

A Estatística, Molière e Henry Adams*

William Kruskal

Chicago

I

Um dos perigos profissionais ao qual nós, os estatísticos, estamos expostos vislumbra-se no decorrer de um acontecimento social, por exemplo, um jantar elegante. Suponhamos que fiquei sentado junto de uma jovem encantadora com quem travei conhecimento nessa reunião, que ela se volta para mim, com um

* O presente artigo foi originariamente publicado em língua inglesa pela Revista *The Centennial Review*, Vol IX, N.º 1, pp. 79-96, de 1965. Trata-se de uma lição apresentada dentro do âmbito do «Programa de Professores-Visitantes de Estatística, em 1963-64», financiado pela Fundação Nacional para a Ciência dos Estados Unidos da América (National Science Foundation) e patrocinado pela COPSS. A preparação desta lição foi financiada em parte por um Subsídio para Fins de Investigação (Research Grant No. NSF-GP 609) da Divisão de Engenharia e de Ciências Matemáticas e Físicas da referida Fundação Nacional para a Ciência.

Nota do Tradutor: A COPSS (Committee of Presidents of Statistical Societies), da qual fazem parte os Presidentes e Secretários das principais associações profissionais de Estatística dos Estados Unidos — nomeadamente a Associação Americana de Estatística (American Statistical Association), o Instituto de Estatística Matemática (The Institute of Mathematical Statistics) e a Sociedade de Biometria, Região Oriental da América do Norte (Biometric Society, Eastern North American Region) — tem, entre outras missões, as de coordenar o trabalho destas sociedades de Estatística e a elaboração anual do «Programa de Professores-Visitantes de Estatística», patrocinado pela COPSS.

sorriso atraente e, à guiza de quem pretende iniciar uma conversa que destrua a barreira de gelo ainda existente entre nós, me pergunta: «Agora diga-me: qual é a sua profissão?»

Ora, é forçoso que lhe diga a verdade; de modo que lhe respondo que sou estatístico. Isso, no entanto, é normalmente a ruína do que poderia vir a ser uma bela conversa, pois em 86% dos casos o sorriso da jovem desaparece, enquanto ela se volta para o meu «rival», sentado do outro lado dela e eu, abandonado e ferido na minha dignidade incompreendida, não tenho outra solução do que atacar o frango assado que jaz à minha frente.

Outros há que, embora pertencendo ao mundo da Estatística, descobriram métodos mais efectivos, embora menos verdadeiros, de ladear este problema. Assim, há alguns colegas estatísticos que, quando se lhes pergunta o que fazem, respondem que são matemáticos; isto é apenas um começo de conversa um pouco melhor do que o anterior, mas tem o mérito de não ser uma mentira descarada. Outros ainda evadem-se cobardemente à questão e respondem: «Não me ocupo de nada com interesse. Fale-me antes de si». A primeira parte desta declaração é uma mentira manifesta e hoje vou tentar explicar alguma coisa do que é a Estatística, por que nós, os estatísticos, a consideramos uma ciência fascinante e a razão pela qual muitos mais estatísticos convenientemente treinados são de necessidade premente.

Aquele contratempo do jantar acima referido nunca sucede aos meus conhecidos, que são poetas, físicos, exploradores em África ou vendedores de carros usados. O que é que haverá acerca da palavra «Estatística» que, com tanta frequência, provoca um silêncio desmedido?

A algumas pessoas esta palavra faz lembrar uma espécie de feiticeiros-alquimistas, sentados em bancos enormes, com uma coloração esverdeada nos olhos, passando toda a vida a adicionar colunas intermináveis de números. Como há quem possa interessar-se por tal actividade?

Outros, imaginam o estatístico como um fanático de factos que se exprimem por números, um dos tais que, sem pestanejar, é capaz de dizer imediatamente a extensão do rio Amazonas, a população de Viena de Áustria em 1892, o número médio de crianças numa família americana, arredondado até ao centésimo, etc.

Ainda para outros, o estatístico é o que analisa as cotações da bolsa ou estuda balancetes ou contas de ganhos e perdas, no intuito de elaborar conclusões que provavelmente não serão tomadas em consideração por um em cada cinquenta capitalistas.

Claro que há pessoas de cada um destes tipos; são cidadãos bastante úteis cujo trabalho está, de facto, muitas vezes relacionado com problemas estatísticos. Mas o que é certo é que normalmente *não* são estatísticos no sentido em que vamos aqui usar o termo.

Dar definições exactas e concisas de uma certa disciplina é tarefa cada vez mais difícil: de uma maneira típica ou se trata de aforismos, que podem divertir, mas não dão informação alguma⁽¹⁾, ou são verdadeiros relatos

de generalidades vagas. Mas temos de tentar uma: a Estatística, pelo menos como a consideramos, é o estudo e a aplicação, baseados em informações, de métodos, com vista a alcançar conclusões acerca do mundo real, a partir de observações incertas. Esta definição é muito ampla e o seu conteúdo terá de ser limitado de várias maneiras, mas, para começar, será o suficiente. Nesta definição as palavras-chave são «métodos» e «incertas». Na Estatística ocupamo-nos de métodos, das suas características, de como os escolher no caso de poder haver alternativas, de conceitos de optimização. Na nossa profissão também estamos ligados à incerteza dos dados observados, e tal incerteza é saliente. Por esta razão, muitos dos estudos estatísticos modernos são feitos em termos de modelos probabilísticos para obter os dados, isto é, estudamos métodos de chegar a conclusões baseados numa variedade de suposições acerca da maneira como os dados podiam ter sido observados, no decorrer de um mecanismo casual.

A Estatística teórica é o estudo abstracto de tais métodos; a Estatística aplicada é a aplicação dos métodos, feita à luz da teoria.

Muitos se recordam certamente dum personagem famoso de uma das peças de MOLIÈRE, um tal Monsieur JOURDAIN. Pois bem, este Monsieur JOURDAIN acreditava plenamente na cultura, mas era tão ingénuo e ignorante que as suas tentativas para se educar a si próprio eram caricatas, ainda que tocadas por uma certa humanidade. Uma das partes mais interessantes da peça de MOLIÈRE é, na nossa opinião, aquela em que o nosso Monsieur JOURDAIN descobre, com satisfação incalculável, que durante toda a vida falou sempre em prosa.

Do mesmo modo, cada um de nós esteve, desde que nasceu ocupado durante toda a vida com a Estatística, na medida em que cada um de nós procurou alcançar conclusões, baseando-se em observações empíricas. Con-

(1) Exemplos: 1) A Estatística é a arte de declarar em termos precisos o que não se sabe; 2) Um estatístico é aquele que dum suposição injustificada chega a uma conclusão correcta através de um modelo matemático rigoroso.

tudo, há diferenças importantes entre as inferências estatísticas que fazemos durante a nossa vida e as que são feitas pelos profissionais da Estatística. São estas que vamos tentar descrever durante esta conversa. Talvez a diferença mais importante resida no uso, feito pelos estatísticos profissionais, de métodos explícitos, que foram analisados teoricamente e cujas propriedades são conhecidas.

A linguagem que falamos durante toda a nossa vida é trabalhada profissionalmente por novelistas, poetas e actores; é analisada profissionalmente por linguistas e gramáticos. Do mesmo modo, o pensamento estatístico que usamos durante a nossa vida é praticado profissionalmente pelos estatísticos práticos. Estes, consoante a sua actividade específica, podem chamar-se psicometristas, analistas da saúde pública, econometristas, etc., enquanto que a análise e o desenvolvimento da Estatística são aprofundados pelos estatísticos teóricos. Todas estas pessoas têm uma função importante, e, pelo menos no caso da Estatística, o seu número não é, nem de perto, suficiente.

II

A nossa conversa vai ser estruturada a partir de dois exemplos, o primeiro representando uma situação experimental genérica e o segundo acerca de um inquérito estatístico para fins oficiais. Em ambos os casos apontaremos algumas maneiras de mostrar que os dados são incertos e diremos algo acerca dos métodos de chegar a conclusões a partir destes dados incertos.

Suponhamos, no nosso primeiro exemplo, que depois de realizar uma experiência, comparámos dois tratamentos médicos e que os resultados se resumem no quadro seguinte:

Tratamento 1 Taxa de cura: 70%.
 Tratamento 2 Taxa de cura: 90%.

A conclusão imediata é que o tratamento 2 é bastante melhor do que o tratamento 1. Mas qual o grau de confiança que podemos atribuir a esta afirmação? Quais as espécies de incerteza que envolvem as taxas de cura e que espécies de inferências podem ser feitas com segurança? Talvez o tratamento 2, que parece ser muito melhor do que o tratamento 1, possua, de facto, a mesma taxa fundamental de cura; ou talvez não.

Para começar, é claro que as próprias taxas de cura são inadequadas para uma comparação. Além disso, é necessário saber, pelo menos, os números de doentes, ou seja, a chamada dimensão das amostras. Talvez tenham sido dez, em cada caso, os doentes aos quais foram ministrados os tratamentos; daqueles aos quais foi aplicado o tratamento 1, 7 curaram-se, enquanto, no caso do tratamento 2, o número de doentes curados foi de 9. Pode também suceder que o número de doentes aos quais foi aplicado cada um dos tratamentos tivesse sido de 100 ou de 1000. Ou, talvez o tratamento 1 tenha sido aplicado a 1000 doentes e o tratamento 2 a 100. No entanto, para sermos específicos, suponhamos que era de 20, para cada caso, o número de doentes aos quais foi aplicado cada um dos tratamentos. Deste modo, os resultados numéricos da experiência podem exprimir-se na tabela seguinte:

	Número dos curados	Não curados	Total
Tratamento 1	14	6	20
Tratamento 2	18	2	20

De passagem, notemos que o esquema apresentado neste exemplo é aplicável a muitos outros fenómenos diferentes deste das taxas de cura por tratamentos médicos. Podíamos, do mesmo modo, trabalhar com taxas de germinação relativas a duas qualidades de sementes, ou considerar o número de estudantes aprovados num exame,

com o intuito de comparar as taxas de aprovação relativas a rapazes e raparigas. No entanto, será conveniente pensar concretamente em termos da situação médica.

O estatístico classifica os erros e as fontes de variabilidade em duas espécies. Em primeiro lugar, há os erros sistemáticos ou de excentricidade; estes são factores que não dependem essencialmente da dimensão da amostra mas introduzem uma distorsão dos resultados conforme o número de doentes é de 20, 200 ou 2000 por tratamento. Em segundo lugar há a considerar os erros aleatórios ou casuais, cujos efeitos podem ser reduzidos, tomando amostras maiores, tornando assim médios os efeitos da variabilidade devida ao acaso.

Quais são os erros sistemáticos possíveis no caso do nosso exemplo médico genérico? Uma espécie importante de erros sistemáticos resulta das *definições* adoptadas. Qual será, por exemplo, o sentido a atribuir à palavra «curado»? Aqui o problema mais importante depende da opinião subjectiva da pessoa que examina o doente e o considera, ou não, como curado. Suponhamos, para concretizar, que o tratamento 1 é constituído pela aplicação de um medicamento convencional e clássico, enquanto o Tratamento 2 é constituído por um outro medicamento, recente, e proposto pelo cientista que executa a experiência. Suponhamos ainda que é o cientista o próprio a decidir se o doente está curado ou não. Uma primeira causa de excentricidade deriva da tendência muito humana de considerar um caso marginal como curado se fôr do grupo a que é aplicado o tratamento 2, e como não curado se fôr do do tratamento 1. Tal prática parece disparatada quando descrita em termos tão simples, mas o facto é que é muito corrente. Este problema reveste-se de dificuldades especiais em casos particulares, por exemplo na avaliação da terapia psicológica. Uma segunda causa de excentricidade ocorre quando o experimenta-

dor toma a consciência da possibilidade de não poder controlar o seu entusiasmo pelo seu próprio tratamento e actua em sentido contrário, ou seja, procede à classificação dos resultados, penalizando o seu próprio tratamento. Em qualquer dos casos os resultados que se obtêm apresentam-se distorsidos.

O único meio que conhecemos de evitar este problema das excentricidades devidas às classificações é o da *classificação casual*: trata-se de arranjar as coisas de modo que o classificador não tenha qualquer possibilidade de saber qual o tratamento que foi aplicado a cada um dos doentes. Muitas vezes isso é difícil, mas sempre absolutamente essencial para a validade científica da experiência.

Um erro sistemático relacionado é o da *selecção para tratamento*. Como é que os quarenta doentes foram distribuídos pelos dois tratamentos? Mais uma vez é claro o facto de o experimentador — ainda que inconscientemente — poder facilmente introduzir excentricidades nos resultados, seleccionando para o tratamento que subjectivamente favorece os doentes mais saudáveis, mais novos, mais optimistas, etc. Foram-nos referidos casos em que a selecção para o tratamento era feita por uma enfermeira ou secretária, admitindo-se que a sua escolha era feita de maneira imparcial, mas como ela, de facto, sabia qual o tratamento favorecido pelo seu chefe, seleccionava, nessas condições os doentes em melhor estado de saúde. Do mesmo modo, no outro caso da classificação, a selecção dos doentes para tratamento pode ser prejudicada em sentido inverso, no intuito de eliminar excentricidades no sentido subjectivamente favorecido criando excentricidades em direcção oposta.

O único método conhecido para eliminar este problema é o da *escolha casual*, ou seja, a selecção dos pacientes de modo que ela seja completamente independente do conhecimento do tratamento a aplicar. Por motivos que, por falta de tempo, não podemos

aprofundar aqui, a melhor maneira de conseguir uma selecção adequada é usar um mecanismo aleatório semelhante ao de lançar uma moeda ao ar ou de tirar cartas de um baralho.

Outra fonte de erros sistemáticos é a exclusão de doentes que pareçam constituir caso à parte. Este problema é particularmente insidioso porque quando os registos científicos não contêm nenhuma menção de um caso fora do normal, o leitor não tem qualquer possibilidade de compensar a omissão.

A literatura médica está cheia de novos tratamentos apontados inicialmente como descobertas maravilhosas; mais tarde, verificou-se que muitos deles *não* proporcionavam uma cura efectiva, estando as razões do fracasso relacionadas muitas vezes (na classificação inicial) com causas de excentricidade como as que foram apontadas atrás.

Há muitas outras fontes de erros sistemáticos. Poderia, por exemplo, apontar-se que os tratamentos diferem apenas na diferença médica considerada. Se os dois tratamentos são aplicados em dois hospitais diferentes podem existir outros aspectos do regime hospitalar — muito diferentes das divergências de tratamento — a levar em consideração por afectarem os resultados. Mas temos de resistir à tentação de procurar dar uma lista completa dos erros sistemáticos. Passemos, por isso, aos *erros aleatórios*.

Duma maneira geral, as causas de excentricidade não dependem do número de doentes; apresentam-se quer se trate de 20 ou 2000 doentes por tratamento, e não são eliminadas pelo facto de se tomarem os valores médios no caso da dimensão da amostra ser grande. Na realidade, os experimentadores são, muitas vezes, enganados por um sentimento de certeza errado quando analisam experiências muito grandes, devido exactamente ao número elevado de provas a efectuar. Os erros aleatórios, pelo contrário, dependem, em grande parte, do número de experiências efectuadas.

Os erros aleatórios surgem especialmente,

de duas maneiras embora a primeira seja provavelmente menos frequente, no caso concreto do nosso exemplo com os tratamentos médicos. Em primeiro lugar, temos os erros devidos às medições e aos cálculos efectuados... podem ser simplesmente lapsos ocorridos no momento de decidir se o paciente está curado ou não ou ser erros de carácter burocrático ou de cálculo, que surgem na transmissão da informação obtida através dos registos e nas computações. É fácil esquecer estes pormenores, mas os estatísticos sabem bem que é necessário tomá-los sempre em consideração. Ainda não há muito, foi tornado público por um grupo de médicos eminentes o facto de ser perigoso, para mulheres de certas idades, o uso muito generalizado de um medicamento (Enovid). Naturalmente, este anúncio causou quase um tumulto e foi com um sentimento misto de vexame e alívio que os médicos anunciaram pouco depois que a prevenção tinha sido feita com base num erro proveniente de um simples engano nos cálculos.

Em segundo lugar, vamos considerar os *erros de amostragem*. Suponhamos que estávamos suficientemente satisfeitos com o facto dos erros sistemáticos, de medição e de cálculo serem desprezáveis nos dois tratamentos. No entanto, se pudéssemos repetir a experiência, não obteríamos provavelmente os mesmos resultados. Em vez de 14 e 18 doentes curados, respectivamente, podíamos obter 13 e 19, ou 16 e 16, ou até mesmo 17 e 15, sendo o resultado neste último caso completamente oposto ao da diferença aparente inicial. Como é que podemos considerar o erro da amostra, ou, antes, para usar uma expressão mais exacta, a flutuação da amostra?

Um método importante da análise estatística baseia-se no cálculo das probabilidades de obter vários resultados, especialmente na hipótese de que não há de facto diferença alguma entre os tratamentos. Relembremos que os resultados observados foram :

	Curados	Não curados
Tratamento 1	14	6
Tratamento 2	18	2

Usando a técnica estatística corrente podemos calcular a probabilidade de observar este resultado se, de facto, os dois tratamentos não diferissem na taxa de cura. O valor numérico desta probabilidade é de cerca de 1 para 10, ou, mais exactamente, de 0,0958.

De modo semelhante podemos calcular as probabilidades dos outros resultados possíveis, ainda na hipótese de que os tratamentos tivessem a mesma taxa fundamental de cura. Podíamos ter observado uma tabela ligeiramente diferente, a seguinte, por exemplo :

	Curados	Não curados
Tratamento 1	13	7
Tratamento 2	19	1

As taxas de cura, neste caso, são de 65% e 95%, respectivamente. Na hipótese de as taxas fundamentais de cura serem iguais, esta tabela seria observada, pouco mais ou menos, uma vez em cada 50 casos; mais exactamente, a probabilidade é de 0,0202.

O método geral que estamos a descrever serve para basear as inferências estatísticas em cálculos de probabilidades duma variedade de tabelas possíveis observadas, baseadas numa variedade de hipóteses acerca das verdadeiras taxas fundamentais de cura. Estas taxas verdadeiras, embora desconhecidas, podem imaginar-se como sendo as que obteríamos se aplicássemos os tratamentos a um número extraordinariamente grande de doentes, por exemplo um milhão, sempre nas mesmas circunstâncias da experiência real.

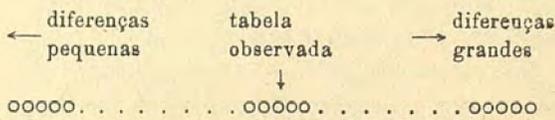
E tradicional e muitas vezes útil concentrar a atenção na hipótese acima mencionada de que os dois tratamentos têm exactamente as mesmas taxas de cura. Imaginemos agora que era possível alinhar lado a lado todas as tabelas que fosse possível observar, de modo que as tabelas do lado direito mostrassem as diferenças máximas observáveis entre as taxas de cura, enquanto as tabelas do lado esquerdo não mostrassem diferença alguma ou só diferenças muito pequenas entre as taxas de cura. Deste modo, na extremidade do lado esquerdo teríamos tabelas como :

0 20	8 12	17 3
0 20	8 12	17 3

etc. Na extremidade do lado direito teríamos os casos extremos

0 20	20 0
20 0	0 20

Imaginemos estas tabelas possíveis alinhadas e representemos cada uma delas por um pequeno círculo :



A tabela que observámos será representada por um destes círculos. Suponhamos que é o que está indicado por uma seta. (A análise completa deste tópico incluiria uma descrição precisa deste alinhamento dos resultados possíveis observáveis e esclareceria vários pontos que só podemos mencionar muito por alto nesta nossa exposição condensada, p. ex., a distinção entre um teste condicionado por todos os valores marginais

e um teste condicionado só por uma margem).

Um método importante da Estatística é o seguinte: calculemos a probabilidade de observar o que realmente observámos ou *casos mais extremos*: os casos cuja probabilidade total é calculada estão indicados no diagrama por um colchete inferior. Calculemos então esta probabilidade, na hipótese de que os dois tratamentos possuem de facto a mesma taxa fundamental de cura. No caso de esta probabilidade ser pequena, então de duas uma: 1) ou, na hipótese de as taxas de cura serem iguais, observámos um resultado muito surpreendente; ou, 2) a hipótese é falsa. Por outro lado, se a probabilidade calculada é de valor médio ou grande, então observámos um resultado que é compatível com a hipótese de não existir diferença. Deste modo, uma probabilidade calculada pequena faz pôr em dúvida a hipótese de não haver diferença, enquanto uma probabilidade moderada ou grande não lança dúvidas sobre a hipótese.

No nosso caso particular, encontramos para a probabilidade de observar a tabela que realmente observámos, ou uma outra correspondente a um caso mais extremo, na hipótese de não haver diferença (nível de significância da amostra com o «teste exacto» de duas abas de FISHER e simétrico) um valor que é aproximadamente de 24/100 ou, mais exactamente, de 0,2352. Quer isto dizer que em 24 de cada 100 repetições da experiência, encontraríamos uma diferença das taxas de cura igual à que encontramos de facto, ou uma diferença maior, embora as taxas fundamentais de cura sejam iguais.

Ora, um resultado que sob uma determinada hipótese pode acontecer 24 em cada 100 vezes, não pode considerar-se como desacreditando seriamente essa mesma hipótese, e deste modo concluímos que as nossas taxas de cura observadas, de 70% e 90%, podiam muito bem representar os resultados

de uma amostragem casual efectuada no grupo dos pacientes tratados por dois tratamentos que têm efeitos iguais. Isto não quer evidentemente dizer que os dois tratamentos sejam igualmente efectivos, mas apenas que a equivalência é completamente compatível com a amostra observada.

Suponhamos a seguir que as duas amostras já não são de dimensão 20, mas antes de dimensão 40, enquanto as taxas observadas de cura permanecem as mesmas. Os resultados observados seriam então:

	Curados	Não curados	Totais
Tratamento 1	28	12	40
Tratamento 2	36	4	40

Se procedermos a cálculos semelhantes aos anteriores achamos que se os tratamentos realmente não diferissem na taxa de cura então a probabilidade de observar um tal desvio, ou outro ainda maior, nas taxas observadas de cura varia dos anteriores 24 em cada 100 casos para apenas cerca de 5 em 100. Quer dizer: os resultados observados são bastantes surpreendentes e seriam frequentemente tomados como uma boa evidência de que os tratamentos são realmente diferentes.

Continuando deste modo, se as amostras fossem aumentadas para 100 cada, ou seja, 5 vezes a dimensão original, e se as taxas observadas de cura continuassem as mesmas, a probabilidade decisiva desceria a menos de 1 em cada 1000 casos e estaria praticamente fora de questão haver realmente uma diferença entre os dois tratamentos.

Repare-se como a interpretação das mesmas taxas observadas de cura muda rapidamente à medida que vamos passando de amostras de dimensão 20 para 40 e depois para 100. Se considerassemos amostras de dimensão ainda maior, a probabilidade decisória continuaria ainda a decrescer.

Uma área importante da Estatística é a

que ajuda a planear ou a delinear as experiências e, no contexto presente, a maior parte do delineamento *antes de executar a experiência*, seria decidir a dimensão da amostra. Devemos considerar 20, 40 ou 100 pacientes para cada tratamento? Os custos monetários podem desempenhar um papel importante em tal decisão; além disso o estatístico pode dar uma contribuição, calculando qual a probabilidade de que uma diferença real entre os tratamentos, de uma dada magnitude, se evidencie, dando origem a uma pequena probabilidade de surpresa, baseada nos resultados a serem observados durante a experiência.

Vimos com um exemplo simples como surgem diferentes espécies de erros e como funciona um método de alcançar conclusões. Mal fizemos uma introdução à análise completa do método e não discutimos sequer outros métodos da análise estatística que podiam ter sido usados.

III

O segundo exemplo que vamos discutir surge num domínio completamente diferente, o das estatísticas oficiais. Para introdução, vamos fazer uma citação da clássica autobiografia americana, *The Education of Henry Adams* (A Educação de Henry Adams). Depois de a lermos, faremos algumas considerações de carácter descritivo sugeridas pela transcrição e só depois nos vamos ocupar com o nosso segundo exemplo.

Henry Adams, a certa altura da sua interminável e frívola busca para uma verdade última, decidiu que a deveria descobrir estudando estatísticas, mas estatísticas no sentido plural de colunas de números. Falando de si próprio na terceira pessoa, Adams diz:

Supondo que a alternativa relativamente à arte era a aritmética, embrenhou-se profundamente nas estatísticas, imaginando que a edu-

cação teria aí a sua base mais segura; e o estudo provou ser o mais fácil que ele alguma vez havia tentado. Até o Governo se oferecia para produzir estatísticas em número ilimitado, colunas de números sem fim, médias insondáveis, apenas pelo prazer de perguntar. No Departamento de Estatística [em Washington: District of Columbia], Worthington Ford fornecia qualquer material que a curiosidade pudesse imaginar, preenchendo as vastas lacunas da ignorância, e métodos para aplicar a argamassa do facto. Dir-se-ia por um momento que se ganhava terreno e que as médias que se calculavam eram projectadas no futuro como se fossem leis. Talvez a parte mais embaraçosa do estudo resida na atitude dos estatísticos, que não mostravam nenhuma confiança entusiástica nos seus próprios números. Eles deviam ter alcançado a certeza, mas falavam como os outros homens que sabiam menos. O método não resultou em crença. Na verdade, todo o aumento de massa (de volume e velocidade), parecia trazer novos elementos e, por último, um escolar, presumido em aritmética e ignorante de álgebra, caiu num supersticioso terror de complexos com o desmoronar dos factos (*The Education of Henry Adams*, Cap. 23)

Esta passagem tem vários aspectos que bem merecem ser discutidos. Primeiro, consideremos a frase-chave «ignorante de álgebra». Adams constata que os números só por si fazem normalmente pouco sentido. Têm que ser trabalhados e condensados — ou numa palavra: analisados. Mas a análise é inevitável em termos de conceitos abstractos e os conceitos abstractos acerca dos números são normalmente expressos matematicamente. O pensamento estatístico requer um certo grau de talento matemático, bem como de treino, dependendo o montante apropriado da espécie de estatísticas que se desejam fazer. É necessário algo de Matemática para a Estatística; mas por outro lado, pode fazer-se um belo trabalho estatístico, tanto teórico como aplicado, sem que seja necessário um Doutoramento em Matemática.

Um segundo aspecto interessante da narra-

tiva de Henry Adams é a sua queixa de que os estatísticos não mostravam nenhuma «confiança entusiástica», e de que o seu método não «resultava em crença». Às vezes ouve-se dizer que os estatísticos, do ponto de vista profissional, são cínicos e destrutivos. Uma variante é a queixa de que a Estatística é subtil para a análise rotineira dos dados, mas que a construção de teorias científicas tem pouco a lucrar com a Estatística. Parece-me que estas queixas reflectem alguma coisa de verdadeiro, mas que são também caricaturas grosseiras. Se os estatísticos tendem a ser profissionalmente cínicos é porque muitas vezes vêem os resultados de optimismo indevido, de credulidade científica e de crença infantil em números escritos num pedaço de papel. Quanto à construção de teorias grandiosas acerca da Natureza, a Estatística propriamente interpretada desempenha um papel real e positivo nesta actividade.

Vamos considerar agora um exemplo específico de estatísticas oficiais. Há alguns anos, de colaboração com um colega — um economista — interessámo-nos pelo Índice dos Preços de Consumo do Departamento de Estatística do Trabalho (Bureau of Labor Statistics), um respeitável organismo do governo federal (U. S. A.). Este índice é um número oficial do governo, publicado periodicamente com o acompanhamento de publicidade da imprensa. Este número-índice é expresso com um impressionante nível de certeza aparente, e nós, os consumidores, aplaudimos ou lamentamos conforme ele decresce ou aumenta uns décimos de ponto. Os seus efeitos económicos são consideráveis, através dos contratos de trabalho que lhe estão ligados e através das decisões de negócios que nele se baseiam. Além disso, o índice e as suas componentes detalhadas — por cidade ou região, ou por espécie de bens económicos — são usados extensivamente por economistas para uma grande variedade de análises.

Quando começámos a estudar o Índice dos Preços de Consumo, tanto o nosso colega como nós próprios estávamos interessados especialmente na estrutura dos seus erros: quais as espécies de erros que o afectam e qual a espécie de justeza que o Índice possui? Depressa verificámos que uma das qualidades dos estatísticos governamentais tinha mudado desde os tempos de HENRY ADAMS. Os estatísticos de Washington que ele conheceu há uns setenta anos podem ter mostrado uma falta de confiança, mas os que actualmente escrevem acerca do Índice dos Preços de Consumo mostram um montante de confiança surpreendente. O material escrito que encontramos exprimiua grande confiança na precisão do Índice e dizia que os erros eram geralmente pequenos e desprezáveis. Seria fascinante especular sobre esta alteração no culto das estatísticas oficiais — poderíamos imaginar por exemplo, que o aumento na confiança era derivado de uma sensibilidade crescente do governo para as relações públicas e a imagem apresentada — mas deixemos esta investigação ao cuidado dos sociólogos e dos cultores da ciência política.

Com surpresa nossa achámos que havia pouca evidência acerca da estrutura dos erros do Índice. Estudámos a evidência que pudemos encontrar e publicámos um artigo sobre o assunto⁽²⁾. Este artigo foi apenas um pequeno contributo num movimento geral em torno de investigações ulteriores acerca da estrutura dos erros do Índice e consta-nos que o Departamento de Estatística do Trabalho procede actualmente a estudos detalhados sobre os erros, como aliás é o seu dever.

O que é que entendemos por estrutura dos erros? Tentemos dar uma ideia dessa estru-

(2) WILLIAM H. KRUSKAL and LESTER G. TELSER, «Food Prices and the Bureau of Labor Statistics», *J. Business*, Univ. of Chicago, 33 (1960), 258-279; comentado por EWAN CLAGUE e réplicas, 280-5.

tura, considerando apenas o preço dos produtos alimentares, que constituem aliás uma componente importante do Índice. A componente constituída pelo preço dos produtos alimentares é em si própria um sub-índice constituído pelos preços de muitos alimentos específicos. Suponhamos que estavamos interessados no preço por quilo de um corte bem definido de carne de vaca de uma certa qualidade vendida em Chicago num determinado dia. Sem dúvida que nesse dia há na cidade de Chicago milhares de talhos, oferecendo a carne do corte e qualidade que seleccionámos a vários preços. Claro que nem todos estes talhos podem ser levados em conta para efeitos de observação e, seja como for, tem de se proceder a uma amostra. A prática corrente é trabalhar com uma amostra mais ou menos fixada dos estabelecimentos de géneros alimentícios, de modo que o Departamento de Estatística do Trabalho envia representantes seus aos gerentes de cada um dos estabelecimentos incluídos na amostra, que se encarregam de perguntar ao gerente respectivo o preço corrente da carne (bem como, evidentemente, dos seus preços para muitos outros alimentos). Em Chicago, em 1960, foram pedidos os preços em todas as grandes cadeias de estabelecimentos e em cerca de 90 retalhistas de géneros alimentícios. Todo este processo de amostragem está sujeito a uma variedade de erros, muitos sem dúvida alguma desprezáveis, mas outros talvez grandes.

No caso dos dois tratamentos médicos foi conveniente classificar os erros em sistemáticos e aleatórios. Vamos considerar agora primeiro um erro sistemático possível, um erro de excentricidade, ou seja, um erro que tende a prosseguir numa direcção pouco mais ou menos com a mesma intensidade, e que é relativamente insensível à dimensão da amostra. É o caso de alguns estabelecimentos não estarem dispostos a cooperar com o Departamento de Estatística do Trabalho, na me-

da em que se recusam sistematicamente a fornecer as suas listas de preços aos representantes do Departamento. (Claro que os representantes podiam simplesmente deslocar-se a outro estabelecimento e anotar os preços, mas esta operação requer muito mais tempo do que tirar simplesmente os números de uma lista). Pode muito bem ser que estes estabelecimentos menos cooperantes tendam para possuir um padrão de preços e uma variação dos preços diferentes dos estabelecimentos cujos gerentes têm um espírito mais aberto. Se, por exemplo, se trata de estabelecimentos que usam processos primitivos, ou mais recentes, ou ainda de estabelecimentos em áreas mais pobres, é possível que sejam menos cooperantes. Deste modo podem surgir excentricidades ou erros sistemáticos resultantes da diferença entre os estabelecimentos que cooperam e os que não cooperam. Claro que podemos pensar em muitas outras fontes de excentricidades; uma das funções do estatístico é estar sempre alerta para elas e sugerir programas experimentais especiais para avaliar as suas grandezas. O estatístico podia, por exemplo, planejar um programa especial para observar alguns preços em estabelecimentos que não estivessem dispostos a colaborar, com o fim de comparar os movimentos destes preços com os movimentos dos preços dos estabelecimentos cooperantes.

E o que é que teremos acerca dos erros aleatórios no exemplo presente? Como anteriormente, pudemos subdividi-los em erros de medição e de cálculo e em erros de amostragem propriamente ditos. Os erros aleatórios de medição podem imaginar-se facilmente no exemplo presente; o gerente de um dos estabelecimentos pode simplesmente citar um preço errado, o empregado do Departamento de Estatística do Trabalho pode não ter ouvido o preço correctamente ou pode enganar-se ao escrevê-lo. Pode haver erros burocráticos no processamento posterior destes

números. Tais erros parecem à primeira vista ridículos, mas são mais frequentes do que se pode pensar, especialmente numa operação grande e complexa. Muitos de nós temos contas de depósitos à ordem nos bancos e quantas vezes sucede que temos dificuldades em ajustar as nossas contas com o extracto mensal enviado pelo banco, apenas devido a um simples lapso aritmético! Acontecerá isto uma vez em cada cinco, em dez, em vinte? Possivelmente, pelo menos, o último caso. Agora imaginemos o que pode suceder num processo complexo que requer milhares de fases de carácter burocrático e aritmético.

Voltemos agora à segunda subdivisão dos erros aleatórios: os erros de amostragem propriamente ditos. Suponhamos que era possível registar exactamente cada uma das vendas da espécie particular de carne de vaca que seleccionámos durante o dia específico em que procedemos à operação. Então podíamos conhecer sem erro o preço médio verdadeiro daquele dia. Se apenas procedemos à *recolha de amostras* dos preços e calculamos uma média para a amostra, obtemos em geral um número diferente da média verdadeira. Mas qual será o valor provável desta diferença? Um centimo, dois, cinco? Há algumas diferenças, consoante o caso. Este é o erro de amostragem.

Para os nossos fins o aspecto mais importante do erro de amostragem é provavelmente o que resulta das amostragens de estabelecimentos alimentares. Suponhamos que os registos do Departamento de Estatística do Trabalho tinham desaparecido inesperadamente e que se tornava necessário seleccionar uma nova amostra, ignorando os resultados da anterior, mas usando os mesmos métodos. Os estabelecimentos incluídos na nova amostra seriam duma maneira geral diferentes dos que estavam na amostra anterior e deste modo os seus preços — e o preço médio da amostra — seriam diferentes. Mas qual o valor desta diferença? Como é que

podemos discutir esta espécie de erros da amostragem relativos aos estabelecimentos?

Suponhamos que sabemos o preço médio de cada amostra possível que poderíamos ter escolhido. Suponhamos que estes preços médios variavam entre 77 e 89 centimos, com a maioria dos preços médios nas proximidades dos 83 centimos, com uma diferença de dois centimos para mais ou para menos. A distribuição dos preços mostraria a *variabilidade do preço médio de amostra para amostra* e dessa distribuição poderíamos calcular a probabilidade de que a média de uma amostra individual, escolhida ao acaso, apresentasse um desvio do valor verdadeiro da ordem de um centimo, dois centimos, etc.

Infelizmente não há nenhuma maneira prática de obter o preço médio de cada amostra possível, e deste modo achar a variabilidade do preço médio de amostra para amostra. É aqui que entra um dos ramos fundamentais da Estatística. Se o método usado para a amostragem é o de escolher os estabelecimentos segundo a chamada *amostra casual*, isto é, fazer com que um estabelecimento entre ou não na amostra segundo um determinado mecanismo aleatório conhecido, podemos normalmente dizer que a variabilidade da média dos preços de amostra para amostra seria obtida olhando apenas para a variabilidade dos preços de uns estabelecimentos para os outros dentro da amostra *escolhida de facto*. O conceito de amostra casual é de importância fundamental e seria desejável se nos pudessemos ocupar dele durante mais tempo.

A noção básica de passar da variabilidade *observável* dentro da amostra para a variabilidade entre as várias amostras *que não se pode observar* pode apreciar-se facilmente. Exactamente como podemos estimar o preço médio de todos os estabelecimentos a partir da média duma amostra convenientemente escolhida, também podemos estimar a variabilidade dos preços entre todos os estabele-

cimentos a partir da variabilidade dos preços de uns estabelecimentos para os outros dentro de uma amostra convenientemente escolhida. Uma vez que estejamos de posse de uma estimativa da variabilidade de estabelecimento para estabelecimento, podemos modificá-la, tomando em consideração o nosso interesse na variabilidade entre as médias das amostras.

Assim, se presumirmos que a nossa amostra foi escolhida por métodos probabilísticos, podemos estimar a variabilidade do preço médio de amostra para amostra e deste modo fazer afirmações acerca da probabilidade de que a média da nossa amostra apresente um desvio, relativamente à média verdadeira geral, de uma certa amplitude.

Por outro lado, se os estabelecimentos são seleccionados pelos chamados especialistas-avaliadores, ou pedindo voluntários, ou por qualquer outro método não probabilístico, não se pode a partir da estrutura interna da amostra predizer a variabilidade de amostra para amostra. Podemos ainda notar que os estabelecimentos que entram nas amostras do Índice dos Preços de Consumo *não* são escolhidos segundo o processo da amostra casual (ou, pelo menos, não o eram em 1960), e este facto torna as discussões acerca do erro da amostragem consideravelmente mais complicadas.

E, deste modo, aqui temos de novo exemplos de duas grandes classes de erros: sistemáticos e aleatórios. Há algumas espécies de erros que não são fáceis de classificar deste modo. Em inquéritos que incluem perguntas sobre a idade, surgem fenómenos muito interessante. Senhoras mais idosas tendem a decrescer as suas idades, enquanto que jovens bastante novas tendem a aumentar as suas. Outra espécie de erro que não cai estritamente dentro das nossas rúbricas é um erro aleatório que afecta grupos inteiros de observações. Sabe-se que características pessoais dos entrevistadores podem influenciar as res-

postas em certas amostras feitas por inquérito, de modo que todas as respostas obtidas por intermédio de um determinado entrevistador ficarão sujeitas à mesma influência. Tal erro pode ter uma componente sistemática, juntamente com uma componente aleatória, derivada do facto de a escolha dos entrevistadores ser feita por um processo aleatório. O efeito da componente aleatória decrescerá, não à medida que o número total de observações cresce, mas antes com o aumento do número de entrevistadores.

Deste modo completámos a nossa discussão acerca do Índice dos Preços de Consumo, com ênfase especial na estrutura dos erros que o influenciam e em questões do método estatístico, especialmente o uso de médias dos resultados obtidos por amostragem para estimar médias verdadeiras e a estimação da variabilidade entre as amostras com base na amostra casual.

IV

Neste espaço limitado de tempo não pudemos descrever o trabalho dos estatísticos teóricos, que comparam métodos alternativos de inferência em termos abstractos e que inventam novos métodos. Contudo, quase todos os estatísticos teóricos se socorrem duma motivação e interesse associados permanentemente com problemas da Estatística aplicada. Uma das fontes de beleza da Estatística está na variedade quase ilimitada de investigações que por seu intermédio se podem realizar. Apenas para mencionar algumas, tiradas da nossa própria experiência pessoal, indicamos as seguintes:

Alterações meteorológicas baseadas na sedimentação das nuvens.

Estabelecimento da autoria dos disputados *Federalist Papers* (Estudos Federalistas).

Modelos de livros usados numa biblioteca universitária.

Uso dos raios X no tratamento de úlceras.
Ensaio de hipóteses acerca da resistência à
fractura de matérias frágeis.

Esperamos ter traçado alguns aspectos do pensamento e do trabalho dos estatísticos. Embora, de certo modo, todos nós nos ocupemos com a Estatística, a prática da Estatística como uma profissão requer um estudo disciplinado, do mesmo modo que a prática profissional do Direito ou da Medicina. A Estatística de que, tal como Monsieur JOURDAIN temos andado a falar durante toda

a vida, é justamente como as discussões jurídicas ou médicas que fazemos como leigos: muitas vezes, embora imperfeitas, são de algum mérito; mas sempre que surge qualquer problema real, temos de nos socorrer dos especialistas. Como nota HENRY ADAMS, os profissionais da Estatística são muitas vezes profundamente cépticos, tal como os especialistas do Direito e da Medicina, mas é o cepticismo que, esperamos, nos irá conduzir até à verdade, pela análise dos erros e pelo estudo quantitativo dos métodos de inferência.