 39 anos	Controle
		Tratamento
	De 40 a 49 anos	Controle
		Tratamento
	Mais de 50 anos	Controle
		Tratamento



Aleatorização

Planejamento de pares combinados

Os sujeitos são pareados de acordo com a similaridade. Um sujeito do par é selecionado para receber um tratamento enquanto o outro recebe um tratamento diferente



Replicação

Replicação é a repetição de um experimento usando um grande número de sujeitos

Para testar uma vacina contra uma variação do vírus da gripe, 10.000 pessoas recebem a vacina e outras 10.000 recebem um placebo. Por causa do tamanho da amostragem, muito provavelmente poderá ser observada a efetividade da vacina



Exemplo: planejamento experimental

Uma empresa deseja testar a efetividade de um novo chiclete para ajudar as pessoas a parar de fumar.

Identifique um problema em potencial com o desenho experimental escolhido e sugira um modo de melhorá-lo.

A empresa identifica mil adultos que sejam fumantes inveterados. As cobaias são divididas em blocos de acordo com o gênero. Depois de dois meses, o grupo feminino tem um número significativo de pessoas que deixaram de fumar.



Solução: planejamento experimental

Problema

Os grupos não são similares. O novo chiclete pode ter um efeito melhor em mulheres ou vice-versa.

Correção

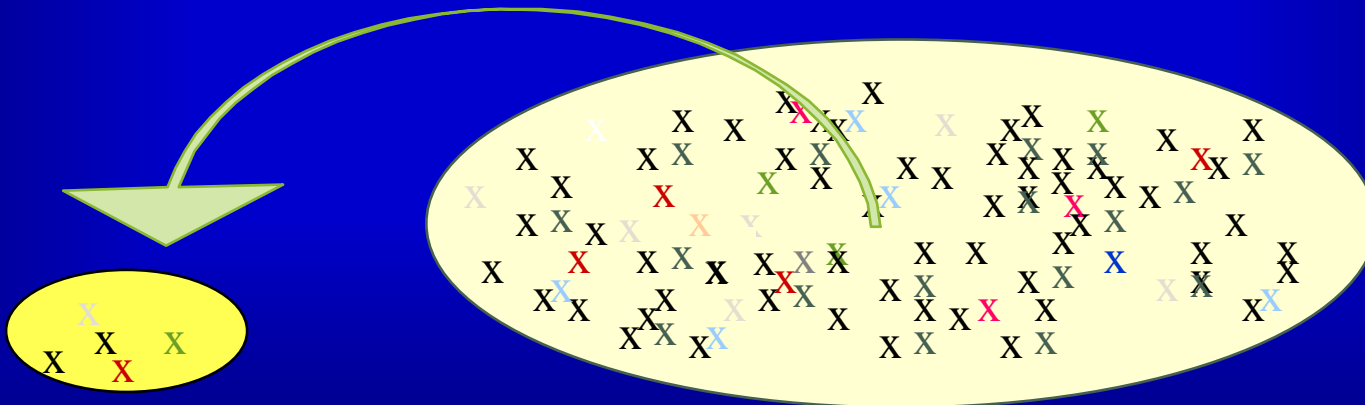
Os sujeitos podem ser divididos em blocos de acordo com o gênero, mas, dentro de cada bloco, precisam ser aleatoriamente distribuídos para estar no grupo de tratamento ou no grupo de controle.



Técnicas de amostragem

Amostra aleatória simples

Todo membro da amostra de mesmo tamanho tem a mesma chance de ser selecionado.





Amostra aleatória simples

Números aleatórios podem ser gerados por uma tabela de números aleatórios, um software ou uma calculadora

Dê um número a cada membro da população

Membros da população que correspondam a esses números tornam-se membros da amostragem

Exemplo: amostra aleatória simples



Há 731 estudantes que se inscreveram no curso de estatística em sua faculdade. Você deseja formar uma amostra de oito estudantes para responder às questões de uma pesquisa. Selecione os estudantes que pertencerão à amostra aleatória simples

- Dê números de 1 até 731 para cada estudante de estatística

Solução: amostra aleatória simples



Digamos que foram sorteados os números: 719, 662, 650, 4, 53, 589, 403 e 129. Estes seriam a amostra.



Outras técnicas de amostra

Amostra estratificada

Divida a população em grupos (estrato) e selecione uma amostra aleatória de cada grupo

- Para coletar uma amostragem estratificada do número de pessoas que vivem em casas em um certo estado, você poderia dividir as casas em níveis socioeconômicos e então selecionar casas aleatoriamente em cada nível

Grupo 1:
renda baixa

Grupo 2:
renda média

Grupo 3:
renda alta

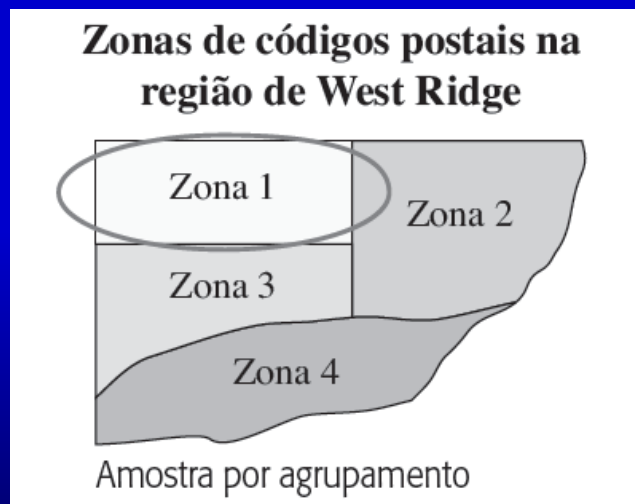
Outras técnicas de amostra



Amostra por agrupamento

Divida a população em grupos (blocos) e separe todos os membros em um ou mais deles, mas não em todos.

- No exemplo anterior você poderia dividir as casas em blocos de acordo com seus CEPs e, então, escolher todas as casas em um ou mais CEP, mas não em todos

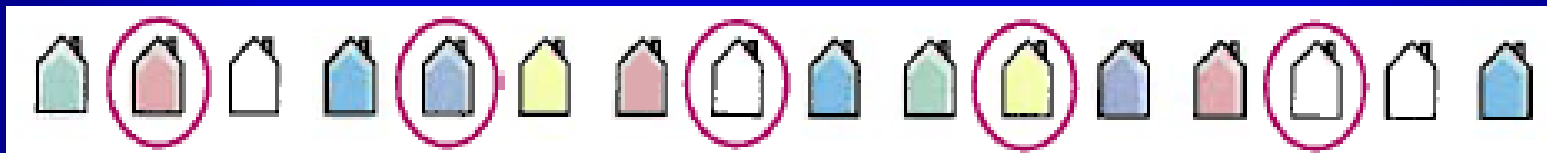




Amostra sistemática

Escolha um valor inicial aleatoriamente. Então, separe todo x-ésimo membro da população.

- Por exemplo, você poderia dar um número diferente para cada casa, escolher um número inicial aleatório e, então, escolher toda 100ª casa.





Exemplo: identificando técnicas de amostra

Você está conduzindo um estudo para determinar a opinião dos alunos da sua escola à respeito das pesquisas com células-tronco. Identifique a técnica de amostra usada.

1. Você divide a população de estudantes por cursos e, aleatoriamente, escolhe e questiona alguns dos estudantes de cada graduação.

Solução:

Amostra estratificada [os alunos são divididos em strata (graduações) e uma amostra é selecionada de cada graduação].



2. Você dá um número para cada aluno e gera números aleatórios. Então você pergunta quais estudantes tiveram seus números escolhidos aleatoriamente.

Solução:

Amostra aleatória simples (cada amostragem de mesmo tamanho tem chance igual de ser selecionada e cada estudante tem chance igual de ser selecionado).



Sumário da Seção 1.3

Discutimos como planejar um estudo estatístico

Discutimos técnicas de coleta de dados

Discutimos como planejar um experimento

Discutimos técnicas de amostra