

**1**

INTRODUÇÃO

1.1 INTRODUÇÃO À METODOLOGIA DA ESTATÍSTICA

O último quarto de século tem visto um aumento drástico no uso de métodos estatísticos nas ciências sociais. Existem várias razões para isso. A pesquisa nas ciências sociais tem, cada vez mais, empregado a orientação quantitativa. Como a pesquisa nas outras ciências, a nas ciências sociais geralmente estuda questões de interesse que analisam evidências fornecidas pelos dados empíricos. O crescimento da internet resultou em um aumento da informação quantitativa prontamente disponível. Finalmente, com o desenvolvimento de computadores ainda mais poderosos, *softwares* e metodologia estatística, os novos métodos disponíveis podem tratar mais realisticamente de questões que surgem na pesquisa na ciência social.

Por que estudar estatística?

O crescimento no uso da estatística é evidente nas mudanças dos conteúdos dos artigos publicados nas revistas de pesquisas nas ciências sociais e nos relatórios preparados pelo governo e pelas indústrias privadas. Uma rápida olhada nos assuntos recentes de periódicos como o *American Political Science Review* e *American Sociological Review* revela o papel fundamental da estatística na pesquisa. Por exemplo, para aprender sobre quais fatores têm o maior impacto no desempenho dos estudantes

na escola ou para investigar quais fatores afetam as crenças políticas das pessoas ou a qualidade do seu sistema de saúde ou sua decisão sobre quando se aposentar, os pesquisadores coletam informações e as processam usando análises estatísticas. Pelo papel da estatística em muitos estudos e pesquisas, cada vez mais os departamentos acadêmicos requerem que seus graduandos cursarem disciplinas de estatística.

Nos dias de hoje, os cientistas sociais trabalham em uma ampla variedade de áreas que usam métodos estatísticos, como agências governamentais, organizações de comércio e centros de saúde. Por exemplo, cientistas sociais atuando nas agências do governo que lidam com políticas de bem-estar social, com problemas ambientais ou de saúde pública precisam usar métodos estatísticos ou pelo menos ler relatórios que contenham estatística. Os geriatras geralmente devem avaliar recomendações de estudos que contenham investigações quantitativas de novas terapias ou novas formas de tratar os idosos. Alguns cientistas sociais ajudam os gerentes a avaliar o desempenho dos empregados usando medidas quantitativas e a determinar fatores que ajudem a prever a venda dos produtos. Na verdade, de modo crescente, muitos empregos para os cientistas sociais esperam que eles tenham um conhecimento sobre métodos estatísticos como uma ferramenta básica de trabalho. Como diz a piada: "O que um sociólogo que foi apro-

vado em estatística diz para um sociólogo que foi reprovado? ‘Eu vou querer um Big Mac, fritas e uma Coca.’”

Não obstante, entender de estatística é importante mesmo se você nunca usar os métodos estatísticos na sua carreira. Todos os dias você está exposto a uma explosão de informações, de propagandas, novos relatórios, campanhas políticas, pesquisas de opiniões sobre assuntos controversos e outras comunicações contendo argumentos estatísticos. A estatística ajuda você a dar sentido a toda essa informação e a entender melhor o mundo. Assim, você vai achar os conceitos deste livro úteis para julgar a informação que encontrará no seu dia a dia.

Sabemos que você não está lendo este livro esperando se tornar um estatístico. Além do mais, você deve sofrer de fobia da matemática e tem medo do que está por vir. Por favor, tenha certeza de que você pode ler este livro e aprender os conceitos básicos e os métodos estatísticos com pouco conhecimento de matemática. Só porque pode ter tido dificuldades no curso de matemática, não significa que estará em desvantagem aqui. Para entender este livro, o pensamento lógico e a perseverança são mais importantes do que a matemática. Na nossa experiência, o fator mais importante para ir bem em um curso de estatística é o tempo que você gasta no curso – assistindo às aulas, fazendo as tarefas de casa, lendo e relendo este livro, estudando suas anotações de aula, trabalhando junto com os seus colegas de aula, pedindo ajuda ao seu professor ou assistente – não o seu conhecimento matemático ou o seu gênero ou a sua raça ou se você tinha medo de estatística no início do curso.

Não fique frustrado se o aprendizado for lento e você tiver que ler o capítulo algumas vezes até que ele comece a fazer sentido. Da mesma forma que você não esperaria fazer um único semestre de uma língua estrangeira e ser capaz de falar aquela língua fluentemente, o mesmo é verdadeiro com a linguagem da estatística. Uma

vez que você completou apenas uma parte deste livro, irá entender melhor como lidar com a informação estatística.

Dados

A coleta de informação é o coração de todas as ciências, é o que fornece as observações usadas na análise estatística. As observações coletadas sobre as características de interesse são chamadas coletivamente de dados.

Por exemplo, um estudo pode conduzir um levantamento de dados com 1000 pessoas para observar características como opinião sobre a legalização da maconha, filiação política, ideologia política, quantas vezes participam de serviços religiosos, número de anos de escolaridade, renda anual, estado civil, raça e sexo. Os dados para uma pessoa em particular consistiriam de observações como (opinião = não é a favor da legalização, partido = republicano, ideologia = conservador, religiosidade = uma vez por semana, educação = 14 anos, estado civil = casado, raça = branca, sexo = feminino). Olhar os dados da forma correta nos ajuda a aprender sobre como essas características estão relacionadas. Podemos, então, responder a perguntas como: “As pessoas que frequentam mais a igreja tendem a ser mais conservadoras politicamente?”

Para obter dados, as ciências sociais usam uma série de métodos, incluindo inquéritos, experimentos e observações diretas do comportamento em ambientes naturais. Além disso, os cientistas sociais geralmente analisam dados já coletados para outros propósitos, como relatórios da polícia, materiais do censo e arquivos de hospital. Coleções de dados existentes já coletados são denominados de bancos de dados. Muitos bancos de dados estão, agora, disponíveis na internet. Um banco de dados muito importante para os cientistas sociais contém resultados a partir de 1972 da General Social Survey (Pesquisa Social Geral).

EXEMPLO 1.1 A Pesquisa Social Geral (PSG)

A cada dois anos, o Centro Nacional de Pesquisa de Opinião da Universidade de Chicago conduz a Pesquisa Social Geral (PSG). Este levantamento de dados de aproximadamente 2000 adultos fornece dados sobre opiniões e comportamentos do público norte-americano. Os cientistas sociais a utilizam para investigar como adultos americanos respondem a uma grande variedade de perguntas como: "Você acredita na vida após a morte?", "Você estaria disposto a pagar preços mais altos para proteger o meio ambiente?" e "Você acha que uma criança da pré-escola provavelmente sofrerá se sua mãe trabalhar?". Levantamentos de dados similares também ocorrem em outros países, como a Pesquisa Social Geral administrada pela Statistics Canada, a Pesquisa de Atitudes Sociais Britânica e o levantamento de dados da Eurobarometer e a Pesquisa Social Europeia para os países da União Europeia.

É fácil obter resumos dos dados do banco de dados da PSG. Mostraremos isso usando uma pergunta que foi feita num levantamento de dados: "Aproximadamente quantos bons amigos você tem?".

- Vá para o site <http://sda.berkeley.edu/GSS/> da Universidade da Califórnia, Berkeley.
- Clique em [*GSS – with 'no weight' as the default weight selection*].
- No painel à direita na linha *Row*, abaixo de "*REQUIRED Variable names to specify*", digite "NUMFRIEND". Clique em [*Run the table*] abaixo, quase ao final da página.

Como resultado você terá uma tabela que mostra os valores possíveis para o "número de bons amigos" e o percentual de cada resposta possível bem como o número de pessoas (abaixo do percentual). As respostas mais comuns foram 2 e 3 (com aproximadamente 16% para cada uma delas). ■

O que é estatística?

Neste livro, usamos o termo "estatística" em um sentido amplo para remeter aos métodos para obter e analisar dados.

Estatística

A estatística consiste em um conjunto de métodos para obter e analisar dados.

Especificamente, a estatística fornece métodos para:

1. **Delineamento:** Planejar como coletar os dados para estudos de pesquisa.
2. **Descrição:** Resumir os dados.
3. **Inferência:** Fazer previsões baseadas nos dados.

O **delineamento** se refere ao planejamento de como obter os dados. Para um levantamento de dados, por exemplo, os aspectos do delineamento especificariam como selecionar as pessoas a serem entrevistadas e como construir o questionário a ser administrado.

A descrição se refere ao resumo dos dados, para ajudar a entender a informação que eles fornecem. Por exemplo, uma análise do número de bons amigos baseada nos dados da PSG pode iniciar com uma lista do número relatado para cada pessoa que respondeu àquela pergunta naquele ano. Os dados brutos são uma lista completa de observações, pessoa por pessoa. Entretanto, eles não são fáceis de compreender. Ficamos atolados em números. Para apresentar os resultados, em vez de listar *todas* as informações, podemos resumir os dados com um gráfico (diagrama) ou uma tabela mostrando os percentuais relatando 1 bom amigo, 2 bons amigos, 3, ... , e assim por diante. Ou poderíamos apresentar o número médio de bons amigos, que foram 6, ou a resposta mais comum, que foi 2 amigos. Os gráficos, tabelas e resumos numéricos são chamados de **estatísticas descritivas**.

A **inferência** se refere a fazer previsões baseadas nos dados. Por exemplo, para os dados da PSG sobre o número apresentado de bons amigos, 6,2% relataram ter somente 1 bom amigo. Podemos usar essa informação para prever o percentual de mais de 200 milhões de adultos nos Estados Unidos naquele período que tinham somente 1 bom amigo? Um método apresentado neste livro nos permite prever que aquele percentual não é maior do que 8%. As previsões feitas usando dados são chamadas de inferências estatísticas.

* A **descrição e a inferência** são os dois tipos de **análise estatística** – maneiras de se analisar dados. Os cientistas sociais usam a estatística descritiva e a inferencial para responder a perguntas sobre fenômenos sociais. Por exemplo, “Ter a pena de morte disponível para punição está associado à redução de crimes violentos?”, “O desempenho dos estudantes na faculdade depende da quantia de dinheiro gasta por estudante, do tamanho das turmas ou dos salários dos professores?”.

1.2 ESTATÍSTICA DESCRITIVA E ESTATÍSTICA INFERENCIAL

A Seção 1.1 explicou que a estatística consiste em métodos para *delinear* estudos e *analisar* dados coletados nos estudos. Os métodos para analisar dados incluem métodos descritivos para resumir os dados e métodos inferenciais para fazer previsões. Uma análise estatística é classificada como **descritiva** ou **inferencial** se seu propósito principal for descrever ou fazer previsões. Para melhor explicar essa distinção, a seguir definiremos a população e a amostra.

Populações e amostras

As entidades que um estudo observa são chamadas de sujeitos para o estudo. Fre-

quentemente, os sujeitos são pessoas, como na PSG, mas podem ser famílias, escolas, cidades ou empresas, por exemplo.

✓ População e amostra

A população é todo o conjunto de sujeitos de interesse em um estudo. Uma amostra é um subconjunto da população sobre o qual o estudo coleta os dados.

Na PSG de 2004, a amostra foram os 2813 norte-americanos adultos que participaram do levantamento de dados. A população era formada por todos os norte-americanos adultos daquele período – mais de 200 milhões de pessoas.

O objetivo final de qualquer estudo é aprender mais sobre as populações. Mas, é quase sempre necessário e mais prático observar somente amostras destas populações. Por exemplo, a PSG e organizações de pesquisas de opinião pública como a Gallup geralmente selecionam amostras de aproximadamente 1000 a 3000 norte-americanos para coletar informações sobre opiniões e crenças da população de *todos* os norte-americanos.

✓ Estatística descritiva

A estatística descritiva resume a informação em uma coleção de dados.

A estatística descritiva consiste em gráficos, tabelas e números tais como médias e percentuais. O principal objetivo da estatística descritiva é reduzir os dados a formas mais simples e de fácil entendimento sem distorcer ou perder muita informação.

Embora os dados estejam geralmente disponíveis apenas para uma amostra, estatísticas descritivas são úteis, também, quando os dados estão disponíveis para toda a população, como em um censo. Ao contrário, a estatística inferencial só é aplicada quando os dados fizerem parte de

parâmetro → população
estatístico → amostra

uma amostra e se quer fazer uma previsão para toda a população.

✓ Estatística inferencial

A estatística inferencial fornece previsões sobre a população, baseada nos dados de uma amostra desta população.

✓ Parâmetros e estatística

Um parâmetro é um resumo numérico da população. Uma estatística é um resumo numérico dos dados amostrais.

EXEMPLO 1.2 Crença no paraíso

Em dois dos seus levantamentos de dados a PSG perguntou: "Você acredita no paraíso?". A população de interesse era o grupo de todos os adultos dos Estados Unidos. No levantamento de dados mais recente no qual esta pergunta foi feita, 86% dos 1158 sujeitos amostrados respondeu *sim*. Estaríamos interessados, entretanto, não somente nestas 1158 pessoas, mas em toda a população de adultos dos Estados Unidos.

A estatística inferencial fornece uma previsão para a população usando os dados amostrais. Um método inferencial apresentado no Capítulo 5 prevê que o percentual da população que acredita no paraíso está entre 84% e 88%. Isto é, o valor amostral de 86% tem uma "margem de erro" de 2%. Embora o tamanho da amostra fosse muito pequeno comparado com o tamanho da população, podemos concluir que uma grande parte da população acredita no paraíso. ■

As análises inferenciais podem prever características de toda a população muito bem, selecionando amostras que são pequenas em relação ao tamanho da população. Por isso que muitas pesquisas amostram somente mil pessoas, mesmo quando a população tem milhões de pessoas. Neste livro, veremos porque isto funciona.

No último quarto de século, os cientistas sociais têm reconhecido de modo crescente o poder dos métodos da estatística inferencial. A apresentação destes métodos ocupa grande parte deste livro, iniciando no Capítulo 5.

O Exemplo 1.2 estimou o percentual de norte-americanos que acreditavam no paraíso. O parâmetro era o percentual da população que acreditava no paraíso. Seu valor era desconhecido. A inferência sobre este parâmetro foi baseada em uma estatística – o percentual dos 1158 sujeitos do levantamento de dados que responderam *sim*, a saber, 86%. Visto que este número descreve uma característica da amostra, ele é uma estatística descritiva.

Na prática, o interesse principal são os valores dos parâmetros, não os valores da estatística da amostra específica selecionada. Por exemplo, observando os resultados de uma pesquisa de opinião pública antes de uma eleição, estamos mais interessados nos percentuais da população em favor dos vários candidatos do que nos percentuais da amostra das pessoas entrevistadas. É importante que a amostra e a estatística a descrevam somente enquanto elas nos ajudam a fazer inferências sobre os parâmetros desconhecidos da população.

Um aspecto importante da inferência estatística envolve relatar a precisão provável da amostra estatística que estima o parâmetro da população. Para o Exemplo 1.2 da crença no paraíso, um método de estatística inferencial previu quão próximo o valor da amostra de 86% era provável de ser o percentual desconhecido da população que acredita no paraíso. A margem de erro relatada foi de 2%.

Quando os dados existem para toda a população, como em um censo, é possível encontrar valores reais dos parâmetros de interesse. Neste caso, não existe necessidade de se usar os métodos de estatística inferencial.

Definindo populações: real e conceitual

Frequentemente, a população à qual a inferência se aplica é um conjunto real de sujeitos. No Exemplo 1.2, foram os residentes adultos dos Estados Unidos. Algumas vezes, a generalização se refere a uma população conceitual – uma que não existe na realidade, ou seja, é hipotética.

Por exemplo, suponha que uma organização de consumidores avalia o consumo de combustível para um novo modelo de automóvel, observando o número médio de milhas por galão para cinco carros amostrados dirigidos em um percurso padrão de 100 milhas. Suas inferências se referem ao desempenho neste percurso para a população conceitual de *todos* os carros deste modelo que serão ou poderiam hipoteticamente ser manufaturados.

1.3 O PAPEL DOS COMPUTADORES NA ESTATÍSTICA

Ao longo do tempo, computadores cada vez mais poderosos são comercializados e *softwares* mais potentes e fáceis de usar são, também, desenvolvidos para métodos estatísticos. Esses *softwares* trazem um grande benefício para o uso da estatística.

Software estatístico

O SPSS (Pacote Estatístico para as Ciências Sociais), SAS, MINITAB e Stata são os *softwares* estatísticos mais populares nos *campi* universitários. É mais fácil aplicar os métodos estatísticos usando estes *softwares* do que fazer os cálculos à mão. Além disso, muitos métodos apresentados neste livro são muito complexos para serem feitos à mão ou com calculadoras.

A maioria dos capítulos deste livro, incluindo todos aqueles que apresentam métodos que requerem muitos cálculos, mostram exemplos da saída de um *soft-*

ware estatístico. Um dos objetivos deste livro é ensiná-lo o que procurar em uma saída computacional e como interpretá-la. O conhecimento de programação não é necessário para usar um *software* estatístico ou para ler este livro.

Um dos apêndices do livro, organizado por capítulos, explica como usar o SPSS e o SAS. Você pode consultar esse apêndice à medida que ler cada capítulo para aprender como usá-los para executar a análise daquele capítulo.

Arquivos de dados

A Figura 1.1 mostra um exemplo de dados organizados em um **arquivo de dados** para a análise por um *software* estatístico. Um arquivo de dados tem a forma de uma planilha:

- Cada linha contém as observações para um sujeito da amostra em particular.
- Cada coluna contém as observações de uma característica em particular.

A Figura 1.1 é uma janela para editar dados no SPSS. Ela mostra dados para os primeiros dez sujeitos em uma amostra, para as características: sexo, grupo racial, estado civil, idade e renda anual (em milhares de dólares). Alguns dos dados são numéricos e outros consistem em rótulos. O Capítulo 2 introduz os tipos de dados utilizados nos arquivos de dados.

Usos e usos impróprios de um software estatístico

Uma nota de alerta: o acesso fácil a métodos estatísticos usando *softwares* tem riscos assim como vantagens. É comum a aplicação de métodos inapropriados. Um computador executa a análise solicitada não se importando se as suposições requeridas para o uso adequado do método foram satisfeitas.

Análises incorretas ocorrem quando os pesquisadores não tomam tempo sufi-

chapter_1_data.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	subject	sex	race	married	age	income
1	1	female	white	yes	23	18.3
2	2	female	black	no	37	31.9
3	3	male	white	yes	47	64.0
4	4	female	white	yes	61	46.2
5	5	male	hispanic	yes	30	16.5
6	6	male	white	no	21	14.0
7	7	male	white	yes	55	26.1
8	8	female	white	no	27	59.8
9	9	female	hispanic	yes	61	21.5
10	10	male	black	no	47	50.0

Data View Variable View

SPSS Processor is ready

Figura 1.1 Exemplo de parte de um arquivo de dados do SPSS.

ciente para entender o método estatístico, as suposições para o seu uso ou a sua adequabilidade para um problema específico. É vital entender o método antes de usá-lo. Apenas saber como usar o *software* estatístico não garante uma análise apropriada. Você irá precisar de um bom conhecimento de estatística para entender qual método selecionar, quais opções escolher daquele método e como tirar conclusões válidas do resultado. O objetivo principal deste livro é dar a você este conhecimento.

1.4 RESUMO DO CAPÍTULO

O campo da estatística inclui métodos para:

- delinear levantamentos amostrais,
- descrever os dados e
- fazer inferências (previsões) usando os dados.

Os métodos estatísticos são, normalmente, aplicados às observações de uma amostra coletada de uma população de interesse. A estatística resume os dados amostrados, enquanto os parâmetros resumem toda a população. Existem dois tipos de análise estatística:

- A **descritiva**, que resume a amostra ou os dados da população com números, tabelas e gráficos.
- A **inferencial**, que faz previsões sobre os parâmetros da população, baseada nos dados amostrais.

Um arquivo de dados tem uma linha separada de dados para cada sujeito e uma coluna separada para cada característica. Os métodos estatísticos são fáceis de aplicar aos arquivos de dados usando um *software*. Isto nos livra do trabalho computacional penoso e nos ajuda a focar na aplicação e interpretação adequada dos métodos.