

Capítulo 1 | Introdução à estatística



O que é estatística?

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

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: Whole Grains Council.*)
- “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.*)

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, 1.708 adultos nos EUA foram questionados se achavam que o aquecimento global era um problema que requeria ação governamental imediata. 939 dos adultos disseram que sim.

Identifique a população e a amostra. Descreva o conjunto de dados. (*Adaptado de: Pew Research Center.*)

- A população consiste nas respostas de todos os adultos nos EUA
- A amostra consiste nas respostas dos 1.708 adultos estadunidenses da pesquisa
- A amostra é uma parcela das respostas de todos os adultos nos EUA
- O conjunto de dados consiste de 939 sim e 769 não

Respostas dos adultos
nos EUA (população)

Respostas dos
adultos na pesquisa
(amostragem)

Parâmetro e estatística

Parâmetro

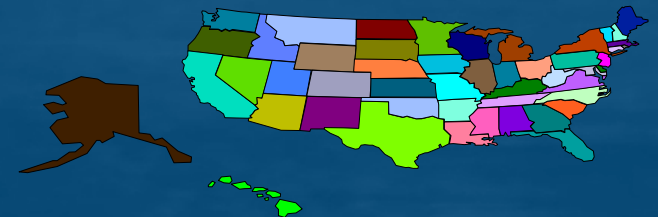
Um número que descreve uma característica populacional.

Idade média de todas as pessoas nos Estados Unidos

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.*)

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.

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%

Segundo turno para prefeitos 2020

LARSON | FARBER
estatística aplicada
4ª edição

57 cidades tiverem segundo turno, sendo 18 capitais

1. São Paulo

Em São Paulo, o cenário se encaminha para a reeleição de **Bruno Covas** (PSDB) – ao menos de acordo com a **pesquisa do Ibope**. O levantamento mostra o tucano com 57% dos votos válidos, contra 43% de **Guilherme Boulos** (Psol).

Pesquisa Datafolha, também **divulgada neste sábado (28)**, aponta Covas como favorito – mas em uma vantagem menor à registrada no levantamento do Ibope. Segundo o Datafolha, o tucano tem 55% das intenções de votos válidos, contra 45% de Boulos.

Dados sem nbi obtidos 02/Dez/2020 em
<https://www.gazetadopovo.com.br/eleicoes/2020/pesquisas-ibope-favoritos-segundo-turno-10-cap>

Metodologia das pesquisas citadas

São Paulo

- Sob encomenda da TV Globo e do jornal O Estado de S. Paulo, o Ibope ouviu 1.204 eleitores de São Paulo entre os dias 27 e 28 de novembro de 2020. O levantamento tem nível de confiança de 95%, com margem de erro de 3 pontos percentuais. A pesquisa está registrada no Tribunal Superior Eleitoral (TSE) sob a identificação SP-02990/2020.
- Sob encomenda da Folha de S. Paulo e da TV Globo, o Datafolha ouviu 3.047 eleitores de São Paulo entre os dias 27 e 28 de novembro de 2020. O levantamento tem nível de confiança de 95%, com margem de erro de 2 pontos percentuais para mais ou para menos. A pesquisa está registrada no Tribunal Superior Eleitoral (TSE) sob a identificação SP-02945/2020.

Segundo turno para prefeitos 2020 (candidatos ordem alfabética)

LARSON | FARBER
estatística aplicada
4ª edição

cidade	IBOPE (28/11)	DATAFOLHA (28/11)	Resultado (29/11)
São Paulo	57% a 43%	55% a 45%	
Rio de Janeiro	68% a 32%	68% a 32%	
Porto Alegre	51% a 49% (e 3pp)	-	
Recife	50% a 50% (en)	50% a 50% (en)	
Fortaleza	39% a 61%	-	
Goiânia	59% a 41%	-	
Vitória	50% a 50% (en)	-	
Belém	42% a 58%	-	
João Pessoa	58% a 42%	-	
Teresina	68% a 32%	-	

Segundo turno para prefeitos 2020 (candidatos ordem alfabética)

LARSON | FARBER
estatística aplicada
4ª edição

cidade	IBOPE (28/11)	DATAFOLHA (28/11)	Resultado (29/11)
São Paulo	57% a 43%	55% a 45%	59,4% a 30,6%
Rio de Janeiro	68% a 32%	68% a 32%	• 64,1% a 35,9%
Porto Alegre	51% a 49% (e 3pp)	-	45,4% a 54,6%
Recife	50% a 50% (en)	50% a 50% (en)	• 56,3% a 43,7%
Fortaleza	39% a 61%	-	48,3% a 51,7%
Goiânia	59% a 41%	-	• 52,6% a 47,4%
Vitória	50% a 50% (en)	-	• 58,5% a 41,5%
Belém	42% a 58%	-	• 48,2% a 51,8%
João Pessoa	58% a 42%	-	• 53,2% a 46,8%
Teresina	68% a 32%	-	• 62,3% a 37,7%

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.

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.

Observação

Experimentação

Simulação

Pesquisa

3. Descreva os dados usando as técnicas de estatística descritiva.
4. Interprete os dados e tome decisões sobre a população usando estatística inferencial.
5. Identifique quaisquer erros possíveis.

Coleta de dados

Estudo observacional

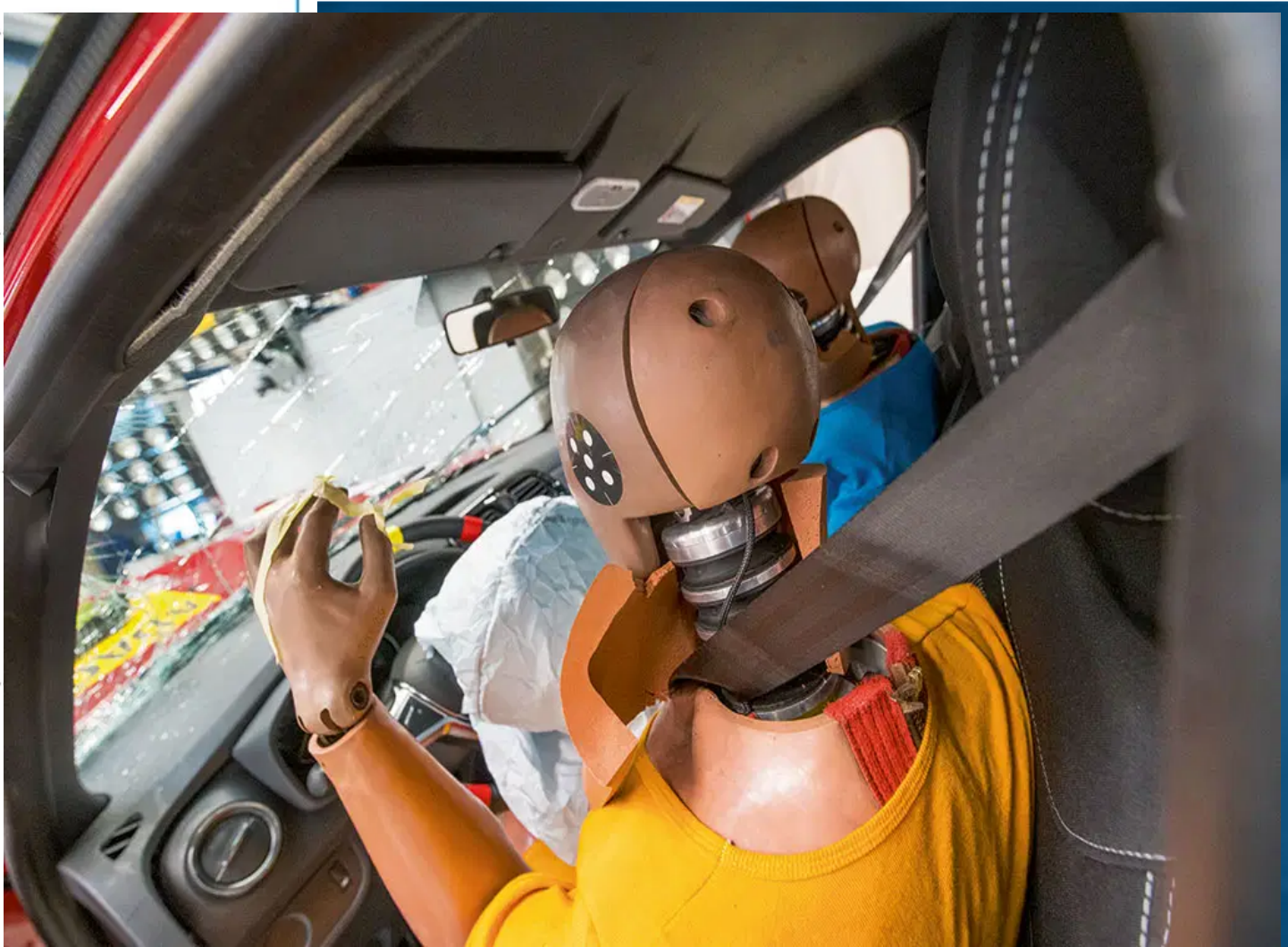
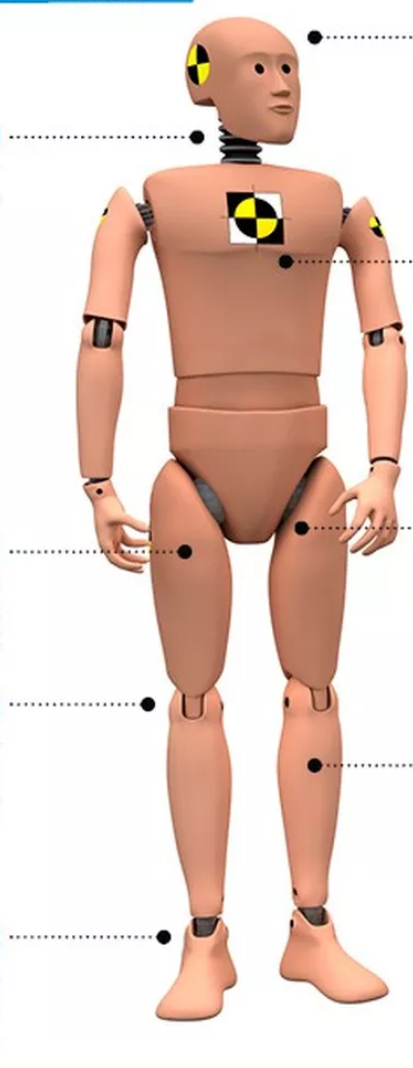
- Um pesquisador observa e mede as características interessantes de uma parte da população
- Pesquisadores observaram e registraram o comportamento oral de crianças acima de três anos de idade com objetos não comestíveis (*Fonte: Pediatric Magazine.*)

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
- Modelos imitam o real
- **Empresas automobilísticas usam simulações com bonecos para estudar os efeitos de acidentes em pessoas**
- Frequentemente envolve o uso de computadores



Fonte Cesvi Brasil

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 médicas mulheres para determinar se a razão primária na escolha de suas carreiras é 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.

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.



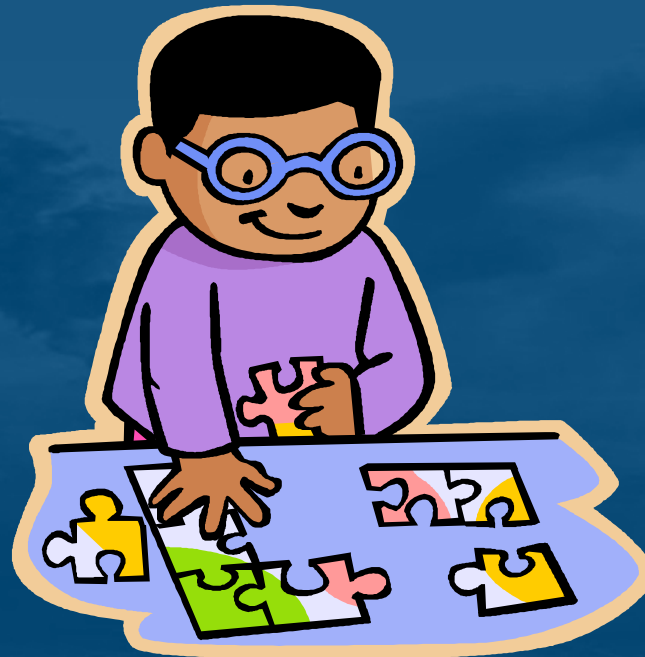
2. Um estudo sobre o efeito da aveia para abaixar a pressão sanguínea.

Solução:

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



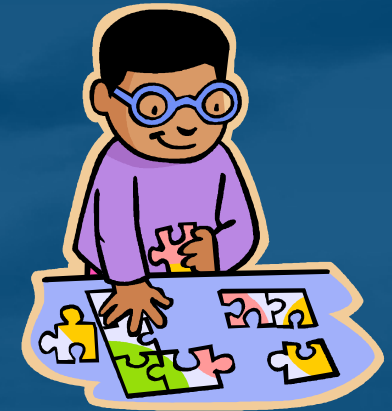
3. Um estudo sobre como alunos da quarta série montam um quebra-cabeça.



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 nos EUA aprovam seu presidente ou não.



4. Um estudo dizendo se os residentes nos EUA aprovam seu presidente ou não.

Solução:

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



Classificação dos dados

- **Tipo:** Qualitativos x Quantitativos
- **Nível de medição:** nominais, ordinais, intervalos e razões

Tipos de dados

- Qualitativos: atributos, rótulos e entradas não numéricas
- Quantitativos: entradas numéricas (medidas, contagens ...)

Model	Suggested retail price
Focus Sedan	\$15,995
Fusion	\$19,270
Mustang	\$20,995
Edge	\$26,920
Flex	\$28,495
Escape Hybrid	\$32,260
Expedition	\$35,085
F-450	\$44,145

Níveis de medição

- **Nominais:** nomes, rótulos e qualidades.
- **Ordinais:** podem ser organizados em ordem ou ranqueados. Diferenças entre entradas não trazem um significado importante.
- **Intervalos e proporções:** entradas podem ser ordenadas e a diferença entre elas é importante.

	Example of a Data Set	Meaningful Calculations
Nominal Level (Qualitative data)	<i>Types of Shows Televised by a Network</i> Comedy Documentaries Drama Cooking Reality Shows Soap Operas Sports Talk Shows	<i>Put in a category.</i> For instance, a show televised by the network could be put into one of the eight categories shown.
Ordinal Level (Qualitative or quantitative data)	<i>Motion Picture Association of America Ratings Description</i> G General Audiences PG Parental Guidance Suggested PG-13 Parents Strongly Cautioned R Restricted NC-17 No One Under 17 Admitted	<i>Put in a category and put in order.</i> For instance, a PG rating has a stronger restriction than a G rating.
Interval Level (Quantitative data)	<i>Average Monthly Temperatures (in degrees Fahrenheit) for Denver, CO</i> Jan 29.2 Jul 73.4 Feb 33.2 Aug 71.7 Mar 39.6 Sep 62.4 Apr 47.6 Oct 51.0 May 57.2 Nov 37.5 Jun 67.6 Dec 30.3 <i>(Source: National Climatic Data Center)</i>	<i>Put in a category, put in order, and find differences between values.</i> For instance, $57.2 - 47.6 = 9.6^{\circ}\text{F}$. So, May is 9.6° warmer than April.
Ratio Level (Quantitative data)	<i>Average Monthly Precipitation (in inches) for Orlando, FL</i> Jan 2.4 Jul 7.2 Feb 2.4 Aug 6.3 Mar 3.5 Sep 5.8 Apr 2.4 Oct 2.7 May 3.7 Nov 2.3 Jun 7.4 Dec 2.3 <i>(Source: National Climatic Data Center)</i>	<i>Put in a category, put in order, find differences between values, and find ratios of values.</i> For instance, $\frac{7.4}{3.7} = 2$. So, there is twice as much rain in June as in May.

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 (nocebo)**

- Um sujeito reage favoravelmente a um placebo quando na verdade ela 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.

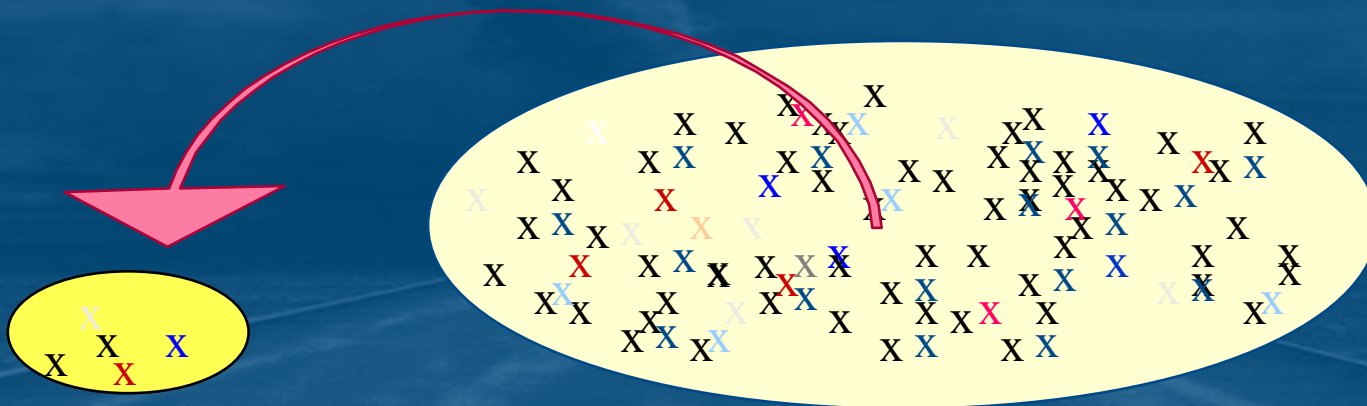
Correção

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

Toda amostra possível de mesmo tamanho tem a mesma chance de ser selecionada.



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 uma certa 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
- Na tabela de números aleatórios, escolha um espaço inicial aleatoriamente (suponha que você comece na terceira linha da segunda coluna)

Solução: amostra aleatória simples



Tabela 1 – Números aleatórios

92630 78240 19267 95457 53497 23894 37708 79862
79445 78735 71549 44843 26104 67318 00701 34986
59654 71966 27386 50004 05358 94031 29281 18544
31524 49587 76612 39789 13537 48086 59483 60680
06348 76938 90379 51392 55887 71015 09209 79157

- Leia os dígitos em grupos de três
- Ignore números maiores que 731

719 | 66 2 | 738 | 6 50 | 004 | 053 | 58 9 | 403 | 1 29 | 281 | 185 | 44

Os estudantes de número 719, 662, 650, 4, 53, 589, 403 e 129 comporiam a amostra.

Outras técnicas de amostra

Amostra estratificada

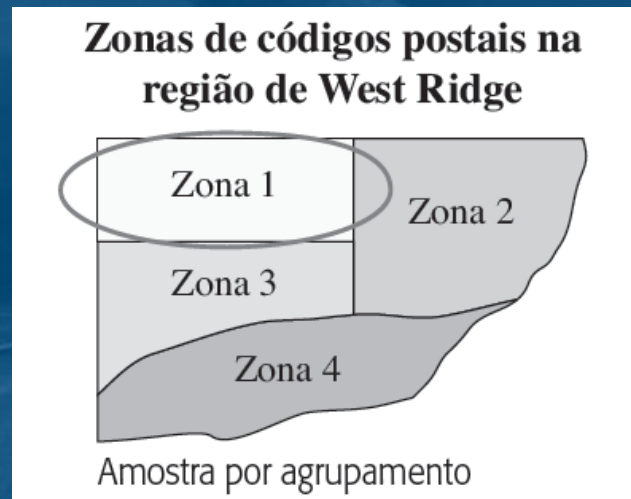
- Divida a população em grupos (estrato) e selecione uma amostra aleatória de cada grupo
- Para coletar uma amostragem estratificada do número de pessoas que vivem em casas no condado de West Ridge, você poderia dividir as casas em níveis socioeconômicos e então selecionar casas aleatoriamente em cada nível

Grupo 1:	Grupo 2:	Grupo 3:
renda baixa	renda média	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 do condado de West Ridge 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° membro da população.
- No condado de West Ridge, 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.

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.

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).

- Estatística descritiva
- Probabilidade
- Teste de hipótese e intervalo de confiança
- Correlação e Regressão
- Análise de Variância