IAN STEWART

17 EQUAÇÕES QUE MUDARAM O MUNDO

autor de
Mania de matemática
e *Almanaque das curiosidades matemáticas*

ZAHAR

Ian Stewart

Dezessete equações que mudaram o mundo

Tradução:
George Schlesinger

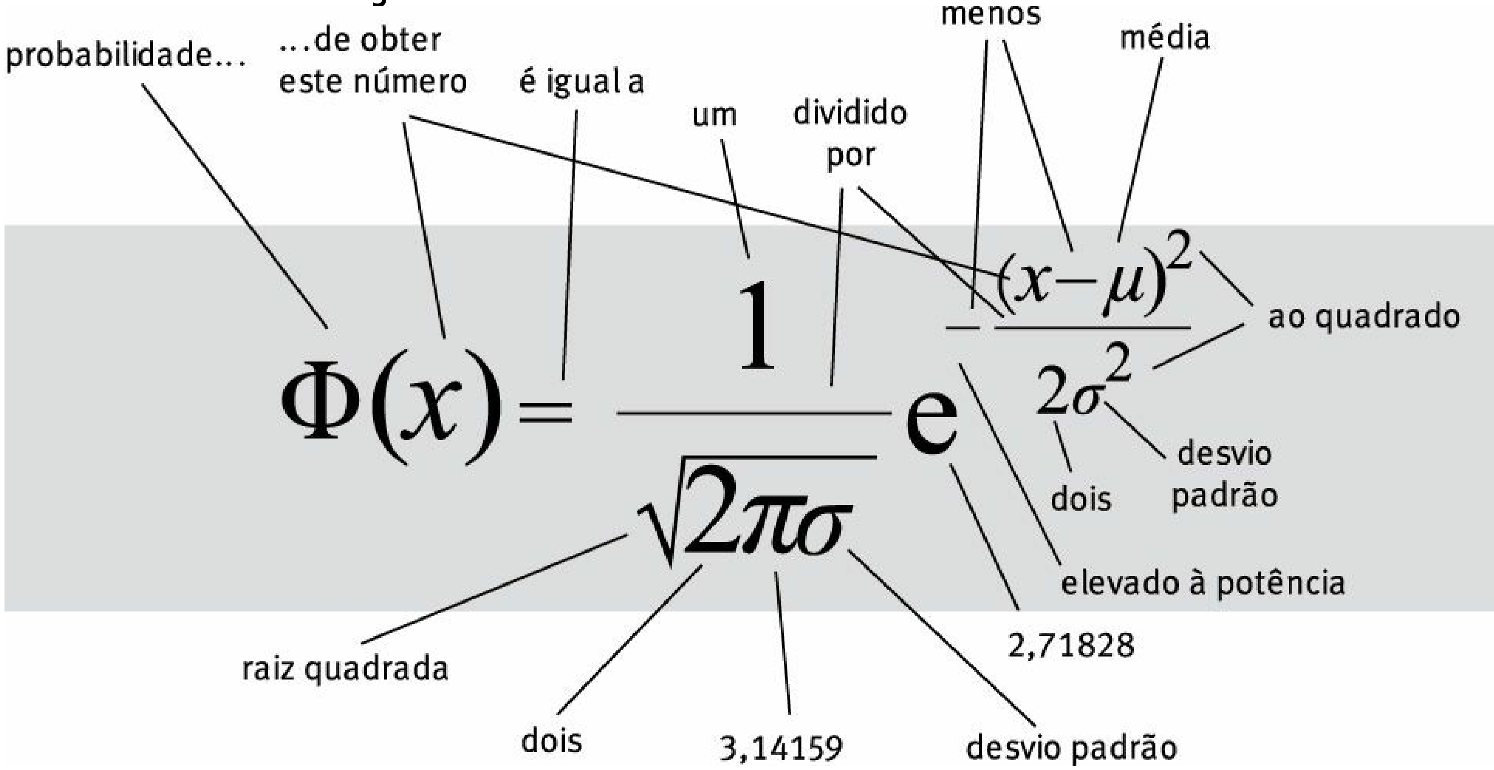
Revisão técnica:
Diego Vaz Bevilaqua

ZAHAR

Sumário

7. Padrões de probabilidade

Distribuição normal



$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

O que diz?

A probabilidade de observar um valor específico entre dados é maior perto do valor médio – a média – e decresce rapidamente à medida que aumenta a diferença em relação à média. A rapidez desse decréscimo depende de uma grandeza chamada desvio padrão.

Por que é importante?

Define uma família especial de distribuição de probabilidade em forma de sino, que são geralmente bons modelos das observações comuns do mundo real.

Qual foi a consequência?

O conceito de "homem médio", testes de significância de resultados experimentais, tais como testes médicos, e uma infeliz tendência de assumir a curva do sino como se nada mais existisse.

A MATEMÁTICA LIDA com padrões. O funcionamento aleatório da probabilidade parece estar tão distante de padrões quanto se possa imaginar. Na verdade, uma das definições em voga de "aleatório" se refere a "falta de padrões discerníveis". Matemáticos vinham investigando padrões em geometria, álgebra e análise por séculos antes de perceber que mesmo a aleatoriedade tem seus próprios padrões. Mas os padrões de probabilidade não entram em conflito com a ideia de que eventos aleatórios não possuem padrão porque as regularidades dos eventos aleatórios são estatísticas. São características de toda uma série de eventos, tais como o comportamento médio em uma longa série de tentativas. Elas nada nos dizem sobre qual evento ocorre em qual instante. Por exemplo, se você jogar um dado repetidamente, então um sexto das vezes cairá 1, e o mesmo vale para 2, 3, 4, 5 e 6 – um claro padrão estatístico. Mas isso não nos diz nada sobre o número que deverá cair na próxima jogada.

Foi apenas no século XIX que os matemáticos e cientistas se deram conta da importância dos padrões estatísticos em eventos sujeitos a probabilidade. Mesmo atos humanos, tais como suicídio e divórcio, estão sujeitos a leis quantitativas, em médio e em longo prazo. Levou algum tempo para que se acostumasse àquilo que de início parecia contradizer o livre-arbítrio. Mas hoje essas regularidades estatísticas formam a base para testes médicos, política social, prêmios de seguros, avaliações de risco e esportes profissionais.

E jogo, que foi onde tudo começou.

SIGNIFICATIVAMENTE, tudo começou com um erudito jogador, Girolamo Cardano. Tendo o hábito de desperdício, Cardano conseguia seu necessitado dinheiro fazendo apostas em partidas de xadrez e jogos de azar. E aplicava seu poderoso intelecto a ambos os casos. O xadrez não depende de sorte: vencer depende de uma boa memória para posições e jogadas padronizadas, e um senso intuitivo de como corre a partida de forma geral. Em jogos de azar, no entanto, o jogador está sujeito aos caprichos da Dona Sorte. Cardano percebeu que podia aplicar seus talentos matemáticos de forma efetiva mesmo nessa relação tempestuosa. Ele podia melhorar seu desempenho em jogos de azar possuindo uma compreensão melhor das chances – a probabilidade de ganhar ou perder – do que tinham seus oponentes. E organizou um livro sobre o assunto, *Liber de Ludo Aleae* ("O livro dos jogos de azar"). A obra permaneceu inédita até 1633. Seu conteúdo erudito é o primeiro tratamento sistemático da matemática da probabilidade. Seu conteúdo menos respeitável é um capítulo sobre como trapacear e se dar bem.

Um dos princípios fundamentais de Cardano era que numa aposta justa, as chances seriam proporcionais ao número de maneiras conforme as quais cada jogador pode ganhar. Por exemplo, suponhamos que os competidores joguem um dado, e o primeiro jogador ganha se tirar um 6, ao passo que o segundo ganha se tirar qualquer outro número. O jogo seria muitíssimo injusto se cada um apostasse a mesma quantia, porque o primeiro jogador tem apenas uma maneira de ganhar, ao passo que o segundo tem cinco. Porém, se o primeiro jogador apostar uma libra e o segundo apostar cinco libras, as chances tornam-se equivalentes. Cardano estava ciente de que este método de calcular chances justas depende de que as várias formas de ganhar sejam igualmente prováveis, mas em jogos de dados, cartas ou moedas era claro como assegurar que tais condições se aplicassem. Lançar uma moeda tem dois resultados, cara ou coroa, e ambos são igualmente prováveis numa moeda honesta. Se a moeda tende a dar mais caras do que coroas, ela é, com certeza, viciada – desonesta. Da mesma forma, os seis resultados de um dado são igualmente prováveis, assim como os 52 resultados de uma carta tirada de um baralho.

A lógica por trás do conceito de objeto “viciado” é ligeiramente circular, porque inferimos um viés, uma tendência, a partir do fracasso no ajuste às condições numéricas óbvias. Mas essas condições são respaldadas por mais do que simples contagem. Elas se baseiam numa sensação de simetria. Se a moeda é um disco chato de metal, de densidade uniforme, então os dois resultados estão relacionados com uma simetria da moeda. Para os dados, os seis resultados estão relacionados por simetrias do cubo. E para cartas, a simetria relevante é que uma não difere significativamente da outra, exceto pelo valor impresso na face. As frequências $\frac{1}{2}$, $\frac{1}{6}$ e $\frac{1}{52}$ para qualquer resultado dado se assentam sobre essas simetrias básicas. Pode-se criar uma moeda ou um dado viciado pela inserção de pesos; uma carta viciada é criada usando-se marcas sutis no verso, que revelam seu valor àqueles que conhecem o segredo.

Há outros modos de trapacear, envolvendo prestidigitação – digamos, introduzir um dado viciado no jogo e retirá-lo antes que alguém perceba que ele sempre dá 6. Contudo, o meio mais seguro de “trapacear” – ganhar mediante um subterfúgio – é ser perfeitamente honesto, mas conhecer as probabilidades melhor do que seu oponente. Num certo sentido, você está seguindo os princípios morais, e pode melhorar ainda mais suas chances de encontrar um adversário adequadamente ingênuo recorrendo não às probabilidades, mas às expectativas que ele tem das probabilidades. Há diversos exemplos em que as chances reais de um jogo de azar são muito diferentes do que a maior parte das pessoas imaginaria.

Um exemplo é o jogo da coroa e âncora, muito jogado por marinheiros britânicos no século XVIII. O jogo utiliza três dados, cada um trazendo não os números de 1 a 6 e sim seis símbolos: uma coroa, uma âncora e os quatro naipes, ouros, espadas, paus e copas. Esses símbolos também estão marcados numa esteira. Os jogadores apostam pondo dinheiro sobre a esteira e jogando os três dados. Se der algum dos símbolos em que ele apostou, a banca paga a aposta multiplicada pelo número de dados que deram esse símbolo. Por exemplo, se o jogador aposta 1 libra na coroa e os dados dão duas coroas, ele ganha duas libras além da sua aposta; se forem três coroas, ganha três libras além da aposta. Tudo parece muito razoável, mas a teoria da probabilidade nos diz que a longo prazo o jogador pode esperar perder 8% do dinheiro apostado.

A TEORIA DA PROBABILIDADE começou a decolar quando atraiu a atenção de Blaise Pascal. Pascal era filho de um coletor de impostos de Rouen e menino-prodígio. Em 1646 converteu-se ao jansenismo, uma seita do catolicismo romano que o papa Inocêncio X declarou herética em 1655. No ano anterior, Pascal havia vivenciado o que chamou de sua “segunda conversão”, provavelmente ocasionada por um acidente quase fatal quando seus cavalos despencaram da ponte de Neuilly e a carruagem quase fez o mesmo. A maior parte da sua produção, daí em diante, foi de filosofia religiosa. Mas pouco antes do acidente, ele e Fermat estavam se correspondendo acerca de um problema matemático que tinha a ver com jogos de azar. Chevalier de Meré, um escritor francês que se autodenominava cavaleiro mesmo não sendo, era amigo de Pascal, e quis saber como as apostas numa série de jogos de azar deveriam ser divididas se a competição tivesse de ser abandonada no meio do caminho. Não era uma questão nova: ela remonta à Idade Média. Nova foi a solução. Numa troca de cartas, Pascal e Fermat acharam a resposta correta. No decorrer do processo, criaram um novo ramo da matemática: a teoria da probabilidade.

Um conceito central na sua solução foi o que atualmente chamamos de “expectativa”. Num jogo de azar, há um retorno médio do jogador a longo prazo. Seria, por exemplo, 92 centavos de libra para uma aposta de uma libra na coroa e na âncora. Após sua segunda conversão, Pascal deixou para trás seu passado de jogatina, mas registrou sua contribuição em um famoso argumento filosófico, a aposta de Pascal.[1] Pascal assumia, fazendo o papel de advogado do diabo, que alguém pudesse considerar a existência de Deus altamente improvável. Em seus *Pensées*, de 1669, Pascal analisou as consequências do ponto de vista das probabilidades:

> Vamos pesar o ganho e a perda na aposta de que Deus é [existe]. Vamos estimar essas duas probabilidades. Se você ganhar, você ganha tudo; se perder, não perde nada. Aposte então, sem hesitar, que Ele é … Há aqui uma infinidade de uma vida infinitamente feliz a ganhar, uma chance de ganho contra um número finito de riscos de perda, e o que você aposta é finito. Assim, nossa proposição é de força infinita, quando existe a aposta finita num jogo em que há riscos iguais de ganho e perda, e o infinito a ganhar.

A teoria da probabilidade chegou como área plenamente reconhecida da matemática em 1713, quando Jacob Bernoulli publicou seu *Ars Conjectandi* (“A arte de conjecturar”). Ele começa com a definição prática usual de probabilidade de um evento: a proporção de vezes em que ele ocorre, a longo prazo, quase o tempo todo. Eu digo “definição prática” porque esta abordagem de probabilidades vai gerar problemas se você quiser fazer dela um princípio exato. Por exemplo, suponha que tenho uma moeda honesta e a lanço seguidas vezes. Na maior parte do tempo eu obtenho uma sequência aparentemente aleatória de caras e coroas, e se continuar lançando-a por tempo suficiente, obterei cara em cerca de metade das vezes. No entanto, é raro eu obter cara exatamente metade das vezes: isso é impossível, por exemplo, se eu lançar a moeda um número ímpar de vezes. Se eu tentar modificar a definição inspirando-me no cálculo, de modo que a probabilidade de obter cara seja o limite da proporção de caras quando o número de lançamentos tende ao infinito, preciso provar que esse limite existe. Mas às vezes ele não existe. Por exemplo, suponha que a sequência de caras e coroas seja, com C para cara e K para coroa,

KCCKKKCCCCCCKKKKKKKKKKKK…

com uma coroa, duas caras, três coroas, seis caras, doze coroas, e assim por diante – os números dobrando em cada fase após as três coroas. Após três lançamentos a proporção de caras é $\frac{2}{3}$, após seis lançamentos é de $\frac{1}{3}$, após doze lançamentos volta a ser de $\frac{2}{3}$, após 24 lançamentos é de $\frac{1}{3}$…, de modo que a proporção oscila para a frente e para trás, entre $\frac{2}{3}$ e $\frac{1}{3}$, e portanto não tem limite bem definido.

Concordo que tal sequência de lançamentos é muito improvável, mas para definir “improvável” temos de definir probabilidade, que é o que se esperava que o limite possibilitasse. Assim, trata-se de uma lógica circular. Ademais, mesmo que o limite exista, poderia não ser o valor “correto” de ½. Um caso extremo ocorre quando a moeda sempre dá cara. Agora o limite é 1. Mais uma vez, isso é altamente improvável, mas…

Bernoulli resolveu abordar toda a questão da direção oposta. Começar simplesmente *definindo* a probabilidade de cara ou coroa como sendo um número p entre 0 e 1. Digamos que a moeda é honesta se $p = ½$ e viciada se não for ½. Agora Bernoulli prova um teorema básico, a lei dos grandes números. Vamos introduzir uma regra razoável para atribuir probabilidades a uma série de eventos repetidos. A lei dos grandes números afirma que a longo prazo, com exceção de uma fração de tentativas que se torna arbitrariamente pequena, a proporção de caras tem, sim, um limite, e o limite é p. Filosoficamente, este teorema mostra que atribuindo probabilidades – ou seja, números – de uma forma natural, a interpretação "proporção de ocorrências no longo prazo, ignorando raras exceções" é válida. Assim, Bernoulli assume o ponto de vista de que os números atribuídos como probabilidades fornecem um modelo matemático consistente do lançamento de uma moeda vezes e vezes repetidas.

Sua demonstração depende de um padrão numérico que era muito familiar a Pascal. É geralmente chamado de triângulo de Pascal, ainda que ele não tenha sido o primeiro a notá-lo. Historiadores traçaram suas origens até o *Chandas Shastra*, um texto em sânscrito atribuído a Pingala, escrito em algum momento entre 500 e 200 a.C. O original não sobreviveu, mas o trabalho é conhecido por meio de comentários hindus do século X. O triângulo de Pascal tem o seguinte aspecto:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1

onde todas as linhas começam e terminam com 1, e cada número é a soma dos dois imediatamente acima. Atualmente chamamos esses números de coeficientes binomiais, porque surgem na álgebra da expressão binomial $(p + q)^n$. Ou seja,

$(p + q)^0 = 1$

$(p + q)^1 = p + q$

$(p + q)^2 = p^2 + 2pq + q^2$

$(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$

$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

e o triângulo de Pascal é visível como os coeficientes dos termos separados.

A percepção-chave de Bernoulli é que ao lançarmos uma moeda n vezes, com uma probabilidade p de obter cara, então a probabilidade de um número específico de lançamentos dar cara é o termo correspondente de $(p + q)^n$, em que $q = 1 - p$. Por exemplo, suponha que eu lance a moeda três vezes. Então os oito resultados possíveis são:

CCC
CCK CKC KCC
CKK KCK KKC
KKK

em que agrupei as sequências de acordo com o número de caras. Então, em oito sequências possíveis, temos:

1 sequência com 3 caras
3 sequências com 2 caras
3 sequências com 1 cara

1 sequência com 0 cara

A ligação com os coeficientes binomiais não é coincidência. Se você expandir a fórmula algébrica $(C + K)^3$ mas não agrupar os termos, obterá:

CCC + CCK + CKC + KCC + CKK + KCK + KKC + KKK

Agrupando os termos conforme o número de Cs, obtemos:

$C^3 + 3C^2K + 3CK^2 + K^3$

Depois disso, é uma questão de substituir cada C ou K por sua probabilidade, p ou q, respectivamente.

Mesmo neste caso, cada extremo CCC e KKK ocorre somente uma vez em oito tentativas, e números mais equitativos ocorrem nas outras seis. Um cálculo mais sofisticado usando propriedades padrões de coeficientes binomiais prova a lei de Bernoulli para os grandes números.

PROGRESSOS EM MATEMÁTICA muitas vezes acontecem por causa da ignorância. Quando os matemáticos não sabem como calcular algo importante, descobrem um modo de desviar-se e abordá-lo indiretamente. Nesse caso o problema é calcular esses coeficientes binomiais. Há uma fórmula explícita, mas se, por exemplo, quisermos saber a probabilidade de obter exatamente 42 caras ao lançar cem vezes uma moeda, é preciso fazer duzentas multiplicações e então simplificar uma fração complicadíssima. (Existem atalhos; ainda assim, é uma grande trapalhada.) Meu computador me diz numa fração de segundo que a resposta é

28.258.808.871.162.574.166.368.460.400 $p^{42}q^{58}$

mas Bernoulli não podia contar com esse luxo. Ninguém o fez até 1960, e os sistemas algébricos de computadores não se tornaram acessíveis em larga escala até o final dos anos 1980.

Uma vez que este tipo de cálculo direto não era viável, os sucessores imediatos de Bernoulli tentaram achar boas aproximações. Por volta de 1730, Abraham De Moivre deduziu uma fórmula aproximada para as probabilidades envolvidas em lançamentos repetidos de uma moeda viciada. Isso levou à função de erro, ou distribuição normal, muitas vezes chamada de "curva do sino" por causa de seu formato. O que ele provou foi o seguinte: define-se *distribuição normal* $\Phi(x)$ com média μ e variância σ^2 pela fórmula

$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Assim, para um n grande a probabilidade de obter m caras em n lançamentos de uma moeda viciada é muito próxima de $\Phi(x)$, quando

$$x = {}^{m}\!/_{n} - p \qquad \mu = np \qquad \sigma = npq$$

Lembrando que "média" é a nossa habitual de valor médio, "variância" é a medida de quanto os dados estão espalhados – a largura da curva do sino. A raiz quadrada da variância, o próprio σ, chama-se desvio padrão. A Figura 32 (*esquerda*) mostra como o valor de $\Phi(x)$ varia com x. A curva se parece um pouco com um sino, daí seu nome informal. A curva do sino é um exemplo de distribuição de probabilidade; isso significa que a probabilidade de obter dados entre dois valores dados é igual à área abaixo da curva e entre as linhas verticais correspondentes a esses valores. A área total sob a curva é 1, graças àquele inesperado fator $\sqrt{2\pi}$.

A ideia pode ser captada mais facilmente por meio de um exemplo. A Figura 32 (*direita*) mostra um gráfico de probabilidades de se obter vários números de caras lançando uma moeda honesta quinze vezes (barras retangulares), junto com a curva do sino aproximada.

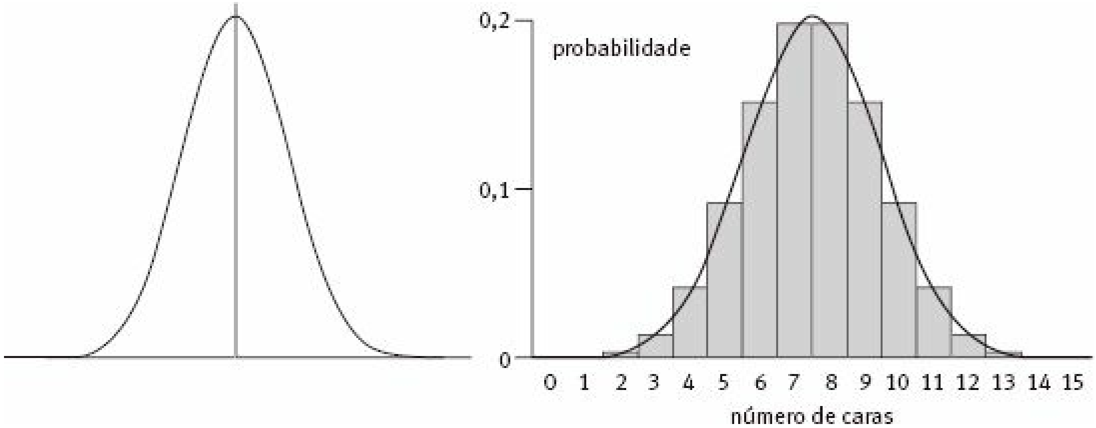


FIGURA 32 *Esquerda:* Curva do sino. *Direita:* Como ela aproxima o número de caras em quinze lançamentos de uma moeda honesta.

A curva do sino adquiriu status icônico quando começou a aparecer em dados empíricos nas ciências sociais, e não só na matemática teórica. Em 1835 Adolphe Quetelet, um belga que, entre outras coisas, foi pioneiro nos métodos quantitativos em sociologia, coletou e analisou grande quantidade de dados sobre criminalidade, taxa de divórcios, suicídios, nascimentos, mortes, altura humana, peso, e assim por diante – variáveis que ninguém esperava que se adequassem a qualquer lei matemática porque suas causas subjacentes eram complexas demais e envolviam escolhas humanas. Considere, por exemplo, o tormento emocional que leva alguém a cometer suicídio. Parecia ridículo pensar que isso podia ser reduzido a uma simples fórmula.

Essas objeções faziam sentido se você quisesse predizer exatamente quem iria se matar, e quando. Mas a partir do momento em que Quetelet concentrou-se em questões estatísticas, tais como a proporção de suicídios em vários grupos de pessoas, vários locais e diferentes anos, começou a enxergar padrões. Estes se mostraram controversos: se se predissesse que haveria seis suicídios em Paris no ano seguinte, como isso podia fazer algum sentido se cada uma das pessoas envolvidas possui livre-arbítrio? Elas podiam todas mudar de ideia. Mas a população formada por aqueles que de fato se matam não é especificada de antemão; ela vem junto como consequência das escolhas feitas não apenas por aqueles que cometem suicídio, mas também aqueles que pensaram na possibilidade e não fizeram. As pessoas exercem o livre-arbítrio no contexto de muitas outras coisas, que influenciam aquilo que elas decidem livremente: aqui os constrangimentos incluem problemas financeiros, problemas de relacionamento, estado mental, passado religioso… Em todo caso, a curva do sino não faz predições exatas; ela simplesmente afirma qual cenário é mais verdadeiro. Podem ocorrer cinco ou sete suicídios, deixando espaço de sobra para qualquer um que queira exercer o livre-arbítrio e mudar de ideia.

Os dados acabaram ganhando a parada: qualquer que fosse o motivo, pessoas em massa se comportavam mais previsivelmente do que indivíduos. Talvez o exemplo mais simples seja altura. Quando Quetelet registrou a proporção de pessoas com certa altura, obteve uma bela curva do sino (Figura 33). Ele obteve o mesmo formato de curva para muitas outras variáveis sociais.

FIGURA 33 Gráfico de Quetelet de quantas pessoas (eixo vertical) têm determinada altura (eixo horizontal).

Quetelet ficou tão impressionado com os resultados que escreveu um livro, *Sur l'homme et le développement de ses facultés* ("Sobre o homem e o desenvolvimento de suas faculdades"), publicado em 1835. Nele, Quetelet introduziu a noção de "homem médio", um indivíduo fictício, médio em todos seus aspectos. Há muito já se percebeu que isso não funciona inteiramente: o "homem" médio – ou seja, a pessoa, para o cálculo incluir também homens e mulheres – tem (um pouco menos de) um seio, um testículo, 2,3 filhos, e assim por diante. Não obstante, Quetelet via seu homem médio como a meta da justiça social, não apenas uma ficção matemática sugestiva. E não é tão absurdo quanto parece. Por exemplo, se a riqueza humana for dividida igualmente entre todos, então todos terão a média da riqueza. Não é uma meta prática, já que exige enormes mudanças sociais, mas alguém com fortes visões igualitárias poderia defendê-la como algo desejável.

A CURVA DO SINO RAPIDAMENTE se tornou um ícone na teoria da probabilidade, em especial seu braço aplicado, a estatística. Houve duas razões básicas: a curva do sino era relativamente simples de calcular, e havia um motivo teórico para que ela ocorresse na prática. Uma das principais fontes dessa maneira de pensar era a astronomia do século XVIII. Dados observacionais estão sujeitos a erros, causados por ínfimas variações no equipamento, erros humanos ou meramente movimento de correntes de ar na atmosfera. Os astrônomos do período queriam observar planetas, cometas e asteroides, bem como calcular suas órbitas. Isso exigia descobrir que órbita se encaixava melhor nos dados. O encaixe jamais seria perfeito.

A solução prática para este problema surgiu primeiro. Residia no seguinte: passe uma linha reta através dos dados e escolha esta linha de maneira que o erro total seja o mínimo possível. Os erros aqui devem ser considerados positivos, e a forma mais fácil de consegui-lo mantendo uma álgebra limpa é elevá-los ao quadrado. Assim, o erro total é a soma dos quadrados dos desvios das observações em relação à linha reta que serve de modelo, e a reta desejada é aquela que minimiza essa soma. Em 1805, o matemático francês Adrien-Marie Legendre descobriu uma fórmula simples para esta reta, tornando-a fácil de calcular. O resultado é chamado de método dos mínimos quadrados. A Figura 34 ilustra o método em dados artificiais relacionando estresse (medido por um questionário) e pressão sanguínea. A reta da figura, encontrada usando-se a fórmula de Legendre, é a que melhor se encaixa nos dados segundo as medições dos erros ao quadrado. Em dez anos o método dos mínimos quadrados passou a ser o padrão entre astrônomos na França, na Prússia e na Itália. Em mais vinte anos já era padrão na Inglaterra.

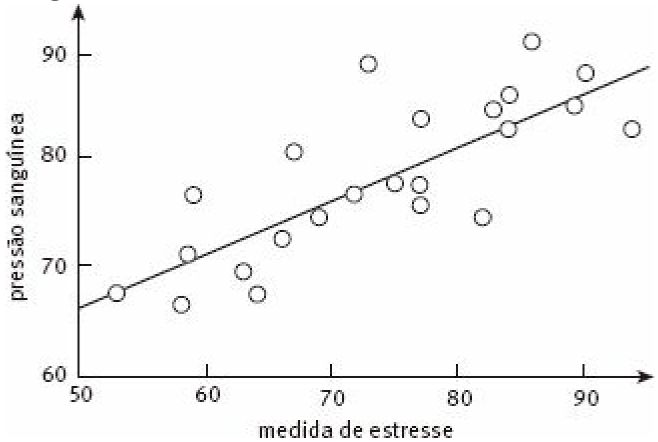


FIGURA 34 Usando o método dos mínimos quadrados para relacionar pressão sanguínea com estresse. Pequenos círculos: dados. Reta contínua: melhor reta de encaixe.

Gauss fez do método dos mínimos quadrados a pedra angular de seu trabalho em mecânica celeste. Ele entrou nessa área em 1801 predizendo com sucesso o retorno do asteroide Ceres depois de ficar oculto pelo brilho do Sol, quando a maioria dos astrônomos julgava que os dados disponíveis eram limitados demais. Esse triunfo selou sua reputação matemática entre o público e o estabeleceu como professor vitalício de astronomia na Universidade de Göttingen. Gauss não empregou os mínimos quadrados para essa predição: seus cálculos se concentraram em resolver uma equação algébrica de oitavo grau, o que ele fez por meio de um método numérico especialmente inventado. Mas em seu trabalho posterior, de 1809, culminando com seu *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum* ("Teoria do movimento dos corpos celestes em secções cônicas em torno do Sol") ele colocou grande ênfase no método dos mínimos quadrados. E declarou também ter desenvolvido a ideia, e a utilizado, dez anos antes de Legendre, o que causou um pouco de rebuliço. No entanto, isso era bastante provável, e a justificativa de Gauss para o método era bem diferente. Legendre o havia encarado como um exercício de adequação de uma curva, ao passo que Gauss o via como um meio de ajustar a distribuição de probabilidades. Sua justificativa da fórmula pressupunha que os dados subjacentes, aos quais a linha reta estava sendo ajustada, seguiam uma curva de sino.

Faltava justificar a justificativa. Por que erros observacionais haveriam de estar distribuídos normalmente? Em 1810 Laplace forneceu uma estarrecedora resposta, também motivada pela astronomia. Em muitos ramos da ciência é padrão fazer a mesma observação diversas vezes, e então calcular a média. Então é natural modelar esse procedimento matematicamente. Laplace usou a transformada de Fourier, ver Capítulo 9, para provar que a média de muitas observações é descrita por uma curva de sino, ainda que as observações individuais não sejam. Seu resultado, o teorema do limite central, foi um ponto de virada em probabilidade e estatística, pois forneceu uma justificativa teórica para o uso da distribuição favorita dos matemáticos, a curva do sino, na análise de erros observacionais.[2]

O TEOREMA DO LIMITE CENTRAL escolheu a curva do sino como a única distribuição probabilística adequada para a média de muitas observações repetidas. Ela adquiriu, portanto, o nome de "distribuição normal", e era vista como a escolha de referência para a distribuição de probabilidades. A distribuição normal não só tinha agradáveis propriedades matemáticas, como havia uma razão sólida para admitir que ela modelasse dados reais. Essa contribuição de atributos provou ser muito atraente para cientistas desejosos de ganhar compreensão dos fenômenos sociais que haviam interessado a Quetelet, porque oferecia um meio de analisar dados de registros oficiais. Em 1865, Francis Galton estudou como a altura de uma criança se relaciona com as alturas dos pais. Isso era parte de um objetivo maior: entender a hereditariedade – como traços humanos passam de pai para filho. Ironicamente, o teorema do limite central de Laplace levou Galton inicialmente a duvidar que essa herança existisse. E mesmo que existisse seria difícil prová-lo, porque o teorema do limite central era uma faca de dois gumes. Quetelet descobrira uma bela curva de sino para alturas, mas isso parecia influenciar muito pouco nos diferentes fatores que afetavam a altura, porque o teorema do limite central predizia de qualquer modo uma distribuição normal, quaisquer que fossem as distribuições desses fatores. Mesmo se os traços dos pais estivessem entre esses fatores, poderiam ser sobrepujados por todos os outros – tais como nutrição, saúde, posição social, e assim por diante.

Em 1889, porém, Galton já tinha descoberto uma forma de sair desse dilema. A prova do maravilhoso teorema de Laplace residia em adicionar à média os efeitos de muitos fatores distintos, mas estes precisavam satisfazer algumas condições rigorosas. Em 1875, Galton descreveu essas condições como "altamente artificiais", e notou que as influências a serem trazidas para a média

> precisam ser (1) todas independentes em seus efeitos, (2) todas iguais [tendo a mesma distribuição de probabilidades], (3) todas admitidas como simples alternativas "acima da média" ou "abaixo da média", e (4) … calculadas com a suposição de que as influências variáveis sejam infinitamente numerosas.

Nenhuma das condições se aplicava à hereditariedade humana. A condição (4) corresponde à premissa de Laplace de que o número de fatores somados *tende* ao infinito, então "infinitamente numeroso" é um pouco de exagero; porém, o que a matemática estabelecia era que para se obter uma boa aproximação de uma distribuição normal era preciso combinar um grande número de fatores. Cada um desses fatores contribuía em pequena dose para a média: com, digamos, cem fatores, cada um contribuía com um centésimo do seu valor. Galton referiu-se a tais fatores como "insignificantes". Cada um em si não tinha efeito significativo.

Havia uma saída potencial, e Galton se agarrou a ela. O teorema do limite central oferecia condição satisfatória para uma distribuição ser normal, não uma condição necessária. Mesmo quando suas premissas falhavam, a distribuição em questão ainda poderia ser normal *por outras razões*. A tarefa de Galton era descobrir quais podiam ser essas razões. Para ter alguma esperança de ligá-la à hereditariedade, teriam de recorrer a uma combinação de poucas influências grandes e disparatadas, e não a um número enorme de influências insignificantes. Ele lentamente tateou o caminho rumo a uma solução, e a encontrou mediante dois experimentos, ambos datados de 1877. Um deles era um dispositivo chamado quincunx, no qual bolas descem um plano inclinado batendo num arranjo de pinos, com igual possibilidade de ir para a direita ou para a esquerda. Em teoria, as bolas deveriam se amontoar na base segundo uma distribuição binomial, uma aproximação discreta pra uma distribuição normal, de modo que deveriam formar – e formaram – uma pilha com formato aproximado de um sino, como na Figura 32 (*direita*). Sua grande sacada foi imaginar parar temporariamente as bolas quando estivessem a meio caminho da descida. Ainda assim formariam uma curva de sino, porém mais estreita que a curva final. Imagine liberar apenas um compartimento de bolas. Elas cairiam até a base, espalhando-se numa minúscula curva de sino. E o mesmo valia para qualquer outro compartimento. E isso significava que a curva grande, final, podia ser vista como a soma de uma porção de curvas pequenas. A curva do sino reproduz a si mesma quando diversos fatores, cada um seguindo sua própria e separada curva do sino, são combinados.

O elo decisivo veio quando Galton criou ervilhas-de-cheiro. Em 1875, ele distribuiu sementes para sete amigos. Cada um recebeu setenta sementes, mas um recebeu sementes muito leves, outro, ligeiramente mais pesadas, e assim por diante. Em 1877, ele mediu os pesos das sementes da progênie resultante. Cada grupo tinha distribuição normal, mas o peso médio diferia em cada caso, sendo comparado ao peso de cada semente do grupo original. Ao combinar os dados para todos os grupos, os resultados tiveram novamente distribuição normal, mas a variância era maior – a curva do sino era mais larga. Mais uma vez, isso sugeria que combinar diversas curvas de sino levava a outra curva de sino. Galton foi buscar a razão matemática para isso. Suponha que duas variáveis matemáticas tenham distribuição normal, não necessariamente com as mesmas médias nem as mesmas variâncias. Então sua soma também terá uma distribuição normal; sua média é a soma das duas médias, e sua variância é a soma das duas variâncias. Obviamente o mesmo vale para somas de três, quatro ou mais variáveis aleatórias normalmente distribuídas.

Este teorema funciona quando um número pequeno de fatores é combinado, e cada fator pode ser multiplicado por uma constante, de modo que, na verdade, funciona para qualquer combinação linear. A distribuição normal é válida mesmo quando o efeito de cada fator é abrangente. Agora Galton podia ver esse resultado aplicado à hereditariedade. Supondo que a variável aleatória dada pela altura de uma criança seja alguma combinação das correspondentes variáveis aleatórias para as alturas dos pais, e que essas estejam distribuídas normalmente. Supondo que os fatores hereditários funcionem por adição, a altura da criança também terá uma distribuição normal.

Galton anotou suas ideias em 1889 sob o título *Herança natural*. Em particular, ele discutia a ideia que chamou de regressão. Quando um genitor alto e um baixo têm filhos, a altura média dos filhos deveria ser intermediária – na verdade, deveria ser a média das alturas dos pais. Da mesma forma, a variância deveria ser a média das variâncias, mas as variâncias dos pais pareciam ser aproximadamente iguais, então a variância não mudava muito. Com o passar de gerações sucessivas, a altura média "regrediria" a um valor no-meio-do-caminho, enquanto a variância se manteria praticamente inalterada. Então a bela curva do sino de Quetelet podia sobreviver de uma geração à seguinte. Seu pico rapidamente se assentaria num valor fixo, a média geral, enquanto sua largura seria a mesma. Assim, cada geração teria a mesma diversidade de alturas, apesar da regressão à média. A diversidade seria mantida pelos raros indivíduos que falhassem em regredir à média e seria automantenedora numa população suficientemente grande.

COM O PAPEL CENTRAL DA CURVA do sino firmemente cimentado naquilo que na época foram consideradas fundações sólidas, os estatísticos puderam sistematizar as percepções e os raciocínios de Galton, e profissionais de outras áreas puderam aplicar os resultados. A ciência social foi uma das primeiras beneficiárias, mas foi logo seguida pela biologia, e as ciências físicas já estavam avançadas no jogo graças a Legendre, Laplace e Gauss. Em pouco tempo toda uma caixa de ferramentas estatísticas estava disponível a qualquer um que quisesse extrair padrões a partir de dados. Focalizarei apenas uma técnica, por ser rotineiramente utilizada para determinar a eficácia de drogas e procedimentos médicos, além de muitas outras aplicações. É chamada de teste de hipótese, e sua meta é avaliar a importância de padrões aparentes em dados. Foi concebida por quatro pessoas: os ingleses Ronald Aylmer Fisher, Karl Pearson e seu filho Egon, juntamente com um polonês nascido na Rússia que passou a maior parte de sua vida nos Estados Unidos, Jerzy Neyman. Vou me concentrar em Fisher, que desenvolveu as ideias básicas ao trabalhar como estatístico em agricultura na Estação Experimental de Rothamstead, analisando novas cepas de plantas.

Suponha que você está criando uma nova variedade de batata. Seus dados sugerem que essa variedade é mais resistente a alguma praga. Mas todos esses dados estão sujeitos a algumas fontes de erro, de modo que você não pode ter total confiança de que os números sustentem essa conclusão – certamente não a confiança de um físico que pode fazer medições muito precisas e eliminar a maioria de seus erros. Fisher percebeu que a questão-chave é distinguir uma diferença genuína de outra que surge por mero acaso, e que a maneira de fazer isso é perguntar-se qual seria a probabilidade dessa diferença se apenas o acaso estivesse envolvido.

Admita, por exemplo, que a nova espécie de batata parece conferir o dobro de resistência, no sentido de que a proporção dela que sobrevive à praga é o dobro da proporção da velha espécie. É concebível que esse efeito ocorra por puro acaso, e pode-se calcular essa probabilidade. Na verdade, o que se calcula é a probabilidade de um resultado pelo menos tão extremo quanto o que é observado nos dados. Qual é a probabilidade de que a proporção da nova espécie sobrevivente à praga seja ao menos o dobro do que era a da velha espécie? Proporções ainda maiores são aqui permitidas porque a probabilidade de se obter *exatamente* o dobro da proporção está sujeita a ser muito pequena. Quanto maior a gama de resultados que são incluídos, mais prováveis se tornam os efeitos do acaso, então pode-se ter maior confiança na conclusão se o cálculo sugerir que não se trata de acaso. Se a probabilidade deduzida deste cálculo for baixa, digamos 0,05%, então é improvável que o resultado seja provocado pelo acaso; diz-se que ele é significativo num nível de 95%. Se a probabilidade for ainda menor, digamos 0,01%, então é extremamente improvável que o resultado seja provocado pelo acaso, e diz-se que ele é significativo num nível de 99%. A porcentagem indica que somente por acaso o resultado não seria tão extremo quanto o observado em 95% das tentativas, ou em 99% das tentativas.

Fisher descreveu seu método como uma comparação entre duas hipóteses distintas: a hipótese de que os dados são significativos no nível citado e a assim chamada hipótese nula, a de que os resultados se devem ao acaso. Ele insistia que seu método não devia ser interpretado como uma confirmação de que os dados são significativos; devia ser interpretado como uma rejeição da hipótese nula. Ou seja, ele fornece evidência contra os dados *não* serem significativos.

Isso pode parecer uma distinção muito sutil, uma vez que evidência contra os dados não serem significativos seguramente conta como evidência a favor de serem significativos. No entanto, isso não é totalmente verdade: o motivo é que a hipótese nula tem uma premissa extra embutida. Para calcular a probabilidade de que um resultado tão extremo seja devido ao acaso, precisa-se de um modelo teórico. O meio mais simples de consegui-lo é presumir uma distribuição de probabilidade específica. Esta premissa aplica-se apenas em relação à hipótese nula, porque é isso que você usa para fazer as somas. Você não está presumindo que os dados estejam normalmente distribuídos. Mas a distribuição de referência para a hipótese nula é normal: a curva do sino.

Este modelo embutido tem uma consequência importante, que a expressão "rejeita a hipótese nula" tende a ocultar. A hipótese nula é "os dados se devem ao acaso". Logo, é fácil demais ler essa afirmação como "rejeitar que os dados se devam ao acaso", o que, por sua vez, significa que você aceita que eles *não* são devidos ao acaso. Na realidade, porém, a hipótese nula é "os dados se devem ao acaso *e* os efeitos do acaso têm distribuição normal", logo, pode haver duas razões para rejeitar a hipótese nula: os dados não se devem ao acaso *ou* não estão com uma distribuição normal. A primeira sustenta a significância dos dados, mas a segunda, não. Ela só diz que você pode estar usando o modelo estatístico errado.

No trabalho de Fisher em agricultura, geralmente havia uma profusão de evidências para a distribuição normal dos dados. Assim, a distinção que estou fazendo não importava de fato. Em outras aplicações de testes de hipóteses, porém, ela pode ter importância. Dizer que o cálculo rejeita a hipótese nula tem o mérito de ser verdade, mas pelo fato de a premissa da distribuição normal não ser mencionada explicitamente, é muito fácil esquecer que é necessário verificar a normalidade da distribuição dos *dados* antes de concluir que os resultados são estatisticamente significativos. Como o método vai sendo usado por mais e mais pessoas, que foram treinadas em como fazer somas mas não nas suposições por trás dessas somas, há um crescente perigo de assumir de maneira errônea que o teste mostra que seus dados são significativos. Sobretudo quando a distribuição normal se tornou a premissa de referência automática.

NA CONSCIÊNCIA DO PÚBLICO, o termo "curva do sino" está indelevelmente associado ao controvertido livro de 1994 *A curva do sino*, de dois americanos, o psicólogo Richard J. Herrnstein e o cientista político Charles Murray. O tema central do livro é um suposto vínculo entre inteligência, medido pelo quociente de inteligência (QI), e variáveis sociais tais como renda, emprego, índices de gravidez e criminalidade. Os autores argumentam que os níveis de QI são melhores para prever tais variáveis do que o status econômico e social dos pais ou seu nível de educação. Os motivos para a controvérsia, e os argumentos envolvidos, são complexos. Um esboço rápido não pode de fato fazer justiça ao debate, mas os temas remetem imediatamente a Quetelet e merecem ser mencionados.

A controvérsia era inevitável, não importando os méritos ou deméritos acadêmicos que o livro pudesse ter, porque tocava um nervo sensível: a relação entre raça e inteligência. Reportagens de mídia tendem a salientar a proposta de que diferenças de QI possuem uma origem predominantemente genética, mas o livro foi mais cauteloso acerca desse vínculo, deixando em aberto a interação entre genes, meio ambiente e inteligência. Outro tópico controverso era uma análise sugerindo que a estratificação social nos Estados Unidos (e, de fato, em qualquer outra parte) aumentou significativamente ao longo do século XX, e que a principal causa eram as diferenças de inteligência. Outro, ainda, era uma série de recomendações quanto à política para se lidar com este alegado problema: uma era reduzir a imigração, que o livro mencionava estar reduzindo o QI médio. Talvez a sugestão mais controversa fosse que as políticas de bem-estar social supostamente incentivando mulheres pobres a ter filhos deveriam ser interrompidas.

Ironicamente, esta ideia remonta ao próprio Galton. Seu livro de 1869, *Gênio hereditário*, composto de escritos anteriores para desenvolver a ideia de que "as habilidades naturais do homem são derivadas por herança, sob exatamente as mesmas limitações que a forma e os traços físicos de todo o mundo orgânico. Em consequência … seria bastante prático produzir uma raça de homens altamente dotada por casamentos judiciosos durante várias gerações consecutivas". Ele afirmava que a fertilidade era maior entre os menos inteligentes, mas evitou qualquer sugestão de seleção deliberada em favor da inteligência. Em vez disso, expressava a esperança de que a sociedade pudesse mudar de modo que as pessoas mais inteligentes compreendessem a necessidade de ter muitos filhos.

Para muitos, a proposta de Herrnstein e Murray de uma reengenharia do sistema de bem-estar social estava desconfortavelmente próxima ao movimento da eugenia do início do século XX, no qual 60 mil americanos foram esterilizados, conforme alegado por causa de doenças mentais. A eugenia se tornou amplamente desacreditada quando ficou associada à Alemanha nazista e ao Holocausto, e muitas de suas práticas são agora consideradas violações de direitos humanos, em alguns casos culminando em crimes contra a humanidade. Propostas para se criar seres humanos seletivamente são vistas de modo geral como racistas. Um número de cientistas sociais endossou as conclusões científicas do livro, mas questionou a acusação de racismo; alguns deles tinham menos certeza quanto às políticas propostas.

A curva do sino deu início a um extenso debate sobre os métodos usados para compilar dados, os métodos matemáticos para analisá-los, a interpretação dos resultados e as sugestões de políticas baseadas nessas interpretações. Uma força-tarefa criada pela Associação Americana de Psicologia concluiu que alguns pontos apresentados no livro eram válidos: resultados de QI são bons para prever desempenho acadêmico, e isto está correlacionado com status de emprego, não havendo diferença significativa no desempenho de homens e mulheres. Por outro lado, o relatório da força-tarefa reafirmou que tanto genes como ambiente influenciam o QI e não encontrou evidência significativa de que diferenças raciais em resultados de QI sejam determinadas geneticamente.

Outros críticos argumentaram que existem falhas na metodologia científica, tais como ignorar dados inconvenientes, e que o estudo e algumas respostas a ele podem ter sido, até certo ponto, motivados por questões políticas. Por exemplo, é verdade que a estratificação social cresceu dramaticamente nos Estados Unidos, mas pode-se argumentar que a principal causa é a recusa dos ricos de pagar impostos, ao invés de diferenças na inteligência. Parece haver também uma inconsistência entre o problema alegado e a solução proposta. Se a pobreza leva as pessoas a ter mais filhos, e você acredita que isso é uma coisa ruim, por que raios você haveria de querer torná-las ainda mais pobres?

Uma parte importante do histórico frequentemente ignorada é a definição de QI. Em vez de ser algo mensurável de forma direta, como peso ou altura, o QI é inferido estatisticamente a partir de testes. Os indivíduos submetidos a eles respondem a questionários, e os resultados são analisados usando uma ramificação do método dos mínimos quadrados chamada análise de variância. Da mesma forma que os mínimos quadrados, esta técnica pressupõe que os dados estejam normalmente distribuídos, e busca isolar os fatores que determinam a maior quantidade de variabilidade nos dados – são, portanto, os mais importantes para moldar os dados. Em 1904, o psicólogo Charles Spearman aplicou esta técnica em diversos testes de inteligência diferentes. Ele observou que os resultados obtidos em testes diversos estavam altamente correlacionados; ou seja, se alguém se saísse bem em um teste, tendia a se sair bem em todos. Intuitivamente, eles pareciam estar medindo a mesma coisa. A análise de Spearman mostrou que um único fato comum – uma variável matemática que ele chamou de *g*, significando “inteligência geral” – explicava quase todas as correlações. O QI é uma versão padronizada do *g* de Spearman.

Uma pergunta-chave é se *g* é uma grandeza real ou uma ficção matemática. A resposta é complicada pelos métodos empregados para escolher testes de QI. Estes assumem que a distribuição “correta” da inteligência na população é normal – a paradigmática curva do sino – e calibram os testes manipulando matematicamente os resultados de modo a padronizar a média e o desvio padrão. Um perigo potencial aqui é que você obtém o que espera porque adota medidas para filtrar qualquer coisa que o contradiga. Stephen Jay Gould fez uma crítica extensiva de tais riscos já em 1981, no livro *A falsa medida do homem*, assinalando, entre outras coisas, que os resultados brutos de testes de QI frequentemente não obedecem em absoluto a uma distribuição normal.

O principal motivo para que se pense que *g* representa uma característica legítima da inteligência humana é que ele é fator *único*: matematicamente, ele define uma só dimensão. Se diversos testes parecem medir um mesmo aspecto, é tentador concluir que este mesmo aspecto deve ser real. Se não, por que os seriam tão próximos? Parte da resposta pode ser que os resultados de testes de QI se reduzem a uma contagem numérica, o que sintetiza uma gama multidimensional de questões e atitudes em potencial numa simples resposta unidimensional. Além disso, o teste foi desenvolvido de modo que a contagem se relacione fortemente com as respostas consideradas inteligentes por quem o criou – caso contrário, ninguém o levaria em conta.

Por analogia, imagine-se coletar dados em vários aspectos diferentes de “tamanho” no reino animal. Uma pessoa pode medir a massa, outra a altura, outros o comprimento, a largura, o diâmetro da pata traseira esquerda, o tamanho dos dentes, e assim por diante. Cada uma dessas medidas seria um número único. Estariam, de forma geral, intimamente correlacionados: animais altos tendem a pesar mais, ter dentes maiores, patas mais grossas… Se você correr os dados através de uma análise de variância, muito provavelmente encontrará que uma única combinação desses dados é responsável pela vasta maioria da variabilidade, assim como o *g* de Spearman faz com diferentes medições de elementos tidos como relacionados com inteligência. Será que isso necessariamente implicaria que todas essas características dos animais teriam a mesma causa subjacente? Que *uma coisa* controla todas? Possibilidade: o nível do hormônio de crescimento, talvez? Mas é provável que não. A riqueza da forma animal não se comprime de maneira confortável num único número. Muitas outras características não se correlacionam em absoluto com o tamanho: habilidade de voar, ter listras ou manchas, comer carne ou vegetais. A combinação especial única de medições responsável pela maior parte da variabilidade poderia ser uma consequência matemática dos métodos utilizados para achá-la – especialmente se essas variáveis foram escolhidas, como aqui, para começar por terem muita coisa em comum.

Voltando a Spearman, vemos que seu tão alardeado *g* pode ser unidimensional porque testes de QI são unidimensionais. O QI é um método estatístico para quantificar certos tipos específicos de capacidade de resolução de problemas, matematicamente prático mas que não necessariamente corresponde a um atributo real do cérebro humano, e não necessariamente representa o que quer que seja aquilo a que nós nos referimos como “inteligência”.

Focalizando um tópico, o QI, e usando-o para estabelecer uma política, *A curva do sino* ignora o contexto mais amplo. Mesmo que fosse sensato projetar geneticamente a população de um país, por que restringir o processo aos pobres? Mesmo que em média os pobres tenham QI mais baixo que os ricos, a qualquer hora uma criança pobre inteligente pode superar uma criança rica boba, a despeito das óbvias vantagens sociais e educacionais de que as crianças ricas desfrutam. Por que recorrer a cortes na política de bem-estar social quando se pode concentrar esforços no que se alega ser o verdadeiro problema: a inteligência em si? Por que não melhorar a educação? Na verdade, por que direcionar a política ao aumento de inteligência? Há muitos outros traços humanos desejáveis. Por que não reduzir a ignorância, a agressividade ou a cobiça?

É um erro pensar num modelo matemático como se fosse a realidade. Em ciências físicas, quando o modelo muitas vezes se encaixa muito bem na realidade, este pode ser um modo conveniente de pensar, um modo que traz pouco prejuízo. Mas em ciências sociais modelos geralmente são pouco melhores que caricaturas. A escolha do título *A curva do sino* aponta para essa tendência de misturar modelo com realidade. A ideia de que o QI seja alguma medida precisa de capacidade humana, apenas porque tem *pedigree* matemático, comete o mesmo erro. Não é sensato basear políticas sociais abrangentes e altamente duvidosas em modelos matemáticos falhos e simplistas. O verdadeiro ponto relevante em *A curva do sino*, um que o livro toca de forma extensiva mas inadvertida, é que esperteza, inteligência e sabedoria não são a mesma coisa.

A TEORIA DA PROBABILIDADE é muito usada em experimentos médicos de novas drogas e tratamentos, objetivando testar o significado estatístico de dados. Os testes são frequentemente, mas nem sempre, baseados na premissa de que a distribuição subjacente seja normal. Um exemplo típico é a detecção de aglomerados de câncer. Um aglomerado, para determinada doença, é um grupo dentro do qual a doença ocorre com mais frequência do que o esperado na população geral. O aglomerado pode ser geográfico, ou pode se referir mais metaforicamente a pessoas com um estilo de vida particular, ou um período específico de tempo. Por exemplo, lutadores profissionais aposentados, ou meninos nascidos entre 1960 e 1970.

Aglomerados aparentes podem ser devidos inteiramente ao acaso. É raro que números aleatórios se distribuam de forma aproximadamente uniforme; ao contrário, com frequência se aglomeram juntos. Em simulações da Loteria Nacional do Reino Unido, em que seis números entre 1 e 49 são retirados ao acaso, mais da metade parece mostrar algum tipo de padrão regular, tal como dois números consecutivos e três números separados pelo mesmo valor, por exemplo, 5, 9 e 13. Ao contrário da intuição comum, a aleatoriedade é agregada. Quando se encontra um aglomerado aparente, as autoridades médicas procuram avaliar se ele se deve ao acaso ou se poderia haver alguma possível conexão causal. Numa época, a maioria das crianças geradas por pilotos de caça israelenses eram meninos. Seria fácil pensar numa possível explicação – pilotos são muito viris e homens viris geram mais meninos (o que, aliás, não é verdade), pilotos são expostos a mais radiação que o normal, experimentam forças-*g* mais altas – mas tal fenômeno teve breve duração, apenas um aglomerado ao acaso. Em dados posteriores, desapareceu. Em qualquer população de pessoas, sempre é provável que haja mais crianças de um sexo ou de outro; a igualdade exata é muito improvável. Para avaliar a significação do aglomerado, dever-se-ia continuar observando e ver se ele persiste.

No entanto, essa procrastinação não pode prosseguir de maneira indefinida, especialmente se o aglomerado envolve uma doença séria. A aids foi detectada no início como um aglomerado de casos de pneumonia em homens homossexuais americanos na década de 1980, por exemplo. As fibras de amianto como causa de um tipo de câncer pulmonar, o mesotelioma, foi um fato que se revelou inicialmente entre ex-operários que trabalhavam com o produto. Assim, métodos estatísticos são usados para avaliar qual a probabilidade do surgimento de aglomerados se fosse por razões aleatórias. Os métodos de Fisher para testar a significância e métodos correlacionados são largamente usados com este propósito.

A teoria da probabilidade também é fundamental para a nossa compreensão do risco. Esta palavra tem um significado técnico específico. Refere-se ao potencial de que uma determinada ação provoque um resultado indesejável. Por exemplo, voar de avião pode resultar em um acidente aéreo, fumar cigarros pode causar câncer de pulmão, construir uma usina nuclear pode levar a vazamento de radiação num acidente ou ataque terrorista, construir uma represa para uma usina hidrelétrica pode provocar mortes se a represa se romper. Aqui "ação" pode se referir a não fazer alguma coisa: deixar de vacinar uma criança pode ocasionar sua morte por doença, por exemplo. Nesse caso, há também um risco associado a vacinar a criança, tal como uma reação alérgica. Na população como um todo o risco é menor, mas para grupos específicos pode ser maior.

Muitos conceitos diferentes de risco são empregados em contextos diversos. A definição matemática usual é que o risco associado a alguma ação ou inação é a probabilidade de um resultado adverso multiplicada pela perda que então seria provocada. Por esta definição, uma chance de um em dez de matar dez pessoas tem o mesmo nível de risco de uma chance de um em 1 milhão de matar 1 milhão. A definição matemática é racional no sentido de que há um raciocínio específico por trás, mas isso não significa que ele seja necessariamente sensato. Já vimos que "probabilidade" refere-se ao longo prazo, mas para raros eventos o longo prazo é de fato muito longo. Os humanos, e suas sociedades, podem se adaptar a repetidos números baixos de mortes, mas um país que subitamente perde um milhão de pessoas de uma só vez está em sérios problemas, porque todos os serviços públicos e a indústria ficariam ao mesmo tempo sob severa pressão. Não serviria de consolo dizer que nos próximos 10 milhões de anos o total de mortes, nos dois casos, seria comparável. Assim, novos métodos estão sendo desenvolvidos para quantificar o risco em tais casos.

Métodos estatísticos, derivados de perguntas sobre jogo, têm uma imensa variedade de usos. Fornecem ferramentas para analisar dados sociais, médicos e científicos. Como todas as ferramentas, o resultado depende de como são usadas. Qualquer um que use métodos estatísticos precisa estar ciente das premissas subjacentes aos métodos, e de suas implicações. Alimentar cegamente um computador com números e aceitar o resultado como um evangelho, sem entender as limitações do método usado, é uma receita para o desastre. O uso legítimo da estatística, porém, tem melhorado nosso mundo além de qualquer reconhecimento. E tudo começou com a curva do sino de Quetelet.