Estatística, Ciência e Filosofia das Ciências[1]

Dinis Pestana

CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa
CFCUL — Centro de Filosofia das Ciências da Universidade de Lisboa
Instituto de Investigação Científica Bento da Rocha Cabral

2013 — Ano Internacional da Estatística

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

> *The claim of General Education is that the history of science is part of science. So are its philosophy, its great literature, and its social and intellectual context. The contribution of science instruction to the life of the university and to society should include these elements, since science includes them...*
>
> *Harvard committee on general education*

Transformar a informação em conhecimento

As Ciências da Informação conquistaram um merecido prestígio. Mas por vezes parece que está esquecida a importância de transformar informação em conhecimento.

A Filosofia das Ciências tem neste campo um relevo muito especial na visão crítica, progresso e aperfeiçoamento da metodologia da investigação científica, campo em que também a Estatística intervém.

Não se limita a isso a forte ligação da Filosofia das Ciências com as ciências. Por exemplo, questões éticas são incontornáveis na investigação bio-médica (e deveriam

[1] Trabalho financiado por fundos nacionais através da FCT — Fundação para a Ciência e a Tecnologia, no âmbito do projecto PEst-OE/MAT/UI0006 /2011.

ser também determinantes na avaliação e controle da investigação em muitas outras áreas, da Educação ao Ambiente).

T. Greenfield, no Prefácio de *Research Methods for Postgraduates* — um livro que *The New Scientist* classificou como "*The most useful book any postgraduate could ever buy*" — começa com esta frase:

> *The* [UK] *government proposed in 1994*[(2)] *in their White Paper* Realising Our Potential *that all graduates who wish to study for doctorate should first take a one-year master's course in research methods. Several universities have since introduced such course and more are planned. This book is a response to that development.*

Research Methods for Postgraduates é de facto um livro notável, que aborda temas desde como procurar financiamento e fontes documentais, uma discussão sobre creatividade, descrição de vários tipos de planeamento experimental, uma interessante abordagem ao problema das medições (tantos cursos se alhearam completamente da Metrologia!). Trata também da análise dos dados, modelação matemática (determinista e estocástica, incluindo uma discussão de simulação).

Finalmente, ensina princípios gerais sobre como apresentar resultados, com capítulos especiais sobre redacção da tese, e preparação de apresentações e relatórios, e como apresentar-se em páginas pessoais na *net*, e termina com indicações gerais sobre como preparar o futuro profissional, quer no que respeita a oportunidades de emprego, quer no que que convém fazer para protecção da propriedade intelectual.

Da descrição sumária das diversas partes daquele interessante livro percebe-se que a Estatística tem um papel de relevo na preparação da formação avançada de trabalhadores científicos.

De facto, mais do que um terço das suas quase 400 páginas correspondem a capítulos sobre diversos aspectos da Estatística:

11: Choosing and using software for statistics
19: Randomized trials
20: Laboratory and industrial experiments
21: Agricultural experiments
22: Survey research
23: Principles of sampling
24: Sampling in human studies
25: Sources of population statistics
26: Interviewing
29: Elementary statistics
30: Further statistical methods

(2) *Realising Our Potential* é, de facto, datado de 1993, e pode actualmente ser consultado em http://www.official-documents.gov.uk/document/cm22/2250/2250.pdf.

31: Computer support for data analysis
35: Stochastic models and simulation

A Estatística na quantificação da incerteza

Acaso e necessidade são duas abstracções matemáticas; de facto, tudo aquilo que nos rodeia é uma fusão, em partes variáveis, daqueles dois extremos inatingíveis. Um dos fundadores da Probabilidade, Abraham de Moivre, no seu notável livro *The Doctrine of Chances*, escreveu com admirável lucidez:

> "*Further, The same Arguments which explode the notion of Luck may, on the other side, be useful in some Cases to establish a due comparison between Chance and Design: We may imagine Chance and Design to be as if it were in Competition with each other, for the production of some sorts of Events, and may calculate what Probability there is, that those Events should be rather owing to one than to the other.*"

Indica assim que um dos papeis da probabilidade é o dilucidar entre vários padrões de aleatoriedade possível. Aponta isso para a utilidade de **modelos** — devendo ter-se sempre em mente que os modelos são para ser úteis, e não mais do que isso, e que não é correcto dizer "o modelo para estes dados", pois o rigor exige-nos que digamos "um dos modelos para estes dados".

Os modelos são úteis porque a Ciência se ocupa de um colectivo (a população), ainda que o faça usando uma parcela desse colectivo (uma amostra). Há modelos que são construídos como consequência lógica de algumas premissas simplificadoras (por exemplo, o modelo binomial e o binomial negativo, para contagens assumindo os pressupostos de Jacques Bernoulli — experiências aleatórias independentes com resultado dicotómico, sendo a probabilidade de cada um dos resultados possíveis sempre o mesmo; ou o modelo hipergeométrico, para contagens em condições semelhantes às descritas mas em que em vez de independência se tem permutabilidade, quando se procede a extracções simultâneas ou, o que tem o mesmo efeito, sem reposição); por outro lado, há modelos que surgem como aproximações justificadas por sofisticados resultados assintóticos, por exemplo contagens usando o modelo de Poisson quando se admite que em termos médios se pode admitir estabilidade.

Um dos mais extraordinários resultados assintóticos é uma consequência da simples observação da estabilidade da média das observações, em amostras $(x_1, x_2, \dots)$ de dimensão crescente. De facto, denotando a soma das n primeiras observações s_n, isto é, $s_n = \sum_{k=1}^{n} X_n$, e a correspondente média $\overline{x}_n$, ou seja $\overline{x}_n = \frac{S_n}{n}$, tem-se

$$\overline{x}_{n+1} = \frac{n}{n+1}\,\overline{x}_n + \frac{x_{n+1}}{n+1} \approx \overline{x}_n$$

Isto levou Jacques Bernoulli a estabelecer que a sucessão de médias amostrais converge para a média populacional.

Uma consequência notável deste resultado é a sua aplicação a contagens binomiais. Admitindo que cada experiência aleatória tem como resultado o acontecimento A ou o seu complementar $\overline{A}$, e "rebaptizando" A como 1 e $\overline{A}$ como 0, assumindo que a probabilidade $\mathbb{P}(A) = p$ é constante, temos para o resultado de cada uma das experiências aleatórias réplicas independentes X_k do modelo

$$X = \begin{cases} 0 & 1 \\ 1-p & p \end{cases},$$

sendo portanto a média populacional $\mathbb{E}(X) = 0 \times (1-p) + 1 \times p = p$. Consequentemente, a sucessão de frequências relativas de observação de A converge para a probabilidade p desse acontecimento.

Este resultado notável libertou a probabilidade das limitações do pressuposto de equiprobabilidade dos acontecimentos elementares, que naturalmente se assume no caso de jogos de azar, admitindo que dados e moedas não são viciados, que cartas são bem baralhadas (na prática, sete ou mais inserções de uma parte do baralho na outra parte), o que evidentemente é mais um desiderato do que uma realidade. Esta descoberta, que Jacques Bernoulli começou a comunicar aos seus correspondentes nos finais do século XVII (o seu livro *Ars Conjectandi* só foi publicado postumamente, pelo seu sobrinho Nicolau Bernoulli, em 1713) tem consequências tão notáveis que o seu autor sempre se lhe referiu como o seu *teorema de ouro*(3); actualmente é referido como Lei dos Grandes Números, apodo dado por Poisson quando chamou Lei dos Pequenos Números ao seu teorema de convergência de sucessões de binomiais com probabilidade de sucesso evanescente mas valor médio estável.

é na esteira da descoberta da Lei dos Grandes Números que Halley observa uma coorte de indivíduos para construir a tabela de mortalidade que a *Royal Society* lhe encomendou, para satisfazer o que seria o embrião das companhias de seguros, que desejavam estabelecer anuidades competitivas mas que assegurassem lucro nos seguros de vida que tinham começado a contratar. é também esta base da Estatística frequencista que leva John Abthurnot a inventar os testes de hipóteses, como David and Edwards (2001) anotam, ao observar que nos 72 anos de registos de nascimentos, existentes na sua paróquia de Londres, ano a ano tinham consistentemente nascido mais rapazes do que raparigas, o que o levou a concluir que a probabilidade de um nasciturno ser do sexo masculino é certamente maior do que $\frac{1}{2}$(4).

(3) Este teorema de ouro continua uma carreira gloriosa, e nomeadamente no século XX inspiraria Ulam a inventar o Método de Monte Carlo, para obter rapidamente aproximações numéricas tão precisas quanto necessário de valores cujo cálculo usando os instrumentos analíticos clássicos seria demorado ou mesmo impossível.

(4) Se alguém lançar 72 vezes uma moeda ao ar, e nessas 72 experiências observar sempre que o lado que fica virado para cima é Face, só se for tolo é que admite que a Probabilidade de sair Face é $\frac{1}{2}$. De facto, em termos modernos, considerando que a hipótese nula é que $\mathbb{P}(M) = \mathbb{P}(F) = \frac{1}{2}$, observando que em 72 "sucessos"em setenta e duas provas independentes o nível de significância

Claro que a utilidade de qualquer resultado limite depende da *velocidade de convergência*, no caso da Lei dos Grandes Números das médias parciais para o valor médio populacional. Foi o já referido Abraham de Moivre que mostrou que no caso de contagens binomiais essa velocidade de convergência é da ordem de $\frac{1}{\sqrt{n}}$[(5)].

Este resultado de Abraham de Moivre teve consequências tais que o seu papel como avaliador da velocidade de convergência na Lei dos Grandes Números fica muitas vezes omisso. De facto, o resultado de de Moivre mostra que centrando e reduzindo convenientemente a soma de variáveis independentes cujo padrão de aleatorieda é binomial, o resultado pode ser aproximado por uma variável aleatória contínua, com função densidade de probabilidade[(6)] $f(x) = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$.

Este resultado aparentemente modesto foi sucessivamente generalizado (o bem humorado texto de Le Cam (1986) descreve a cristalização deste resultado), usando instrumentos cada vez mais sofisticados (transformadas integrais, momentos, semigrupos de convolução), mostrando que a lei com densidade de probabilidade $f(x) = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}$ aparecia como limite de somas centradas e reduzidas, sob condições muito gerais. Como ainda por cima se pode trabalhar eficientemente com essa lei, usando uma tabela que cabe numa página A_5, tornou-se uma favorita, e no século XIX Pierce chamou-lhe "lei normal", um apodo que foi vulgarizado quando Karl Pearson o adoptou. A convergência de somas para a "lei normal"tornou-se o resultado limite mais importante da Teoria da Probabilidade, e em 1920 Pólya chamou-lhe Teorema Limite Central. Para perceber a importância do Teorema Limite Central em inferência, basta observar que as média são somas para se perceber a importância deste padrão de aleatoriedade (e recordando que a variância é uma soma de quadrados, vê-se imediatamente a importância sa soma de quadrados de normais, uma família de leis com o nome de qui-quadrado, um nome que surgiu por na altura se usar habitualmente o símbolo χ para denotar a lei normal padrão).

A acessibilidade a meios de cálculo rápidos e baratos levou, na segunda metade do século XX, ao uso de modelos que se adaptavam melhor à amostra efectivamente observada, mas o modelo "normal", devido ao Teorema Limite Central, continua a ser uma referência incontornável em aproximações quando as amostras têm uma dimensão razoável (o que evidentemente é vago, mas não faz sentido estar aqui a especificar mais); por outro lado, assumir um modelo normal na análise de amos-

---

descritivo (*p-value*), ou seja, a probabilidade de observar um resultado tão ou mais discrepante com a hipótese nula do que aquele que foi observado é $\left(\frac{1}{2}\right)^{72} \approx 2.11758 \times 10^{-22}$.

(5) Modernamente, é possível melhorar a velocidade de convergência, estabelecendo que é da ordem de $\frac{1}{n}$, usando sofisticadas técnicas de *tilting* e *back-tilting* e aproximações em sela, que os interessados podem consultar no trabalho seminal de Daniels (1954) ou na interessante monografia de Barndorff-Nielsen and Cox (1989).

(6) Claro que então não existia a noção de densidade de probabilidade, o que de Moivre mostrou foi que $\mathbb{P}[a < \frac{S_n - np}{\sqrt{np(1-p}} < b] \approx \frac{1}{\sqrt{2\,\pi}} \int_a^b \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$.

tras de dimensão reduzida tem vantagens muito grandes, uma vez que há resultados distribuicionais exactos para média amostral e variância amostral, ou no que respeita a estatísticas (isto é, funções da amostra, que não dependem de parâmetros desconhecidos, e que são um veículo apropriado quer em estimação de parâmetros quer em testes de hipóteses) para comparação de valores médios ou comparação de variâncias.

A capacidade de medir, cada vez com maior precisão, foi um dos factores que mais alterou a vida humana, e seria deveras interessante uma história da Ciência que explorasse essa perspectiva. A Probabilidade e a Estatística trouxeram progressos notáveis à nossa capacidade de medir, quer de forma directa, quer de forma indirecta (desta, a regressão, falaremos mais adiante). De facto, observe-se que a Lei dos Grandes Números mostra que a média de medições repetidas (admitindo que não há erros sistemáticos) fornece uma excelente avaliação da verdadeira medida que se pretende, e o Teorema Limite Central quantifica a probabilidade dessa avaliação, com uma precisão fixada pelo experimentador, ser a que que se pretende. De uma outra perspectiva, o Teorema Limite Central (e em particular a sua extensão para o caso de uma dependência fraca, a permutabilidade, que é comum quando se faz observações sempre distintas), indica-nos quantos dados temos que observar para, com um grau de confiança tão alto quanto desejarmos fixar, podermos estimar um parâmetro com uma precisão pré-fixada conveniente para os fins em vista, indicando um “intervalo de confiança”suficientemente apertado. Assim, o Teorema Limite Central é, entre muitas outras coisas, um dos fundamentos da moderna teoria da amostragem, e por extensão do planeamento de experiências.

Lidar com a incerteza e o risco

Against the Gods, The Remarkable Story of Risk, do economista P. L. Bernstein, começa desta forma:

> *What is that distinguishes the thousands of years of history from what we think of as modern times? The answer goes way beyond the progress of science, technology, capitalism, and democracy.* [...]
>
> *The revolutionary idea that defines the boundary between modern times and the past is the mastery of risk.* [...] *This book tells the story of group of thinkers whose remarkable vision revealed* [...] *how to understand risk, measure it, and weigh its consequences.* [...]

E Max Born não hesitou em escrever

> *The conception of chance enters into the very first steps of scientific activity in virtue of the fact that no observation is absolutely correct. I think chance is a more fundamental conception than causality; for*

> *whether in a concrete case, a cause-efect relation holds or not can only be judged by applying the laws of chance to the observation.*

Um precursor genial do ideário estatístico foi Galton (são suas as primeiras ideias sobre correlação ordinal, inventou a regressão, financiou o primeiro Laboratório Estatístico do mundo, em Londres, nomeando Karl Pearson para o dirigir).

De facto, apercebeu-se de que a Estatística (e a Probabilidade, que é a "linguagem"com que a Estatística exprime os seus conceitos e análises) podia desenvolver as ferramentas necessárias para a investigação de fenómenos complexos. Em *Natural Inheritance* (1889) escreveu

> *Some people hate the very name of Statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of dificulties that bar the path of those who pursue the Science of man.*

... só lhe faltou dizer que a Estatística é um canivete suisso virtual, apropriado para lidar com as questões complexas que indagamos, ou relativamente às quais temos que tomar decisões.

De facto, todas as ciências experimentais são construídas com base em informação muito parcelar da realidade. A indução para a generalidade (população) do que se pode estabelecer com base nessa informação parcial (amostra) é um salto no desconhecido.

Por isso a Estatística desempenha um papel de grande relevo, pois apoiando-se no rigor do raciocínio dedutivo nos seus desenvolvimentos matemáticos, quantifica o **erro** a que o raciocínio indutivo nos pode conduzir, em qualquer processo de inferência, de previsão ou de controle.

O que torna a Estatística específica é lidar com a **incerteza** (*domesticar a incerteza*), apoiar decisões avaliando **riscos** competitivos. O "preço"é ter que assumir explicitamente que os resultados que fornece têm uma margem de erro — mas ao quantificá-la, fornece de facto excelentes bases para tomar decisões racionais ao lidar com fenómenos complexos.

Causalidade *vs*. associação estatística?

O diálogo entre a Probabilidade e Estatística e a Filosofia das Ciências tem acompanhado o desenvolvimento destas disciplinas. Leibniz e Hume têm textos notáveis sobre conceitos da Probabilidade; o princípio da razão insuficiente, de Jacques Bernoulli (depois retomado por Laplace), a inversão da probabilidade e a quantificação da possibilidade das causas, de Bayes e posteriormente explorado por Laplace, a dialéctica subjacente à reavaliação da probabilidade por incorporação de nova informação, mostrando os limites do conceito de probabilidade objectiva, a discussão de probabilidade e certeza de Poincaré, são apenas exemplos de meditação filosófica por alguns dos homens que levaram Probabilidade e Estatística a saltos qualitativos memoráveis.

Aquele diálogo parece ocasionalmente interrompido, mas é sempre reatado pelos grandes criadores de novas ideias sobre como lidar com o risco e a incerteza. Entre estes, no despontar da moderna Estatística Matemática, singularizamos de imediato as contribuições de Karl Pearson, de 'Student' e de Sir Ronald Fisher.

Karl Pearson, criador da correlação cruzada, impulsionador da investigação do efeito da assimetria e curtose no afastamento do modelo normal — originando assim os estudos de robustez e resistência —, e inspirado inventor do teste do qui-quadrado, publicou em 1896 *The Grammar of Science*, um livro que mudou de forma indelével o paradigma de Ciência, defendendo que a busca de causalidade não pode ser um objectivo obsessivo e limitador, pois em muitos casos a mera identificação de associações estatísticas entre fenómenos diversos aumenta, de facto, o nosso conhecimento. O princípio da razão suficiente de Leibniz é decerto um padrão de ouro que se deseja atingir, mas quando esse óptimo não se pode alcançar a meta de estabelecer causalidade pode ser abrandada para uma pesquisa de associação estatística.

A associação estatística pode ser avaliada de diversas formas, dependendo inclusivamente das "escalas de medição"[(7)] das variáveis. Galton parece ter sido o primeiro a ter a ideia de definir uma "correlação"entre variáveis ordinais (uma ideai que viria a impor-se com a correlação ordinal de Spearman), mas foi a "correlação produto"de Pearson, ou covariância entre as versões padronizadas (i. e., centradas em 0 e com escala unitária) que impôs a ideia de medir a associação entre variáveis quantitativas. Foi também Pearson que abriu caminho, com o teste do qui-quadrado, para decidir se variáveis qualitativas eram independentes ou se, pelo contrário, havia indícios significativos de que existe associação entre elas.

A argúcia de medir uma variável através de outra

Medir o valor assumido por uma variável através da medição de outra a que está associada é um truque notável, que contribui para a actual qualidade de vida de forma tão comum e subtil que nem damos por isso. Por exemplo, o instrumento que uso para medir diariamente o valor da glicemia capilar mede, de facto, o ângulo de refracção da luz na amostra de sangue que deposito na ponta de uma tira, e transforma essa medição no que me interessa, afinal, medir.

O geneticista Galton, muito interessado em quantificar características hereditárias, observou que os filhos de pais muito altos tendiam a ser altos, mas menos do que os progenitores, e que os filhos de pais baixos tendiam a ser baixos, mas me-

---

[(7)] A referência incontornável continua a ser Steven (1946).

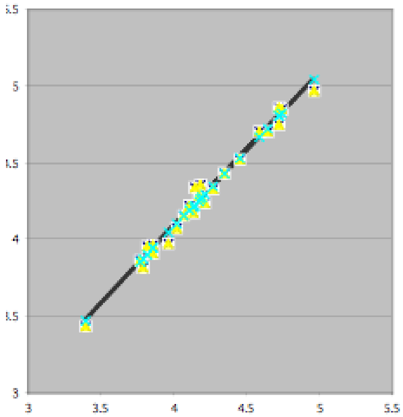
nos do que os progenitores. Era como se a média populacional funcionasse como um imã, observando-se uma tendência generalizada para "regressar" a valores próximos da média. Daí a palavra *regressão* para designar genericamente a disciplina que se ocupa da predição do valor de uma "variável resposta", difícil de medir directamente, através das medições de uma "variável controlada", mais fácil de medir.

Por exemplo, para avaliar a riqueza florestal de uma região, a observação e medição directa é praticamente impossível, enquanto a avaliação através de fotografia aérea é fácil; por isso, escolhem-se ao acaso alguns pontos de amostragem, em que se procede a uma medição directa, para comparar com a medição indirecta fornecida pela fotografia aérea nesse pontos, e depois um simples *estimador de razão* pode estimar-se o valor total real através do valor "fictício" obtido com a fotografia aérea.

As figuras que se seguem (três fotografias amavelmente cedidas por Miss Levinska, com um charuto de tamanho ergonómico na boca, e com as medições que abaixo se explicam para determinar o ângulo de abertura dos maxilares, e o ajustamento de uma regressão linear entre abertura assim medida e abertura entre os dentes) mostram um exemplo típico do uso da regressão: Na reconstrução maxilo-facial de acidentados, o maxilar tem que ser fixado de forma que o paciente venha a poder fechar a boca devidamente, para trincar os alimentos (e articular palavras, e morder, se for agressivo), e abrir o suficiente para ser cómodo introduzir na boca a colher de sopa (mas não a concha de serviço), ou um grosso charuto, ou qualquer outra coisa que se deseje. Num indivíduo não acidentado, é fácil medir a distância entre os dentes, com a boca totalmente aberta (claro que medições repetidas com um instrumento de precisão vão ser quase certamente distintas, mas é uma situação em que podemos usar todos os valores, por exemplo através da sua média). Mas num acidentado cujos dentes ficaram destruídos, que tem a cara inchada, e eventualmente os maxilares fracturados, essa medição não faz sentido.

Porém, num indivíduo em bom estado podemos, cada vez que vamos medir a distância entre os dentes, medir também a distância entre um ponto marcado no nariz e um ponto marcado no queixo, com a boca fechada e com a boca aberta.

Ficamos assim a dispor de uma sucessão de pares de observações fortemente correlacionadas, a distância entre os dentes e a diferença entre as duas medições de distâncias entre aqueles pontos. Isto pode ser feito para uma amostra adequadamente dimensionada de indivíduos saudáveis. A disposição dos pontos num diagrama de dispersão sugere, neste caso, que os pontos pouco se afastam de uma recta intermédia (que é obtida calculando o declive e ordenada na origem da recta que minimiza o quadrado das distâncias verticais entre os pontos observados e os pontos dessa recta). Essa "regressão linear" permite assim, medindo a distância entre pontos marcados no nariz e no queixo de um acidentado, pedindo-lhe que faça o seu melhor para emular o que, quando saudável, correspondia a boca fechada e boca totalmente aberta, predizer como deve ser feita a reconstrução maxilo-facial, para que os maxilares possam fazer um ângulo conveniente.



Observacional *vs.* **Experimental**

A velha frase “contra factos não há argumentos”parece aceitar que os factos são sempre bons, enquanto os argumentos podem não o ser. Mas basta um pouco de reflexão para perceber que os factos podem ser uma caricatura grotesca da realidade. Em geral, podemos observar apenas um escasso número de observações, mas a nossa ambição é induzir do que observamos nessa “amostra”para uma população muito mais vasta. Ora nem todas as colecções de dados têm a qualidade necessária para podermos limitar devidamente os riscos inerentes ao processo de indução.

Para que uma colecção de dados seja representativa, vários requisitos devem ser preenchidos. Intuitivamente percebe-se que quanto maior for a variabilidade, mais elementos deve a amostra conter, para poder representar essa variabilidade.

A situação mais corrente é querermos estimar uma média populacional μ com base na média amostral $\overline{x}$. Admita-se que o número de elementos da população é N, e que retiramos n elementos da população para estimar μ. O mais comum é fazer as extrações em simultâneo (ou sequencialmente, sem reposição, para assegurar que qualquer das observações traz nova informação). Qual é o número de elementos que devemos ter na amostra por forma a poder estimar μ com um erro inferior a B, correndo um risco de nos enganarmos de apenas 5%?

A média empírica $\overline{x}$ é o valor observado de uma variável aleatória $\overline{X}$ cuja variabilidade (variância) é $\frac{\sigma^2}{n}\frac{N-n}{N-1}$, onde σ^2 é a variância da população, que avaliamos pelas variância amostral s^2. Em extrações sem reposição não temos independência entre as observações, mas o tipo de dependência fraca que existe neste caso — permutabilidade — permitiu a Erdös e a Rényi estabelecerem uma extensão do Teorema Limite Central. Assim, vamos aproximar usando os quantis da variável "normal", $z_{0.025} = -1.96$ e $z_{0.975} = 1.96$, e consequentemente com um coeficiente de confiança 0.95 podemos supor que μ está no "intervalo de confiança"$\overline{x} - 1.96\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}, \overline{x} + 1.96\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ Assim, para conseguir os nossos objectivos, basta determinar o tamanho n da amostra resolvendo a inequação

$$2 \times 1.96\,\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < B.$$

Com amostras de dimensão inferior não podemos garantir a precisão pretendida. Mas note-se que é um erro usar amostras de dimensão muito superior, porque a estimativa da variância decresce como $\frac{1}{\sqrt{n}}$, e isso pode levar-nos a concluir que diferenças irrelevantes têm importância.

Fazer o número requerido de observações garante que temos uma amostra representativa? Claro que não, se as observações não forem casuais não reflectem a variabilidade da população. Medindo a altura de jogadores de basquetebol não ficamos com uma amostra representativa de alturas da população em geral. Não vale a pena multiplicar exemplos, penso que se percebe facilmente que uma amostra "de conveniência"pode falsear fatalmente todas as conclusões. A recolha de uma amostra tem que obedecer a uma disciplina rigorosa, para garantir que as observações foram extraídas "ao acaso", "aleatoriamente", e que isso confere probabilidade equilibrada de qualquer elemento da população ser seleccionado para a amostra(8).

Há uma enorme quantidade de estudos "científicos"que em algum ponto dizem qualquer coisa do género: "*Uma equipa da universidade de ... pediu a 46 voluntários que usassem a mesma roupa durante 24 horas, e que ao tirá-la a colocassem num saco de plástico hermeticamente fechado. O saco era depois dado a cheirar a cada um dos outros 45 indivíduos, pedindo-se que cada um deles classificasse o cheiro de 0 (abominável) a 10 (agradável, inebriante, despertando todos os desejos, mesmo os mais secretos e inconfesséveis), concluindo-se que ...*"Claro que usando amostras de conveniência o que se produz é ciência da treta.

(8) A amostragem aleatória sem reposição garante que todos os subconjuntos de tamanho n têm igual probabilidade de ser seleccionado como amostra; outras disciplinas amostrais não atingem este equilíbrio perfeito, mas quem as usa conscientemente deve saber qual é o tipo de representatividade que proporcionam.

No início do século XX Lord Rutherford emitiu a opinião "*Se a sua experiência precisar de Estatística, faça uma experiência melhor*".

Creio que, perante a complexidade crescente de tudo aquilo que nos rodeia e nos leva a indagar, um cientista do início do século XXI que se quisesse pronunciar sobre a questão diria antes "*A sua experiência precisa de Estatística, mas faça a experiência mais adequada e recolha os melhores dados que puder*". Por outras palavras, controle tudo o que puder controlar, o que não puder controlar, aleatorize, e faça do acaso o seu melhor aliado.

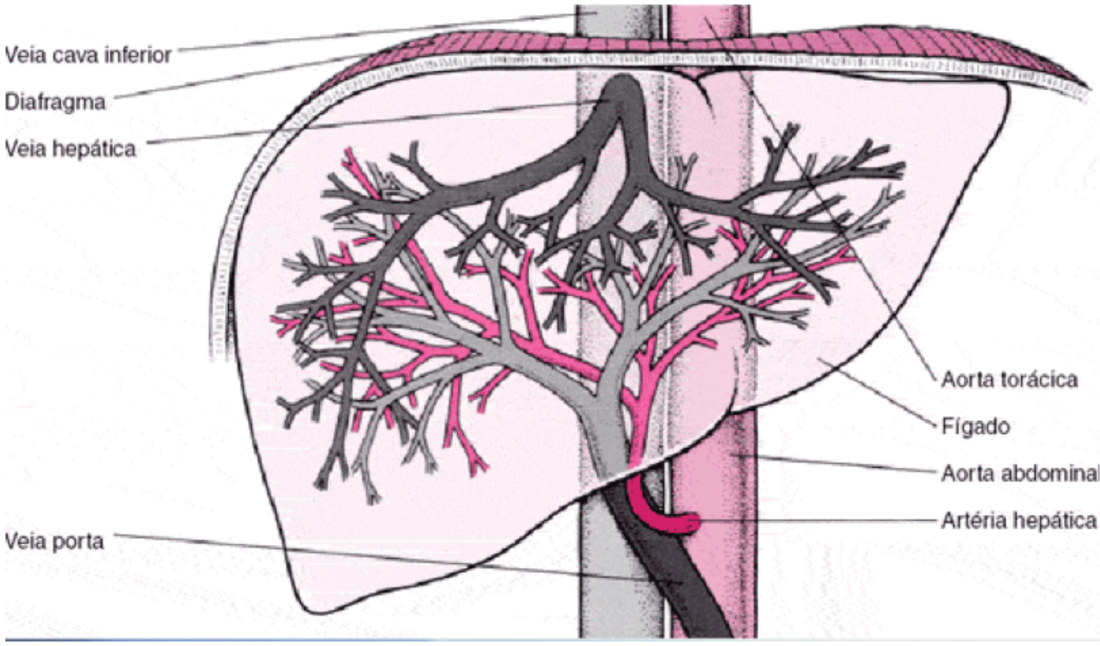
Sabemos hoje que não podemos observar sem modificar o que observamos. No entanto, em muitas situações esperamos que essa modificação seja irrelevante, e a amostragem é o paradigma de recolha de dados em *estudos observacionais*.

Mas a partir dos anos 30, Ronald Fisher estabeleceu uma nova área da Estatística, o Planeamento de Experiências, que tem consequências importantes na aquisição de dados. No *estudos experimentais* uma parte (pelo menos) da população é modificada, para se comparar o que nesse subconjunto é observado com o que acontece num outro subconjunto de controle. Tal como "amostras de conveniência"não são amostras no sentido estrito que deve prevalecer na construção da ciência, "experiências sem controle"não são admissíveis, porque estudam efeitos que tanto podem resultar do "tratamento"como da intervenção do acaso. Para os estatísticos a comparação com um "grau zero"do conhecimento (num sentido inspirado no "grau zero da escrita", de Barthes), muitas vezes cotejando a resposta a um tratamento com a resposta à administração de um placebo, e a comparação de diversos tratamentos entre si e com um placebo, são um padrão (quase um vício, quando se pensa no estatístico que quando lhe perguntaram "*A sua mulher está boa?*"retorquiu "*comparada com quem?*".)

De facto, qualquer investigação sem controle pode levar a conclusões muito apelativas, mas efectivamente descabidas. é célebre a experiência de ligação directa (*shunt*) da veia porta à veia cava, chamada operação porta-cava, sem controle, que levou a tantas intervenções cirúrgicas destinadas a "tratar"doentes com cirrose hepática, e que foi aliás propagandeada pelos intervencionados, que durante algum tempo sentiam melhoras.

Porém, esta operação obviamente impedia metabolização adequada de nutrientes, e a sobrevivência e qualidade de vida dos doentes eram dramaticamente diminuídas, a tal ponto que os médicos portugueses que participaram na avaliação retrospectiva chamavam *operação porta-à-cova* a esta operação porta-cava (a história foi-me contada pelo Professor Nuno Grande, que pertenceu a esse grupo).

Sir Ronald Fisher foi, sem dúvida, o estatístico mais importante, introduzindo conceitos e desenvolvendo resultados que provocaram saltos qualitativos enormes na estatística teórica, e consequentemente no uso sofisticado e eficiente da Estatística na análise de dados. Os seus livros *The Design of Experiments* (1935) e *Statistical Methods and Scientific Inference* (1956) são, também, reflexão profunda sobre os



caminhos da Ciência.

Muitos consideram que o seu livro *Statistical Methods for Research Workers* (1925, com mais do que uma dúzia de edições alargadas com novos desenvolvimentos), pela forma pragmática como expõe metodologias adequadas para lidar com dados, foi o livro mais influente na aceitação da Estatística como instrumento indispensável na indagação científica, em todas as áreas experimentais. Foi esse livro que definitivamente divulgou a *ANalysis Of VAriance* (ANOVA), que, para usar uma notável apreciação de Gilbert, é a técnica estatística "*mais usada, e mais mal usada*"na análise dos dados.

De facto, a ANOVA é o instrumento que nos permite comparar o efeito médio de diversos tratamentos. Antes desta notável abordagem de Fisher, havia o hábito de comparar os efeitos médios de tratamentos dois a dois, usando desde 1908 o teste de 'Student'. Ora Fisher notou que se tomarmos decisões correndo um risco de 5% de errar, quando fazemos isso repetidamente o risco de errar cresce, atingindo valores reais muito superiores aos riscos nominais indicados (repare-se que admitindo errar 5% das vezes — uma vez em cada vinte, em média — se compararmos catorze tratamentos, dois a dois, estamos a testar $\binom{14}{2} = 91$, e é de esperar tomar a decisão errada cerca de quatro ou cinco vezes, e nem sabemos em quais das 91 comparações; por outras palavras, o risco de erro na comparação global é quase 100%).

Desde os seus passos iniciais, usa-se em Probabilidade uma forma de calcular uma probabilidade como média ponderada: se particionarmos o universo em acontecimentos $A_k,\ k \in \mathbb{K}$, podemos calcular a probabilidade de qualquer acontecimento B como $\mathbb{P}(B) = \sum_{k\in\mathbb{K}} \mathbb{P}(B|A_k)\,\mathbb{P}(A_k)$. Este "teorema da probabilidade total", que em certo sentido traduz em linguagem matemática o método cartesiano de resolver um problema complexo particionando-o em subproblemas simples, resolvendo cada um deles, e globalizando depois para resolver o problema complexo que nos ocupa, é um recurso óptimo em inúmeras situações, e é a base de estratégias tais como a amostragem estratificada. Em certo sentido, pode dizer-se que a ANOVA de Fisher é uma primeira constatação de que o método de Descartes nem sempre é adequado, e que a complexidade nem sempre pode ser tratada através da resolução de subproblemas parciais.

Falsear hipóteses

Fisher desenvolveu também "testes de significância", em que se avaliava a probabilidade de se obter um resultado tão mau como ou pior do que o observado. Teve no entanto o desgosto de a teoria de os "testes de hipóteses", em que uma hipótese nula é avaliada em contraponto com uma alternativa(9), ter sido uma criação de Neyman e Egon Pearson (o filho de Karl Pearson). Os testes de hipóteses são um instrumento incontornável em Estatística, e em particular no seu serviço ao progresso científico. A ideia de base é que conhecimento científico é o que não é falseado pelos factos, pelo que quando se quer estabelecer H_A se contrapõe esta hipótese alternativa com uma "hipótese nula" H_0, sendo em geral o objectivo "falsear" H_0.

A notável resposta de Linus Pauling quando um entrevistador lhe perguntou como se conseguia um prémio Nobel, "*é preciso ter muitas ideias, e a coragem de deitar quase todas fora*", ilustra a importância da Lei dos Grandes Números e o recurso aos testes de hipóteses para rejeitar hipóteses que são provavelmente falsas.

Abusos da Estatística e Ciência da Treta

Anualmente um grupo de foliões de Harvard atribui prémios IgNobil a resultados que fazem rir (embora actualmente os mentores do evento reconheçam que alguns deles fazem, numa segunda abordagem, pensar). Em muitas circunstâncias, esses prémios chamam a atenção para a falta de preparação metodológica que é evidente muita ciência da treta que vai sendo publicada. Por exemplo, há alguns anos o prémio IgNobil da Química foi partilhado por dois trabalhos, um deles mostrando que a coca-cola é um espermicida, outro demonstrando que não é um espermicida, numa clara ignorância que a *repetibilidade* é uma das pedras angulares da Ciência. Trabalhos com um aparato grandioso — por exemplo, imagens médicas "demons-

(9) A forma como se realizam os testes de hipóteses — tentar falsear o que se admite inicialmente ser verdadeiro indagando a que ponto uma função dos dados observados (estatística de teste) é improvável face ao observado — é curiosamente similar ao funcionamento da justiça: a hipótese nula de que qualquer um é inocente até prova em contrário é o "preconceito" de base, que peritagens, testemunhos, etc., põem em causa; se esta hipótese nula for rejeitada, a hipótese alternativa de culpabilidade prevalece.

trando"que a actividade cerebral é estimulada pela respiração apenas por uma das narinas ("*uninostril breathing*"), parecem irrelevantes quando se atenta na diminuta amostra, constituída usando voluntários, e sem atender ao confundimento (será que a maior actividade cerebral que os autores reclamam para as ocasiões em que se faz respiração unilateral se deve a estimulação do raciocínio? ou ao simples esforço de atenção para respirar apenas por uma das narinas? — e isto para não referir que dificilmente se acredita que a capacidade de controlar esse respirar unilateral é a mesma em todos os indivíduos) — foram premiados, levando alguns autores ao rubro da indignação, o que é estranho, pois um dos meus desejos mais secretos e inconfessados é ser galardoado com um prémio IgNobil.

O mesmo grupo publica o *Annals of Improbable Research*, onde aparecem caricaturas da ciência da treta, como a resolução da famosa questão, *O que apareceu primeiro, a galinha ou o ovo?*

Este mistério foi finalmente resolvido: escolheram-se 100 aviários ao acaso, e dois dos experimentadores fizeram telefonemas simultâneos, encomendando por via postal um deles uma galinha, devidamente assassinada, depenada, esventrada, e outro encomendando uma dúzia de ovos, a serem enviados por correio para uma dada morada. E a Estatística não deixa qualquer dúvida: em 97% dos casos a dúzia de ovos apareceu primeiro!

A Ciência é considerada por muitos economistas o empreendimento humano de maior retorno. Mas, porventura devido ao mito americano "*publish or perish*"cada vez se publicam mais tretas, e há grupos organizados que põem em causa a integridade, veja-se por exemplo o notável editorial *Integrity Under Attack:The State of Scholarly Publishing*, de D. N. Arnold (`http://www.ima.umn.edu/~arnold//siam-columns/\\integrity-under-attack.pdf`). Artigos como *Why Most Published Research Findings Are False*, de John P. A. Ioannidis (`http://www.plosmedicine.org/article/info:doi/10.1371/\\journal.pmed.0020124` alertam para o descrédito em que a Ciência pode cair por esta ânsia de publicação incentivar o desbrochar de muita treta, resultado de má preparação, ingenuidade, incompetência, ou mesmo competência em cometer fraudes científicas. O fascínio de com *software* se conseguir sempre uma resposta é outra das causas de muitos pensarem que sabem usar Estatística, quando o que estão a fazer efectivamente é abusar da Estatística.

E o muito que fica por dizer ....

Claro que muito mais fica por dizer do que aquilo que é dito. A intervenção da Probabilidade e da Estatística na construção da Ciência, e influenciando a reflexão sobre probabilidade e certeza, e genericamente sobre o conhecimento, é um tema virtualmente inesgotável. Por outro lado, a Filosofia da Ciência é uma alavanca para o progresso da Probabilidade e da Estatística, como a de qualquer outra ciência.

Entre o que ficou por dizer, selecciono para breve menção dois assuntos:

- Teorema da Probabilidade Inversa, de Thomas Bayes, a probabilidade ao serviço da causalidade.

  A "inversão"do condicionamento, $\mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$, foi um controverso instrumento de indagação das probabilidades das causas, uma vez que associado ao Teorema da Probabilidade Total permite calcular a probabilidade das possíveis causas A_k de um efeito B:

  $$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\,\mathbb{P}(A_j)}{\sum_{k\in\mathbb{K}} \mathbb{P}(B|A_k)\,\mathbb{P}(A_k)}$$

  se as possíveis "causas"$A_k,\ k \in \mathbb{K}$, forem uma partição do universo em acontecimentos.

  Sendo matematicamente válido, é usado em muitas situações em que a sua aplicação é controversa — a tal ponto que antes do formalismo de Kolmogoroff foi repetidamente posto em causa, por exemplo por Maynard Keynes. A causalidade é uma indagação sempre renovada, veja-se por exemplo a monografia de Pearl (2000). E, em última instância, a questão: "quando é que uma probabilidade é suficientemente elevada para promovermos uma relação altamente provável ao estatuto de lei (certeza)?"coloca a questão prévia do que é a probabilidade, questão inesgotável, que os interessados podem explorar lendo "passo e apontado"o livro de Barnett (1999) ou, mais recentemente, no notável livro póstumos, inacabado — mas não é por isso que a é inconclusivo — de Jaynes (2003).

- Lei dos Grandes Números, Teorema da Transformação Uniformizante, e o Método de Monte Carlo.

  Ulam, quando trabalhou no projecto Manhattan, viu-se confrontado com a necessidade de cálculo rápido de expressões analiticamente muito difíceis (ou mesmo impossíveis) de calcular. Teve a ideia notável de usar a Lei dos Grandes Números, de convergência para a média, e as novas possibilidades de simulação de milhares de resultados pelo computador, para criar um método de cálculo de soluções aproximadas, com o grau de aproximação que se pretender (basta usar amostras simuladas tão grandes quanto necessário). Um dos seus colaboradores propôs o nome "Método de Monte Carlo", que von Neuman apadrinhou.

  Qualquer computador gera de forma simples números pseudo-aleatórios (deterministas, mas que imitam bem a intervenção do acaso) uniformes em (0,1). Como o teorema da transformação uniformizante estabelece que a aplicação da função de distribuição F_X à variável aleatória X a transforma numa uniforme padrão, isto é, $F_X(X) = Y \frown Uniforme(0,1)$, basta usar a função inversa (eventualmente inversa generalizada) para se obter simplesmente conjuntos de pseudo-aleatórios com a distribuição pretendida. O método da transformação uniformizante foi o primeiro a ser usado, mas nem sempre é expedito, e têm sido inventados inúmeros algoritmos para gerar pseudo-aleatórios apropriados para a simulação que se pretende efectuar.

O êxito da Estatística Computacional ocasionou novos processos de reamostragem, que oferecem interessantes problemas para reflexão filosófica, nomeadamente se informação pleonasticamente acumulada usando como base uma determinada quantidade de informação pode ser considerada informação aumentada. Mas o seu êxito prático parece um dado adquirido.

À guisa de conclusão

Finalmente, parece-me interessante anotar que a evolução da capacidade técnica de lidar com a aquisição e análise de dados foi alterando qualitativamente a Estatística Matemática, uma disciplina que apenas nos fins do século XIX começou a ganhar ímpeto. Até fins do primeiro quartel do século XX, o modelo "normal"foi dominante, pois era possível obter resultados exactos com amostras de dimensão diminuta, e o Teorema Limite Central justificava o seu uso como aproximação num vasto leque de situações. A crítica ao uso imoderado — e por vezes mal justificado — desse modelo criou uma abordagem radical não-paramétrica, em que não se assumia explicitamente nenhum modelo; na mesma altura, as ideias de Fisher sobre planeamento experimental ganhavam cada vez mais adeptos.

Tukey exprimia claramente que a inteligência natural é nobre e rara, pelo que se deve usar inteligência artificial sempre que esta baste. De facto, o desenvolvimento da computação proporcionou novas capacidades na análise dos dados, sendo possível uma "análise exploratória", por contraposição às clássicas estatística descritiva e estatística confirmatória, em que se procura deixar os dados guiarem a análise. Foi o início de uma vasta área de estatística computacional, promovendo experimentação com simulação, técnicas de reamostragem, que se pode caricaturalmente descrever como o milagre da multiplicação dos pães e dos peixes em Estatística, pois em vez de resultados exactos com pequenas amostras assumindo normalidade, o que se assume é que se pode usar a distribuição empírica associada a dados escassos para gerar quantos quisermos do mesmo tipo.

Por outro lado, as possibilidades de aquisição automática de amostras enormes, trouxe novos problemas, pois amostras muito grandes trazem muita ganga, e nelas é necessário minerar o escasso ouro entre tanta lama. As novas áreas de pesquisa estocástica e *data mining* são a resposta a este problema de excesso de informação, com muito ruído e pouco sinal, para usar imagens da linguagem cara aos engenheiros.

Quando fiz o meu doutoramento, telefonava a dar notícias à família, o que implicava deslocar-me aos correios com algumas horas de antecedência para encomendar a chamada, o que deve parecer incrível a pessoas com menos uma vintena de anos do que eu. As telecomunicações revolucionaram a nossa vida, e naturalmente também a Estatíistica. A comunicação entre grupos de investigadores, a necessidade de colaborar, a vantagem de combinar amostras ou a análise de dados de fontes diversas, originou a nova área de Meta-Análise, que começa a ser um padrão em muita publicação nas áreas de saúde e farmacologia.

Todas estas alterações trazem novas necessidades de reflectir sobre os fundamentos da ciência, e sobre os limites de criar conhecimento. Não serão os filósofos da ciência o motor da mudança, mas por muito questionarem muito inspiram, e fertilizam o espírito, preparando-o para novas culturas.

Referências

- Arbuthnott, J. (1710). "An argument for divine providence, taken from the constant regularity observed in the birth of both sexes", *Philosophical Transactions* **27**, 186–90; reeditado em Kendall, M. G., and Plackett, R. L. (1977). *Studies in the History of Statistics and Probability*, Griffin, London.

- Barndorff-Nielsen, O., and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London and New York.

- Barnett, V. (1999). *Comparative Statistical Inference*, Wiley, New York.

- Bayes, T. (1763). An Essay Towards Solving a Problem in the Doctrine of Chances, reeditado em *Biometrika* **45**, 296–315.

- Bernoulli, J.(1713). *Ars Conjectandi*, reedição moderna (Editions Culture et Civilisation, 1968, ou Editions Jacques Gabay, Paris, apenas a primeira das quatro partes da obra), e tradução inglesa preparada por Bing Sung, Harvard University, Cambridge, MA, 1966.

- Bernstein, P. L. (1996). *Against the Gods, The Remarkable Story of Risk*, Wiley, New York.

- Daniels, H. E. (1954). Saddlepoint approximations in Statistics, *Ann. Math. Statist.* **25**, 631–650.

- David, H. A., and Edwards, A. W. F. (2001). *Annotated Readings in the History of Statistics*, Springer, New York.

- Erdös, P., and Rényi, A. (1959). On a central limit theorem for samples from a finite population, *Publ. Math. Instit. Hungar. Acad. Sci.* Vol. 4, p. 49–61.

- de Moivre, A. (1712) "De Mensura Sortis", *Phil. Trans.* **27**, 213–264, que foi traduzido em inglês e publicado na *Intern. Statist. Rev.* **52** (1984), 237–262.

- de Moivre, A. (1718). *The Doctrine of Chances*, edição moderna da terceira edição, revista e aumentada, de 1756, Chelsea, New York, 1967. A *American Mathematical Society* tem também edição inglesa moderna desta obra, bem como as Editions Jacques Gabay, Paris. Mais barato: `http://www.ibiblio.org/chance/`.

- Fisher, R. A. (1995). *Statistical Methods, Experimental Design and Scientific Inference* (re-issue of *Statistical Methods for Research Workers*, *The Design of Experiments*, and *Statistical Methods and Scientific Inference*), Oxford Univ. Press, Oxford.

- Galton, F. (1989). *Natural Inheritance*, MacMillan, London, acessível em `http://www.mugu.com/galton/books/natural-inheritance/pdf/galton-nat-inh-1up-clean.pdf`

- Gilbert, N. (1989). *Biometrical Interpretation Ñ Making Sense of Statistics in Biology*, 2nd ed., Oxford Univ. Press, Oxford.

- Greenfield, T. (2002). *Research Methods. Guidance for Postgraduates*, 2nd ed., Arnold, London.

- Ioannidis, J. P. A. (2005). Why most published research findings are false, *PLoS Med.* **2**(8):e124, 696–701.

- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*, Cambridge Univiversity Press, Cambridge.

- Laplace (1812) *Théorie Analytique des Probabilités*, reedição recente das Editions Jacques Gabay, Paris. A segunda edição, de 1814, contém um extenso e célebre prefácio, que também teve tiragem à parte, *Essai Phylosophique sur les Probabilités*, tradução inglesa em `http://books.google.pt/books?id=vDZzuGcM4DUC&printsec=frontcover&dq=Laplace+Essai+Philosophique+sur+les+probabilits&hl=en&sa=X&ei=QFWWUajOBpKoOAG2pIGQAg&ved=0CDQQ6AEwAQ#v=onepage&q=Laplace\%20Essai%20Philosophique%20sur%20les%20probabilits&f=false`

- Le Cam, L. (1986). The central limit theorem around 1935, *Statistical Science* Vol. 1, p. 78–96.

- Pearl, J. (2000). Causality — Models, Reasoning, and Inference, Cambridge University Press, Cambridge.

- Pearson, K. (1896). *The Grammar of Science*, reedição moderna em 2012, pela módica quantia de $12.82, da Forgotten Books.

- Pólya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Math. Zeit.* **8**, 171–181.

- Steven, S. S. (1946). On the theory of scales of measurement, *Science* **103**, 677–680.

- Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass.