

## **Abolir os parâmetros**

– David Salsburg ;

UMA SENHORA TOMA CHÁ... como a estatística revolucionou a ciência no século XX, capítulo 16 ; pag 137 - 143

Durante os anos 1940, Frank Wilcoxon, químico da American Cyanamid, foi importunado por um problema estatístico. Estivera fazendo testes de hipótese comparando os efeitos de diferentes tratamentos, usando os testes t de Student e as análises de variância de Fisher. Esse era o modo padrão de analisar os dados experimentais àquela época. A revolução estatística tinha dominado o laboratório científico, e os livros e tabelas de interpretação desses testes de hipótese estavam presentes nas estantes de todo cientista. Wilcoxon, porém, estava preocupado com o que frequentemente parecia ser uma falha nesses métodos.

Podia fazer uma série de experimentos em que era óbvio para ele que os tratamentos seriam diferentes no efeito. Algumas vezes os testes t indicariam significância, outras vezes, não. Frequentemente, ao fazer um experimento em engenharia química, o reator químico em que a reação ocorre não está aquecido o bastante no começo da sequência de ensaios experimentais. Pode acontecer de uma enzima particular começar a variar em sua capacidade de reagir. O resultado é um valor experimental que parece errado. Muitas vezes o número é demasiadamente grande ou pequeno. Algumas delas, é possível identificar a causa desse resultado fora de padrão. Outras, o resultado é discrepante, diferindo drasticamente de todos os demais resultados, sem razão óbvia para isso.

Wilcoxon examinou as fórmulas para calcular testes t e análises de variância e compreendeu que esses valores discrepantes, extremos e incomuns, influenciaram enormemente os resultados, causando valores de t de Student menores que deveriam ser (em geral, valores grandes do teste t levam a valores de p pequenos). Era tentador eliminar o dado discrepante do conjunto de observações e calcular o teste t a partir dos demais valores. Isso introduziria problemas na derivação matemática dos testes de hipóteses. Como o químico poderia saber se um número realmente era discrepante? Quantos valores discrepantes teriam de ser eliminados? O químico poderia continuar utilizando as tabelas de probabilidade para estatísticas-padrão de teste se os valores discrepantes tivessem sido eliminados?

Wilcoxon pesquisou o assunto na literatura. Certamente os encontrou uma solução para o problema. Ela envolvia cálculos tediosos baseados em combinações e permutações (as combinatórias de F.N, David forem mencionados no capítulo anterior). Wilcoxon começou a elaborar um método para calcular esses valores combinatórios.

Ah, mas isso era bobagem! Por que um químico como Wilcoxon teria de elaborar cálculos simples, mas tediosos? Sem dúvida alguém da estatística já havia feito isso antes! Mais uma vez ele voltou à literatura estatística a fim de localizar algum artigo prévio sobre o assunto. Nada encontrou. Sobretudo para verificar sua própria matemática, ele submeteu o artigo à revista *Biometrics* (não confundir com a *Biometrika* de Pearson). Ainda acreditava que seu trabalho não poderia ser original e contava com que os pareceristas da revista soubessem onde artigos sobre o assunto houvessem sido publicados antes assim, que rejeitassem seu artigo. Ao rejeitar, eles também o notificariam sobre as outras referências. No entanto, tanto quanto pareceristas e editores da revista puderam determinar, aquele era um trabalho original. Ninguém tinha mesmo pensado naquilo antes, e o artigo de Wilcoxon foi publicado em 1945.

O que nem Wilcoxon nem os editores de *Biometrics* sabiam é que um economista chamado Henry B. Mann e um estudante de pós-graduação em estatística na Universidade do Estado de Ohio, chamado D. Ransom Whitney, estavam trabalhando em problema correlato. Eles tentavam ordenar distribuições estatísticas para que se pudesse dizer, por exemplo, se a distribuição de salários de 1940 era menor do que a distribuição de salários em 1944, e criaram um método de ordenar que envolvia uma sequência simples, embora tediosa, de métodos de contagem.

Isso levou Mann e Whitney a uma estatística-teste cuja distribuição podia ser calculada por aritmética combinatória - o mesmo tipo de computação que Wilcoxon usava. Eles publicaram um artigo descrevendo sua nova técnica em 1947, dois anos depois que o artigo de Wilcoxon aparecera. Logo se verificou que os testes de Wilcoxon e de Mann-Whitney estavam relacionados e produziam os mesmos valores de  $p$ . Os dois testes envolviam algo novo. Até a publicação do artigo de Wilcoxon, pensava-se que todas as estatísticas-teste teriam de ser baseadas em estimativas de parâmetros de distribuições. Esse era um teste, no entanto, que não estimava nenhum parâmetro. Ele comparava a dispersão de dados observados com o que se poderia esperar de uma dispersão puramente aleatória. Era um teste não-paramétrico?

Dessa forma, a revolução estatística avançou um passo além das idéias originais de Pearson. Ela agora podia lidar com distribuições de medições sem lançar mão de parâmetros. Bastante desconhecido no Ocidente, no final dos anos 1930, Andrei Kolmogorov investigou, na União Soviética, com um aluno seu, N.V. Smirnov, um enfoque diferente para a comparação de distribuições que não utilizava parâmetros. Os trabalhos de Wilcoxon e de Mann-Whitney tinham aberto uma nova janela de investigação matemática, ao atrair a atenção para a natureza subjacente de níveis ordenados, e o trabalho de Smirnov-Kolmogorov foi logo acrescentado à lista.

### **Desenvolvimentos posteriores**

Uma vez que uma nova janela tinha sido aberta na pesquisa matemática, os investigadores começaram a olhar através dela, de diferentes maneiras. O trabalho original de Wilcoxon foi logo seguido por enfoques alternativos. Herman Chernoff e I. Richard Savage descobriram que o teste de Wilcoxon poderia ser considerado em termos dos valores médios esperados de estatísticas ordenadas; e eles foram capazes de ampliar o teste não-paramétrico em um conjunto de testes envolvendo diferentes distribuições subjacentes, nenhuma das quais requerendo a estimativa de um parâmetro. No começo dos anos 1960, essa classe de testes (agora conhecidos como "testes livres de distribuição") era o máximo em pesquisa. Estudantes de doutorado preenchiam pequenos nichos da teoria para defender suas teses. Faziam-se reuniões exclusivamente para discutir essa nova teoria. Wilcoxon continuou a trabalhar na área, ampliando o alcance dos testes ao desenvolver algoritmos extremamente inteligentes para os cálculos combinatórios.

Em 1971, Jaroslav Hájek, da Tchecoslováquia, produziu um livro definitivo que forneceu uma visão unificadora para todo esse campo. Hájek, que morreu em 1974, aos 48 anos, descobriu uma generalização subjacente para todos os testes não-paramétricos e ligou esse enfoque geral às condições de Lindeberg - Lévy do teorema central do limite. Muitas vezes é esse o caminho da pesquisa matemática. Em certo sentido, toda a matemática está interligada, mas a natureza exata desses vínculos e a perspicácia para explorá-los podem levar vários anos para aparecer.

Enquanto buscava as implicações de sua descoberta estatística, Frank Wilcoxon deixou seu campo original, a química, e passou a dirigir o grupo de serviços estatísticos da American Cyanamid e de sua divisão, os Lederle. Em 1960, ele ingressou no Departamento de Estatística da Universidade do Estado da Flórida, onde se revelou professor e pesquisador admirado e orientou diversos candidatos a doutorado. Quando morreu, em 1965, deixou um legado de alunos e inovação estatística que continua a ter feito notável sobre a área.

### **Problemas não resolvidos.**

O desenvolvimento de procedimentos não-paramétricos pode ter levado à explosão de atividade nesse novo campo. No entanto, não havia um vínculo óbvio entre os métodos paramétricos usados até então e os métodos não-paramétricos. Havia duas questões a resolver:

1. Se os dados têm distribuição paramétrica conhecida, como a distribuição normal, o que aconteceria de ruim com a análise se usássemos métodos não-paramétricos?
2. Se os dados não se ajustam bem a um modelo paramétrico, quão afastados daquele modelo os dados precisam estar para que os métodos não-paramétricos sejam os melhores?

Em 1948, os editores de *Annals of Mathematical Statistics* receberam o artigo de um desconhecido professor de matemática da Universidade da Tasmânia, ilha na costa sul da Austrália. Esse artigo notável resolveu os dois problemas. Edwin James George Pitman tinha publicado três artigos anteriores no *Journal of the Royal Statistical Society* e um no *Proceedings of the Cambridge Philosophical Society*, que, da perspectiva atual, lançaram os fundamentos de seu trabalho posterior, mas que tinham sido ignorados ou esquecidos. Além desses quatro artigos, Pitman, que tinha 52 anos quando submeteu seu artigo à revista *Annals* nada mais publicara e era desconhecido.

E.J.G. Pitman nasceu em Melbourne, Austrália, em 1897. Frequentou a Universidade de Melbourne, mas teve de interromper os estudos durante a Primeira

Guerra Mundial, quando serviu dois anos no Exército. Voltou para completar seu curso. "Naqueles dias", escreveu depois, "não havia pós-graduação de matemática nas universidades australianas". Algumas universidades ofereciam bolsas de estudo a seus melhores estudantes para se doutorarem na Inglaterra, mas a de Melbourne não. "Quando deixei a Universidade de Melbourne, depois de quatro anos, não tinha treinamento em pesquisa; acreditava, porém, que aprendera a estudar e usar a matemática, e que poderia enfrentar qualquer problema que aparecesse." O primeiro deles foi ganhar a vida.

A Universidade da Tasmânia procurava alguém para ensinar matemática. Pitman candidatou-se e foi nomeado professor de matemática. O Departamento inteiro consistia no novo professor e um conferencista em regime de meio expediente. O departamento precisava dar cursos de matemática para os estudantes de graduação dos demais departamentos, e o novo professor estava ocupado com a carga horária de ensino, que lhe tomava quase todo tempo. O conselho diretor decidiu contratar um professor de matemática em tempo integral; como um dos membros do conselho ouvira falar que havia um novo ramo da matemática chamado estatística, pediram ao novo candidato que preparasse um curso de estatística (fosse isso o que fosse).

Pitman respondeu: "Não posso afirmar que tenha qualquer conhecimento especial de teoria da estatística; mas, se nomeado, estarei preparado para dissertar sobre esse assunto em 1927. "Ele não tinha conhecimento especial nem qualquer outro tipo de conhecimento sobre teoria estatística. Em Melbourne, fez um curso de lógica avançada, durante o qual o professor dedicou um par de palestras à estatística. Pitman afirmaria depois: "Decidi, ali e naquele instante, que estatística era o tipo de coisa em que eu não estava interessado e com a qual nunca teria de me preocupar."

O jovem E.J.G. Pitman chegou a Hobart, na Tasmânia, no outono de 1926, apenas com o diploma universitário para justificar seu cargo de professor em uma pequena escola provincial que estava tão longe quanto se podia chegar da fermentação intelectual de Londres e Cambridge. "Nada publiquei até 1936", escreveu. "Havia duas razões

principais para o atraso na publicação;a carga de trabalho que tinha e a natureza de minha educação, declarou referindo-se a sua falta de treinamento em métodos de pesquisa matemática.

Em 1948, quando enviou seu notável artigo para a revista *Annals of Mathematical Statistics*, o Departamento de Matemática da Universidade da Tasmânia crescera. Tinha agora um professor (Pitman), um professor associado, dois professores convidados e dois tutores. Eles ensinavam uma ampla gama de matemática, tanto aplicada como teórica. Pitman proferia 12 palestras por semana e ainda dava aula aos sábados. Tinha então algum apoio para sua pesquisa. Em 1936 o governo da Comunidade Britânica começou a fornecer 30 mil libras por ano para a promoção da pesquisa científica nas universidades da Austrália. Essa quantia era alocada em diferentes estados, de acordo com a população; como a Tasmânia era um dos estados menores, sua cota era de 2,400 libras por ano para toda a universidade. Quando disso chegou a ele, Pitman não informa.

Pitman empenhou-se gradualmente em diferentes tipos de pesquisa. Seu primeiro artigo publicado lida com um problema de hidrodinâmica. Os três artigos seguintes investigam aspectos altamente específicos da teoria dos testes de hipótese. Esses artigos não eram notáveis em si mesmos, mas representavam suas teses de aprendizagem. Ele estava explorando como desenvolver ideia se relacionar estruturas matemáticas umas com as outras.

Quando começou a trabalhar no artigo de 1948, Pitman tinha desenvolvido clara linha de raciocínio sobre a natureza dos testes estatísticos de hipótese e as inter-relações entre os velhos testes (paramétricos) e os novos (não-paramétricos). Com seus métodos, atacou os dois importantes problemas.

O que ele descobriu surpreendeu a todos. Mesmo quando as suposições originais eram verdadeiras, os testes não-paramétricos eram quase tão bons quanto os paramétricos. Ele foi capaz de responder à primeira pergunta: quão mau será usarmos testes não-paramétricos em uma situação na qual conhecemos o modelo paramétrico e em que deveríamos utilizar um teste paramétrico específico? Nada mau, diria Pitman.

A resposta à segunda pergunta era ainda mais surpreendente. Se os dados não se ajustam ao modelo paramétrico, quão longe desse modelo eles devem estar para que os testes não-paramétricos sejam melhores? Os cálculos de Pitman mostraram que bastava um ligeiro desvio do modelo paramétrico para que os testes não-paramétricos se mostrassem muito melhores do que os paramétricos.

Parecia que Frank Wilcoxon, o químico que estava certo de que alguém já fizera essa descoberta antes dele, tinha tropeçado numa verdadeira pedra filosofal. Os resultados de Pitman sugerem que todos os testes de hipótese deveriam ser não-paramétricos. A descoberta de Pearson das distribuições estatísticas baseadas em parâmetros era apenas o primeiro passo. Agora os estatísticos eram capazes de lidar com distribuições estatísticas sem se preocupar com parâmetros específicos.

Existem sutilezas dentro das sutilezas em matemática. De modo bem profundo, em seus enfoques aparentemente simples, Wilcoxon, Mann e Whitney tinham premissas sobre as distribuições dos dados. Ainda seriam necessários outros 25 anos para que essas premissas fossem entendidas. O primeiro problema perturbador foi descoberto por uma pessoa entendida ("jimmie") Savage, na Universidade de Chicago, em 1956.

Quando mostrei o artigo de Bahadur e Savage a um amigo meu da Índia, ele observou a congruência de seus nomes. Bahadur significa "guerreiro" em hindi. Foi preciso um guerreiro e um selvagem para dar o primeiro golpe na teoria dos testes estatísticos não-paramétricos. Os problemas que Savage e Bahadur revelaram originavam-se do problema que primeiramente sugeriu os testes não-paramétricos a Wilcoxon: o problema dos dados discrepantes. Se as discrepâncias são observações raras e completamente "erradas", então os métodos não-paramétricos reduzem sua influência sobre a análise. Se as discrepâncias são parte de uma sistemática contaminação de dados, mudar para métodos não-paramétricos só agrava a situação. Investigaremos o problema das distribuições contaminadas no Capítulo 23.