

1º Texto

1. Qual é o objetivo?

Texto extraído do livro "Estatística: O que é, para que serve, como funciona" de Charles Whellan

TENHO NOTADO UM fenômeno curioso. Os alunos se queixam de que a estatística é confusa e irrelevante. Aí os mesmos alunos saem da classe e conversam alegremente durante o almoço sobre médias de rebatidas (durante o verão) ou sensação térmica (no inverno) ou médias de notas (sempre). Reconhecem que o “índice de passes” da Liga Nacional de Futebol Americano (NFL, na sigla em inglês) – uma estatística que condensa o desempenho de um *quarterback* num único número – é uma medida um tanto falha e arbitrária do desempenho do *quarterback* num dia de jogo. Os mesmos dados (índice de passes completos, média de jardas por tentativa de passe, porcentagem de passes para *touchdown* por tentativa de passe e índice de interceptações) poderiam ser combinados de maneira diferente, tal como dar um peso maior ou menor para qualquer uma dessas informações, de modo a gerar uma medida de desempenho diferente, mas igualmente confiável. Todavia, qualquer um que já tenha assistido a uma partida de futebol americano reconhece que é conveniente ter um número único que possa ser usado para abranger a performance do *quarterback*.^a

A avaliação do *quarterback* é perfeita? Não. A estatística raramente oferece um único modo “certo” de fazer qualquer coisa. Ela fornece informação significativa de uma maneira facilmente acessível? Com toda certeza. É uma boa ferramenta para fazer uma comparação rápida entre os desempenhos de dois *quarterbacks* em determinado dia. Eu sou fã dos Chicago Bears. Durante os *playoffs* (jogos eliminatórios, ou mata-mata) de 2011, os Bears jogaram contra os Packers; os Packers ganharam. Há uma porção de recursos que eu poderia usar para descrever o jogo, inclusive páginas e páginas de análise e dados brutos. Mas eis uma análise mais sucinta: Jay Cutler, *quarterback* dos Chicago Bears, teve um índice de passes de 31,8. Em contraste, o *quarterback* dos Green Bay Packers, Aaron Rodgers, teve um índice de passes de 55,4. De forma similar, podemos comparar a atuação de Jay Cutler com a que ele teve num jogo anterior contra os Packers na mesma temporada, quando teve um índice de passes de 85,6. Esses dados revelam muito do que se precisa saber para compreender por que os Bears venceram os Packers mais cedo nessa temporada, mas perderam nas eliminatórias.

Essa é uma sinopse muito útil do que aconteceu em campo. Ela simplifica as coisas? Sim, e esse é ao mesmo tempo o ponto forte e a fragilidade de qualquer descrição estatística. Um número nos diz que Jay Cutler foi

superado por Aaron Rodgers na derrota dos Bears no mata-mata. Por outro lado, o número não diz se tudo deu errado para o *quarterback*. Não tem como saber, por exemplo, se ele fez um passe perfeito que não foi agarrado pelo recebedor e que depois foi interceptado, ou se ele “compareceu” em certas jogadas fundamentais (já que todo passe completo tem o mesmo peso, seja numa terceira descida crucial ou numa jogada sem propósito no fim do jogo), ou se a defesa esteve terrível. E assim por diante.

O curioso é que as mesmas pessoas que se sentem perfeitamente à vontade discutindo estatísticas no contexto de esportes, ou do clima, ou de notas escolares, reagem com ansiedade quando um pesquisador começa a explicar algo como o índice de Gini (ou coeficiente de Gini), que é uma ferramenta-padrão em economia para medir a desigualdade de renda. Explicarei o que é o índice de Gini daqui a um instante, mas por enquanto *o mais importante é reconhecer que este índice é exatamente a mesma coisa que o índice de passes*. Trata-se de uma conveniente ferramenta para reduzir informações complexas a um único número. Como tal, ele tem os pontos fortes da maioria das estatísticas descritivas, notadamente ao prover um meio fácil de comparar a distribuição de renda em dois países, ou num só país em diferentes períodos de tempo.

O coeficiente de Gini mede o quanto a riqueza (ou renda) é partilhada equitativamente dentro de um país numa escala de zero a um. A estatística pode ser calculada para a riqueza ou para a renda anual e em nível individual ou familiar. (Todas essas estatísticas serão bastante correlatas, mas não idênticas.) O coeficiente de Gini, assim como o índice de passes, não tem significado intrínseco. Ele é um instrumento de comparação. Um país em que todas as famílias tivessem uma riqueza idêntica teria um índice de Gini de zero. Em contraste, um país em que uma única família detivesse toda a riqueza teria um coeficiente de Gini de um. Como você provavelmente pode imaginar, quanto mais perto o país estiver de um, mais desigual é sua distribuição de riqueza. Os Estados Unidos têm um índice de Gini de 0,45, segundo a Agência de Inteligência Americana (CIA, na sigla em inglês) (aliás, uma grande coletora de estatísticas!).¹ E daí?

Uma vez colocado em contexto, esse número pode nos dizer muita coisa. Por exemplo, a Suécia tem um coeficiente de Gini de 0,23; o Canadá, 0,32; a China, 0,42; o Brasil, 0,54; e a África do Sul, 0,65.^b Ao passarmos os olhos por esses números, temos uma noção de onde os Estados Unidos se encontram em relação ao resto do mundo quando se trata de desigualdade de renda. Também podemos comparar momentos diferentes no tempo. O índice de Gini para os Estados Unidos era de 0,41 em 1997 e aumentou para 0,45 durante a década seguinte. (Os dados mais recentes da CIA são de 2007.) Isto nos diz de forma objetiva que, ao mesmo tempo em que os Estados Unidos ficaram mais ricos nesse período, a distribuição da riqueza ficou mais

desigual. Mais uma vez, podemos comparar as mudanças no coeficiente de Gini entre países ao longo de praticamente o mesmo intervalo de tempo. A desigualdade no Canadá permaneceu praticamente inalterada durante esse mesmo período. A Suécia teve um crescimento econômico significativo ao longo das últimas duas décadas, mas seu coeficiente de Gini na verdade caiu de 0,25 em 1992 para 0,23 em 2005, o que significa que a Suécia ficou mais rica e mais igualitária nesse período.

Será que o coeficiente de Gini é o instrumento perfeito para medir a desigualdade? Absolutamente não – exatamente da mesma maneira que o índice de passes não é uma medida perfeita para aferir o desempenho de um *quarterback*. Mas com certeza nos fornece alguma informação valiosa sobre um fenômeno socialmente significativo num formato conveniente.

Ao mesmo tempo, fomos lentamente pavimentando o caminho para responder à pergunta formulada no título do capítulo: qual é o objetivo? O objetivo é mostrar que a estatística nos ajuda a processar dados, que na verdade é apenas um nome pomposo para informação. Às vezes os dados são triviais no grande esquema das coisas, como no caso de estatísticas esportivas. Às vezes servem para revelar algum aspecto da natureza da existência humana, como no caso do coeficiente de Gini.

Mas, como qualquer propaganda de produtos na TV ressaltaria, *Isso não é tudo!* Hal Varian, economista-chefe do Google, disse ao *New York Times* que a carreira de estatístico será considerada “sexy” na próxima década.² Sou o primeiro a reconhecer que os economistas às vezes têm uma definição deturpada de “sexy”. Ainda assim, considere as seguintes perguntas disparatadas:

Como podemos surpreender escolas que estejam trapaceando em seus testes padronizados?

Como a Netflix sabe o tipo de filme que você gosta?

Como podemos descobrir que substâncias ou comportamentos causam câncer, considerando que não podemos conduzir experimentos cancerígenos em seres humanos?

Rezar por pacientes que passaram por cirurgia melhora sua recuperação?

Existe de fato algum benefício econômico em ter o diploma de uma faculdade ou universidade altamente seletiva?

O que está provocando o aumento do índice de autismo?

A estatística pode ajudar a responder a essas perguntas (ou, esperamos, em breve poderá). O mundo está produzindo mais e mais dados, cada vez mais rápido. Contudo, como observou o *New York Times*, “dados são meramente a matéria-prima do conhecimento”.^{3c} A estatística é a ferramenta mais poderosa que temos para usar as informações com finalidade significativa, seja para identificar jogadores de beisebol subestimados ou conceder um pagamento mais justo aos professores. Eis uma rápida exposição de como a estatística

pode dar significado a dados brutos.

Descrição e comparação

Um placar de boliche é uma estatística descritiva. Uma média de rebatidas também. A maioria dos fãs americanos de esporte com mais de cinco anos já está familiarizada com o campo da estatística descritiva. Usamos números, nos esportes e em qualquer outra área da vida, para resumir informações. Até que ponto um jogador de beisebol como Mickey Mantle foi bom? Seu índice de rebatidas na carreira foi de 0,298. Para um fã de beisebol, essa é uma constatação significativa, o que é notável se você pensar no assunto, pois abrange uma carreira de dezoito temporadas.⁴ (Há algo, suponho eu, de deprimente em ter o trabalho de uma vida inteira reduzido a um único número.) É claro que os fãs de beisebol também vieram a reconhecer que estatísticas descritivas diferentes da média de rebatidas podem dar uma visão geral melhor do valor do jogador no campo.

Nós avaliamos o desempenho acadêmico de estudantes do ensino médio e superior por meio de uma pontuação baseada na média de notas, chamada GPA – *grade point average*.^d A cada letra que simboliza uma nota é dado um valor numérico; geralmente um A vale quatro pontos, um B vale três, um C vale dois, e assim por diante. Em vias de se formar, por exemplo, quando alunos de ensino médio estão se candidatando à faculdade e alunos de faculdade estão procurando emprego, a GPA é um instrumento prático para avaliar o potencial acadêmico. Um estudante que tem GPA 3,7 é claramente mais forte do que um na mesma escola com GPA 2,5. Isso constitui uma bela estatística descritiva. É fácil de calcular, fácil de entender e fácil de fazer comparações entre estudantes.

Mas não é perfeita. A GPA não reflete a dificuldade dos cursos que diferentes alunos possam ter cursado. Como podemos comparar um estudante com GPA 3,4 em aulas que parecem ser relativamente não desafiadoras com um estudante com GPA 2,9 que cursou cálculo, física e outras matérias difíceis? Eu frequentei um colégio de ensino médio que tentou solucionar esse problema dando um peso extra a aulas difíceis, de modo que um A numa aula “honrosa” valia cinco pontos em vez dos quatro habituais. Isto gerava seus próprios problemas. Minha mãe foi rápida em reconhecer a distorção causada por esse “ajuste” de GPA. Para um aluno que cursasse uma porção de matérias honrosas (eu), qualquer A num curso não honroso, tal como educação física ou ginástica, na realidade puxaria o meu GPA para baixo, mesmo sendo impossível sair-se melhor do que tirar A nessas matérias. Como resultado, meus pais me proibiram de tomar aulas de direção no ensino

médio, pois até mesmo um desempenho perfeito diminuiria minhas chances de entrar numa faculdade competitiva e seguir adiante para escrever livros populares. Em vez disso, pagaram uma autoescola para eu cursar à noite durante o verão.

Foi uma maluquice? Foi. Mas um dos temas deste livro é que se basear exageradamente em qualquer estatística descritiva pode levar a conclusões errôneas, ou gerar comportamentos indesejáveis. Meu rascunho original da sentença acima usava a expressão “estatística descritiva ultrassimplificada”, mas risquei a palavra “ultrassimplificada” porque é redundante. A estatística descritiva existe para simplificar, o que sempre implica alguma perda de nuance ou detalhe. Qualquer um que trabalhe com números tem que reconhecer isso.

Inferência

Quantas pessoas sem teto vivem nas ruas de Chicago? Com que frequência pessoas casadas fazem sexo? Essas podem parecer perguntas de tipos brutalmente diferentes, mas, na verdade, ambas podem ser respondidas (não perfeitamente) pelo uso de ferramentas estatísticas básicas. Uma função-chave da estatística é usar os dados que temos para fazer conjecturas informadas sobre perguntas mais amplas para as quais não temos informação completa. Em suma, podemos usar dados do “mundo conhecido” para fazer inferências informadas sobre o “mundo desconhecido”.

Começamos com a pergunta sobre os sem-tetos. É custoso e logisticamente difícil contar a população sem-teto numa grande área metropolitana. Todavia, é importante ter uma estimativa numérica dessa população a fim de prover serviços sociais, obter elegibilidade para recursos estaduais e federais e garantir representação no Congresso. Uma importante prática estatística é a amostragem, que é o processo de coletar dados de uma área pequena, digamos, um punhado de setores censitários, e aí usar esses dados para fazer um juízo informado, ou inferência, sobre a população sem-teto da cidade como um todo. A amostragem requer muito menos recursos do que tentar contar uma população inteira; se feita adequadamente, ela pode ser igualmente acurada.

Uma pesquisa de opinião política é uma forma de amostragem. Uma instituição de pesquisa entra em contato com uma amostragem de lares que seja amplamente representativa da população total e pergunta suas opiniões sobre um determinado assunto ou candidato. Obviamente, esse procedimento é muito mais barato e rápido do que tentar contatar cada lar num estado ou no país inteiro. Para a empresa de pesquisa de opinião Gallup, uma pesquisa

metodologicamente sólida de mil lares produz aproximadamente os mesmos resultados que uma pesquisa que tente contatar cada lar americano.

Foi assim que calculamos com que frequência os americanos fazem sexo, com quem e de que tipo. Em meados dos anos 1990, o Centro Nacional de Pesquisa de Opinião da Universidade de Chicago realizou um ambicioso estudo sobre o comportamento sexual americano. Os resultados baseavam-se em levantamentos detalhados conduzidos em pessoa com uma amostragem grande e representativa de adultos americanos. Se você avançar neste livro, no Capítulo 10 saberá o que descobriram. *Quantos outros livros de estatística podem lhe prometer isso?*

Avaliação de risco e outros eventos relacionados com probabilidade

Cassinos ganham dinheiro no longo prazo – sempre. Isso não quer dizer que estão ganhando dinheiro em qualquer dado momento. Quando os sinos e apitos disparam, algum apostador de peso acabou de ganhar milhares de dólares. Toda a indústria do jogo é constituída de jogos de azar, o que significa que o resultado de qualquer lance de dados ou virada de carta é incerto. Ao mesmo tempo, as probabilidades subjacentes para fatos relevantes – fazer 21 pontos no *blackjack* ou acertar no vermelho na roleta – são conhecidas. Quando as probabilidades subjacentes favorecem os cassinos (como sempre ocorre), podemos estar cada vez mais certos de que a “casa” vai sair na frente à medida que o número de apostas cresce mais e mais, mesmo que aqueles sinos e apitos continuem disparando.

Acontece que esse é um poderoso fenômeno em áreas da vida muito além dos cassinos. Muitos negócios precisam avaliar os riscos associados a uma variedade de resultados adversos. Não é possível fazer esses riscos desaparecerem inteiramente, da mesma forma que um cassino não pode garantir que você não vá ganhar toda mão de *blackjack* que jogar. No entanto, qualquer negócio sujeito a incertezas pode administrar esses riscos fazendo uso de processos de engenharia, de modo que a probabilidade de um resultado adverso, qualquer coisa desde uma catástrofe ambiental até um produto defeituoso, torne-se aceitavelmente baixa. Empresas de Wall Street avaliam com frequência os riscos apresentados aos seus portfólios em diferentes cenários, sendo que cada um desses cenários é pesado com base na sua probabilidade. A crise financeira de 2008 foi precipitada em parte por uma série de eventos de mercado que haviam sido considerados extremamente improváveis, como se cada jogador do cassino fizesse 21 pontos a noite inteira. Mais adiante no livro argumentarei que esses modelos de Wall Street

eram falhos e que os dados utilizados para avaliar os riscos subjacentes eram limitados demais, mas o ponto aqui é que qualquer modelo para lidar com riscos deve ter a probabilidade como alicerce.

Quando indivíduos e empresas não conseguem eliminar riscos inaceitáveis, buscam proteção de outras maneiras. Toda a indústria de seguros se baseia em cobrar clientes para protegê-los contra algum infortúnio, tal como uma batida de carro ou incêndio na casa. A indústria de seguros não ganha dinheiro eliminando esses acontecimentos; carros batem e casas pegam fogo todos os dias. Às vezes carros chegam a colidir contra casas, provocando incêndios. Em vez disso, o ramo de seguros ganha dinheiro cobrando prêmios que são mais que suficientes para pagar pelos prejuízos de batidas de carro e incêndios de casas. (A companhia de seguros pode tentar também reduzir seus prejuízos estimulando práticas de condução segura, a colocação de cercas ao redor de piscinas, a instalação de detectores de fumaça em cada quarto, e assim por diante.)

A probabilidade pode ser usada inclusive para detectar trapagens em algumas situações. A empresa Caveon Test Security é especializada naquilo que ela descreve como “dados forenses” para descobrir padrões que sugiram trapagens.⁵ Por exemplo, num colégio ou local de testes, a companhia (que foi fundada por um ex-elaborador do Teste de Raciocínio SAT nas escolas) formula exames em que a quantidade de *respostas erradas* idênticas é bastante improvável, geralmente um padrão que ocorreria por acaso apenas uma vez em 1 milhão. A lógica matemática provém do fato de que não podemos descobrir muita coisa quando um grupo grande de alunos inteiro responde à questão corretamente. É isso que se espera deles; eles podem estar colando, ou podem ser inteligentes. Mas quando esses mesmos alunos que fazem o teste assinalam uma resposta errada, não poderiam todos consistentemente dar *a mesma resposta errada*. Se isso acontecer, é possível que estejam copiando um do outro (ou compartilhando as respostas via mensagem de texto). A companhia também procura identificar exames em que alguém se saia significativamente melhor em questões difíceis do que em questões fáceis (o que sugere que ele ou ela tiveram acesso às respostas com antecedência) e exames nos quais o número de rasuras “errado para certo” seja muito maior que o de rasuras “certo para errado” (sugerindo que o professor ou o administrador alteraram a folha de respostas após o teste).

É claro que você pode ver limitações no uso da probabilidade. Um grupo grande de pessoas fazendo um teste pode dar as mesmas respostas erradas por coincidência; na verdade, quanto mais escolas avaliamos, mais provável é que observemos tais padrões como mero fruto do acaso. Uma anomalia estatística não comprova que houve má-fé. Delma Kinney, um homem de cinquenta anos de Atlanta, ganhou US\$1 milhão num jogo lotérico instantâneo em 2008 e depois outro milhão num jogo instantâneo em 2011.⁶ A probabilidade de

isso acontecer com a mesma pessoa é algo em torno de uma em 25 trilhões. Não podemos prender o sr. Kinney por fraude com base apenas nesse cálculo (embora possamos inquirir se ele tem parentes que trabalham para a loteria estadual). A probabilidade é uma arma num arsenal que requer bom julgamento.

Identificando relações importantes (Trabalho de detetive estatístico)

Fumar cigarros causa câncer? Temos uma resposta para essa pergunta – mas o processo de respondê-la não foi nem um pouco direto como se poderia pensar. O método científico determina que, se estamos testando uma hipótese científica, devemos conduzir um experimento controlado no qual a variável de interesse (por exemplo, o ato de fumar) seja a única coisa diferente entre o grupo experimental e o grupo de controle. Se observarmos uma diferença acentuada em algum resultado entre os dois grupos (por exemplo, câncer de pulmão), podemos inferir com segurança que a variável de interesse é o que causou o resultado. Não podemos fazer esse tipo de experimento com seres humanos. Se nossa hipótese de trabalho é a de que fumar causa câncer, seria antiético dividir recém-formados na faculdade em dois grupos, fumantes e não fumantes, e aí ver quem desenvolveu câncer no encontro de vinte anos de formatura. (Podemos conduzir experimentos controlados em humanos quando a nossa hipótese é que uma nova droga ou tratamento pode melhorar sua saúde; não podemos conscientemente expor sujeitos humanos quando esperamos um resultado adverso.)^e

Agora, você poderia retrucar que não precisamos conduzir um experimento eticamente duvidoso para observar os efeitos do fumo. Não poderíamos simplesmente pular toda a rebuscada metodologia e, no encontro de vinte anos, comparar os índices de câncer entre os que fumaram desde a graduação e os que não fumaram?

Não. Fumantes e não fumantes provavelmente serão diferentes em outros aspectos da vida além do seu comportamento em relação ao fumo. Por exemplo, fumantes podem ter propensão a outros hábitos, tais como beber demais ou comer mal, que provocam resultados de saúde adversos. Se os fumantes tiverem uma saúde particularmente ruim no encontro de vinte anos, não saberíamos se devemos atribuir essa condição ao fumo ou a outras práticas não saudáveis que muitos fumantes costumam adotar. Também teríamos um sério problema com os dados nos quais estamos baseando a nossa análise. Fumantes que ficaram gravemente doentes de câncer têm menos probabilidade de comparecer ao encontro de vinte anos. (Os fumantes

mortos definitivamente não vão aparecer.) Como resultado, qualquer análise da saúde dos participantes do encontro de vinte anos (relacionada com o fumo ou qualquer outra coisa) terá sérias falhas pelo fato de que os membros mais saudáveis da classe são os mais propensos a comparecer. Quanto mais tempo passar desde a época da formatura, digamos, num encontro de quarenta ou cinquenta anos, mais sério será esse viés.

Não podemos tratar seres humanos como ratos de laboratório. Sendo assim, a estatística se aproxima muito de um bom trabalho de detetive. Os dados geram pistas e padrões que, em última instância, podem levar a conclusões significativas. Você provavelmente já assistiu a alguma dessas impressionantes séries policiais, como *CSI: New York*, na qual detetives muito atraentes e peritos forenses se debruçam sobre pistas mínimas – DNA de uma guimba de cigarro, marcas de dentes numa maçã, uma única fibra do carpete do piso de um carro – e então usam a evidência para capturar um criminoso violento. O apelo da série é que esses peritos não dispõem de evidências convencionais usadas para encontrar um bandido, tal como uma testemunha ocular ou a fita de vídeo de um sistema de vigilância. Então, em vez disso, recorrem à inferência científica. A estatística faz basicamente a mesma coisa. Os dados apresentam pistas desorganizadas – a cena do crime. A análise estatística é o trabalho de detetive que processa os dados brutos para obter uma conclusão significativa.

Após o Capítulo 11, você irá apreciar o programa de TV que eu espero que estoure: *CSI: análise de regressão*, que se afastaria apenas um pouco daqueles outros procedimentos policiais carregados de ação. A análise de regressão é a ferramenta que possibilita aos pesquisadores isolar a relação entre duas variáveis, tais como fumo e câncer, ao mesmo tempo em que mantém constantes (ou “controla”) os efeitos de outras variáveis importantes, como dieta, exercício, peso, e assim por diante. Quando você lê no jornal que comer um bolinho de farelo de trigo por dia reduz o risco de se ter câncer de cólon, não precisa temer que algum grupo infeliz de cobaias humanas tenha sido forçado a comer bolinhos de farelo de trigo no porão de algum laboratório federal enquanto o grupo de controle no prédio ao lado comia ovos com bacon. Em vez disso, os pesquisadores coletam informações detalhadas de milhares de pessoas, inclusive sobre a frequência com que comem bolinhos de farelo de trigo, e então usam a análise de regressão para fazer duas coisas cruciais: (1) quantificar a associação observada entre comer bolinho de farelo de trigo e contrair câncer de cólon (por exemplo, um achado hipotético de que pessoas que comem bolinhos de farelo de trigo têm uma incidência de câncer de cólon 9% menor, controlando outros fatores que possam afetar a incidência da doença); e (2) quantificar a probabilidade de que a associação entre bolinhos de farelo de trigo e menor taxa de câncer de cólon observada nesse estudo seja uma mera coincidência – um acaso peculiar

nos dados para essa amostragem de pessoas –, e não uma percepção significativa sobre a relação entre dieta e saúde.

É claro que *CSI* será estrelado por atores e atrizes muito mais atraentes que os acadêmicos que tipicamente se debruçam sobre tais dados. Esses charmosos agentes (todos com PhD, apesar de terem apenas 23 anos) estudariam extensos conjuntos de dados e usariam as mais modernas ferramentas estatísticas para responder a importantes questões sociais: quais são as ferramentas mais eficazes para combater o crime violento? Que indivíduos são mais propensos a se tornarem terroristas? Mais adiante no livro, discutiremos o conceito de uma descoberta “estatisticamente significativa”, o que quer dizer que a análise revelou uma associação entre duas variáveis sem probabilidade de ser produto do mero acaso. Para pesquisadores acadêmicos, esse tipo de descoberta estatística configura uma “arma fumegante”, ou prova irrefutável. Em *CSI: análise de regressão*, posso visualizar uma pesquisadora trabalhando até tarde da noite no laboratório de informática por causa dos seus compromissos diurnos como membro do time olímpico de vôlei de praia. Quando ela pega o relatório impresso da sua análise estatística, vê exatamente o que vinha procurando: uma vasta e estatisticamente significativa relação em seu conjunto de dados entre alguma variável que ela supunha ser importante e os primeiros sintomas de autismo. Ela precisa relatar essa descoberta imediatamente!

Ela pega o relatório e corre pelo corredor, obrigada a andar um pouco mais devagar por estar de salto alto e com uma saia relativamente curta, justa e preta. Ela encontra seu parceiro, inexplicavelmente em boa forma e bronzeado para um sujeito que trabalha catorze horas por dia num laboratório de computadores no porão, e lhe mostra os resultados. Ele corre os dedos pelo cavanhaque cuidadosamente aparado, tira sua pistola Glock 9mm da gaveta da escrivaninha e a coloca no coldre de ombro sob seu terno Hugo Boss de US\$5 mil (o que também é inexplicável, dado que seu salário acadêmico inicial não deve passar de US\$38 mil por ano). Juntos, os peritos em análise de regressão caminham decididamente para ver seu chefe, um veterano grisalho que superou relacionamentos fracassados e o alcoolismo...

Tudo bem, você não precisa entrar na onda do drama televisivo para apreciar a importância desse tipo de pesquisa estatística. Quase todo desafio social a que damos importância tem sido informado mediante a análise sistemática de grandes conjuntos de dados. (Em muitos casos, a coleta de dados relevantes, que é um procedimento caro e que consome tempo, desempenha um papel crucial nesse processo, como será explicado no Capítulo 7.) Posso ter glamorizado meus personagens em *CSI: análise de regressão*, mas não o tipo de questões significativas que eles poderiam examinar. Há uma literatura acadêmica sobre terroristas e atentados suicidas – um tema que seria difícil de estudar por meio de sujeitos humanos (ou mesmo

ratos de laboratório). Um desses livros, *What Makes a Terrorist*, foi escrito por um dos meus professores de estatística da graduação. O livro chega a suas conclusões a partir de dados reunidos sobre ataques terroristas ao redor do mundo. Uma descoberta amostral: terroristas não são desesperadamente pobres, nem têm nível escolar baixo. O autor, o economista de Princeton Alan Krueger, conclui: “Terroristas tendem a vir de famílias com bom nível educacional, de classe média ou alta.”⁷

Por quê? Bem, isso expõe uma das limitações da análise de regressão. Podemos isolar uma associação forte entre duas variáveis usando análise estatística, mas não necessariamente podemos explicar por que essa relação existe e, em alguns casos, não podemos saber ao certo se essa relação é causal, significando que uma mudança numa das variáveis esteja realmente causando mudança na outra. No caso do terrorismo, o professor Krueger parte da hipótese de que, uma vez que os terroristas são motivados por metas políticas, aqueles que são mais cultos e abastados têm o incentivo mais forte de mudar a sociedade. Esses indivíduos também podem ser particularmente inflamados pela supressão da liberdade, outro fator associado ao terrorismo. No estudo de Krueger, países com altos níveis de repressão política têm atividade terrorista mais intensa (mantendo outros fatores constantes).

Essa discussão me leva de volta à pergunta apresentada pelo título: qual é o objetivo? O objetivo não é brincar de matemática, nem impressionar amigos e colegas com técnicas estatísticas avançadas. O objetivo é aprender coisas que informam as nossas vidas.

Mentiras, mentiras deslavadas e estatística

Mesmo nas melhores circunstâncias, a análise estatística raramente revela “a verdade”. Em geral construímos um caso circunstancial baseado em dados imperfeitos. Como resultado, há numerosas razões para que indivíduos intelectualmente honestos discordem acerca de resultados estatísticos ou suas implicações. No nível mais básico, podemos discordar sobre a questão que está sendo respondida. Fãs de esportes discutirão por toda a eternidade sobre quem é “o melhor jogador de beisebol de todos os tempos” porque não há definição objetiva de “melhor”. Estatísticas descritivas rebuscadas podem fornecer informações sobre essa questão, mas jamais a responderão em definitivo. Como será ressaltado no próximo capítulo, questões mais significativas socialmente viram presas do mesmo desafio básico. O que está acontecendo com a saúde econômica da classe média americana? A resposta depende de como se define “classe média” e “saúde financeira”.

Há limites sobre os dados que podemos reunir e os tipos de experimentos

que podemos realizar. O estudo de Alan Krueger sobre terroristas não acompanhou milhares de jovens ao longo de múltiplas décadas para observar quais deles se tornaram terroristas. Simplesmente isso não é possível. Tampouco podemos criar duas nações idênticas – exceto que uma é altamente repressora e a outra não é – para então comparar o número de homens-bomba que surgem em cada uma. Mesmo quando podemos conduzir experimentos amplos, controlados, com seres humanos, eles não são fáceis nem baratos. Pesquisadores fizeram um estudo em larga escala sobre se preces reduzem ou não complicações pós-cirúrgicas, que foi uma das perguntas levantadas neste capítulo. *Esse estudo custou US\$2,4 milhões.* (Quanto aos resultados, você terá de esperar até o Capítulo 13).

O ex-secretário de Defesa dos Estados Unidos Donald Rumsfeld fez uma declaração que ficou famosa: “Você vai à guerra com o exército que tem – não o exército que gostaria ou desejaria ter posteriormente.” Qualquer que seja sua opinião sobre Rumsfeld (e a guerra no Iraque que ele estava explicando), esse aforismo aplica-se também à pesquisa. Nós conduzimos análises estatísticas usando os melhores dados, metodologias e recursos à nossa disposição. A abordagem não é como uma soma ou uma longa divisão, na qual a técnica correta produz a resposta “certa” e um computador é sempre mais preciso e menos falível que um humano. A análise estatística é mais como um bom trabalho de detetive (daí o potencial comercial de *CSI: análise de regressão*). Gente inteligente e honesta com frequência discorda sobre o que os dados estão tentando nos dizer.

Mas quem foi que disse que todo mundo que usa estatística é inteligente e honesto? Conforme mencionado, este livro começou como uma homenagem a *Como mentir com a estatística*, publicado pela primeira vez em 1954 e que vendeu mais de 1 milhão de exemplares. A realidade é que você *pode* mentir com estatística. Ou pode cometer erros inadvertidos. Em qualquer um dos casos, a precisão matemática vinculada à análise estatística pode encobrir algum sério absurdo. Este livro percorrerá muitos dos erros e más interpretações estatísticas mais comuns (de modo que você possa reconhecê-los, não para usá-los).

Então, voltando ao título do capítulo, qual é o objetivo de aprender estatística?

- Sintetizar enormes quantidades de dados.

- Tomar decisões melhores.

- Responder a questões sociais importantes.

- Reconhecer padrões capazes de refinar o modo como fazemos tudo, desde vender fraldas até capturar criminosos.

- Pegar trapaceiros e processar criminosos.

- Avaliar a efetividade de políticas, programas, drogas, procedimentos médicos e outras inovações.

E identificar os canalhas que usam essas mesmas ferramentas poderosas para fins nefastos.

Se você puder fazer tudo isso vestindo um terno Hugo Boss ou uma saia preta curtinha, talvez você possa também ser a próxima estrela ou astro de *CSI: análise de regressão*.

^a Neste parágrafo, são usados alguns termos do futebol americano: *quarterback* é o armador e principal jogador do time, responsável em campo pela escolha e distribuição das jogadas; *touchdown* é o “gol” no futebol americano, que vale seis pontos.

Quanto ao termo “índice de passes”, o original *passer rating* significa literalmente “índice do passador”. Há em inglês dois outros termos para o mesmo conceito, *quarterback rating* (“índice do *quarterback*”) e *pass efficiency* (“eficiência de passe”). O nome mais fiel ao sentido, embora não tão usado, seria “eficiência do *quarterback*”, pois o índice é calculado levando-se em conta não apenas os passes, mas o desempenho como um todo: jardas aéreas, jardas corridas, *touchdowns*, passes tentados, passes completos etc. (N.T.)

^b O coeficiente de Gini às vezes é multiplicado por cem, para obter um número inteiro. Nesse caso, os Estados Unidos teriam um coeficiente de Gini de 45.

^c O título da matéria em inglês é “Data is merely the raw material of knowledge”. Como o inglês usa a palavra latina “data”, o autor comenta nesta nota que “data” [“dados”] tem sido historicamente usada no plural. Por exemplo, “*The data are very encouraging*” (“Os dados são muito animadores”). O singular em latim é “datum”, que se referiria a um dado único, tal como a resposta de uma pessoa a uma única pergunta numa pesquisa. O uso do plural “data” é um meio rápido de mostrar a qualquer pessoa que faça uma pesquisa séria que você tem familiaridade com estatística. Dito isto, muitas autoridades em gramática inglesa e muitas publicações, como o *New York Times*, aceitam atualmente que “data” pode ser singular ou plural, como na passagem citada. Em português não temos esse problema, pois geralmente usamos o termo “dados”, já no plural. No entanto, serve para entendermos expressões utilizadas sem tradução, muitas vezes em nomes próprios, como database etc. (N.T.)

^d A explicação do autor é importante em sistemas educacionais cuja avaliação é feita por critérios inicialmente não numéricos, como ocorre nos Estados Unidos, com notas A, B e C, por exemplo. Para um sistema educacional em que a avaliação já é numérica, a média numérica é um conceito mais óbvio, como ocorre no Brasil. (N.T.)

^e Essa é uma simplificação grosseira do fascinante e complexo campo da ética médica.