



THOMAS H. DAVENPORT

Autor do best-seller *Competição analítica*

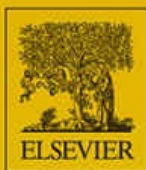
com Jin Ho-Kim

# DADOS DEMAIS!

COMO DESENVOLVER  
HABILIDADES ANALÍTICAS  
PARA RESOLVER  
PROBLEMAS COMPLEXOS,  
REDUZIR RISCOS E  
DECIDIR MELHOR

- ▶ O valor da análise quantitativa para converter dados em decisão
- ▶ BIG DATA: extraindo valor de imensas quantidades de dados desestruturados
- ▶ O pensamento analítico e a lógica quantitativa





THOMAS H. DAVENPORT  
Autor do best-seller *Competição analítica*  
com Jin Ho-Kim

# DADOS DEMAIS!

COMO DESENVOLVER  
HABILIDADES ANALÍTICAS  
PARA RESOLVER  
PROBLEMAS COMPLEXOS,  
REDUZIR RISCOS E  
DECIDIR MELHOR

- ▶ O valor da análise quantitativa para converter dados em decisão
- ▶ BIG DATA: extraíndo valor de imensas quantidades de dados desestruturados
- ▶ O pensamento analítico e a lógica quantitativa

  
CAMPUS

THOMAS H. DAVENPORT  
com Jinho Kim

# DADOS DEMAIS!

COMO DESENVOLVER  
HABILIDADES ANALÍTICAS  
PARA RESOLVER  
PROBLEMAS COMPLEXOS,  
REDUZIR RISCOS E  
DECIDIR MELHOR

Tradução: Afonso Celso da Cunha



Do original: *Keeping Up With the Quants*

Tradução autorizada do idioma inglês da edição publicada por Harvard Business Review Press  
Copyright © 2013, by Harvard Business School Publishing Corporation

© 2014, Elsevier Editora Ltda.

Todos os direitos reservados e protegidos pela Lei no 9.610, de 19/02/1998.

Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.

*Copidesque*: Cynthia dos Santos Borges

*Revisão*: Tássia Hallais Veríssimo

*Conversão para e-book*: Freitas Bastos

Elsevier Editora Ltda.

Conhecimento sem Fronteiras

Rua Sete de Setembro, 111 – 16º andar

20050-006 – Centro – Rio de Janeiro – RJ – Brasil

Rua Quintana, 753 – 8º andar

04569-011 – Brooklin – São Paulo – SP – Brasil

Serviço de Atendimento ao Cliente

0800-0265340

[atendimento1@elsevier.com](mailto:atendimento1@elsevier.com)

ISBN 978-85-352-6401-2

ISBN (versão digital): 978-85-352-7453-0

Edição original: ISBN: 978-1-4221-8725-8

**Nota:** Muito zelo e técnica foram empregados na edição desta obra. No entanto, podem ocorrer erros de digitação, impressão ou dúvida conceitual. Em qualquer das hipóteses, solicitamos a comunicação ao nosso Serviço de Atendimento ao Cliente, para que possamos esclarecer ou encaminhar a questão.

Nem a editora nem o autor assumem qualquer responsabilidade por eventuais danos ou perdas a pessoas ou bens, originados do uso desta publicação.

CIP-Brasil. Catalogação na Publicação  
Sindicato Nacional dos Editores de Livros, RJ

---

D269d Davenport, Thomas

Dados demais!: como desenvolver habilidades analíticas para resolver problemas complexos, reduzir riscos e decidir melhor. / Thomas Davenport, Jinho Kim; tradução Afonso Celso da Cunha. – 1. ed. – Rio de Janeiro: Elsevier, 2014.

240 p. : il.; 23 cm.

Tradução de: Keeping up with the quants: your guide to understanding using analytics

Inclui índice

ISBN 978-85-352-6401-2

1. Tecnologia da informação – Administração. 2. Administração de empresas. 3. Planejamento estratégico. 4. Processo decisório. I. Kim, Jinho. II. Cunha, Afonso Celsa da. III. Título.

13-06364

CDD: 658.4038

CDU: 65.012.4

---

# Agradecimentos

Tom agradece (e dedica sua parte do livro) à esposa adorada de há mais de 30 anos, Joan Powel Davenport. Os filhos Hayes e Chase já deixaram o ninho, mas ainda apoiam o trabalho do pai à distância. A Divisão de Pesquisas da Harvard Business School foi menos amorosa, mas também forneceu importante suporte em pesquisa.

Jinho agradece a John e Bonnie Ries pelo apoio e pelo estímulo entusiástico, inclusive pela leitura do manuscrito original. Também expressa gratidão especial a Jung-Hwa Shin, por motivá-lo e sempre apoiá-lo na elaboração desta obra. Jinho ainda é grato à filha mais velha, Nuri, e a seu genro, Dong-Wook, assim como à filha caçula, Youngri, pelo amor e pela inspiração. Ele dedica sua parcela do livro à mãe, Tae-Hwa Cho, pelo amor e sabedoria.

Tom e Jinho agradecem a Melinda Merino pelo esplêndido trabalho de edição na Harvard Business Review Press, assim como a todo o staff talentoso que faz da HBRP a melhor editora de livro de negócios do mercado. Obrigado, também, a Mark Allen por produzir as ilustrações que tornam esta obra tão mais atraente. Também gostaríamos de transmitir nossa gratidão aos quants desbravadores e aos executivos com índole quantitativa, vivos ou mortos, que encenaram os exemplos deste livro. Somos apenas os cronistas de suas explorações épicas.

## Sobre os autores

**THOMAS H. DAVENPORT** é professor visitante da Harvard Business School, President's Distinguished Professor de Tecnologia da Informação e de Administração na Babson College e Research Fellow do MIT Center for Digital Business. Também é assessor sênior da Deloitte Analytics e cofundador e diretor de pesquisa do International Institute for Analytics. Davenport é coautor de *Competing on Analytics* e *Analytics at Work*. Este é o décimo sétimo livro de que é autor, coautor ou editor.

**JINHO KIM** é professor de Negócios e Estatística na Korea National Defense University e diretor de pesquisa do KNDU Lab for Analytics Research. Tem PhD pela Wharton School e é autor de seis livros publicados na Coreia, inclusive o best-seller *100 Common Senses in Statistics* e *Freak Statistics*. Kim desenvolveu e dirige um programa educacional de desenvolvimento da capacidade analítica. Suas atuais pesquisas se concentram no uso de métodos analíticos para tratar de várias questões empresariais e sociais.

# Sumário

*Agradecimentos*

*Sobre os autores*

- 1** Por que todos precisam de capacidade analítica
- 2** Formulação do problema
- 3** Solução do problema
- 4** Resultados: Comunicação e ação
- 5** Análise quantitativa e criatividade
- 6** Desenvolvimento de recursos de análise quantitativa
- 7** Trabalhando com os quants

*Notas*



# Por que todos precisam de capacidade analítica

Vivemos em um mundo inundado de dados. Eles se proliferam com velocidade espantosa – dispomos de cada vez mais dados, e grande parte deles foi coletada a fim de melhorar as decisões no setor privado, no setor público e na sociedade em geral. Se não formos capazes de explorar esses dados para melhorar a tomada de decisões, por meio de análises quantitativas, não só os desperdiçaremos, mas também, com grande probabilidade, apresentaremos desempenho aquém do ótimo. Portanto, nosso objetivo neste livro é mostrar como funciona a análise quantitativa – mesmo que você não tenha formação quantitativa – e como usá-la para tomar decisões mais eficazes.

## A ascensão da ciência analítica e do *Big Data*

A ascensão do *Big Data* está ocorrendo em praticamente todas as áreas da sociedade. Quem atua em esportes decerto conhece o termo *moneyball*, a transformação do beisebol profissional – e, agora, de quase todos os esportes importantes – mediante a aplicação da *ciência analítica*, ou simplesmente *analítica*, ao *Big Data*, ou grandes volumes de dados. Quem gosta de jogos on-line decerto já se deu conta de que todos os aspectos de seu comportamento durante o jogo estão sendo captados e analisados por empresas como Zynga e Electronic Arts. E os fãs de cinema já tomaram conhecimento dos algoritmos usados pela Netflix para prever os filmes preferidos pelos clientes. Talvez não saibam, porém, que algumas produtoras de Hollywood, como Relativity Media, empregam algoritmos para escolher os filmes a serem financiados.

Evidentemente, há diferentes tipos de dados. Alguns são coletados e gerenciados para fins transacionais – por exemplo, sua empresa ou organização mantém sob controle os atributos de seu contrato de trabalho, como data de admissão, evolução salarial e situação de férias. Com o passar do tempo, contudo, as organizações passam a acumular grandes volumes de dados, a serem processados e interpretados para que se tornem úteis nas decisões. Os dados sobre transações de recursos humanos precisam ser interpretados com base na analítica. As organizações podem fazer perguntas como: “Quantos empregados tendem a aposentar-se no próximo ano?” ou “Há alguma relação entre duração das férias e escore na avaliação do desempenho?”

Os dados e a analítica, entretanto, não se limitam a melhorar o processo decisório interno. Muitas empresas de Internet – Google, Facebook, Amazon, eBay e outras – exploram o denominado *Big Data* de transações on-line não só para tomar decisões, mas também para lançar novos produtos e para oferecer novos recursos aos clientes. Qualquer que seja o propósito – aprimorar a tomada de decisões ou criar valor para os clientes – é aqui que entra a analítica – resume dados, encontra significados e identifica padrões. Interpretar os dados e deles extrair valor exigem análise matemática ou estatística dos dados – em outras palavras, *ciência analítica* ou simplesmente *analítica*. (ver “O que é analítica?”)

## O que é analítica?

*Analítica* significa o uso amplo de dados, de análise estatística e quantitativa, de modelos explanatórios e preditivos e de gestão fatural para orientar decisões e agregar valor.<sup>a</sup>

A analítica pode ser classificada como descritiva, preditiva e prescritiva, de acordo com seus métodos e propósitos. A *analítica descritiva* envolve coleta, organização, tabulação e apresentação de dados, para a exposição das características do que está sendo estudado. Esse tipo de analítica, historicamente, tem sido denominada *reporting* ou *elaboração de relatórios*. Trata-se de algo que pode ser muito útil, mas não explica os resultados ou as ocorrências nem indica o que pode acontecer no futuro.

A *analítica preditiva* vai além da mera descrição das características dos dados e das relações entre variáveis (fatores que podem assumir uma faixa de valores diferentes); usa dados do passado para prever o futuro. Primeiro identifica as associações entre as variáveis e depois prevê a probabilidade de um fenômeno – como, por exemplo, que um cliente reagirá de maneira positiva à propaganda de um produto e o comprará – com base nas relações identificadas. Embora as associações entre variáveis sejam exploradas para propósitos preditivos, não se supõe nenhuma relação de causa e efeito na analítica preditiva. De fato, a constatação de relações causais nem sempre é necessária para fazer previsões exatas.

A *analítica prescritiva*, incluindo métodos como projeto experimental e otimização, se estende ainda mais. Da mesma maneira como a receita de um médico, a analítica

prescritiva sugere um curso de ação. O *projeto experimental* tenta responder às perguntas sobre *por que* algo aconteceu, por meio de experimentos. Para fazer inferências causais com base em pesquisas de causa e efeito, os pesquisadores devem manipular uma ou mais variáveis independentes e efetivamente controlar outras variáveis exógenas. Se o grupo de teste – o que preenche a condição experimental – apresentar desempenho substancialmente melhor que o do grupo de controle – o decisor pode aplicar a condição em âmbito mais amplo.

A *otimização*, outra técnica prescritiva, tenta descobrir o nível ideal de determinada variável em suas relações com outra, como, por exemplo, determinar o preço de um produto tendente a oferecer a mais alta rentabilidade. Do mesmo modo, os métodos de otimização podem identificar os níveis de estoques em que menores seriam as probabilidades de faltas ou de sobras numa organização de varejo.

A analítica pode ser classificada como qualitativa ou quantitativa, de acordo com os processos adotados e os tipos de dados coletados e analisados. A *análise qualitativa* almeja promover a compreensão profunda das razões e motivações subjacentes a um fenômeno. Em geral, coletam-se dados não estruturados de um pequeno número de casos não representativos, que são analisados sob abordagem não estatística. A analítica qualitativa quase sempre oferece ferramentas úteis para a pesquisa exploratória – a primeira fase do processo analítico. A *analítica quantitativa* se presta à investigação empírica sistemática de um fenômeno, por meio de técnicas estatísticas, matemáticas ou computacionais. Coletam-se e analisam-se estatisticamente dados estruturados de um grande número de casos representativos.

Várias são as ferramentas de analítica que atendem a diferentes propósitos dos pesquisadores:

- **Estatística:** Ciência da coleta, organização, análise, interpretação e apresentação de dados.
- **Previsão:** Estimativa do valor de uma variável, em algum tempo futuro, com base em dados do passado.
- **Mineração de dados:** Identificação automática ou semiautomática de padrões interessantes, até então desconhecidos, em grandes massas de dados, por meio do uso de algoritmos de computação e de técnicas estatísticas.
- **Mineração de textos:** Processo de extrair padrões e tendências de um texto, à semelhança da mineração de dados.
- **Otimização:** Uso de técnicas matemáticas na busca de soluções ótimas, observando determinados critérios e satisfazendo a certas restrições.
- **Projeto experimental:** Estudo de grupos de teste e de grupos de controle, com a atribuição aleatória de temas ou casos para cada grupo, a fim de identificar relações de causa e efeito em determinado resultado.

Embora a lista acima apresente várias abordagens analíticas de uso comum, constata-se considerável imbricação no uso das diferentes técnicas. Por exemplo, a *análise de regressão*, talvez a técnica mais comum em analítica preditiva, também é muito usada em estatística, em previsão e em mineração de dados. Do mesmo modo, a *análise de séries temporais*, ferramenta estatística específica para estudar dados que variam ao longo do tempo, é adotada com muita frequência em estatística e em previsão.

a. Thomas Davenport e Jeanne G. Harris. Competição analítica: vencendo através da

O tipo de dados transacionais mencionado acima para decisões sobre recursos humanos é estruturado (captado com facilidade em linhas e colunas), quantitativo, e em volumes relativamente pequenos (1 ou 2 terabytes, mesmo em grandes empresas). Como se trata de um contexto tradicional para a analítica, vamos chamá-lo de *Small Data*. Essa já foi a única opção para a analítica.

Hoje, contudo, grandes empresas, organizações sem fins lucrativos e pequenas startups estão empolgadas com o *Big Data* – dados não estruturados, em grandes volumes. Podem originar-se de discussões on-line na internet, de tomadas em câmeras de vídeo ou de análises de DNA de um grupo de pacientes em um hospital. Esses tipos de dados tendem a aglomerar-se em volumes muito maiores – de vários petabytes ( $10^{15}$  bytes) ou multipetabytes. Por exemplo, a Google processa cerca de 24 petabytes de dados de Internet por dia, e a AT&T transfere cerca de 30 petabytes em telecomunicações de voz e dados por dia. Com o uso de novas tecnologias de hardware e de software, agora podemos analisar enormes volumes de dados e interpretar seu significado. (Ver “O que é *Big Data*?”)

## O que é *Big Data*?

O termo *Big Data* denota volumes de dados inusitadamente grandes, ou tipos de dados não estruturados. Alguns exemplos incluem:

- Trinta bilhões de unidades de conteúdo foram acrescentadas ao Facebook este mês, por mais de 600 milhões de usuários.
- A empresa de jogos sociais Zynga processa mais de 1 petabyte de dados de jogos por dia.
- Os usuários do YouTube veem mais de 2 bilhões de vídeos por dia.
- Os usuários do Twitter executam mais de 32 bilhões de buscas por mês.
- Os usuários do Google realizaram mais de 5 bilhões de buscas por dia em 2011.
- Mais de 2,5 bilhões de mensagens de texto foram enviadas por dia em 2009.
- Cinco bilhões de telefones móveis foram usados em todo o mundo em 2010.

- A análise completa de um genoma humano envolve cerca de 1 terabyte de dados.
- Um sensor sem fio que avalia a saúde de uma única vaca transmite 200 megabytes de dados por ano.
- Em 2008, a quantidade de dispositivos ligados à internet ultrapassou o número de pessoas no planeta.
- A Cisco Systems estimou que, no fim de 2011, 20 domicílios típicos geraram mais tráfego na internet que todos os usuários de internet em 2008.
- A McKinsey & Company estima que, em quase todos os setores da economia dos Estados Unidos, empresas com mais de mil empregados armazenam, em média, mais informações que a Biblioteca do Congresso dos Estados Unidos.

O *Big Data* e a analítica nele baseada prometem mudar praticamente todos os setores de atividade e todas as funções de negócios nos próximos 10 anos. Qualquer organização e, nela, qualquer indivíduo que partir na frente na análise do *Big Data* conquistará vantagem competitiva importante. Da mesma maneira como os primeiros concorrentes analíticos na era do *Small Data* se anteciparam aos demais e construíram vantagem competitiva vultosa, chegou a hora de empresas e organizações em geral explorarem as oportunidades do *Big Data*.

A realização do vasto potencial do *Big Data* é viabilizada pela ubiquidade dos dispositivos de coleta e processamento de dados. O fato é que os sensores e os microprocessadores em breve serão onipresentes. Praticamente todos os dispositivos mecânicos ou eletrônicos podem deixar rastros que descrevem seu desempenho, localização e situação. Esses dispositivos e seus usuários se comunicam por meio da internet – o que gera outra fonte profusa de dados. Quando se combinam todos esses bits com os de outras mídias – telefonia com e sem fio, cabo, satélite e assim por diante – o futuro dos dados parece ainda mais amplo.

A disponibilidade de todos esses dados significa que praticamente todas as atividades empresariais ou organizacionais podem ser consideradas problemas ou iniciativas de *Big Data*. A indústria, em que a maioria das máquinas já tem um ou mais microprocessadores, é cada vez mais ambiente de *Big Data*. O marketing de consumo, com miríades de pontos de contato e fluxos de cliques com os clientes, já é um problema de *Big Data*. A Google até já descreveu seu carro autodirigido como projeto de *Big Data*.

CEOs como Gary Loveman, da Caesars Entertainment (conhecido pela pergunta “Pensamos ou conhecemos?”); Jeff Bezos, da Amazon (“Nunca



descartamos dados”); e Reid Hoffman, da LinkedIn (a “Web 3.0 tem a ver com dados”) são exemplos notórios de que decidir com base na analítica é a chave do sucesso organizacional e da fortuna pessoal.

Todas as organizações em qualquer setor de atividade terão de interpretar a enxurrada de dados. Para tanto, precisarão de pessoas capazes de fazer análises minuciosas dessa massa de informações – indivíduos com os mais diversos nomes, mas todos chamados *quants*. Este livro não foi feito para *quants*. As mesmas organizações, todavia, também precisarão de pessoas aptas a decidir e a agir de maneira eficaz, com base nos resultados dessas análises minuciosas – e é para elas que estamos escrevendo – para usuários não analistas e não quantitativos, em diferentes organizações, que trabalharão e decidirão com o apoio do *Big Data* e à luz da analítica.

## O que esperamos que você extraia deste livro

Como autores, trazemos antecedentes um tanto diferentes para essa questão. Compartilhamos, porém, os objetivos comuns de expandir o uso do raciocínio analítico nas empresas e na sociedade, e, em especial, de ajudar os *não quants* a tornar-se melhores consumidores de dados. Tom, principalmente, não é quant sério – como sociólogo por formação acadêmica, as qualificações dele em estatística são relativamente modestas – mas ele é autoridade em analítica e em suas aplicações nos negócios. Já há uns 20 anos pesquisa, escreve, leciona e presta consultoria em organizações, sobre o desenvolvimento de recursos analíticos. A experiência dele em trabalhar com gestores, ajudando-os a usar a analítica em suas atividades profissionais, é o fundamento deste livro. Além disso, ele é o principal autor dos best-sellers *Competição analítica e Inteligência analítica nos negócios*, que se concentram em como as grandes empresas usam a analítica, na condição de fator estratégico. Neste livro, ele desloca o foco para como os indivíduos podem desenvolver capacidade e abordagem analítica.

Jinho, com formação acadêmica em negócios e em estatística, exerce com naturalidade o papel de quant convicto. As pesquisas dele versam sobre como usar métodos analíticos para tratar de várias questões empresariais e sociais. Além disso, desenvolveu e dirige um programa educativo para o desenvolvimento das qualificações analíticas. Jinho é professor de Gestão Empresarial e de Estatística na Coreia e autor de seis livros, inclusive 100

*Common Senses in Statistics* e *Freak Statistics*, com o objetivo de ajudar não estatísticos a “alfabetizar-se” em Estatística e a digerir informações estatísticas.

Este livro o converterá em melhor consumidor de dados e o tornará mais fluente em analítica. Também lhe facilitará não só o relacionamento com os quants, mas também a compreensão das análises deles. E ainda o preparará para falar no idioma da análise quantitativa e para fazer as perguntas certas. Talvez até o inspire a tornar-se quant!

## Consumidor esclarecido de analítica

Muitas são as maneiras de os gestores, trabalhando em estreito entrosamento com os analistas quantitativos em suas organizações, usar a analítica para melhorar suas decisões. Vamos ver uma escolha recente de Jennifer Joy, na Cigna – importante empresa de serviços médicos. Joy é vice-presidenta de operações clínicas, e dirige um grande call center da empresa. O call center trabalha com clientes, para melhorar a saúde e bem-estar deles, em especial se padecerem de alguma doença crônica, como diabetes ou cardiopatia, que exija tratamento contínuo. Jen é enfermeira por formação, não analista. Mas tem MBA e acredita na importância do raciocínio analítico. O exemplo dela mostra como a definição do problema e a formulação da pergunta certa – dois aspectos básicos do raciocínio analítico – a ajudaram a economizar dinheiro para a empresa e para os clientes.

Importante decisão no trabalho de Joy é quanto tempo dedicar aos clientes no manejo de condições crônicas – instruindo-os sobre como evitar doenças e como limitar seu avanço. Ela quer ter a certeza de não só oferecer valor aos clientes, mas também de controlar custos para a empresa. Variável-chave para ela em termos de resultados é até que ponto os clientes são ou não admitidos ou readmitidos em hospitais. A certa altura, ela recebia dezenas de páginas de relatórios mensais que mostravam a frequência das readmissões em hospitais. As porcentagens subiam e desciam, mas Joy nunca sabia ao certo o que as aumentava ou diminuía. “Eu recebia páginas e páginas de informações”, disse Joy, “mas eram apenas dados – não respondiam à pergunta ‘E daí?’”. A verdadeira questão que ela queria esclarecer era se as conversas com os clientes realmente faziam diferença para a saúde deles e para as taxas de hospitalização.

Para ter melhor ideia das relações de causa e efeito entre as atividades de

manejo de condições crônicas, de um lado, e as taxas de hospitalização, de outro, Joy pediu ajuda a alguns especialistas em analítica da empresa. Michael Cousins, chefe do grupo de analítica, e alguns colegas começaram a trabalhar com Joy para sondar com mais profundidade o problema. Michael observou: “Jen talvez não compreenda os detalhes da metodologia de causa e efeito, mas ela é lógica, inquisitiva e faz perguntas inteligentes. Talvez não tenha índole muito quantitativa, mas o raciocínio dela é muito analítico.” Ela é exatamente o tipo de gestora com que os grupos de analítica amam trabalhar.

O grupo de Cousins enfatiza o alinhamento da metodologia analítica com as necessidades da empresa. Cousins e colegas concordaram com Joy que as informações sobre taxas de hospitalização não eram muito úteis, uma vez que não respondiam à questão de causa e efeito sobre o impacto que os profissionais de saúde exerciam sobre as condições dos pacientes e as taxas de hospitalização. Em outras palavras, os relatórios anteriores que mostravam aumentos e reduções não adotavam metodologia que respondesse às dúvidas sobre correlações entre as variáveis; não havia comparação confiável ou grupo de referência que servisse de base para as decisões. Por exemplo, os relatórios anteriores não consideravam os níveis de risco das condições dos pacientes – alguns pacientes estão mais doentes que outros. O grupo de Cousins desenvolveu uma metodologia de *controle de pares de casos*, confrontando cada paciente com outro, cujo estado de saúde, situação demográfica, estilo de vida e área geográfica fossem semelhantes. Um cliente de cada par de confronto recebia treinamento em manejo de condições crônicas, e o outro não, de modo que Joy pudesse determinar até que ponto o esforço era eficaz. A esse respeito, assim se manifestou Cousin: “Ela precisou de muita coragem para investigar se essas intervenções – o principal propósito da organização de Jen – realmente eram eficazes, mas não hesitou em apurar a verdade.”

Os resultados sugerem que certas intervenções do call center, embora não fossem tão úteis quanto se previra para muitos tipos de doenças, eram muito mais eficazes do que se imaginara para outros tipos de doença. Em relação ao primeiro grupo, Joy partiu para a ação e resolveu abreviar a duração das chamadas para clientes portadores de certas doenças, até que se definisse a extensão em que realmente produziam efeito positivo. Quanto ao segundo grupo, a equipe está ampliando e remanejando o pessoal do call center para atividades mais agregadoras de valor. Hoje, ela trabalha com o grupo de Cousin

em outros projetos de analítica, inclusive experimentos controlados que testam diferentes métodos de treinamento, como envolvimento mais profundo com os médicos dos clientes. Em vez de confiar nos instintos para dizer o que é eficaz, ela adotou com determinação a ideia de aplicar métodos de analítica estruturada, como programas pilotos *teste e aprenda*, realizando de 20 a 30 testes por ano.

Jennifer Joy sempre cultivou raciocínio analítico; depois da parceria com o pessoal de analítica, porém, ela passou a dispor de ferramentas adequadas para comprovar suas ideias e para tomar decisões factuais. E embora a maneira mais eficaz de melhorar as condições de saúde dos pacientes pelo telefone ainda não tenha sido descoberta, a Cigna, agora, está gastando menos com métodos ineficazes e mais com abordagens eficazes. Por seu turno, Michael Cousins e seu grupo também ganharam ao trabalhar com Joy e com outros executivos da empresa, dotados de raciocínio analítico. Eles se tornaram muito mais capazes de expressar os resultados das análises em termos de negócios e de transmitir o que significam para os clientes e para o desempenho financeiro.

Portanto, este livro é para as Jennifer Joys do mundo, não para os Michael Cousins. Não estamos propondo que você se converta em analista quantitativo nem em cientista de dados pleno. Isso exigiria muito mais estudos que ler este livro, e nem todos têm interesse ou qualificações para desempenhar essas funções. Em vez disso, queremos que você se familiarize com dados e análises – que os solicite, que os use no trabalho, que tome decisões com base neles, e que os difunda em sua organização. Queremos que as pessoas comentem sua “paixão por dados”, ou que falem a seu respeito como sobre Jen Joy, na Cigna: “Ela não é nerd em dados, mas compreende e aprecia o trabalho que eles fazem.” Não queremos que você faça análises de dados complexas, mas queremos que você seja inteligente consumidor de dados – ajudando a formular as questões, fazendo perguntas sobre os dados e as metodologias, esforçando-se para compreender os resultados e usando-os para melhorar o desempenho da organização. Parafraseando Xiao-Li Meng, chefe do Departamento de Estatística de Harvard, nossa intenção com este livro não é que você se torne enólogo magistral (termo que ele usa ao referir-se a PhDs em estatística), mas, sim, que se transforme em connoisseur de vinhos finos.<sup>1</sup>

Quase sempre, no passado, era muito mais difícil ser consumidor eficaz de analítica sem compreender em profundidade os métodos e processos de análise

em si. Mas essa situação mudou drasticamente. Agora, da mesma maneira como não se precisa conhecer o funcionamento do motor de combustão interna para ser bom motorista, raramente é necessário saber como se chegou aos números para usá-los na tomada de decisões. Softwares específicos fazem cada vez mais o trabalho pesado, às vezes até escolhendo os métodos estatísticos mais adequados, com base nos atributos dos dados e das variáveis. Novos softwares de analítica (principalmente os da empresa SAS) têm o recurso “What does it Mean?” (“O que significa?”), que explica em texto simples e objetivo o que é correlação ou como fazer previsões estatísticas.

Apesar das necessidades dos grandes consumidores de analítica, até agora ainda não se escreveu obra específica, em linguagem fluente e leiga, com o objetivo de desenvolver as qualificações de novatos em analítica. Este livro descreve o que é analítica, mostra como aplicar a analítica em muitos casos da vida real e sugere maneiras de melhorar suas qualificações na disciplina. Nosso propósito é aprimorar substancialmente sua compreensão de analítica e sua capacidade de discutir com eficácia soluções de analítica para problemas organizacionais. De acordo com relatório de 2011 do McKinsey Global Institute sobre *Big Data*, precisaremos de nada menos que 1,5 milhão de gestores experientes em análise de dados para tirar proveito dos grandes volumes de dados gerados pelas sociedades.<sup>2</sup>

Esperamos que você seja um deles.

## A importância da analítica no processo decisório

As decisões nas organizações, em geral, e nas empresas, em especial, podem basear-se em ampla variedade de fatores, como experiência pessoal, intuição, experimentação ou em analítica e análise de dados. Como no livro *Moneyball*, sobre a aplicação da analítica no beisebol profissional, a adoção de processo decisório analítico não garante resultado positivo; o Oakland A’s não venceu todos os jogos do campeonato descrito no livro, nem em qualquer outro subsequente. A analítica, entretanto, talvez ofereça pequeno trunfo a seus usuários. O Oakland continua a sair-se melhor do que seria de esperar, considerando o total de sua folha de pagamento. Evidentemente, também é possível tomar boas decisões com base apenas na experiência e na intuição – mormente em circunstâncias em que o decisor tem muita experiência no



assunto em foco. Em quase todas as situações da vida, porém, dispõe-se de evidências de que as decisões analíticas são mais adequadas e eficazes, além de produzirem melhores resultados.<sup>3</sup> No beisebol profissional, quase todos os times hoje adotam métodos analíticos introduzidos pelo Oakland A's. Mesmo o New York Yankees – que já se orgulhou de sua abordagem não analítica à seleção de jogadores e à elaboração de estratégias – hoje emprega 21 estatísticos de beisebol.

Em negócios, a analítica tradicional é usada principalmente para apoiar decisões internas nas organizações – questões como “Qual deve ser o preço deste produto?” ou “Que promoção de vendas mais tenderá a levar os clientes a comprar nossos produtos?”. A analítica em ambientes de *Big Data* geralmente se presta a apoiar decisões sobre novos produtos ou novos recursos para os clientes – como o algoritmo PageRank, de busca no Google; o recurso “Pessoas que talvez você conheça”, no site da rede social LinkedIn; ou o desenvolvimento de novos jogos na Zynga. Ainda se trata de forma de decisão – seja da empresa em si, seja dos clientes; por exemplo, “Quem devo incluir em minha rede?” no LinkedIn. (A lista “Que tipos de decisões empresariais podem se basear em métodos analíticos?” oferece mais exemplos.) E estamos apenas arranhando a superfície. Podemos criar listas semelhantes para outras indústrias e setores – como governo, saúde, esportes, para citar apenas uns poucos.

Que tipo de decisões empresariais se podem basear na analítica?

**Marketing**



- Precificação
- Localização de lojas e sucursais
- Alvos de promoções
- Customização de sites da internet
- Colocação de propaganda em mídias digitais

#### **Cadeia de fornecimento**



- Que níveis de estoque manter
- Onde instalar centros de distribuição
- Roteamento de produtos ou veículos
- Carregamento de caminhões

### **Finanças**



- Indutores do desempenho financeiro
- Boletins de desempenho
- Previsões de vários tipos

### **Recursos humanos**



- Quais empregados contratar

- Quais empregados tendem a deixar a empresa
- Qual deve ser o nível de remuneração dos empregados
- Que tipo de educação mais beneficiaria os empregados

### Pesquisa e desenvolvimento



- Que atributos de produtos são mais desejados pelos clientes
- Quão eficaz é determinado produto
- Qual desenho de produto é mais atraente

Ao avaliar as alternativas, o responsável pela decisão considera informações de duas fontes: análise quantitativa e fontes não quantitativas. Fontes não quantitativas são as oriundas de intuição, experiências, regras práticas, boatos e palpites. Estas às vezes podem ser úteis, mas apresentam vários problemas. Mesmo que se tenha muita experiência na área em que se toma a decisão, é possível que não abranja todo o domínio, que não represente situações mais amplas. Os palpites são sempre arriscados. Em geral, deve-se questionar a intuição. Muita gente a superestima como determinante do processo decisório. Toda uma escola do pensamento econômico, por exemplo – denominada *economia comportamentalista* – se baseia no reconhecimento de que, em geral, as decisões intuitivas em economia não são boas.

Apesar das vantagens da analítica, nem sempre é sensato adotá-la como base



para a tomada de decisões. Caso a questão seja pouco importante ou envolva preferências pessoais, não se dê ao trabalho de coletar ou de analisar dados. Se a decisão for urgente, talvez não haja tempo para adotar abordagem analítica. E na hipótese de a decisão ser singular ou não repetitiva, talvez não se justifiquem os custos de reunir dados e de construir um modelo analítico. As decisões que mais se beneficiam com a analítica, portanto, são as que se repetem; que não são prementes, admitindo análises; e que são bastante importantes para justificar o investimento.

## Três estágios do raciocínio analítico e como aplicá-los

O cerne deste livro descreve os três principais estágios do raciocínio analítico. Dedicamos um capítulo a cada estágio, nos quais apresentamos exemplos de trabalhos analíticos em que determinada etapa é sobretudo importante. A [Figura 1.1](#) mostra os três estágios e os passos de cada um deles.

### FIGURA 1.1

---

#### **Os três estágios e os seis passos da análise quantitativa**

### Formulação do problema

1. Reconhecimento do problema
2. Revisão das descobertas anteriores



### Solução do problema

3. Modelagem
4. Coleta de dados
5. Análise dos dados



### Resultados: comunicação e ação

6. Apresentação dos resultados e adoção de providências



---

O [Capítulo 2](#) trata do primeiro estágio, *Formulação do problema*. Formular o problema consiste em definir a questão a ser respondida pela analítica e as decisões a serem tomadas com base nos resultados. Evidentemente, trata-se de

etapa muito importante; caso se formule o problema de maneira equivocada, por mais fartos que sejam os dados e por mais sofisticada que seja a análise, não se chega a conclusões certas. A formulação do problema compõe-se de dois passos – um é o *reconhecimento do problema* e o outro é a *revisão das descobertas anteriores*. Veremos no [Capítulo 2](#) que, quando se supõe já se ter reconhecido e definido o problema a ser resolvido com a analítica, muitas vezes se descobrem aspectos que já foram abordados por alguém – o que ajuda a reformular o problema ainda melhor.

O segundo estágio, objeto do [Capítulo 3](#), é o que, a princípio, se supunha fosse o primeiro estágio – *solução do problema*. É a etapa em que se escolhem as variáveis a serem incluídas no modelo, em que se coletam dados para medir essas variáveis e em que efetivamente se faz a análise dos dados. Supondo que você próprio não seja e não pretenda tornar-se quant, você pode trabalhar com analistas quantitativos, que se incumbirão de muitas dessas atividades. Ainda é muito útil, contudo, saber quais são as atividades e como são executadas. Você mesmo talvez não resolva o problema; suas perguntas e ideias, porém, em muito contribuirão para o desenvolvimento de soluções mais adequadas e mais proveitosas.

O [Capítulo 4](#) trata do terceiro e último estágio – resultados: comunicação e ação – que, embora tão importante quanto as duas etapas anteriores, não raro é negligenciado. A maneira como se divulgam o processo analítico em si e seus resultados é fundamental para a sua difusão e eficácia. Se os usuários potenciais (talvez você) não compreenderem as análises que foram feitas e os resultados a que se chegou, eles dificilmente irão se sentir à vontade para tomar decisões com base nas conclusões do processo. Nessas condições, de pouco adianta executar os dois estágios anteriores. Ainda por cima, neste mundo com deficiência de atenção em que vivemos hoje, a divulgação dos resultados da analítica de maneira interessante e atraente, capaz de reter a atenção, é de extrema importância. Não mais se podem apresentar os resultados em relatórios repletos de linhas e colunas de números em preto e branco, na expectativa de que alguém faça algo com eles.

Os capítulos subsequentes tratam de questões mais específicas, relacionadas com o raciocínio analítico. O [Capítulo 5](#) discute como explorar a criatividade em analítica – criatividade e analítica não são de modo algum incompatíveis! O [Capítulo 6](#) aborda algumas maneiras diferentes de desenvolver a própria

capacidade analítica, caso se esteja inclinado a fazê-lo. E o [Capítulo 7](#) descreve alguns meios eficazes de os usuários não quantos de analítica entrosar-se com os especialistas quantos de analítica para chegar a melhores decisões. Desnecessário dizer, trata-se de via de mão dupla. Ao longo de todo o percurso, numerosos são os exemplos, em diferentes praias da vida, que mostram como usar a analítica para resolver problemas (ou não – ver a seguir “O perigo de *não* pensar analiticamente”). Também oferecemos planilhas com detalhes sobre como aplicar a analítica, além de seções que destacam maneiras simples de aplicar o raciocínio analítico de imediato.

## O Perigo de *não* pensar analiticamente

Aprendemos com exemplos negativos e positivos; como, porém, os exemplos negativos são sempre mais dramáticos, vamos apresentar um deles. E nada poderia ser mais dramático que o caso de Joe Cassano, personagem que, quase sozinho, derrubou uma grande empresa e que quase arrastou com o desmonte a economia dos Estados Unidos e, por incrível que pareça, a economia mundial.

O que fez Cassano, e o que não sabia ele sobre analítica? Talvez você se lembre de algo a respeito dele, ao mencionarmos o cargo que exercia: chefe da AIG Produtos Financeiros (AIG Financial Products – AIGFP), pequena unidade da AIG, ex-seguradora gigante, com 400 pessoas. O que ele fez foi presidir a perda de quantia colossal – ainda se questiona o total exato, mas trata-se de algo em torno de US\$85 bilhões –, obrigando os pagadores de impostos americanos a custear com o próprio bolso a sobrevivência e o pagamento das dívidas da AIG. Cassano não foi a primeira pessoa a perder dinheiro, mas, como observou o jornalista investigativo Matt Taibbi, na revista *Rolling Stone*, foi o “Paciente Zero do colapso econômico global”.<sup>a</sup> Taibbi o retratou como “um careca rechonchudo, com pós-graduação pelo Brooklyn College, com olhos pequenos brilhantes e testa grande demais”. Mas isso não é parte de nossa história. Se, em vez de perder, ele tivesse ganhado muito dinheiro, não temos dúvida de que a aparência dele seria muito melhor.

Nisso tudo, onde entra a analítica e o raciocínio quantitativo – ou, mais exatamente, a falta de ambos? A AIGFP perdeu tanto dinheiro vendendo um produto financeiro denominado *credit default swaps* (CDSs), ou apólices de seguro que garantiam o valor de títulos derivativos lastreados em empréstimos hipotecários. Gretchen Morgenson, repórter do *New York Times*, escreveu sucintamente depois desses eventos: “Embora o colapso do mercado habitacional dos Estados Unidos geralmente seja citado como causa da crise, o sistema ficou vulnerável em consequência de contratos financeiros complexos, conhecidos como derivativos de crédito, que garantiam os credores contra calotes. Esses títulos eram emitidos por instituições financeiras privadas, sem a supervisão dos reguladores – às vezes até além da compreensão dos executivos que os mascateavam.”<sup>b</sup>

Cassano, ao que tudo indica, se incluía entre os executivos que não os compreendiam. Tanto os derivativos quanto os CDSs eram produtos complexos, com base em matemática e em estatística – na verdade, em matemática e em estatística atabalhoadas, como se veio a constatar. Se os devedores hipotecários deixassem de pagar os empréstimos, os derivativos perderiam o valor e a AIG seria obrigada a cobrir as perdas dos credores com a inadimplência, até o valor segurado dos derivativos. Não precisa

dizer que isso foi exatamente o que aconteceu.

Michael Lewis descreve a essência do problema da unidade de negócios da AIGFP na *Vanity Fair*:

*Como e por quê o milagre se converteu em catástrofe, dizem os operadores da AIGFP, é história complicada, mas o começo é simples: mudança na maneira como se tomavam as decisões, resultante de mudança na liderança. Em fins de 2001, seu segundo CEO, Tom Savage, aposentou-se, e seu vice, Joe Cassano, foi promovido. Savage é matemático por formação, que compreendia os modelos usados pelos operadores da AIG para precificar o risco que estavam correndo – e, assim, garantir que receberiam remuneração adequada. Ele gostava de discutir sobre os modelos e sobre os méritos das diferentes operações da AIGFP. Cassano, por seu turno, conhecia muito menos matemática e tinha muito menos interesse pelo debate.<sup>c</sup>*

Cassano não queria saber se os riscos subjacentes e se os modelos de precificação eram exatos ou não. Tampouco fazia perguntas difíceis aos analistas da AIG e de Wall Street, que haviam desenvolvido os modelos. Ele vendia sucessivos lotes de CDSs a quem quer que se dispusesse a comprá-los. A possibilidade de se basearem em premissas falsas – mormente quanto à questão elementar de se os mutuários desses empréstimos hipotecários de baixa qualidade realmente teriam condições de pagá-los – parece que nunca lhe ocorreu. Ou, se ocorreu, ele não se preocupou com o risco de inadimplência.

a. Matt Taibbi. AIG and the Long Con. *Rolling Stone*, 23 de março de 2009.

b. Gretchen Morgenson. Behind Insurer's Crisis, Blind Eye to a Web of Risk. *New York Times*, 27 de setembro de 2008.

c. Michael Lewis. The Man Who Crashed the World. *Vanity Fair*, agosto de 2009, <http://www.vanityfair.com/politics/features/2009/08/aig200908>.

Para os novatos, estudar sozinho a fim de melhorar a capacidade analítica não é fácil. Criamos um site na internet (<http://keepingupwiththequants.weebly.com>), em que os leitores podem fazer qualquer pergunta sobre conhecimentos, técnicas ou questões de analítica com que tenham deparado na solução de problemas específicos. Por favor, não hesitem em postar qualquer pergunta. No mesmo site, também se encontram análises detalhadas de alguns casos descritos neste livro. Vez por outra referimos a esse site no texto.



# 2

## Formulação do problema

Embora muitos sejam os tipos de análise quantitativa, todas têm em comum certos atributos e etapas. Conforme observamos no [Capítulo 1](#), a análise quantitativa, em geral, compõe-se dos seguintes três estágios e seis passos:

### FORMULAÇÃO DO PROBLEMA

- Reconhecimento do problema
- Revisão das descobertas anteriores

### SOLUÇÃO DO PROBLEMA

- Modelagem e seleção de variáveis
- Coleta de dados
- Análise dos dados

### RESULTADOS: COMUNICAÇÃO E AÇÃO

- Apresentação dos resultados e adoção de providências

Neste capítulo e nos [Capítulos 3](#) e [4](#), descreveremos cada estágio e cada passo, um a um, e oferecer um par de exemplos de análises quantitativas que

cobrem os seis passos, mas destacam o estágio específico do processo, que está sendo analisado no capítulo. Ao fim de cada um dos três capítulos, apresentaremos dois exemplos – geralmente um de negócios e outro envolvendo a sociedade mais ampla ou experiências pessoais – que ilustram como a análise abrangeu todos os seis passos, salientando, porém, mais uma vez, determinado estágio. Nosso processo de três estágios e seis passos não é a única maneira de aplicar a analítica (por exemplo, a metodologia Seis Sigma analisa variações na qualidade dos produtos que sejam superiores a 3,4 defeitos por milhão de produtos produzidos), mas esperamos que a maioria dos especialistas em analítica o endosse, mormente por ser bastante amplo para abranger diferentes tipos de problemas e de análises de empresas.

## **Passo do reconhecimento do problema**

A análise quantitativa começa com o reconhecimento do problema a ser resolvido ou da decisão a ser tomada, daí partindo para a busca de solução. Em análise de decisões, esse estágio é denominado *formulação*, e é um dos componentes mais importantes do bom processo decisório. Várias são as situações que levam a esse primeiro estágio, como:

### **1. Reconhecimento do problema**



- Pura curiosidade (bom senso, observação de acontecimentos).
- Experiências no trabalho.
- Necessidade de decisão ou ação.
- Situações em andamento, exigindo atenção (de uma pessoa, de uma organização, de um país).
- Desenvolvimento, ou contestação, de teorias existentes ou de pesquisas passadas.
- Aceitação de ofertas de projetos ou estimativa da disponibilidade de

financiamento.

Observe que, nessa fase, ainda não se iniciou a analítica em si. A decisão de prosseguir com algum tipo de análise pode emergir de um palpite ou de uma intuição. O padrão de evidências a essa altura é baixo. Evidentemente, toda a questão da análise quantitativa consiste em analisar alguns dados e testar o palpite. Essa é a diferença entre pensadores analíticos e outros: os pensadores analíticos testam os palpites com base em dados e análises.

O fator mais importante no estágio de reconhecimento do problema é compreender plenamente a questão e sua importância. As respostas a essas duas perguntas não só esclarecem o que se conseguirá com a solução do problema, mas também facilitam os estágios subsequentes.

## Identificação dos *stakeholders* na análise

Talvez não haja dúvida de que as pessoas envolvidas nesse primeiro passo sejam, basicamente, gestores e decididores – os “donos” do negócio ou do problema organizacional. Mesmo nesse passo, entretanto, o trabalho desses protagonistas será muito enriquecido pela presença de analistas quantitativos que compreendam o problema de negócios, o processo decisório, e os prováveis métodos quantitativos a serem empregados. Caso nenhum profissional detenha sozinho todos esses conhecimentos, talvez seja preciso constituir uma equipe que garanta abordagem mais ampla.

Nesse passo, vale a pena refletir com seriedade sobre quem são os stakeholders (partes interessadas) da análise a ser empreendida e como se sentem em relação ao problema (ver “Planilha de análise dos stakeholders”). Poderiam alguns stakeholders participar das ações a serem adotadas com base nos resultados? Estariam céticos quanto à existência do problema? Haveria como convencê-los a fazer algo?

## Planilha de análise dos stakeholders

---

Se não for possível responder afirmativamente à maioria das perguntas, o projeto talvez enfrente dificuldades desde o começo:

1. Resta alguma dúvida sobre que executivos têm interesse no sucesso do projeto de

análise quantitativa?

2. Foram os executivos informados sobre o problema e sobre o esboço da solução?
  3. Estão eles preparados para oferecer os recursos imprescindíveis e para promover as mudanças indispensáveis ao sucesso do projeto?
  4. Em geral, apoiam eles o uso de analítica e de dados como ferramentas na tomada de decisões?
  5. Coincidem a abordagem analítica e o método de comunicação propostos com a maneira típica de pensar e de decidir dos executivos?
  6. Dispõe você de um plano para promover feedback regular e para divulgar resultados intermediários aos executivos?
- 

A tendência dos analistas quase sempre é precipitar-se direto na análise sem pensar nos stakeholders. Quanto mais confiantes forem eles em relação às suas quantificações analíticas, menos se preocuparão com os destinatários finais dos resultados e com quem decidirá sobre agir ou não agir e sobre o que fazer.

Se você estiver convencido da necessidade de manejo dos stakeholders em seu projeto analítico, alguns dos componentes comuns desse processo são:

1. Identificar todos os stakeholders.
2. Registrar as necessidades dos stakeholders.
3. Avaliar e analisar os interesses e influências dos stakeholders.
4. Administrar as expectativas dos stakeholders.
5. Tomar providências.
6. Revisar o status e repetir.<sup>1</sup>

A análise dos stakeholders destina-se a identificar quem são os principais responsáveis pelas decisões e qual é a maneira mais provável de convencê-los quanto aos resultados do trabalho. Mesmo os projetos analíticos mais rigorosos e mais blindados serão pouco proveitosos se não induzirem à ação os responsáveis pelas decisões. Com efeito, talvez até seja sensato adotar abordagem

questionável sob o ponto de vista metodológico se esse for o único argumento em que confiarão os decididores.

Por exemplo, Rob Duboff dirige uma empresa de pesquisa e de estratégia de marketing, denominada HawkPartners. Em geral, ele confia no valor da pesquisa quantitativa, sempre que necessária, mas também sabe que alguns executivos não compreendem os métodos quantitativos como meios para melhor interpretar os desejos e as necessidades dos clientes, acreditando muito mais em abordagens qualitativas, como grupos de foco – reunir pequeno grupo de clientes atuais ou potenciais, perguntar-lhes o que acham dos produtos e serviços da empresa, ouvir e registrar suas respostas. Duboff, contudo, também está consciente de que, em termos metodológicos, os grupos de foco são suspeitos. É de conhecimento geral na área de pesquisa de mercado que os clientes tendem a dizer-lhe o que você pretende ouvir, e que o fato de afirmarem que gostam de algo não significa que de fato venham a comprá-lo. Esses problemas podem ser mitigados por um líder de discussão habilidoso, mas os resultados dos grupos de foco não são projetáveis para universo mais amplo. Duboff, todavia, sente que qualquer pesquisa é melhor que nenhuma; e se concluir que os resultados de um grupo de foco merecerão mais confiança e serão mais considerados pelos executivos, como base de ação, em comparação com resultados mais quantitativos, ele não hesitará em adotar essa metodologia menos quantitativa.

No mesmo sentido, os stakeholders podem ajudar a definir a forma de apresentação dos produtos e resultados. Os seres humanos diferem em suas preferências quando se trata de ver resultados quantitativos. Alguns preferem fileiras e colunas de números; outros gostam mais de gráficos; e ainda há quem opte por textos que descrevem os números. É importante identificar essas preferências em fase relativamente incipiente. Se os resultados se destinarem não a seres humanos, mas a computadores – e essa é a tendência, na medida em que as decisões, cada vez mais, são automatizadas, no todo ou em parte – faz pouco sentido preocupar-se com formatos visuais. Basta plantar os números para que floresçam!

Também é possível que certas abordagens analíticas ajudem a envolver os stakeholders em todo o processo de análise. Por exemplo, na Cisco Systems, um projeto tratava da possibilidade de conseguir previsões muito mais exatas por meio de métodos estatísticos (descreveremos os seis passos desse exemplo no fim do [Capítulo 7](#)). Alguns gestores da Cisco apoiavam o projeto, mas outros



questionavam a factibilidade de melhores previsões. Anne Robinson, que gerenciava o projeto, adotou metodologia “ágil”, promovendo sucessivas entregas de resultados aos stakeholders, a intervalos de poucas semanas. Essa abordagem mais incremental à solução do problema ajudou os stakeholders a comprar o novo método. Por fim, ficou claro mesmo para os gestores céticos que o novo método de previsão era muito mais exato, muito mais rápido e muito mais abrangente que a abordagem anterior, não analítica.

Constatamos ser útil no passo de reconhecimento do problema concentrar-se nas decisões específicas a serem tomadas em consequência da análise. Muitas são as razões para tanto. A principal é levar todos os participantes a concluir que esse é o objetivo mais importante da análise quantitativa; não se trata de exercício inútil. Outra é identificar um importante stakeholder: a pessoa ou grupo que tomará a decisão com base na análise. Uma terceira razão é que, caso não se tenha em vista, com muita clareza, uma decisão específica, talvez não valha a pena fazer a análise.

Por exemplo, Mike Thompson, chefe da First Analytics, empresa de serviços de análise, descreve uma reunião que teve com um cliente no passo de reconhecimento do problema. O cliente, uma cadeia de restaurantes, acreditava que o propósito básico da análise era a lucratividade dos produtos. Os executivos do cliente queriam que a First Analytics determinasse quão lucrativo era cada item do menu. Como também concordava com a ideia de concentrar-se nas decisões, Mike perguntou aos gestores do cliente que decisões tomariam com base nas análises de lucratividade. Seguiu-se um longo silêncio. Um executivo sugeriu que a decisão básica seria manter ou não os itens no menu. Outro observou, no entanto, que a cadeia não havia eliminado um único item do menu ao longo dos últimos 20 anos. Depois de algum debate, a equipe do cliente concluiu que talvez o foco da análise devesse convergir para a precificação dos itens do menu, em vez de para a lucratividade. “Desde nossa fundação mudamos os preços”, observou um executivo.

## Que tipo de história analítica se está contando?

Depois de decidir que decisões tomar, começa-se a pensar em como desenvolver respostas ou ideias para a decisão. Conversaremos no [Capítulo 4](#) sobre *contar uma história com os dados*, que é a melhor maneira de comunicar os resultados a

público não analítico. A essa altura, porém, já é preciso começar a pensar sobre o tipo de história a ser contada e como narrá-la, embora muitos detalhes da história se manifestem mais adiante, no processo de análise. Pelo menos seis são os tipos de histórias em análise quantitativa; cada uma das quais é descrita a seguir, junto com um ou dois exemplos.

**HISTÓRIAS CSI (INVESTIGAÇÃO CRIMINAL).** Algumas análises quantitativas são como programas de televisão sobre métodos policiais: tentam resolver problemas de negócios com análises quantitativas. De repente despontam alguns problemas operacionais e se usam dados para confirmar a natureza da questão e para desenvolver uma solução adequada. Essas situações geralmente não exigem análises estatísticas profundas, bastando a compilação de dados e a elaboração de relatórios. Geralmente são comuns em negócios on-line, onde a sequência de cliques fornece muitos dados – não raro até demais – para os analistas.

Um especialista praticante do método história CSI é Joe Megibow, vice-presidente e gerente geral nos Estados Unidos da agência de viagens on-line Expedia. Joe já trabalhara como especialista em analítica da internet – o que ainda faz – mas seus métodos de solução de problemas com base em dados e análises lhe propiciaram impressionante sucessão de promoções.

Muitas das investigações da Expedia consistem na identificação das causas de quedas nas vendas on-line. Determinada história CSI tratava da perda de receita com transações de pagamentos de hotéis. A análise dos dados sugeriu que, tendo o cliente escolhido o hotel, preenchido as informações para faturamento e clicado no botão “Compre Agora”, algumas das transações não eram concluídas com êxito. A equipe de Megibow investigou as causas das falhas, mais uma vez usando dados de internet e arquivos de servidores durante todo o processo.

Aparentemente, o campo “empresa”, abaixo do nome do cliente, era a causa do problema. Alguns usuários entendiam tratar-se do nome do banco que fornecia o cartão de crédito, o que os levava a incluir o endereço do banco nos campos referentes a endereço para faturamento. A transação não era efetuada pelo processador do cartão de crédito. A simples remoção do campo “empresa” imediatamente aumentou o lucro da Expedia em US\$12 milhões. Megibow diz que a Expedia explora muitas dessas histórias CSI e que elas sempre rendem benefícios operacionais e financeiros significativos.

Às vezes, as histórias CSI envolvem análises quantitativas e estatísticas mais

profundas. Um membro da equipe de Megibow estava investigando que pontos de contato com os clientes impulsionavam as transações de vendas on-line. Os analistas usavam o *modelo de regressão Cox* – método desenvolvido de início para determinar que pacientes morreriam ou sobreviveriam ao longo de determinados períodos – de “análise da sobrevivência”. A análise levou à conclusão de que os modelos anteriores mais simples não estavam de modo algum corretos sobre as abordagens de marketing que realmente geravam vendas. Megibow lamentou: “Não sabíamos que estávamos deixando dinheiro na mesa.”<sup>2</sup>

**HISTÓRIAS EURECA.** São semelhantes às histórias CSI, a não ser pelo fato de, em geral, consistir em abordar de propósito determinado problema (em vez de deparar com o problema), no intuito de decidir importante mudança na estratégia ou no modelo da organização. Elas tendem a ser mais longevas, envolvendo mais análises ao longo do tempo. Às vezes, as histórias Eureka se associam a outros tipos de histórias de analítica, pelo simples fato de os resultados serem tão importantes para a organização.

Novamente na Expedia, por exemplo, uma história Eureka tratava de eliminar a cobrança de taxas de mudança ou cancelamento de reservas em hotéis, cruzeiros ou alugueis de carros. Até 2009, a Expedia e os concorrentes cobravam até US\$30 pelas mudanças ou cancelamentos – acima e além dos encargos exigidos pelo hotéis. Os encargos da Expedia e de outros corretores on-line eram sempre mais baixos que os de transações feitas diretamente com os hotéis, e os clientes se mostravam tolerantes com as taxas de mudança e de cancelamento.

Em 2009, porém, já se tornara notório que as taxas de mudança ou cancelamento haviam ficado onerosas demais. Os preços da Expedia se aproximavam dos cobrados pelos hotéis, e o principal apelo da Expedia agora era conveniência – e as taxas de mudança ou cancelamento não eram de modo algum convenientes. Os analistas examinaram os índices de satisfação dos clientes e constataram ser demasiado baixos os dos clientes que tinham de pagar esses encargos adicionais. Em consequência, os atendentes do call center da Expedia foram autorizados a isentar da taxa de mudança ou cancelamento apenas os clientes que alegavam caso de morte em família. A partir de então, as análises dos números de dispensas passaram a mostrar porcentagens de crescimento de dois dígitos nos casos de mortes nos três anos anteriores. Ou

estava ocorrendo alguma epidemia letal ou os clientes tinham descoberto o atalho para conseguir isenção ou para receber o dinheiro de volta.

Os executivos da Expedia perceberam que o mercado já não era o mesmo, mas as taxas de mudança ou cancelamento representavam importante fonte de receita. Diante do dilema, ficaram pensando se as vendas realmente aumentariam caso o encargo fosse eliminado. Em abril de 2009, anunciaram dispensa temporária da taxa de mudança ou cancelamento durante todo o mês (um pouco como nos métodos de cientista maluco, descritos mais adiante). O aumento das vendas foi imediato e substancial. Os executivos, então, concluíram que já dispunham de evidências suficientes para eliminar a cobrança da taxa de mudança ou cancelamento, no que a Expedia foi acompanhada pelo resto do setor.

Na cidade de Seattle situa-se a sede da Zillow, empresa que fornece informações sobre o mercado habitacional. Ela talvez seja mais bem conhecida entre os quants pelo “Zestimates”, algoritmo proprietário que gera estimativas de valores de moradias. Como na Expedia, entretanto, toda a cultura da Zillow se baseia em dados e análises – algo nada surpreendente, porquanto foi fundada por Richard Barton, que também criou a Expedia.

Uma das histórias Eureka da Zillow envolveu importante decisão sobre como o relacionamento com corretores de imóveis gerava receita. Até 2008, a empresa lidava apenas com consumidores; só a partir de então passou a trabalhar com corretores. Uma característica desse modelo de negócios voltado para corretores era vender anúncios através desses intermediários, além de fornecer-lhes pistas de negócios. A Zillow cobrava dos corretores pelo fornecimento de pistas, mas, na opinião dos executivos, o preço não era suficiente. Chloe Harford, gestora da Zillow, responsável pela gestão de produtos e de estratégia, empenhava-se, em especial, no desenvolvimento de um modelo certo para aumentar o valor e otimizar o preço das pistas.

Harford, PhD em vulcanologia, ou estudo de vulcões, estava afeita a análises matemáticas bastante sofisticadas. Ela e os colegas, contudo, recorriam originalmente ao que denominavam “matemática de guardanapo”, no intuito de explorar outras maneiras de gerar mais pistas e de cobrar o preço justo dos corretores. Em abril de 2010, Zillow criou novo atributo – copiado de imediato pelos concorrentes – com a venda de anúncios aos corretores. Daí resultou muito mais contatos com os clientes, permitindo que os consumidores se

relacionassem diretamente com os corretores. A Zillow também adotou complexo algoritmo de precificação, capaz de estimar o valor econômico da pista, com base na taxa de conversão esperada. Os concorrentes, até certo ponto, também seguem esse procedimento, mas, provavelmente, não com o nível de sofisticação da Zillow. As pistas e seus preços são tão importantes que Harford e colegas frequentemente testam novas abordagens, adotando alguns dos experimentos de cientista maluco, apresentados a seguir. Em síntese, as histórias Eureka da Zillow estão interligadas estreitamente com o modelo de negócios e com o sucesso da empresa.

**HISTÓRIAS DE CIENTISTA MALUCO.** Todos estamos familiarizados com o uso de testes científicos em setores com base científica, como na indústria farmacêutica. As empresas de medicamentos experimentam seus produtos em grupos de testes, enquanto oferecem placebos em grupos de controle. Para garantir a neutralidade dos resultados, empenham-se em distribuir aleatoriamente os participantes dos grupos de teste e dos grupos de controle, para que diferenças de composição importantes não afetem a avaliação da eficácia dos medicamentos. Trata-se de ferramenta analítica poderosa, por permitir tanto quanto possível a estimativa da causalidade – de modo a garantir que o objeto do experimento no grupo de teste realmente esteja produzindo o resultado almejado como causa e efeito.

O rigor dos testes já não é privilégio de cientistas com avental branco; hoje, é método analítico que pode ser adotado por todas as grandes organizações, por meio de softwares de ampla disponibilidade, que orientam gestores e analistas ao longo do processo. Em consequência, as empresas, agora, podem lastrear importantes decisões em experimentos reais, com validade científica. No passado, qualquer incursão em testes aleatórios (a distribuição randômica dos sujeitos entre os grupos acima mencionados) exigia que se recorresse a PhDs em estatística ou a especialistas em “projeto de experimentos”. Atualmente, um MBA com treinamento quantitativo tem condições de supervisionar o processo, com o apoio de softwares que ajudam a definir o tamanho dos grupos, onde realizar os testes e os controles, e se quaisquer mudanças decorrentes do experimento são significativas em termos estatísticos.

As histórias de cientista maluco se ajustam muito bem a certas organizações, como varejistas (com muitas lojas) e bancos (com muitas agências), características que facilitam o uso de algumas localidades para experimentos e de

outras localidades para controles. Também é muito fácil fazer testes em sites da internet, nos quais é possível encaminhar alguns clientes para uma nova versão, no intuito de verificar se os resultados apresentam diferenças significativas (denominados *A/B testing* em análise de internet).

Alguns exemplos de histórias de cientista maluco pretendem responder a perguntas do tipo:<sup>3</sup>

- Os tanques de lagosta nos supermercados Food Lion aumentam as vendas? A resposta é aparentemente sim, caso se trate da loja em que os clientes já compravam lagosta (ou seja, se os clientes já fossem relativamente sofisticados); e não, se a loja, de início, não atraísse clientes compradores de lagostas.
- Kmart com lojas da Sears vendem mais que lojas só Kmart ou só Sears? Eddie Lampert, chairman da Sears Holdings, é grande fã de testes aleatórios e experimentou várias dessas combinações. Não sabemos a resposta para essa pergunta específica, mas achamos que, se fosse um sim definitivo, estaríamos vendo muito mais desses estabelecimentos mistos.
- Os melhores resultados de vendas dos restaurantes da cadeia Red Lobster, especializados em frutos do mar, são produzidos por unidades remodeladas de baixo, médio ou alto custo? E deve a arquitetura do exterior ou do interior ser o principal objeto de atenção? O resultado, de acordo com os executivos da Red Lobster, foi que a arquitetura interior de médio custo apresentou o melhor retorno. Embora a reforma da fachada atraísse muitos clientes novos, eles não repetiam a experiência, ao perceberem que o interior não fora reformado.

**HISTÓRIAS DE PESQUISAS.** Pesquisas são método clássico de análise quantitativa. Os pesquisadores observam um fenômeno que já aconteceu ou que está acontecendo agora, sem manipular o resultado – apenas codificando-o e analisando-o. Tipicamente, os pesquisadores tentam identificar os traços ou variáveis observados na pesquisa que apresentam correlação estatística com



outros traços ou variáveis. O exemplo mais simples seria pedir a uma amostra de clientes de determinados produtos várias informações pessoais, inclusive características demográficas, como gênero e idade. Se também indagarmos que produtos preferem, descobriremos os itens mais atraentes para homens ou para mulheres, assim como para outras categorias demográficas, como jovens.

As pesquisas são populares e relativamente fáceis de executar. É preciso lembrar, todavia, que seus resultados e histórias podem ser muito diferentes, dependendo de como se formulam as perguntas e de como elas variam no tempo. Por exemplo, o US Census trabalha literalmente há décadas com questões sobre a raça dos cidadãos americanos. O número de categorias raciais em pesquisas censitárias continua em expansão: no censo de 2010 havia 15 escolhas, inclusive “alguma outra raça”. Essa foi opção muito popular entre 50 milhões de cidadãos americanos latinos, dentre os quais 18 milhões marcaram essa quadrícula.<sup>4</sup> Se há tanta confusão sobre raça, imagine as dificuldades dos pesquisadores com tópicos mais escorregadios, como política, religião, atitudes sociais e comportamento sexual.

Também é necessário lembrar que, na análise de pesquisas, a correlação de duas variáveis nem sempre significa causalidade. Voltaremos a esse ponto no [Capítulo 6](#), mas, por enquanto, basta lembrar que outras variáveis ignoradas podem ser o fator causal do fenômeno observado.

Nas histórias de pesquisas, geralmente se interrogam pessoas sobre crenças e atitudes, mas também é possível fazer pesquisas sem envolver pessoas. Veja, por exemplo, essa pesquisa sobre aviões, conduzida durante a Segunda Guerra Mundial, relatada em um compêndio clássico de estatística:

Durante a Segunda Guerra Mundial, era necessário manter os aviões em ação tanto tempo quanto possível. Portanto, tomou-se a decisão de verificar a possibilidade de reduzir sem risco a frequência das revisões de motores, muito demoradas. Com base em pesquisa retrospectiva dos aviões perdidos, constatou-se, diferentemente de todas as expectativas, que as perdas resultantes de problemas com o motor eram mais comuns logo depois das revisões e efetivamente diminuía com o decurso do tempo depois da revisão. Esse resultado levou a aumento considerável nos intervalos entre as revisões e, desnecessário dizer, a mudanças importantes na maneira como se faziam as revisões, de modo a assegurar que todos os

parafusos e porcas fossem realmente ajustados da maneira certa.<sup>5</sup>

Antes de executar ou de analisar uma pesquisa, não deixe de refletir com muito cuidado sobre os significados das perguntas e das variáveis. Variável é qualquer característica mensurável, com dois ou mais níveis ou valores, de traços pessoais, de situações e de comportamentos. Gênero, escore em testes, temperatura ambiente, amor, felicidade e coesão das equipes são bons exemplos de variáveis.

Também é importante garantir que a amostra seja representativa da população que se pretende estudar. A maneira de executar a pesquisa pode afetar a amostra. Por exemplo, caso se queira pesquisar atitudes ou comportamentos de jovens, não se deve contratar empresa de pesquisa que use apenas telefones fixos para contatar os membros da amostra. Essa abordagem é muito comum, mas todos sabemos que muitos jovens não têm, e pretendem jamais ter, telefones fixos. Portanto, seriam mal representados na amostra, quando os pesquisadores usam apenas telefones fixos.<sup>6</sup>

**HISTÓRIAS DE PREDIÇÕES.** Têm a ver com antecipar o que acontecerá. Embora seja muito difícil conseguir bons dados sobre o futuro, levantar dados históricos e compreender os fatores que impulsionaram os acontecimentos passados é algo muito objetivo e direto em análise quantitativa. Em geral, o processo é denominado *analítica preditiva* ou *modelagem preditiva*.

Várias são as histórias de previsões a serem desenvolvidas pelos analistas. Encontra-se a seguir uma amostra das possibilidades. Veja como são específicas:

- *Respostas a ofertas:* Que clientes responderão a um e-mail com oferta de entrega gratuita, durante dois dias úteis, para compras de US\$50 ou mais?
- *Venda cruzada e venda incrementada:* Que clientes com saldo em conta bancária superior a US\$2 mil comprarão CDB com juro real acima de 1,5% ao ano, no prazo de um mês, a contar do recebimento da proposta?
- *Rotatividade dos empregados:* Que empregados há mais de seis meses na empresa, que ainda não aderiram ao programa 401(k), pedirão demissão nos próximos três meses?

Muitas são as possibilidades de análise preditiva. Em negócios, demanda muito comum é prever que ofertas tenderão a ser aceitas pelos clientes. As versões mais sofisticadas desse tipo de análise são cada vez mais automáticas; nenhum ser humano precisa analisar a oferta, antes de ser enviada aos clientes, e pode haver centenas ou milhares de ofertas diferentes.

A Microsoft, por exemplo, desenvolveu incrível capacidade de ajustar dinamicamente as “ofertas” de seu motor de busca Bing (como o produto é gratuito, a Microsoft apenas se esforça para difundi-lo). As ofertas o induzem a experimentar o Bing, ou a criar uma barra de pesquisas Bing no seu navegador, ou a experimentar determinado atributo do Bing, e assim por diante. A customização da oferta se baseia em vários fatores – localização, idade, sexo e buscas recentes do cliente – tendências que podem ser rastreadas com base em cookies e em outras fontes. Quem tem o Passaporte Microsoft deixa rastro de informações ainda mais farto, que possibilita melhor direcionamento das ofertas. A Microsoft, com a ajuda do software Infor Epiphany Interaction Advisor, tem condições de compor imediatamente um e-mail específico, no momento em que você clica em uma oferta em sua caixa de entrada; tudo demora cerca de 200 milissegundos. Segundo a Microsoft, trata-se de método bastante eficaz para aumentar as vendas.

Em geral, as histórias de predições podem ser um pouco como “atirar para todos os lados”. Como não sabemos exatamente que fatores nos ajudarão a prever algo, fazemos inúmeras tentativas, em busca da mais eficaz. Às vezes os resultados são inesperados. Por exemplo, nas ofertas do Bing que acabamos de descrever, o número de ligações no Microsoft Messenger se mostrou bom indicador da probabilidade de alguém experimentar o Bing.

Na Google, a empresa queria identificar os traços dos empregados mais sugestivos de alto desempenho. Algumas análises levaram à conclusão de que os fatores de início usados pela Google – notas nas escolas e avaliações nas entrevistas – eram maus previsores de desempenho. Como não sabiam ao certo que fatores eram importantes, pediram aos empregados para responder a um questionário com 300 perguntas. Eis como Laszlo Bock, chefe de Operações com Pessoas da Google, descreveu o experimento: “Queríamos lançar rede muito ampla. Aqui não é incomum andar pelos corredores e topiar com cachorros. Talvez as pessoas que têm cachorros apresentem algum traço de personalidade promissor.”<sup>7</sup>

Levar cachorros para o trabalho não se revelou bom previsor de nada, mas a Google deparou com outras correlações inesperadas. Por exemplo, ter conquistado recordes mundiais ou nacionais ou ter fundado entidades sem fins lucrativos se associavam a alto desempenho. A Google agora faz perguntas sobre essas experiências nas entrevistas de emprego on-line.

Evidentemente, se os fatores que parecem prever algo não fazem nenhum sentido, é bom voltar atrás, verificar os dados e repetir a análise. Seja como for, a análise de dados pode efetivamente superar as previsões intuitivas na maioria das circunstâncias. Como advertência, lembre-se de que as histórias preditivas usam dados do passado para imaginar o futuro. Se algo no mundo mudar depois da análise, as previsões perdem a validade.

**HISTÓRIAS “EIS O QUE ACONTECEU”.** As histórias que contam o que aconteceu com base em dados são talvez as mais comuns. Elas fornecem os fatos – quantos produtos foram vendidos, quando e onde, quais foram os resultados financeiros do último trimestre, quantas pessoas contratamos no ano passado. Como apenas relatam fatos e não recorrem a matemática sofisticada, talvez pareça que é fácil contá-las. Hoje, porém, o grande aumento no volume de dados existentes nas organizações acarretou aumento semelhante nas histórias baseadas em dados. Portanto, às vezes, é difícil atrair a atenção do público almejado para a ampla variedade de relatórios.

Esse tipo de história é muito adequado para a exposição visual de informações. Basta dizer que a apresentação de relatórios em fileiras e colunas de números dificilmente atrai a atenção necessária. Hoje, embora muita gente já esteja cansada até de gráficos e diagramas coloridos, a maioria não hesitaria em afirmar que esses recursos visuais ainda são preferíveis à profusão de algarismos numa página. Como o [Capítulo 4](#) é sobre a divulgação de resultados, falaremos mais sobre como tornar esses tipos de relatórios mais interessantes e mais chamativos.

## Escopo do problema

Por definição, as histórias baseadas em dados e as análises quantitativas subjacentes não raro apresentam escopo um tanto estreito, porque exigem coleta e aplicação de dados a hipóteses sujeitas a testes (ver “Exemplos de

hipóteses sujeitas a testes”). É difícil reunir dados sobre problemas muito amplos, mas é importante, a essa altura, não limitar prematuramente o escopo do problema ou decisão. A reflexão sobre a questão deve ser ampla e é importante desenvolver numerosos percursos mentais alternativos. Por exemplo, se uma empresa identifica um problema de desempenho em determinada unidade de negócios ou área geográfica, é importante manter-se aberto para vasta diversidade de causas – desde insatisfação dos clientes até ineficiências operacionais, passando por deficiências nos produtos e serviços.

No exemplo da Transitions Optical, no fim deste capítulo, o passo de reconhecimento do problema foi precipitado pelo sentimento vago de que as despesas com marketing estavam altas demais, o escopo da decisão, porém, foi ampliado de modo a envolver a otimização geral das despesas com marketing e das escolhas de mídia.

Referimo-nos a esse primeiro passo da análise quantitativa como reconhecimento do problema, mas ele também poderia ser rotulado como identificação de oportunidades. Joseph Jagger, engenheiro inglês, percebeu que havia uma oportunidade de “quebrar a banca” do cassino de Monte Carlo.<sup>8</sup> Depois de desenvolver experiência prática em mecânica nas fábricas de algodão de Yorkshire, Jagger a aplicou ao comportamento da roleta, especulando que os resultados não eram puramente sequências aleatórias, mas, sim, que desequilíbrios mecânicos poderiam gerar vieses para determinados números. E se houvesse imperfeições na roda giratória, suscetíveis de exploração para proveito próprio? Imbuído dessas ideias, partiu para Mônaco, no intuito de testar o conceito.

A roleta franco-europeia tem 37 números: 0 a 36. Quando se gira a roleta uma vez, a probabilidade teórica de cada número é de  $1/37$ . Portanto, a probabilidade de que caia cada número em grande quantidade de giros também é de mais ou menos  $1/37$ . Jagger, entretanto, especulou que eventuais desequilíbrios mecânicos na roda, se houver, acarretariam probabilidade superior a  $1/37$  para determinados números.

## Exemplos de hipóteses sujeitas a testes

- Os tipos de produtos que os clientes compraram no ano passado é o melhor guia das ofertas por e-mail a que responderão positivamente no futuro.

- Tempo de escolaridade é bom previsor do nível de desempenho do empregado em trabalhos do conhecimento.
- Descontos de 10% na semana anterior aos períodos de férias são menos eficazes que os concedidos em outros períodos.
- Os end-cap displays são os lugares mais eficazes para a exposição de nossos produtos em lojas de varejo, com o objetivo de aumentar as vendas semanais.
- Nossos clientes podem ser agrupados em cinco segmentos diferentes quanto aos produtos que compram.
- A flexibilidade para aumentar os preços de certa categoria de produtos de consumo, com baixa diferenciação, sem comprometer a procura é muito mais baixa durante as recessões econômicas.
- Nossas unidades de negócios com gestão de estoques centralizada tendem a apresentar níveis de estoques mais baixos.

Com essas ideias em mente, Jagger contratou seis funcionários para observar as seis roletas no lendário Beaux-Arts Casino, em Monte Carlo, cada um de olho numa roleta, com instruções específicas para registrar os resultados de cada giro. Ao analisar os dados, Jagger constatou que cinco roletas produziam resultados aleatórios, como seria de esperar. Na sexta roda, contudo, nove números (7, 8, 9, 17, 18, 19, 22, 28, 29) apareceram com frequência superior à do mero acaso. Jagger concluiu que a roleta estava enviesada, ou seja, mal equilibrada. E, assim, fez suas primeiras apostas em 7 de julho de 1875, e logo ganhou grande quantia (£14 mil – algo equivalente a 60 vezes esse número em 2012, ou mais de US\$1,3 milhão, depois do ajuste pela inflação). O cassino acabou descobrindo e depois neutralizado a estratégia de apostas de Jagger – mas não antes de ele ter ganhado importância hoje equivalente a US\$6 milhões. Excelente oportunidade propiciada pela analítica!

## Seja específico sobre o que quer descobrir

Embora seja importante raciocinar com amplitude no começo do passo de reconhecimento do problema, ao concluí-lo será necessário ter criado uma descrição clara da questão, com definições concretas dos principais itens ou



variáveis a serem estudados. Eis por quê: em pesquisa quantitativa, faz muita diferença a maneira como as coisas são definidas. Por exemplo, digamos que você seja um executivo de televisão interessado em descobrir os canais preferidos do público. Dois consultores de analítica o procuraram com propostas para descobrir a resposta. Só de brincadeira, você resolveu contratar os dois para comparar os resultados.

Um deles propõe que se peça aos consumidores para registrar (em formulário físico ou on-line) os canais e programas vistos em cada dia da semana. O outro sugere que se peça aos participantes para classificar os canais a que mais assistiram na televisão nos últimos meses. Ambos apresentaram projetos de amostras e de formulários bem concebidos, capazes de representar a população e de captar suas preferências.

Embora os dois consultores estejam tentando resolver os mesmos problemas, é provável que cheguem a resultados muito diferentes. O primeiro, que propôs pedir aos consumidores que registrem os canais e programas efetivamente vistos, tenderá a produzir resultados mais exatos; o trabalho adicional para os participantes, todavia, de anotar diariamente os canais e programas tenderá a acarretar nível de participação mais baixo (A Nielsen Media Research, que monitora continuamente as audiências de canais e programas, apresenta níveis de desistência em torno de 50%, e os registros são automáticos.) Outro problema desse consultor é o fato de os padrões de audiência estarem muito sujeitos a influências excessivas da temporada ou da programação durante a semana específica do estudo.

O outro consultor tenderá a apresentar resultados menos exatos; como o método dele, porém, cobre período mais extenso, suas conclusões estarão menos sujeitas a fatores sazonais. Mais importante, os resultados das duas pesquisas provavelmente serão tão diferentes que haveria dificuldade em reconciliá-los. Daí a importância de concluir o passo de reconhecimento do problema com uma ideia clara do que se pretende estudar.

## **Passo da revisão das descobertas anteriores**

Depois do reconhecimento do problema, devem-se investigar todas as descobertas correlatas anteriores. Esse é o segundo passo do primeiro estágio da análise (formulação do problema), pois essa investigação pode ajudar os analistas

e decididores a refletir sobre como até agora o problema foi estruturado e como poderia ser reconceituado de diferentes maneiras. Com muita frequência, os analistas, neste passo, descobrem algo que acarreta mudança substancial no reconhecimento do problema, o que, por seu turno, altera o conjunto de descobertas anteriores.

Neste passo, basicamente, perguntamos: “Já se contou antes história semelhante a esta?” Em caso positivo, podemos extrair ideias do trabalho realizado no passado. A revisão das descobertas anteriores pode sugerir os seguintes questionamentos:

## **2. Revisão das descobertas anteriores**



- Que histórias contaremos? Envolverão essas narrativas previsões, relatórios, experimentos e pesquisas?
- Que dados deveremos procurar?
- Como as variáveis foram definidas antes?
- Que espécies de análises tenderemos a executar?
- Como contaremos história mais interessante, diferente das anteriores, mais propensa a produzir resultados?

Um dos principais atributos da análise quantitativa (e do método científico, em termos mais amplos) é basear-se em pesquisas e descobertas anteriores. Por exemplo, a busca de informações sobre problemas correlatos, em livros, relatórios e artigos, é muito importante para chegar ao âmago do problema. Os resultados talvez ajudem a identificar variáveis relevantes e associações entre elas.

A revisão completa de quaisquer descobertas anteriores é indispensável em toda análise quantitativa. Em analítica, não se pode fazer algo a partir do nada. Só é possível resolver um problema quando se tem domínio total das descobertas anteriores. Basta lembrar o seguinte: seu problema não é tão singular quanto parece, e é provável que muita gente já tenha feito o que você está tentando fazer. Não reinvente a roda; mas é preciso pesquisar, pesquisar novamente e pesquisar outra vez. Hoje, usando motores de busca como Google, é possível compilar vasto material relacionado com a sua questão. Organizando e avaliando o material, é possível identificar modelo ou abordagem potencial para resolver o problema.

Um exemplo bem-sucedido de revisão de descobertas anteriores ocorreu durante a Segunda Guerra Mundial. Adolf Hitler determinara a produção de poderosa bomba foguete, denominada V-2, e, em 1944, a Luftwaffe começou a aterrorizar os cidadãos de Londres. Nos meses seguintes, 1.358 V-2s, do total de 3.172 foguetes lançados sobre vários alvos aliados, sobrevoaram a Inglaterra e aterrissaram em Londres, acarretando a morte de aproximadamente 7.250 militares e civis.

Durante os ataques a Londres, muitos observadores afirmavam que os pontos

de impacto das bombas formavam aglomerados. Os ingleses queriam descobrir se os alemães realmente podiam direcionar as bombas para os alvos ou se estavam limitados a disparos aleatórios. Caso os alemães só fossem capazes de atingir alvos ao acaso, a dispersão por todo o território nacional de várias instalações de segurança serviria muito bem para proteger o país. Se, porém, os alemães já tivessem condições de efetivamente orientar as ogivas para alvos específicos, os ingleses estariam enfrentando adversário mais poderoso; a redistribuição das unidades militares de pouco adiantaria. O governo inglês contratou, então, o estatístico R. D. Clarke para resolver a questão. Este aplicou um teste simples, na revisão – ou no conhecimento disponível – de descobertas anteriores.

Clarke sabia que a *distribuição de Poisson* poderia ser usada para analisar a distribuição das bombas. A distribuição de Poisson expressa a probabilidade de numerosos eventos acontecerem em determinado período, área ou volume, caso as ocorrências se repitam conforme certos padrões. O único fator que é preciso conhecer para especificar a distribuição de Poisson é o número médio de eventos. No caso do bombardeio de Londres, se as bombas estivessem caindo ao acaso, o número de bombas que atingissem alguma área circunscrita seguiria a distribuição de Poisson. Por exemplo, se o número médio de ocorrências é de uma bomba por área, podemos calcular facilmente a probabilidade de que o número de incidências por área seja 0, 1, 2, 4 ou mais, bastando inserir os números na fórmula de Poisson.

Para estimar as quantidades de bombas que poderiam atingir qualquer área pequena, especificamente definida, Clarke dividiu o Sul de Londres em 576 quadrados de um quarto de quilômetro quadrado cada um e contou as quantidades de quadrados que haviam sido atingidos por 0, 1, 2, 3 etc., bombas voadoras. Se os impactos fossem completamente aleatórios, a probabilidade de um quadrado ser atingido por 0, 1, 2, 3 etc., bombas seria determinada pela distribuição de Poisson. O enquadramento da situação real no padrão de Poisson foi surpreendentemente bom, contrariando a hipótese da formação de aglomerados (ver o site deste livro na internet). Os ingleses ficaram aliviados com a conclusão de Clarke. Felizmente, os alemães se renderam em 1945, antes de as V-2 provocarem muito mais estragos. (*Nota: Apesar da impossibilidade de serem guiados com eficácia, esses foguetes foram a base tecnológica do programa espacial dos Estados Unidos.*)

Da mesma maneira como fez Clarke ao perceber que o problema das bombas

alemãs que atingiam Londres poderia ser descrito pela distribuição de Poisson, é possível retroceder e reexaminar o passo do reconhecimento do problema depois da revisão das descobertas anteriores (ver “Alguns métodos para a revisão das descobertas anteriores”). Talvez se constate que é necessário modificar a história, o escopo do problema, a decisão e até os stakeholders. Em seguida, mudando ou preservando as condições iniciais, o estágio de formulação do problema estará concluído e se avançará para a solução do problema, usando métodos de análise quantitativa.

### Alguns métodos de revisão das descobertas anteriores

- Pesquise na internet os principais termos relacionados com a análise.
- Consulte um livro-texto de estatística sobre análises semelhantes à que se está propondo.
- Converse com analistas que trabalham com sua empresa para ver se eles fizeram algo semelhante.
- Verifique se o sistema de gestão do conhecimento de sua empresa contém informações pertinentes.
- Converse sobre o problema com analistas de outras empresas (mas não concorrentes).
- Participe de seminários e outros eventos (ou pelo menos busque referências on-line) sobre analítica, para ver se alguém já abordou ou pretende abordar temas correlatos.

### Reformulação do problema

Embora tenhamos apresentado o processo analítico de solução do problema como sequência linear de seis passos em três estágios, a abordagem deve ser, acima de tudo, iterativa. Cada passo lança nova luz sobre o problema e é sempre boa ideia pensar em como reconsiderar os passos anteriores sob novo enfoque.

Embora não se deva reexaminar indefinidamente cada passo, sempre vale a pena refletir um pouco sobre o que a revisão das descobertas anteriores sugerem sobre a formulação do problema (a “Planilha de reformulação do problema” pode ser útil).

## **Planilha de reformulação do problema**

---

Foi o problema bem formulado? Em caso positivo, deve ser possível responder afirmativamente à maioria das perguntas a seguir:

1. Você definiu com clareza um problema ou oportunidade a fim de abordar o que é importante para a empresa ou organização?
  2. Você considerou várias maneiras alternativas de resolver o problema?
  3. Você identificou os stakeholders do problema e comunicou-se amplamente com eles sobre a questão?
  4. Você está confiante em que a maneira como pretende resolver o problema será satisfatória para os stakeholders e que eles usarão os resultados para tomar decisões?
  5. Você está seguro quanto às decisões a tomar – e sobre quem as tomará – com base nos resultados de sua análise, depois da solução do problema?
  6. Você começou com uma definição ampla do problema, mas depois o restringiu a aspectos muito específicos, com uma descrição muito objetiva da questão a ser abordada, dos dados a serem aplicados e dos resultados possíveis.
  7. Você é capaz de descrever o tipo de história analítica a ser contada para resolver este problema específico?
  8. Alguém é capaz de ajudá-lo a resolver esse tipo específico de história analítica?
  9. Você procurou sistematicamente descobertas ou experiências anteriores relacionadas com o problema, dentro ou fora da organização?
  10. Você reexaminou a definição do problema com base no que descobriu com a revisão das descobertas anteriores?
- 

Como bom exemplo, Rama Ramakrishnan, especialista em analítica de varejo, e hoje CEO da startup CQuotient, descreve no blog dele situação que demandava reformulação:<sup>9</sup>



Considere o problema de “customer targeting” (definição dos clientes) que ocorre no marketing direto. O propósito é definir que clientes abordar, uma vez que dirigir-se a todos é dispendioso demais. Este é um problema clássico que foi estudado por numerosos pesquisadores e praticantes. O método mais comum é o seguinte:

1. Envie correspondência experimental a uma amostra de clientes.
2. Use os resultados do teste para desenvolver um “modelo de reação” que preveja a propensão de cada cliente a reagir à correspondência em função de seus atributos, história passada etc.
3. Com base no modelo, pontue cada cliente no banco de dados e envie correspondência para os que obtiverem maior pontuação.

A proposta parece razoável, e talvez seja o que importa para as empresas. Mas pode ser que não seja bem assim.

As palavras “modelo de reação” sugerem que a correspondência *levou* os clientes a reagir. Na realidade, o cliente talvez tenha ido à loja e feito uma compra independentemente da correspondência (estou pensando em varejistas multicanais, não em puros varejistas de catálogo. Para estes últimos, sem o catálogo, talvez seja impossível para os clientes fazer a compra, situação em que o termo “reação” seria adequado).

O que esses modelos de reação realmente fazem é identificar os clientes propensos a comprar de qualquer maneira, em vez de os clientes propensos a comprar por terem recebido a correspondência. O que talvez a administração realmente queira, porém, é identificar estes últimos. Para os clientes dispostos a comprar ou a não comprar, não importa o que tenham recebido pelo correio, a correspondência é desperdício de dinheiro e pode comprometer a predisposição positiva dos clientes. O que provavelmente a empresa de fato pretende é identificar os clientes que comprarão se receberem correspondência e não comprarão se não receberem correspondência.

Essa reformulação do problema de “customer targeting” e das abordagens para resolvê-lo é relativamente recente. Já recebeu muitos

nomes – “uplift modeling”, “net lift modeling” – e os trabalhos acadêmicos a respeito são muito poucos, em comparação com os referentes à tradicional modelagem de reações. Para muitos varejistas, contudo, essa é a maneira mais relevante e mais útil de formular e resolver o problema de “customer targeting” que a abordagem tradicional.

Neste exemplo, a cuidadosa revisão das descobertas anteriores poderia ter revelado os trabalhos recentes sobre “uplift modeling” e “net lift modeling”, o que poderia ter ocasionado a reformulação do problema. Ramakrishnan sugere que nessas situações, com métodos de modelagem relativamente novos, “uma vez que o novo problema ainda não foi alvo de atenção suficiente (por definição), algoritmos simples talvez gerem benefícios com mais rapidez”.

Concluiremos este capítulo sobre formulação do problema com dois exemplos, um de negócios e outro de legislação, em que o processo de formulação foi fundamental para os resultados. Um é de formulação certa; o outro, de formulação errada. Você ainda não aprendeu o suficiente sobre os passos além do estágio de formulação, mas estamos confiantes em que será capaz de compreender esses exemplos.

### Exemplo de raciocínio analítico: *Transitions Optical*

Um dos problemas analíticos mais comuns em negócios é decidir quanto gastar com determinada atividade. E essa decisão é ainda mais difícil quando se refere a despesas com marketing. John Wanamaker, criador do conceito de loja de departamentos, e alguns varejistas europeus antes dele, são conhecidos por dizerem: “Metade do dinheiro que gasto com propaganda é desperdício; o problema é que não sei qual é a metade.” Hoje, contudo, as empresas podem recorrer à análise quantitativa para identificar as despesas com marketing eficazes e ineficazes – e para determinar a composição mais produtiva das despesas com marketing. Trata-se do que é tipicamente denominado *análise do mix de marketing*, cada vez mais popular entre empresas que vendem para consumidores.

RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA. A Transitions Optical, que

fornece lentes fotocromáticas para óculos, estava sofrendo pressões das matrizes (PPG e Essilor) quanto aos níveis de despesas com marketing. A PPG, em especial, não atua no negócio de marketing de consumo, razão por que estava muito cética a respeito do valor da propaganda e da promoção, questionando principalmente a eficácia de certas campanhas. Em geral, predominava a percepção intuitiva de que as despesas eram altas demais, embora não se dispusesse de dados empíricos para responder à pergunta de qual seria o nível ótimo de despesas com marketing. Os executivos da Transitions resolveram formular o problema como de otimização de despesas e abordagens de marketing, de maneira a maximizar o aumento das vendas por dólar investido. De acordo com Grady Lenski, que chefiava marketing na época, “estamos confiando demais na arte para tomar decisões de marketing; precisamos recorrer mais à ciência”.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Não havia descobertas anteriores sobre esse tópico; a Transitions dispunha de dados sobre clientes que possibilitariam essa análise, mas estavam espalhados por toda a organização. Lenski e alguns colegas sabiam que era possível analisar a eficácia de diferentes métodos de marketing, mas não conheciam os detalhes.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Os modelos de otimização do mix de marketing, cada vez mais adotados em grandes organizações para otimizar as despesas com marketing, envolvem variáveis de respostas ao marketing, de custos de marketing e de margens de produtos. Os modelos de otimização, que usam métodos de programação linear e não linear, identificam os níveis de propaganda semanal ou mensal, de promoção e de preços que maximizam a receita, a margem de lucro ou ambos. Também determinam os veículos de propaganda específicos mais eficazes para a maximização desses resultados. Em geral, também contêm uma série de variáveis de “controle” que podem afetar os gastos dos consumidores e os comportamentos de compra, como condições climáticas e dados macroeconômicos.

**COLETA DE DADOS.** Esse foi um dos aspectos mais difíceis da análise da Transitions, uma vez que a empresa trabalha com intermediários (laboratórios de ótica, por exemplo) e historicamente tinha pouco contato com dados sobre os clientes finais. Por conseguinte, não tinha condições de medir com exatidão se os anúncios eram vistos pelos clientes nem se geravam aumento das vendas. A

Transitions embarcou em iniciativa multianual para reunir dados sobre os clientes finais de seus parceiros de canais (alguns dos quais eram concorrentes das matrizes). Lenski já fora chefe do canal de varejo, o que facilitou o levantamento de informações. As informações sobre clientes chegavam à Transitions em 30 diferentes formatos, mas a empresa acabou conseguindo reuni-las em armazém de dados (*data warehouse*) integrado para análise. Lenski observou que a organização de marketing precisava convencer diferentes partes da Transitions a fornecer dados. Ao fazer a análise pela primeira vez, a empresa não tinha armazém de dados.

**ANÁLISE DE DADOS.** A Transitions contratou um consultor externo para fazer a análise dos dados, uma vez que não tinha na organização ninguém familiarizado com modelos de otimização de mix de marketing. De início, a análise demorou vários meses, uma vez que foi necessário coletar dados e que o modelo exigia a eliminação de ampla variedade de outros fatores explicativos de qualquer resposta de marketing (como condições climáticas, marketing dos concorrentes e outros). Agora que foi desenvolvido e refinado, o modelo pode ser executado em poucos dias.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** A Transitions sentiu que a atividade de interpretação e apresentação dos resultados era importante o suficiente para dispor de recursos internos, justificando a contratação de pessoal próprio. Esses novos especialistas adotam o modelo dos consultores e o discutem com os executivos para determinar suas implicações e para aprimorá-lo com base nas intuições dos usuários sobre o mercado. No todo, os resultados levaram a aumento das despesas com marketing da Transitions, em especial com anúncios na televisão.

### Exemplo de raciocínio analítico:

#### *Povo versus Collins*

*Povo versus Collins* foi um julgamento na Califórnia que se tornou notório pelo amplo uso forense de matemática e de probabilidade, e é bom exemplo de como a incorreta formulação do problema pode levar a maus resultados.<sup>10</sup> O júri julgou culpados os réus, Malcolm Collins e a esposa, Janet Collins, por roubo em

segundo grau. Malcolm recorreu da decisão, e a Suprema Corte da Califórnia revogou a condenação, criticando o raciocínio estatístico e desautorizando a maneira como a situação fora apresentada ao júri. Examinaremos o caso consoante o referencial de seis passos.

**RECONHECIMENTO DO PROBLEMA.** Mrs. Juanita Brooks tinha feito compras e voltava para casa por um beco, na área de San Pedro. De repente, foi jogada ao chão por alguém que não conseguiu ver. Ficou estonteada com a queda e sentiu alguma dor. Imediatamente depois do incidente, Mrs. Brooks deu pela falta da bolsa, com algo entre US\$35 e US\$40. Uma testemunha da ocorrência depôs que os autores eram um homem negro, com barba e bigode, e uma mulher branca, com cabelos loiros presos em rabo de cavalo. Ambos fugiram em um carro amarelo. Durante o julgamento de sete dias, a acusação teve dificuldade em determinar a identidade dos criminosos. A vítima não conseguiu identificar Janet Collins, e em momento algum viu o assaltante; a identificação pela testemunha foi incompleta. O promotor – talvez afoito para encerrar o caso – resolveu ajudar o júri a determinar a probabilidade de que o casal acusado coincidisse com a descrição das testemunhas.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Sabe-se que o tribunal geralmente não discerne incompatibilidade intrínseca entre direito e matemática e não tende a desaprovar ou menosprezar a matemática como ferramenta para a apuração de fatos à disposição da lei. Em alguns processos penais, a promotoria recorreu à probabilidade matemática como argumento de acusação.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** O modelo sugerido pelo promotor é a probabilidade de que o par de acusados se encaixe na descrição das testemunhas.

**COLETA DE DADOS (MENSURAÇÃO).** O promotor chamou para depor um instrutor de matemática de uma faculdade estadual. Por meio dessa testemunha, sugeriu que o júri poderia estimar com segurança as seguintes probabilidades de encontrar os atributos dos criminosos e do crime:

Homem negro com barba	1 em 10
Homem com bigode	1 em 4

Mulher branca com rabo de cavalo	1 em 10
Mulher branca com cabelos loiros	1 em 3
Carro amarelo	1 em 10
Casal de raças diferentes em carro	1 em 1.000

ANÁLISE DOS DADOS. O instrutor de matemática sugeriu que, no caso de eventos independentes, a oportunidade de ocorrerem juntos pode ser calculada pela multiplicação de cada probabilidade.

$P(A)$  = probabilidade de o casal acusado  
encaixar-se na descrição

$$= \frac{1}{10} \times \frac{1}{4} \times \frac{1}{10} \times \frac{1}{3} \times \frac{1}{10} \times \frac{1}{1.000}$$

$$= \frac{1}{12.000.000}, \text{ ou } 1 \text{ em } 12 \text{ milhões}$$

APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS. O promotor chegou à probabilidade de que a chance era de apenas 1 em 12 milhões de que qualquer casal possuísse as características dos réus. Assim, com base nessa teoria, concluiu-se que só podia existir 1 chance em 12 milhões de que os réus fossem inocentes. Em consequência, o veredito foi culpado.

Os Collinses recorreram do julgamento. A Suprema Corte da Califórnia sentenciou que, sem dúvida, os jurados haviam sido indevidamente pressionados pela mística da demonstração matemática, mas foram incapazes de avaliar sua relevância ou valor. A corte reverteu a condenação, criticando o raciocínio estatístico e desautorizando a maneira como a questão fora exposta ao júri. A Suprema Corte observou que a técnica específica apresentada através do depoimento matemático padeceu de dois defeitos importantes. Primeiro, a acusação não apresentou evidência, qualquer que fosse, da validade das probabilidades nem demonstração da qual fosse possível inferir essas chances.

Segundo, outro defeito flagrante da técnica da acusação: prova inadequada da independência estatística dos seis fatores apresentados como provas (por exemplo, homens que usam barba comumente também usam bigode).

Mais importante, a argumentação e as provas também foram formuladas incorretamente pelo promotor. Mesmo que a conclusão da promotoria fosse aritmeticamente exata, não se podia concluir que os Collins fossem culpados. Não houve absolutamente nenhuma diretriz quanto a uma questão crucial: dentre os poucos casais que, em tese, apresentassem o mesmo conjunto de características, qual, se houvesse algum, teria cometido o roubo?

A variável relevante nesse caso não era a probabilidade de que o casal acusado se encaixasse na descrição das testemunhas, mas a probabilidade de que houvesse outros casais com as mesmas características, uma vez que um deles já se enquadrava. Dependendo exatamente de quantos casais havia na área de Los Angeles, a probabilidade de pelo menos outro casal corresponder à descrição poderia ser de nada menos que 40% (ver <http://keepingupwiththequants.weebly.com>). Portanto, os cálculos da acusação, longe de demonstrar, acima de qualquer dúvida razoável, que os Collinses eram o casal descrito pelas testemunhas de acusação, implicam probabilidade muito significativa de que havia na área mais de um casal com as mesmas características, e que outro casal, não os Collins, foi o visto pela testemunha na cena do crime.

Depois de examinar todo o caso, inclusive as provas, a Suprema Corte revogou a decisão da instância inferior contra os réus. A má formulação do problema sem dúvida pode levar a más decisões.



## Solução do problema

Muita gente considera esse estágio – aquele em que, ao menos até certo ponto, se executa a análise quantitativa e se resolve o problema – o núcleo do processo analítico. Trata-se, evidentemente, de atividade muito importante, embora mais estruturada e mais bem definida que os estágios de (a) formulação do problema e de (b) comunicação e ação, que o precede e o sucede, respectivamente. Para quem não tem muita base em matemática ou estatística, também é possível que seja o estágio com maior probabilidade de ser confiado a terceiros, dotados das habilidades quantitativas necessárias (ver “Como encontrar um quant”). Não importa quem efetivamente processe os números, é bom conhecer algo sobre os três passos da solução do problema.

Algumas observações sobre a sequência dos três passos talvez sejam úteis. Estamos descrevendo um processo analítico baseado em hipóteses. Em outras palavras, começando com os passos que compõem a formulação do problema, no [Capítulo 2](#), e prosseguindo com a modelagem e a seleção de variáveis (primeiro passo da solução do problema), refina-se progressivamente a hipótese que se busca nos dados. Em seguida, o analista coleta dados pertinentes. Cada um desses passos é orientado pelo conhecimento ou, ao menos, pelo palpite de como o mundo funciona. Por fim, o passo da análise dos dados confirma quão correto realmente está o conhecimento ou a presunção.

Em tese, há alguns tipos de análise cujo processo não se baseia em hipóteses prévias. Em alguns casos de mineração de dados e *machine-learning* (em que o software busca modelos para os dados de maneira automática e rápida, na tentativa de encontrar o melhor encaixe), o analista aplica o software de análise aos dados, sem restrições, na tentativa de identificar padrões. As hipóteses surgem depois, quando o analista tenta explicar e comunicar os resultados.

## Como encontrar um quant

Se você precisar de um analista quantitativo para resolver seu problema, eis algumas maneiras de consegui-lo:

- Se você trabalha em uma grande empresa, provavelmente já existem alguns na organização – basta procurá-los em lugares como Pesquisa de Mercado, Business Intelligence ou Pesquisa Operacional.
- Caso você não conheça nenhum, não faltam consultores a quem recorrer. Faça uma pesquisa na internet sobre “consultores em analítica de empresas” ou procure em KDnuggets (<http://www.kdnuggets.com/companies/consulting.html>).
- Caso você prefira consultores em analítica de outros países, os melhores estão na Índia; procure empresas como Mu Sigma, Fractal Analytics e Genpact.
- Você talvez encontre professores de analítica ou alunos de pós-graduação na universidade local; tente falar com o chefe do departamento de estatística, por exemplo.
- Caso pretenda contratar alguém como empregado para executar a tarefa, recorra a sites especializados (por exemplo, Simply Hired oferece numerosos “analistas quantitativos”, enquanto [analyticrecruiting.com](http://analyticrecruiting.com) se concentra em estatísticos) ou a empresas de recrutamento e seleção, como Smith Hanley Associates, que há décadas atua no ramo.

Não gostamos muito dessa abordagem, basicamente por acharmos que não raro ela leva a descobertas inexplicáveis. Como nenhum analista tentava encontrar dados para confirmar presunções sobre o funcionamento do mundo, ninguém se empenha em explicar os resultados ou em convencer alguém a tomar outra decisão com base nos resultados. Em certas circunstâncias, porém, essas abordagens de “caixa-preta” podem alavancar em muito o tempo e a produtividade de analistas humanos. Em ambientes de *Big Data*, onde os dados entram constantemente em grandes volumes, nem sempre é possível criar hipóteses antes da triagem dos dados. No contexto de anúncios digitais em sites, por exemplo, é preciso decidir em milissegundos, por meio de sistemas decisórios automáticos. Para tanto, as empresas devem gerar milhares de modelos estatísticos por semana. Sem dúvida, esse tipo de análise não pode envolver muitas hipóteses e muita reflexão sobre os resultados, tornando o aprendizado por máquinas absolutamente indispensável. Em geral, porém, recomendamos que se atenha às análises baseadas em hipóteses, assim como aos passos e sequências deste livro.

## Passo da modelagem (seleção de variáveis)

Modelo é *representação deliberadamente simplificada* do fenômeno ou problema. *Deliberadamente* significa que o modelo se destina especificamente a resolver determinado problema. *Simplificada* denota que precisamos omitir todos os detalhes desnecessários e triviais e isolar os atributos importantes, úteis e cruciais que fazem diferença. Essa espécie de seleção de variáveis pode ser ilustrada nos seguintes termos:

Os modelos são comparáveis a caricaturas. Estas captam certos traços (um nariz, um sorriso ou um cacho de cabelos) e neles concentram a atenção, negligenciando outros atributos. Boa caricatura é aquela em que esses traços especiais são escolhidos com convicção e eficácia. Da mesma maneira, o modelo se concentra em certos atributos do mundo real. Ao construir um modelo, é preciso ser seletivo. É necessário identificar os aspectos relevantes do mundo real e ignorar o resto. “O importante é que o modelo seja minimalista, que propicie foco exclusivo no problema que se procura resolver.”<sup>1</sup>

Daí decorre que o modelo não reflete exatamente a realidade. Com efeito, George Box, estatístico famoso, observou que “todos os modelos são falhos, mas alguns são úteis”.<sup>2</sup> A chave, evidentemente, é ser capaz de determinar quando o modelo é realmente útil e quando o modelo é tão falho a ponto de distorcer a realidade. Falaremos sobre essa questão no [Capítulo 5](#), mas um dos fatores críticos é determinar que variáveis incluir no modelo.

Como selecionar as variáveis e descobrir como elas se relacionam umas com as outras? Por enquanto, ainda estamos muito no reino do subjetivo. *Hipóteses* – as primeiras histórias que se contam sobre a análise – são apenas palpites elaborados sobre as variáveis realmente importantes para o modelo. Nesse estágio, a construção de modelos exige o uso da lógica, da experiência e das descobertas anteriores para formular hipóteses sobre a variável dependente – aquela que se tenta prever ou explicar – e sobre as variáveis independentes que a afetarão. Evidentemente, as hipóteses serão testadas mais tarde; é isso que diferencia o raciocínio analítico de abordagens menos exatas ao processo decisório, como a intuição.

Por exemplo, caso se trate de um cientista social, tentando prever a renda

familiar (variável dependente), pode-se formular a hipótese de que as variáveis independentes do modelo, referentes aos membros da família, serão idade, educação, estado civil e número dos que trabalham em tempo integral. Esses fatores fazem sentido apenas para explicar diferenças de renda. Mais adiante, no processo de análise quantitativa (especificamente, no passo de análise dos dados), talvez se constate que o modelo não é muito adequado e que talvez seja necessário voltar à prancheta e imaginar alguma nova variável sobre a qual coletar dados.

### 3. Modelagem



Mesmo modelos e variáveis altamente subjetivos podem ser úteis. Por exemplo, Garth Sundem, autor que escreve sobre ciência popular, matemática,

humor e cultura geek, em geral, tratou de numerosas questões práticas por meio de variáveis subjetivas – mas ainda úteis.<sup>3</sup> Uma questão das que abordou foi a de ter animal de estimação e, em caso positivo, que tipo faria mais sentido.

Pense nas variáveis a serem consideradas na decisão sobre ter ou não ter animal de estimação. Sundem selecionou as seguintes variáveis:

- Necessidade de mais amor na vida (D) (1-10, onde 10 é “de dia é carcereiro; de noite é fiscal de impostos”)
- Noção de responsabilidade (R) (1-10, onde 1 é “acha que impostos, filhos e compromissos cuidarão de si mesmos se deixados por conta própria”)
- Maior número de dias que passou viajando nos últimos seis meses (T)
- Tempo ocioso por dia (H)
- Tolerância em relação à maldade alheia (M) (1-10, onde 1 é Cruella de Vil e 10 é Dr. Doolittle)
- Até que ponto você é provedor (N), onde 1 é “meu cacto morreu”)

Trata-se de questões inequivocamente subjetivas, mas, provavelmente, também úteis, além de divertidas. Sundem desenvolveu a seguinte equação (aparentemente um tanto rigorosa) para calcular  $F_{ido}$ , índice que indica a conveniência de ter um animal de estimação.

$$F_{ido} = \frac{(M + N)^2 \bar{D} + HR}{8T^2}$$

O termo mais importante da equação é a necessidade de mais amor na vida (D), que aumenta o índice  $F_{ido}$ . Também é bom sinal dispor de algum tempo ocioso (H), a ser dedicado ao animal de estimação, e destacar-se como pessoa responsável (R). Essas duas variáveis também interagem para aumentar o índice  $F_{ido}$ . Se, no entanto, você viaja muito, seu índice  $F_{ido}$  sofrerá queda substancial.

Sundem sugere os seguintes animais de estimação, conforme os resultados:

Se  $F_{ido}$  for inferior a 1, até artêmia seria muito ambicioso.

Se  $F_{ido}$  estiver entre 1 e 2, você pode ter um peixinho de aquário.

Se  $F_{ido}$  estiver entre 2 e 3, você pode ter um gato.

Se  $F_{ido}$  for maior que 3, você pode ter um cachorro.

Jinho inseriu os próprios números na equação e chegou ao próprio índice  $F_{ido}$  de 0,7, sugestivo de que, no caso dele, até um cacto seria perigoso. É discutível se esse nível de precisão quantitativa seria necessário na decisão de ter ou não ter animal de estimação, mas o exemplo mostra que mesmo decisões relativamente triviais e altamente subjetivas podem ser quantificadas e modeladas.

Que variáveis descartar ou preservar depende do propósito do modelo e da relevância direta da variável para a solução do problema. Por exemplo, caso se esteja desenhando um mapa da Cidade de Nova York, as distâncias entre os distritos são importantes e devem ser proporcionais. Se, no entanto, o mapa for das linhas do metrô, as distâncias entre as estações não precisam ser proporcionais; tudo de que se precisa em mapas de metrô são os elementos que indicam como ir de uma para outra estação.

Outro ótimo exemplo da importância da seleção de variáveis (e, a propósito, também da revisão das descobertas anteriores) é a controvérsia sobre quem escreveu uma série de cartas publicadas em 1861. Dez cartas assinadas por “Quintus Curtius Snodgrass” apareceram no *New Orleans Daily Crescent*, naquele ano. Nessas cartas, Mr. Snodgrass [QCS] descrevia suas aventuras militares enquanto atuava como “High Old Private of the Louisiana Guard”. As cartas QCS, que não chamaram muita atenção na época, mereceram destaque pela primeira vez em 1934, 73 anos depois de seu surgimento, ao serem divulgadas pela autora Minnie Brashear, em *Mark Twain, Son of Missouri*. Nesse livro, ela reimprimiu uma das cartas e descreveu três outras, para depois argumentar: “As cartas QCS são muito importantes na condição de elo no desenvolvimento de Mark Twain como humorista; devem ser reconhecidas

como de Twain; a diferença de estilo reflete os esforços incipientes do escritor para alcançar verve literária mais consciente.”<sup>4</sup> As seis cartas remanescentes foram descobertas por Ernest Leisy e publicadas em 1946.<sup>5</sup> O cuidadoso estudo comparativo de Leisy apresenta argumentos convincentes de que as cartas foram escritas por Twain, mas outros pesquisadores ainda insistem em que elas são de outrem.

Em algumas pesquisas anteriores sobre se Shakespeare foi realmente o autor de todos os trabalhos que lhe são atribuídos, Thomas Mendenhall publicou dois artigos, por volta do começo do século 20, detalhando como adotou a abordagem estatística na controvérsia sobre autoria. Claude Brinegar, executivo de empresa petrolífera, com ótimas credenciais acadêmicas, cujo passatempo era colecionar as primeiras edições de Mark Twain, revisou as descobertas anteriores e seguiu o método de Mendenhall – que veio a ser denominado *estilometria* (*stylometry*), ou análise quantitativa do estilo de escrever – aplicando-a às cartas QCS.

O método presume que todos os autores, inconscientemente, usam com mais frequência certas palavras e mantêm estilo semelhante, ao menos no longo prazo. Sob a perspectiva da análise quantitativa, daí resulta que, nos sucessivos trabalhos, as proporções de palavras com diferentes extensões serão mais ou menos constantes. Grandes diferenças nessas proporções devem ser consideradas forte evidência de que os dois conjuntos não foram escritos pela mesma pessoa. A seleção de variáveis para essa análise, portanto, envolveu, basicamente, a medição da extensão das palavras das cartas QCS e a posterior comparação dos resultados com textos sem dúvida escritos por Twain.

Realizaram-se testes de compatibilidade para determinar se a mesma pessoa escreveu os dois conjuntos de textos. Os resultados de Brinegar mostraram que a discrepância entre as proporções era grande demais para ser atribuída a flutuações aleatórias – ou seja, as cartas não parecem ter sido escritas por Mark Twain (ver o site desse livro).<sup>6</sup>

Falaremos mais neste capítulo sobre análise de textos (em oposição a números), observe, porém, que, na pesquisa de Brinegar, o texto foi convertido em números durante o processo de análise.

## **Passo da coleta de dados (mensuração)**



O passo seguinte é coletar dados e medir as variáveis selecionadas. Medir uma variável é atribuir um número à variável; os dados são mero conjunto desses números. Várias são as maneiras de medir variáveis (ver “Maneiras de medir variáveis”). Primeiro organiza-se o problema reconhecido, por meio de processo de modelagem em variáveis críticas, que, então, se tornam dados, depois da mensuração. Os dados a serem reunidos, evidentemente, devem resultar das variáveis identificadas no passo anterior.

Se as variáveis que estão sendo reunidas foram medidas e analisadas com frequência por outrem (algo que se constata na revisão das descobertas anteriores), esse passo será muito simples; pode-se usar método de mensuração alheio. Em alguns casos, contudo, será preciso medir as variáveis pela primeira vez. Como no processo de selecionar variáveis, mesmo fatores muito subjetivos podem ser medidos de maneira sistemática.

#### **4. Coleta de dados**



Digamos, por exemplo, que você esteja pesquisando a muito comum (a julgar pelos comerciais de televisão, ao menos) disfunção erétil (DE). Felizmente, você está com sorte, pois dispõe-se de medida bem aceita no campo. Se, porém, você tivesse sido dos primeiros pesquisadores, teria sido preciso desenvolver suas próprias medidas.

## Maneiras de medir variáveis

As três maneiras de atribuir medidas a variáveis são:

- **Variáveis binárias:** Têm apenas dois valores, e para propósitos de análises estatísticas, em geral é melhor medi-las como presença ou ausência de algo, com valores de 1 ou 0. Um exemplo seria tratar-se de homem ou mulher (que poderia ser registrado como 0 para “não mulher” e 1 para “mulher”), ou ser ou não ser cidadão

americano.

- **Variáveis categóricas (também denominadas nominais):** Têm várias categorias possíveis como valores, como cor dos olhos, sabor de sorvete e estado ou província de residência. Como não podem ser convertidas facilmente em números, cujos aumentos ou reduções significam qualquer coisa, há um tipo especial de estatística para dados categóricos.
- **Variáveis ordinais:** Aquelas a que se atribuem números, e quanto maior o número, mais a variável está presente. No entanto, a diferença entre 1 e 2 pode não ser igual à diferença entre 5 e 6. Exemplo típico de variável ordinal é o *Item Likert* – batizada com o nome do sociólogo Rensis Likert – que tipicamente envolve respostas em pesquisas do tipo discordo muito, discordo um pouco, não discordo nem concordo, concordo um pouco, concordo muito. Quando se reúnem diversas variáveis como essas, a variável resultante é denominada *escala de Likert*.
- **Variáveis numéricas (intervalo e razão):** Essas variáveis têm números com unidades padronizadas, como peso em libras ou em quilos ou altura em tonelada ou em centímetros. Quanto maior é o número, mais a variável está presente. Assim, as variáveis numéricas são adequadas para métodos estatísticos comuns, como análise de correlação e de regressão.

Na década de 1990, R. C. Rosen e colegas criaram medida breve, confiável e autoadministrável de função erétil, com sensibilidade e especificidade suficientes para detectar mudanças relacionadas com tratamento em pacientes com disfunção erétil.<sup>7</sup> A disfunção erétil é condição relatada pelo próprio paciente e não se dispõe de testes objetivos, o que dificulta a exatidão do diagnóstico. Rosen e colegas definiram que as principais variáveis no diagnóstico da disfunção erétil eram:

Confiança na ereção.

Firmeza da ereção.

Frequência da manutenção.

Capacidade de manutenção.

Satisfação.

Eles mediram cada variável com base nas perguntas apresentadas na [Tabela](#)

### 3.1.

Caso você esteja pensando em como essas perguntas se convertem em diagnóstico, a resposta é que elas geram escores de 5 a 25. A DE foi classificada em cinco níveis de gravidade, com base nos escores: grave (5-7), moderada (8-11), branda a moderada (12-16), branda (17-21) e nenhuma (22-25). Essa ferramenta de diagnóstico de DE, autoadministrada com rapidez, é denominada IIEF-5 (versão de cinco itens do International Index of Erectile Function [Índice Internacional de Função Erétil]). Esse exemplo mostra como é possível reunir dados sobre tópico subjetivo.

Não importa quais sejam os dados disponíveis, sempre há a possibilidade de conseguir mais dados ou de obter dados diferentes dos imaginados, de início, ao refletir sobre o problema. Rama Ramakrishnan, quant talentoso que mencionei no [Capítulo 2](#), numa postagem em blog, descreveu maneira de melhorar o impacto analítico: “Uma de minhas preferidas é levantar *melhores dados*. Não maior quantidade de dados, mas, sim, dados *diferentes* dos aplicados na solução do problema até agora. Se você usou dados demográficos, acrescente dados de compras. Se você já explorou os dois tipos de dados, adicione dados de navegação na internet. Se você tem dados numéricos, agregue dados de texto (em trabalho recente, vimos resultados muito promissores com a complementação de dados tradicionais sobre vendas e promoções de varejo com dados de texto para a modelagem e personalização de clientes).”<sup>8</sup>

**TABELA 3.1**

#### **Principais variáveis do diagnóstico da disfunção erétil**

Nos últimos seis meses...

1. Como você avalia sua confiança em que conseguirá e manterá a ereção?	1 Muito baixa	2 Baixa	3 Moderada	4 Alta	5 Muito alta
2. Quando você tem ereção com estímulo sexual, com que frequência suas ereções são suficientes para penetração?	1 Quase nunca / nunca	2 Poucas vezes (muito menos que metade das vezes)	3 Às vezes (cerca de metade das vezes)	4 A maioria das vezes (muito mais que metade das vezes)	5 Quase sempre / sempre
3. Durante as relações, com que frequência você é capaz	1 Quase	2 Poucas vezes (muito)	3 Às vezes (cerca de)	4 A maioria das vezes	5 Quase sempre

de manter a ereção depois da penetração?	nunca / nunca	menos que metade das vezes)	metade das vezes)	(muito mais que metade das vezes)	/ sempre
4. Durante as relações, até que ponto foi difícil manter a ereção para completar a relação?	1 Extremamente difícil	2 Muito difícil	3 Difícil	4 Um pouco difícil	5 Não difícil
5. Ao tentar manter relações sexuais, com que frequência a experiência é satisfatória para você?	1 Quase nunca / nunca	2 Poucas vezes (muito menos que metade das vezes)	3 Às vezes (cerca de metade das vezes)	4 A maioria das vezes (muito mais que metade das vezes)	5 Quase sempre / sempre

Anand Rajaraman, especialista em mineração de dados, também postou em blog comentário sobre a importância de melhorar a analítica, com a adoção de novas fontes de dados:

Leciono em curso sobre mineração de dados em Stanford. Os alunos devem fazer um projeto que envolve algumas atividades não triviais de mineração de dados. Muitos deles optaram por aceitar o desafio da Netflix: elaborar algoritmo de recomendação de filmes mais eficaz que o desenvolvido pela empresa.

Eis como a competição funciona: A Netflix fornece vasto conjunto de dados que mostra como quase meio milhão de pessoas avaliaram cerca de 18 mil filmes. Com base nessas avaliações, sua tarefa é prever como os mesmos usuários avaliariam filmes ainda não avaliados. A primeira equipe a superar a exatidão do algoritmo proprietário da Netflix por certa margem ganha prêmio de US\$1 milhão!

Diferentes equipes de alunos de meu curso adotaram diversas abordagens para a solução do problema, usando algoritmos conhecidos e aplicando ideias inéditas. Os resultados de duas das equipes ilustram aspecto mais amplo. A Equipe A desenvolveu algoritmo muito sofisticado, usando dados da Netflix. A Equipe B explorou algoritmo muito simples, mas acrescentou novos dados, além do conjunto da Netflix: informações sobre gêneros de filmes extraídas do Internet Movie Database (IMDB). Adivinha que equipe obteve melhores resultados? A Equipe B conseguiu resultados bem superiores, próximos dos melhores do leaderboard da

Netflix!<sup>9</sup>

Rajaraman também observa na mesma postagem que nova fonte de dados – links de hipertexto – foi o principal fator diferenciador do algoritmo de busca da Google, em comparação com serviços similares anteriores, que usavam apenas textos das páginas da internet. Em seu algoritmo de propaganda AdWords, altamente lucrativo, a Google também acrescentou outros dados que ninguém usava na época: a taxa de click-through em cada anúncio.

Rajaraman e Ramakrishnan argumentam que mais e melhores dados quase sempre superam melhor algoritmo. Eles se referem a negócios on-line e de varejo; dispõe-se de muitos outros exemplos, entretanto, em que dados diferentes se mostram mais eficazes. Daryl Morey, gerente geral da NBA Houston Rockets, é um dos gestores mais analíticos do basquete profissional (descreveremos alguns dos trabalhos dele em exemplo do [Capítulo 6](#)). Ele argumenta que “a verdadeira vantagem decorre de dados exclusivos”, e emprega numerosos analistas que classificam as manobras defensivas dos jogadores adversários em todos os jogos da NBA.<sup>10</sup> Morey também se inclui entre os líderes da NBA que estão começando a analisar os vastos arquivos de vídeos gravados em muitos jogos. Em seguros, um dos fatores que há muito tempo distinguem a Progressive em relação a outras seguradoras de automóveis menos analíticas é a diferenciação dos dados. Ela foi pioneira no uso de escores de crédito FICO (como também se descreve em exemplo do [Capítulo 4](#)), entre as variáveis dos modelos de precificação de prêmios de seguros, da mesma maneira como não é de hoje que adota mais variáveis e dados que os concorrentes na análise e precificação dos riscos dos clientes. A Progressive também foi desbravadora na coleta de dados sobre hábitos dos clientes na direção de veículos (com a permissão deles, evidentemente) e na precificação dos prêmios com base nesses dados, por meio de programa hoje denominado “Snapshot”. Muita gente talvez relute em revelar à seguradora seus comportamentos ao volante, mas quem se mostrar bom motorista talvez consiga prêmios mais baixos.

## Valor dos dados secundários

Muitos analistas coletam dados para depois analisá-los. Às vezes, contudo, é possível usar dados coletados por outrem (denominados *dados secundários*). O

uso de dados secundários poupa tempo, que, do contrário, seria consumido no trabalho redundante de reunir dados já disponíveis. Fontes comuns de dados secundários são censos, pesquisas, arquivos de diversas organizações e assim por diante. O mundo está cheio desses dados, à espera de serem analisados.

Em alguns casos, recorre-se a dados secundários para produzir resultados muito importantes. Veja, por exemplo, o trabalho do astrônomo Johannes Kepler. Embora oriundo de família pobre, e tendo crescido em condições adversas, Kepler teve sorte suficiente para reunir dados secundários muito exatos, acumulados cuidadosamente durante várias décadas, sobre os movimentos dos objetos na esfera celeste. Graças à própria sorte e ao talento matemático superior, Kepler resolveu o mistério dos planetas.

Os dados de Kepler foram coletados originariamente por Tycho Brahe (1546-1601), nobre dinamarquês e astrônomo brilhante, que fez as mais exatas observações astronômicas da época, ao conceber os mais precisos instrumentos disponíveis antes da invenção do telescópio. Com o generoso apoio do rei da Dinamarca, Brahe construiu um centro de pesquisa, denominado Uraniborg (castelo dos céus), que se tornou o melhor observatório da Europa. Para tanto, desenvolveu e produziu novos instrumentos, calibrou-os e dedicou-se a meticulosas observações noturnas, ao longo de 20 anos.

Em 1600, Brahe convidou Kepler – professor brilhante, mas desfavorecido – para tornar-se seu assistente. Kepler e Brahe não se davam bem; seus antecedentes e personalidades eram muito diferentes. Brahe receava que o assistente, jovem e inteligente, o eclipsasse como principal astrônomo da época. No ano seguinte, em 1601, Brahe de repente ficou doente e morreu. Seguiu-se uma corrida pelos bens do falecido, e Kepler concluiu que, se não agisse com rapidez, jamais teria acesso a grande parte dos dados de Brahe. Imediatamente se apossou das observações sob seus cuidados (em suas próprias palavras, “usurpou-as”) e as manteve sob controle. Dois dias depois dos funerais, Kepler foi nomeado sucessor do mestre no posto de matemático imperial. Agora, a incomparável coleção de observações astronômicas de Brahe estava nas mãos de Kepler. Utilizando os dados, Kepler acabou descobrindo que as órbitas dos planetas eram elípticas e formulou as três leis do movimento planetário.<sup>11</sup>

Evidentemente, muitos são os casos de uso mais contemporâneo de dados secundários. Veja, por exemplo, a empresa Recorded Future. Sua fonte de dados secundários é muito conhecida – a internet. Essa empresa, fundada por



Christopher Ahlberg, especialista em analítica, analisa a internet para contar e classificar a frequência com que se mencionam entidades e eventos. Seu foco específico é a contagem de previsões – referências ao futuro. Ela vende dados e análises a órgãos de inteligência do governo, cujo interesse óbvio é a frequência com que ocorrem termos como terrorismo ou guerra; e a instituições financeiras, cuja atenção se volta para palavras que indiquem os sentimentos dos investidores e consumidores.

## Dados primários

Se, porém, você não tiver a sorte de Kepler ou da Recorded Future e não herdar dados secundários valiosos nem encontrar dados diretamente relevantes para seu problema, as variáveis podem ser medidas pelo pesquisador (*dados primários*). Vários são os tipos de medidas: as pesquisas incluem projeto e implementação de entrevistas ou de questionários; técnicas de observação envolvem observadores, atuando de maneira direta e irrestrita; experimentos de cientista maluco, cuidadosamente projetados e controlados, não raro se prestam a problemas específicos. Os métodos a serem usados em seu estudo dependem das características do problema reconhecido e das variáveis selecionadas.

**DADOS ESTRUTURADOS E NÃO ESTRUTURADOS.** Durante séculos, quase todas as análises quantitativas eram executadas com dados *estruturados* – ou seja, dados em forma numérica que facilmente podiam ser organizados na forma de fileiras e colunas. Qualquer que fosse o instrumento de análise (planilha eletrônica, poderoso pacote estatístico ou calculadora manual, fileiras e colunas (com as fileiras quase sempre representando casos ou observações; e as colunas, as diversas variáveis) era assim que se estruturavam os dados. Praticamente as únicas perguntas a serem feitas eram a magnitude provável dos números e a quantidade de casas decimais necessárias. Tudo isso começou a mudar com o advento da análise textual, nos últimos anos do século XX. Como descrevemos no exemplo das cartas de Mark Twain, os pesquisadores começaram a buscar padrões em textos e em números. A frequência com que ocorriam certas palavras era a indagação típica. Textos são exemplos de *dados não estruturados* – como geralmente se apresentam em fluxo contínuo, é difícil organizá-los na forma de linhas e colunas.

Só depois de 2000, contudo, os dados não estruturados realmente começaram a explodir em volume e em variedade. Esse ano marcou o começo do uso generalizado da internet, o que levou a volumes maciços de textos, imagens e fluxos de cliques, a serem analisados por organizações como Recorded Future. As telecomunicações e as redes sociais passaram a gerar grandes quantidades de dados de cunho social. A enxurrada de dados de áudio e vídeo que as organizações queriam analisar também experimentou crescimento exponencial por volta dessa época. A revolução genética acarretou enorme massa de dados genéticos e proteômicos.

Agora, entramos oficialmente na era do *Big Data*, em que as organizações lidam no dia a dia com vários petabytes (1.000 terabytes, ou  $10^{15}$  bytes – ou seja, 1.000.000.000.000.000 unidades de dados). A eBay, por exemplo, tem um armazém de dados que abrange mais de 40 petabytes. Sempre que alguém clica numa câmera usada ou num vaso floral, o total aumenta.

A análise desses dados é, em geral, muito diferente da de dados numéricos estruturados nos estágios iniciais. Em muitos casos, é preciso fazer muitas filtrações, classificações e outras formas de triagem, antes de contar os dados. O cientista de dados é especialista não só em analisar dados, mas também em dar-lhes forma suscetível de análise. Ferramentas como Hadoop e MapReduce, já em uso incipiente por muitas organizações que trabalham com *Big Data*, destinam-se a filtrar e a classificar dados a serem submetidos a análise quantitativa. Dados de vídeo e voz também demandam preparação substancial antes de serem analisados por métodos quantitativos. Muitas vezes, depois disso, a organização analisa os dados com software estatístico tradicional. É como Bill Franks, da Teradata, observa numa postagem em blog do International Institute for Analytics:<sup>12</sup>

De uns tempos para cá, dados não estruturados tornaram-se tema muito popular, porquanto muitas fontes de *Big Data* são não estruturadas. Não raro, contudo, se perde uma nuance importante – o fato é que praticamente nenhuma ferramenta de analítica é capaz de analisar dados não estruturados. Dados não estruturados podem ser input para o processo analítico; quando, no entanto, chega a hora de fazer análises reais, não se usam dados não estruturados. “Como isso é possível?”, você perguntaria. Vou explicar...

Comecemos com o exemplo do cotejo de impressões digitais. Quem assiste a programas como CSI, vê os personagens confrontando impressões digitais o tempo todo. A imagem de impressão digital é totalmente não estruturada e talvez seja muito grande se for de alta qualidade. Portanto, na televisão ou na vida real, como age a polícia ao analisar impressões digitais? Confronta ou busca imagens reais ou imagens compatíveis? O que fazem os peritos é primeiro identificar um conjunto importante de pontos em cada impressão digital. Criam, então, um mapa ou polígono com esses pontos. O que confrontam é o mapa ou polígono daí resultante. Mais importante é o fato de o mapa ou polígono ser totalmente estruturado e reduzido, diferentemente das impressões digitais originais. Embora as impressões digitais não estruturadas sejam input do processo, a análise em si, ao confrontá-las, não usa imagens não estruturadas, mas, sim, informações estruturadas, delas extraídas.

Exemplo que todos apreciarão é a análise de textos. Consideremos o método hoje popular de análise de sentimento das mídias sociais. São os tweets, as postagens no Facebook, e outros comentários nas mídias sociais analisados diretamente para captar o sentimento subjacente? Na verdade, não. O texto é dissecado em palavras e frases. Em seguida, esses componentes são sinalizados como bons ou maus. Num exemplo simples, talvez uma palavra “boa” receba “1”; uma palavra “má”; “-1”; e uma palavra “neutra”, “0”. O sentimento da postagem é determinado pela soma da marcação de cada palavra ou frase. Portanto, o escore do sentimento resulta de dados numéricos totalmente estruturados, que foram extraídos de texto a princípio não estruturado. Qualquer análise posterior sobre tendências ou padrões de sentimento se baseia integralmente em resumos numéricos estruturados do texto, não do texto em si.

Do mesmo modo como os exemplos de Frank, acima, muitos programas aplicativos de *Big Data* se concentram de início em dados não estruturados; que, no entanto, depois de serem processados por ferramentas como Hadoop e MapReduce, podem ser analisados como dados estruturados, mediante analítica visual comum ou software de estatística.

## Passo da análise de dados

Uma vez que os dados em si não dizem nada, precisamos analisá-los para decifrar seus significados e relações. A análise de dados resulta na descoberta de padrões consistentes; em outras palavras, as relações entre as variáveis embutidas nos dados. Quando se identifica o surgimento de padrões, fica mais fácil explicar os números. Quando se extraem esses padrões das variáveis, fica mais fácil resolver o problema. Por exemplo, vamos supor que coletamos dados, mediante pesquisa por telefone, de uma amostra de eleitores, sobre as preferências numa eleição presidencial. Ao analisar os dados, tentamos encontrar padrões por *região, educação, renda, gênero, idade e afiliação partidária*, que indiquem apoio provável a um candidato específico. Para descobrir padrões nos dados, dispõe-se de várias técnicas, desde análises básicas, como gráficos, porcentagens e médias, até métodos estatísticos mais elaborados. As características e a complexidade determinam as técnicas específicas a serem adotadas.

### 5. Análise de dados



Lembre-se dos diferentes tipos de histórias analíticas que descrevemos no [Capítulo 2](#). Caso se esteja apenas contando uma história “Eis o que aconteceu”, a única análise provável a ser feita é alguma forma de relatório, ou seja, uma descrição textual ou gráfica relatando o que se descobriu nos dados, em diferentes categorias. Na melhor das hipóteses, se incluirão algumas *medidas de tendência central*, como médias e medianas. Para tanto, serão necessários softwares que se concentrem basicamente na elaboração de relatórios. Painéis de controle, boletins e alertas são formas de relatórios. Em “Principais fornecedores de software para diferentes tipos de análise”, listamos alguns provedores importantes de software de relatórios.

Embora todos esses fornecedores de software também ofereçam recursos para apresentações gráficas, alguns deles se concentram especificamente em *análise visual interativa*, ou no uso de representações visuais de dados e relatórios. Em

geral, essas ferramentas são usadas para gráfico de dados e para *descoberta de dados* – compreender a distribuição dos dados, identificar *outliers* (valores atípicos) e relações visuais entre variáveis. Listamos, portanto, esses fornecedores como categoria à parte.

Também incluímos alguns fornecedores de software para outra categoria de análise, que denominamos *modelagem quantitativa* ou *estatística*, na qual se tenta usar a estatística para compreender as relações entre variáveis e para extrair inferências da amostra, até a população mais ampla. Analítica preditiva, testes aleatórios e diferentes formas de análise de regressão se incluem nesse tipo de modelagem. Os fornecedores de software de modelagem quantitativa ou estatística, de um lado, e de software de relatórios, de outro, tendem a ser diferentes, embora as duas categorias estejam convergindo um pouco com o passar do tempo.

## Principais fornecedores de software para diferentes tipos de análise

### SOFTWARE DE RELATÓRIOS

- BOARD International
- IBM Cognos
- Information Builders WebFocus
- Oracle Business Intelligence (inclusive Hyperion)
- Microsoft Excel/SQL Server/SharePoint
- MicroStrategy
- Panorama
- SAP BusinessObjects

### ANALÍTICA VISUAL INTERATIVA

- QlikTech QlikView
- Tableau
- TIBCO Spotfire

### MODELAGEM QUANTITATIVA OU ESTATÍSTICA

- IBM SPSS
- R (pacote de software de código aberto)
- SAS

O Microsoft Excel, por exemplo, talvez o software analítico mais usado no mundo (embora a maioria das pessoas o veja como planilha eletrônica) pode fazer algumas análises estatísticas (e analítica visual), assim como elaborar relatórios, embora não seja o software estatístico mais robusto quando se têm muitos dados a processar e um modelo estatístico complexo a construir, razão pela qual não foi incluído nessa categoria. O uso do Excel para analítica no contexto empresarial não raro é reforçado por outros produtos da Microsoft, como o SQL Server (basicamente, ferramenta de banco de dados, com algumas funções analíticas) e SharePoint (basicamente, ferramenta de colaboração, com algumas funções analíticas).

## Tipos de modelos

Vários são os modelos adotados por analistas e organizações para aplicar raciocínio analítico e para decidir sobre dados. Realmente não é objetivo deste livro ensinar estatística aos leitores, mas é útil conhecer os critérios usados pelos analistas quantitativos para definir os tipos de modelos a empregar. Acreditamos que esse material ajudará os leitores a dar a partida e a apreender a essência da analítica.

Se quisermos compreender os tipos de modelos mais eficazes em determinada situação, primeiro precisamos esboçar as características da situação com que se defrontam os decididores (e seus analistas). Três são as questões a serem consideradas na identificação do modelo adequado:

- Quantas variáveis serão analisadas ao mesmo tempo? As possibilidades de análise envolvem uma variável (*modelo univariado*), duas variáveis (*bivariado*) e três ou mais variáveis (*multivariado*). Essas hipóteses cobrem todas as opções.
- Queremos responder a questões de descrição ou de inferência? A *estatística*



*descritiva* apenas descreve os dados disponíveis, sem a intenção de generalizar com base neles. Médias, medianas e desvios-padrão são exemplos de ferramentas de estatística descritiva. Geralmente são úteis, mas não são muito interessantes sob perspectiva estatística ou matemática. A *estatística inferencial* extrai uma amostra de dados e tenta inferir ou generalizar as descobertas para uma população mais ampla. Análises de correlação e de regressão (ver a seguir) são exemplos de estatística inferencial, pois incluem estimativas da probabilidade de as relações observadas na amostra se manterem válidas em população mais ampla. Os estatísticos e os analistas quantitativos vibram muito mais com a estatística inferencial que com a estatística descritiva.

- De que níveis de medida se dispõe nas variáveis de interesse? A resposta já foi apresentada em “Maneiras de medir variáveis”, neste mesmo capítulo.

O tipo de modelo a ser adotado é determinado pela espécie de história e pelos dados disponíveis. Alguns atributos das histórias e dos dados – e os tipos de modelos deles resultantes – são descritos a seguir. Eles não representam todas as variações possíveis nos tipos de modelos, mas provavelmente descrevem cerca de 90% do que as organizações fazem com a analítica no dia a dia.

- *Duas variáveis de dados numéricos*: Caso se queira relacionar dois fatores quantificáveis, é provável que se use algum tipo de *análise de correlação*, uma das análises estatísticas mais simples. Basicamente, ela estima se duas variáveis – como, por exemplo, estatura e peso de uma série de pessoas – variam juntos. O peso aumenta à medida que a estatura sobe? Sim, em geral peso e altura variam juntos, razão por que se diz que peso e altura apresentam forte correlação. Como a correlação é inferência estatística, alguns testes determinam a probabilidade de certo nível de correlação ocorrer por acaso. Quando se afirma que uma correlação é “significativa no nível de significância .05”, por exemplo, conclui-se que, em apenas 5 casos em 100 a relação constatada na amostra ocorreria por acaso na população mais ampla.
- *Duas ou poucas variáveis categóricas*: Caso se usem dados de pesquisa que se incluem em *categorias nominais* (por exemplo, gênero masculino ou

feminino; jovem / meia-idade / idoso), deve recorrer-se a uma série de técnicas para análise de dados categóricos. Os resultados dessas análises geralmente são apresentados em tabelas, incluindo-se os números referentes às observações nas células da tabela. Por exemplo, caso se esteja relacionando gênero e idade, é provável que se encontrem números relativamente iguais de homens e de mulheres nas categorias “jovem” e “meia-idade”; como, porém, as mulheres geralmente vivem um pouco mais que os homens, é possível que se encontrem mais mulheres nas células de idosos. Na hipótese de esse ou outro padrão ser nítido nos dados, talvez se conclua que a tabela de dados mostra relação significativa (improvável de resultar do acaso), de acordo com o teste do chi-quadrado (qualidade do ajuste). Novamente, a relação pode ser significativa em nível de significância de .05 ou .01. Variáveis categóricas binárias, como gênero, também podem ser manejadas com a análise de regressão (ver a seguir), usando variáveis “dummy”, ou aquelas que podem ser codificadas como “0”, para a ausência de um fator (por exemplo, “masculinidade”), ou como “1”, para a presença do fator.

- *Mais de duas variáveis de dados numéricos:* A extensão da análise de correlação além de duas variáveis numéricas é denominada *análise de regressão* – às vezes, *regressão múltipla* (porque se usam muitas variáveis para explicar o valor de outra variável) ou *regressão linear* (porque a relação entre as variáveis se mantém semelhante ao longo de toda a faixa de variáveis). Regressão é um método de ajustar uma equação (ou linha, em termos gráficos) a uma distribuição de dados coletados no passado, mas que serve para prever o futuro, desde que se tenha um bom modelo. O modelo de regressão atribui coeficientes a cada variável que explica ou prevê o valor de outra variável.

Como exemplo de regressão linear múltipla, vejamos como Orley Ashenfelter, economista de Princeton, usou a análise de regressão para descobrir o que prevê os preços de leilão de valiosos vinhos franceses. Essa previsão dos preços dos vinhos, com base no clima durante a estação de cultivo, gerou consternação entre os enólogos, e até os enfureceu. (O *New York Times* publicou artigo de primeira página sobre a previsão, com o título “Wine Equation Puts Some Noses Out of Joint”<sup>13</sup> [Equação do vinho torce alguns narizes].) Quando se tem uma

boa equação, quem precisa de especialistas?

A maioria dos observadores pelo menos concorda que se produzem bons vinhos (a) quando o inverno anterior foi úmido, (b) quando a estação de cultivo é quente e (c) quando a estação de colheita é seca. Assim, como variáveis de clima que afetam a qualidade da safra, Ashenfelter escolheu três variáveis independentes: (a) temperatura média no cultivo, (b) pluviosidade na colheita e (c) pluviosidade no inverno anterior. Além disso, uma vez que os vinhos se tornam mais saborosos com o envelhecimento, também usou antiguidade da safra como variável independente.

A qualidade da safra se reflete no preço dos vinhos maduros, a variável dependente que Ashenfelter tentava prever. Coletaram-se, então, os preços de seis Bordeaux Chateaux no mercado de leilões de Londres, de 1960 a 1969. Selecionou-se esse período, uma vez que os vinhos já estavam totalmente maduros e poucas incertezas restavam sobre sua qualidade. As variáveis climáticas foram obtidas nos órgãos meteorológicos locais. Ashenfelter fez a regressão do preço (logaritmo) da safra com a idade da safra e com as variáveis climáticas, obtendo a seguinte equação do vinho:

$$\text{Qualidade do vinho} = 12,145 (\text{constante}) + .0238 \text{ idade da safra} + 0,616 \text{ temperatura média no cultivo} - 0,00386 \text{ pluviosidade na colheita} + 0,00117 \text{ pluviosidade no inverno}.$$

Como indicam os sinais dos coeficientes, antiguidade da safra, temperatura moderada no cultivo e pluviosidade no inverno anterior exercem efeito direto e positivo sobre a qualidade do vinho. Já a pluviosidade na colheita se associa negativamente com a qualidade do vinho. O R ao quadrado ou  $R^2$  (ver mais a seguir) da equação foi 0,828, mostrando que essas variáveis explicam 83% da variação nos preços das safras. Em síntese, elas, sozinhas, exercem papel muito importante na determinação do preço. Vê-se por que os especialistas devem ter achado os resultados um tanto ameaçadores e menos interessantes que falar sobre *terroir*, tonéis de carvalho e uvas maduras demais.

Em “Principais conceitos e técnicas de estatística”, descrevemos os modelos estatísticos inferenciais mais comuns (como observamos, os modelos estatísticos descritivos também são úteis, mas não muito interessantes no sentido quantitativo). Há, evidentemente, livros inteiros sobre a matéria, e o que

apresentamos aqui não passa de mera introdução.

## Principais conceitos e técnicas estatísticas<sup>a</sup>

**Análise de variância:** Teste estatístico para verificar se a média de mais de dois grupos são todas iguais.

**Causalidade:** Relação entre um evento (causa) e um segundo evento (efeito), na qual o segundo evento é considerado consequência do primeiro. No uso comum, causalidade é também a relação entre um conjunto de fatores (causas) e um fenômeno (efeito). As três condições da causalidade são:

- A causa deve preceder o efeito no tempo e no espaço.
- A causa deve estar presente quando o efeito reage.
- A causa deve estar ausente quando o efeito é não reativo.

**Clustering ou análise de cluster:** Reunião de observações (por exemplo, registros em um banco de dados) em grupos (denominados *clusters*) de modo que os objetos dentro dos clusters sejam semelhantes de alguma maneira, enquanto os objetos entre os clusters sejam dessemelhantes entre si. Clustering ou análise de clusters é importante tarefa da mineração de dados exploratória, e técnica comum de análise estatística de dados, usada em muitos campos.

**Correlação:** Extensão em que duas ou mais variáveis se relacionam entre si. O grau de relacionamento se expressa como coeficiente de correlação, que varia de  $-1,0$  a  $+1,0$ .

Correlação =  $+1$  (correlação positiva perfeita, significando que ambas as variáveis sempre se movimentam juntas na mesma direção)

Correlação =  $0$  (não há relação entre as duas variáveis)

Correlação =  $-1$  (correlação negativa perfeita, significando que quando uma variável sobe, a outra tende a cair).

Correlação não implica causalidade. Correlação é condição necessária mas não suficiente para conclusões causais.

**Variável dependente:** Variável cujo valor, por enquanto desconhecido, se quer prever ou explicar. Por exemplo, caso se queira prever a qualidade de uma safra de vinho, usando a temperatura média no cultivo, a pluviosidade na colheita, a pluviosidade no inverno anterior e a antiguidade da safra, a qualidade da safra de vinho seria a variável dependente. Outros nomes são *variável explicada* e *variável de resposta*.

**Análise de fatores:** procedimento estatístico que testa grande número de variáveis ou objetos e descobre as relações subjacentes entre elas. Isso possibilita que numerosas variáveis inter-relacionadas se condensem em menos dimensões, denominadas fatores.

Geralmente é usado para a redução de dados ou identificação de estruturas. Por exemplo, se um pesquisador tivesse mais de 100 variáveis de interesse para o estudo, a análise de fator possibilitaria o desenvolvimento de critério composto capaz de capturar a essência das 100 variáveis em apenas um punhado de medidas ou fatores compostos.

**Teste do chi-quadrado (qualidade do ajustamento):** Teste estatístico que determina se os dados da amostra se encaixam em tipo de distribuição especificados. As medidas de qualidade do ajustamento resumem a discrepância entre valores observados e valores esperados sob a distribuição especificada. O teste de qualidade do ajustamento de uso mais frequente é verificar se as frequências dos resultados seguem determinada distribuição.

**Teste de hipótese:** Método sistemático para avaliar crença experimental (pretensão) a respeito da realidade. Consiste em confrontar a crença ou pretensão com evidências, e decidir, à luz dessas evidências, se a crença pode ser mantida como razoável ou deve ser descartada como insustentável. A pretensão é dividida em duas hipóteses concorrentes: a hipótese nula e a hipótese alternativa. A *hipótese nula* ( $H_0$ ) sugere que não há diferença ou relação estatística significativa entre determinado conjunto de observações. A *hipótese alternativa* ( $H_a$  ou  $H_1$ ) sugere que a mudança ou relação que queremos mostrar de fato existe nos dados. No teste de hipóteses, comparam-se empiricamente descobertas observadas na amostra com descobertas esperadas teoricamente – ou seja, esperadas se a hipótese nula for verdadeira. Por exemplo, caso se queira prever a qualidade de uma safra de vinho com base na antiguidade, uma hipótese nula poderia ser “a antiguidade da safra não é preditor significativo da qualidade do vinho”, enquanto a hipótese alternativa poderia ser “a antiguidade da safra é preditor significativo da qualidade do vinho”. Coletam-se e testam-se dados para ver quão inusitada seria a situação sob a presunção temporária de que  $H_0$  é verdadeira. Dados raros ou insólitos (em geral representados por um *valor-p* abaixo de limiar especificado) são indicadores de que  $H_0$  é falso, o que constitui resultado *estatisticamente significativo* que apoia a hipótese alternativa.

**Variável independente:** Variável cujo valor é conhecido e usado para ajudar a prever ou explicar uma variável dependente. Por exemplo, caso se queira prever a qualidade de uma safra de vinho, usando vários preditores (temperatura média no cultivo, pluviosidade na colheita, pluviosidade no inverno anterior e antiguidade da safra), estes serviriam como variáveis independentes. Nomes alternativos são *variável explanatória*, *variável previsor* e *regressor*.

**Valor-p:** Quando se testa uma hipótese, o valor-*p* dá a probabilidade de ocorrência de dados sob a presunção de que  $H_0$  seja verdadeira. Pequenos valores-*p* são indicadores de dados raros ou inusitados sobre  $H_0$ , o que, por seu turno, reforça que  $H_0$  seja realmente falso (e, em sentido contrário, que a hipótese alternativa seja verdadeira). No teste de hipóteses, “rejeitamos a hipótese nula” quando o valor-*p* é inferior ao nível de significância  $\alpha$  (alfa grego), que, em geral, é 0,05 ou 0,01. Quando se rejeita a hipótese nula, diz-se que o resultado é estatisticamente significativo.

**Regressão:** Qualquer método estatístico capaz de oferecer uma equação para possibilitar a estimativa do valor desconhecido de uma variável dependente a partir do valor conhecido de uma ou mais variáveis independentes. A *regressão simples* usa uma variável independente para prever uma variável dependente. A *regressão múltipla* usa múltiplas variáveis independentes para prever uma variável dependente. A *regressão*

*logística* usa múltiplas variáveis independentes para prever uma variável dependente categórica binária (por exemplo, sim / não, compra / não compra, prós / contras).

**R ao quadrado ou  $R^2$ :** Medida mais popular de quão bem uma linha de regressão estimada ajusta os dados da amostra em que se baseia. Também indica a variabilidade da variável dependente explicada pela linha de regressão. É proporção que varia entre 0 e 1, e se, por exemplo, é 0,52, isto significa que 52% da variância da variável dependente é explicada pelas variáveis independentes usadas na regressão. Em geral, quanto mais alto o valor, melhor o modelo.

**Nível de significância ou alfa/ $\alpha$ :** Entre todos os resultados possíveis da amostra quando a hipótese nula é verdadeira, a proporção máxima (arbitrária) desses resultados considerada inusitada o suficiente para rejeitar a hipótese nula denomina-se nível de significância. Em outras palavras, o nível de significância indica a quantidade de evidências necessárias para aceitar a improbabilidade de certo evento ter ocorrido por acaso (contradizendo, assim,  $H_0$ ). O nível de significância tradicional é 5% (0,05); podem-se usar, contudo, valores mais rigorosos em situações que demandem evidências mais convincentes para se aceitar a hipótese alternativa (por exemplo,  $\alpha = 1\%$  [0,01]). O valor de 5% significa que precisamos de dados que decorram menos que 5% das vezes de  $H_0$  (se  $H_0$  for de fato verdadeira) para se duvidar de  $H_0$  e rejeitá-la como verdadeira. Na prática, estima-se essa condição calculando o valor- $p$ , que, quando inferior a alfa, indica que se deve rejeitar  $H_0$  e apoiar a hipótese alternativa.

**Teste t ou teste de Student:** Teste estatístico que verifica se as médias de dois grupos são iguais, ou se a média de um grupo tem valor específico.

**Erro do tipo I ou erro  $\alpha$ :** Esse erro ocorre quando a hipótese nula, embora verdadeira, é rejeitada. No teste de hipótese tradicional, rejeita-se a hipótese nula se o valor- $p$  for menor que o nível de significância  $\alpha$ . Portanto, a probabilidade de rejeitar incorretamente uma hipótese nula verdadeira é igual a  $\alpha$  e, portanto, esse erro também é denominado erro  $\alpha$ .

a. Para as descrições desta seção, recorremos às definições em *Wikipedia; Statistics for Business and Economics* (2002) de Heinz Kohler; e *Analytics Cheat Sheet* (2012, Tabelas 6 e 8), de Dell.

## Mudando o modelo

Como é de supor, nenhum modelo dura para sempre. Se o mundo mudou de alguma maneira relevante, grandes são as chances de que o modelo não mais seja representação adequada da realidade. Mais adiante, neste livro, mostraremos a importância das premissas nos modelos, assim como da divulgação delas, para que todos os interessados tenham condições de avaliar até que ponto ainda são aplicáveis. Por enquanto, basta dizer que quaisquer organizações ou indivíduos que lidem com modelos quantitativos devem revê-los com regularidade, para

certificar-se de que ainda fazem sentido e que ainda são compatíveis com os dados – mudando-os, caso contrário. Com regularidade, para nós, significa pelo menos uma vez por ano, ou algo parecido, se não houver motivos para maior frequência.

Em alguns contextos, os modelos devem ser atualizados com muito mais frequência. Por exemplo, caso o modelo sirva de base para operações financeiras, provavelmente será necessário examiná-los com muita assiduidade. James Simons, proprietário da Renaissance Technologies, que dirige um dos maiores fundos de hedge do mundo, muda seus modelos o tempo todo. Para tanto, contrata professores, decifradores de códigos, cientistas e engenheiros com mentalidade estatística. Desde sua criação, em março de 1988, a nau capitânia de Simon, o Medallion Fund, de US\$3,3 bilhões, que opera com todos os ativos, desde futuros de soja até bônus do governo francês, gerou retornos anuais de 35,6%. Durante 11 anos ininterruptos, findos em dezembro de 1999, o retorno acumulado do Medallion chegou a espantosos 2.478,6%. Estima-se que Simon tenha auferido US\$2,5 bilhões, em 2008. Com patrimônio líquido avaliado em US\$8,7 bilhões, ele foi classificado pela *Forbes* como a 80ª pessoa mais rica do mundo e a 29ª mais rica dos Estados Unidos. Em 2006, o *Financial Times* o elegeu “o bilionário mais inteligente do mundo”.<sup>14</sup>

Simon reconhece que as oportunidades de operações de mercado, pela própria natureza, são pequenas e fugazes. Eis como se manifestou num seminário: “A teoria dos mercados eficientes está certa na suposição de que neles não ocorrem grandes ineficiências. Buscamos, porém, anomalias pequenas no potencial e breves na duração. Fazemos nossas previsões. Pouco depois, reavaliamos a situação e revisamos as previsões e nosso portfólio. Fazemos isso durante todo o dia. Estamos sempre entrando e saindo, saindo e entrando. Portanto, dependemos dessa atividade constante para ganhar dinheiro.”

Para manter-se à frente, Simons muda seus modelos todas as semanas. As coisas mudam, e essa capacidade de ajustar-se é que fez de Mr. Simons tamanho sucesso. Diz ele: “Os indicadores estatísticos se desgastam ao longo dos anos; podem ser 5 ou 10 anos. É preciso acompanhar as novidades, porque o mercado atua contra nós. Quem não continua melhorando começa a piorar.”

## Exemplo de raciocínio analítico: Modelo Black-Scholes de precificação de opções



Fischer Black e Myron Scholes resolveram um problema de avaliação de ações que durante muito tempo atormentou os investidores.<sup>15</sup> Black, PhD por Harvard em Matemática Aplicada, trabalhava, na época, na empresa de consultoria Arthur D. Little, Inc.; Sholes, PhD em Economia, por Chicago, era membro, havia pouco, do corpo docente do Departamento de Finanças do MIT.

É vasta a terminologia especializada em precificação de opções. *Opção* é título mobiliário que dá o direito, mas não gera a obrigação, de comprar ou vender um ativo, sob certas condições, dentro de determinado período. O preço pago pelo ativo ao se exercer a opção é denominado *preço de exercício*. O último dia para o exercício da opção é chamado *data de vencimento*. O tipo mais simples de opção, geralmente denominada *opção de compra*, dá o direito de comprar uma única ação ordinária. *Prêmio de risco* é a quantia que o investidor paga por uma ação ou por outro ativo acima do preço do investimento sem risco.

Em geral, quanto mais alto for o preço da ação, maior será o valor da opção. Quando o preço da ação é muito maior que o preço de exercício, é quase certo que a opção será exercida. Por outro lado, se o preço da ação for muito menor que o preço de exercício, é quase certo que a opção vencerá, sem ser exercida, razão por que seu valor será quase zero. Se a data de vencimento da opção estiver muito longe no futuro, o valor da opção será aproximadamente igual ao preço da ação. Normalmente, o valor da opção diminui à medida que se aproxima a data de vencimento, desde que o valor da ação não mude. O valor do prêmio de risco, porém, é incerto.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Pré-requisito para a gestão eficiente do risco de opções e de outros títulos derivativos é que se avaliem, ou que se precifiquem, corretamente esses instrumentos. Tentativas anteriores de avaliar derivativos foram falhas sob vários aspectos; precisava-se de novo método, teoricamente rigoroso e empiricamente eficaz, para determinar o valor de derivativos.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Longa é a história das tentativas de avaliar derivativos, remontando aos idos de 1900. A maioria dos trabalhos anteriores sobre avaliação de opções foi sobre bônus de subscrição (opções de compra emitidas por empresas que conferem ao titular o direito de comprar ações da emissora a preços especificados), todos propondo fórmulas de avaliação do mesmo tipo genérico. Estas, contudo, não eram completas, pois envolviam



um ou mais parâmetros arbitrários e padeciam de deficiência fundamental: os prêmios de risco não recebiam tratamento adequado. Infelizmente, parecia não haver modelo de precificação de títulos mobiliários sob condições de equilíbrio do mercado de capitais, que lhes propiciaria oferecer procedimento adequado para a determinação do valor de um bônus de subscrição. Black e Scholes, pela primeira vez na história, tentaram desenvolver fórmula de avaliação teórica, usando a condição de equilíbrio.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Identificaram-se as seguintes cinco variáveis, a afetarem a avaliação de opções:

Prazo de vencimento.

Preço a vista do ativo-objeto.

Preço de exercício.

Taxa de juros livre de risco.

Volatilidade dos retornos do ativo-objeto.

Observe que não se incluiu a atitude do investidor em relação ao risco. A contribuição de Black e Scholes foi vital, ao mostrar que, na verdade, não é necessário considerar prêmio de risco para avaliar uma opção. Isso não significa eliminação do prêmio de risco; ao contrário, ele já está incluído no preço da ação.

**COLETA DE DADOS (MENSURAÇÃO).** O modelo de Black e Scholes se baseia em premissas técnicas e em relações presumidas entre as variáveis. Não se fez nenhuma avaliação no estágio de desenvolvimento do modelo. Black e Scholes, no entanto, realizaram testes empíricos do modelo teórico, aplicando-o a amplo conjunto de dados sobre opções de compra, no trabalho “The Pricing of Options and Corporate Liabilities”.<sup>16</sup>

**ANÁLISE DE DADOS.** Black e Scholes desenvolveram equação diferencial parcial baseada em argumentos e premissas técnicas (modelo de cálculo, não de estatística). A solução dessa equação foi a fórmula Black-Scholes, que sugere como o preço de uma opção de compra podia ser calculado como função da taxa

de juros livre de risco, da variação do preço do ativo sobre o qual a opção foi emitida, e os parâmetros da opção (preço de exercício, prazo e preço de mercado do ativo-objeto). A fórmula introduz o conceito de que quanto mais altos forem o preço da ação hoje, maior a volatilidade do preço da ação e a taxa de juros livre de risco; quanto mais longo for o prazo de vencimento; e quanto mais baixo for o preço de exercício, mais alto será o valor da opção. A avaliação de outros títulos derivativos segue orientação semelhante.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Black e Scholes, de início, tentaram publicar o trabalho, apresentando-o primeiro ao *Journal of Political Economy*, mas ele foi rejeitado de imediato. Ainda convencidos de que o trabalho tinha méritos, submeteram-no à *Review of Economics and Statistics*, onde recebeu o mesmo tratamento. A ideia de que a avaliação de opções podia basear-se exclusivamente em métodos matemáticos, sem considerar as atitudes do investidor em relação ao risco, parecia, na época, estranha e inadmissível para a maioria dos revisores. Depois de alterar o trabalho com base em extensos comentários de economistas famosos, os autores o reapresentaram ao *Journal of Political Economy*, e, finalmente, conseguiram publicá-lo. Em seguida, Robert Merton, então professor do MIT, publicou outro trabalho, ampliando a compreensão matemática do modelo Black-Scholes.

Apesar da dificuldade para publicar o trabalho, milhares de operadores e investidores, hoje, usam a fórmula todos os dias para avaliar opções sobre ações em mercados de todo o mundo. É fácil calcular e modelar explicitamente as relações entre todas as variáveis. Trata-se de aproximação útil, em especial ao analisar a direção dos preços ao transporem pontos críticos. Mesmo quando não são completamente usados, os resultados servem como primeira abordagem a ser ajustada depois.

O modelo de Black e Scholes tornou-se indispensável não só na precificação de opções, mas também na análise de muitos problemas econômicos e é reconhecido como a teoria mais bem-sucedida de toda a economia. Merton e Scholes receberam o Prêmio Nobel de Economia de 1997 pelo desenvolvimento de novo método para determinar o valor de derivativos. Embora inelegível para receber o prêmio, pois havia morrido em 1995, Black foi mencionado como importante colaborador pela academia sueca.

## Exemplo de Raciocínio Analítico: Marido Desconfiado

Em 1973, o seguinte item apareceu na coluna de conselhos “Dear Abby” de um jornal:<sup>17</sup>

*Dear Abby: Você escreveu em sua coluna que a gravidez da mulher dura 266 dias. Quem disse isso? Fiquei grávida durante 10 meses e 5 dias, e não há dúvida a esse respeito, porque sei exatamente a data em que meu bebê foi concebido. Meu marido está na Marinha, e a concepção não poderia ter ocorrido em nenhum outro dia, pois estive com ele durante apenas uma hora, e não o vi até a véspera do nascimento. Não bebo nem ando por aí, e não há como esse bebê não ser dele; portanto, por favor, publique uma retratação sobre a gravidez de 266 dias, pois, do contrário, estarei em maus lençóis.*

*– Leitora de San Diego*

A resposta de Abby foi algo reconfortante, mas não muito quantitativa:

*Cara leitora: O período médio de gestação é de 266 dias. Alguns bebês nascem mais cedo; outros, mais tarde. O seu atrasou.*

Se Abby fosse mais quantitativa, teria escrito uma resposta à “Leitora de San Diego” com mais números. Apresentar números sempre se mostra mais convincente, e este é um problema relativamente simples, envolvendo probabilidade. Analisemos este problema dentro do referencial de seis passos.

**RECONHECIMENTO DO PROBLEMA.** A questão aqui não é se o bebê estava atrasado; esse fato já é conhecido. Dez meses e cinco dias equivale a aproximadamente 310 dias – bem mais que os 266 dias que Abby corretamente sugeriu como média. O ponto é a probabilidade dessa ocorrência ou a extensão em que é inusitada. Em outras palavras: é a situação anormal demais a ponto de sugerir que a mulher não está dizendo a verdade?

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Podemos afirmar com segurança que a distribuição do período de gestação é mais ou menos normal (ou seja, segue a curva normal ou em sino). A probabilidade de que a gestação dure pelo menos 310 dias pode ser calculada com facilidade com base no *escore z* (a distância em

desvios-padrão entre a ocorrência e a média) da distribuição normal, o básico da estatística elementar.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Probabilidade de que a gravidez dure pelo menos 310 dias.

**COLETA DE DADOS (MENSURAÇÃO).** Os dados disponíveis sugerem que o período médio de gestação é de efetivamente 266 dias, com desvio-padrão de 16 dias.

**ANÁLISE DOS DADOS.** Quando a duração média das gestações é de 266 dias, com desvio-padrão de 16 dias, a probabilidade de que uma gestação dure pelo menos 10 meses e 5 dias (300 dias ou mais) é de 0,003, com base na distribuição normal.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Isso indica que três bebês em mil nascerão com esse atraso. A quantidade parece muito pequena, mas não quando se aplica essa probabilidade a grandes números. Como cerca de quatro milhões de bebês nascem todos os anos nos Estados Unidos, mais ou menos 12 mil bebês nascerão com esse atraso. Abby poderia ter respondido nos seguintes termos: “Algo em torno de 12 mil bebês nascem com esse atraso todos os anos nos Estados Unidos. O seu foi um deles.” Essa resposta não só teria sido reconfortante para a esposa, mas também convincente para o marido.

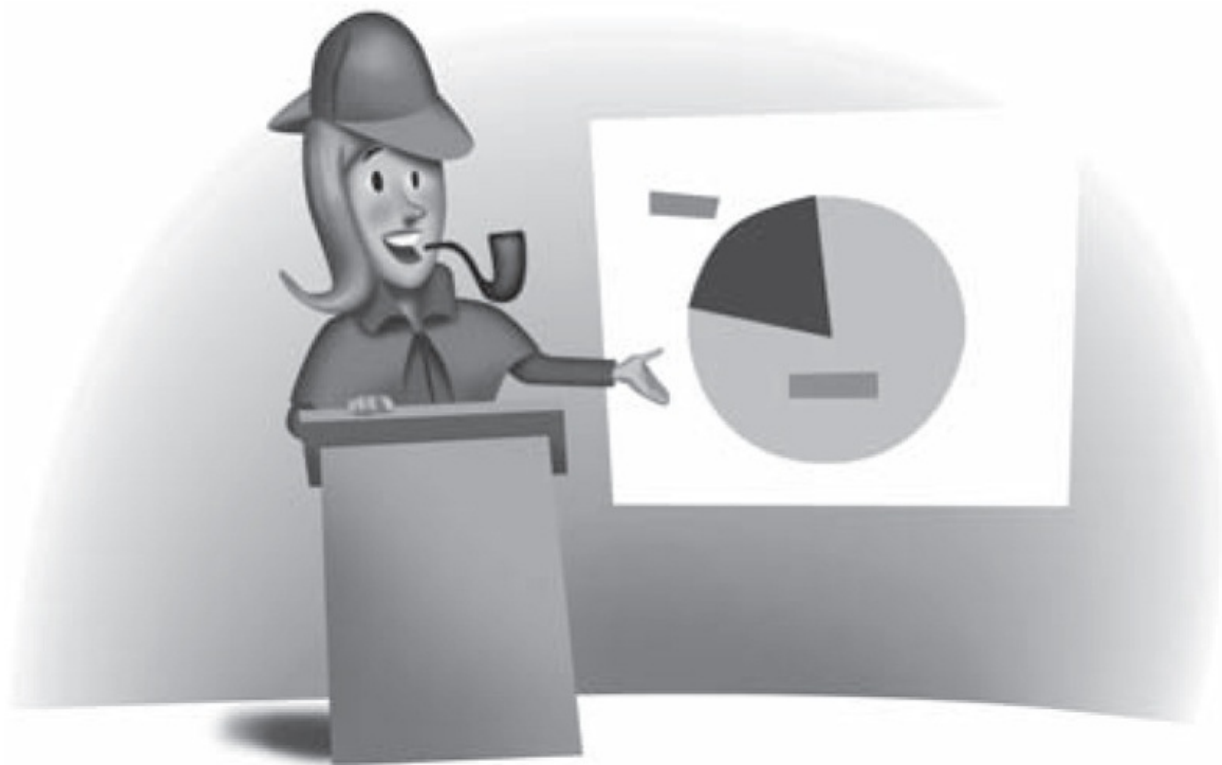
No teste estatístico de hipóteses, a probabilidade de 0,003 calculada acima é chamada valor- $p$  – a probabilidade de obter resultado (por exemplo, valor- $Z$  de 2,75, neste caso) pelo menos tão extrema quanto a ocorrência efetivamente observada (gestação que dura 10 meses e 5 dias ou mais), presumindo que a hipótese nula seja verdadeira. Nesse exemplo, a hipótese nula ( $H_0$ ) é “Este bebê é de meu marido”. Na metodologia tradicional de teste estatístico de hipóteses, rejeita-se a hipótese nula se o valor- $p$  for menor que o nível de significância. Nesse caso, um valor- $p$  de 0,003 resultaria na rejeição da hipótese nula, mesmo no nível de significância de 1% – tipicamente o nível mais baixo que se usa. Normalmente, então, rejeitaríamos a hipótese nula de que o bebê seja filho do marido da Leitora de San Diego. Como interpretar esse resultado de teste (errado)? Trata-se de exemplo típico de um erro tipo I (ou erro  $\alpha$ ), decisão

errada que se toma quando um teste rejeita uma hipótese nula verdadeira ( $H_0$ ). Em outras palavras, às vezes a vida não se enquadra nas previsões da teoria da probabilidade.

## Resultados: Comunicação e ação

A comunicação dos resultados da análise aos stakeholders é o último estágio de nosso referencial de três estágios e seis passos, e é extremamente importante. Mesmo que se tenham executado os outros estágios com brilhantismo, nada de bom acontece se esse passo não for bem conduzido. Os analistas que se preocupam com o aproveitamento de seu trabalho se importam muito com esse estágio e dedicam a ele muito tempo e esforço. Os que não se preocupam com essas considerações – que são, a nosso ver, maus analistas – acreditam que os resultados “falam por si mesmos”, e não dão muita importância a esse estágio.

### **6. Apresentação dos resultados e adoção de providências**



Historicamente, esse não tem sido tema considerado adequado para a educação formal. Os acadêmicos – em especial aqueles com forte orientação analítica nas atividades de pesquisa e de magistério – quase sempre se concentram demais nos métodos analíticos em si e não o suficiente em como comunicá-los com eficácia. Felizmente essa situação está começando a mudar. Xiao-Li Mengt, chefe do departamento de estatística de Harvard (recentemente nomeado reitor da Escola de Pós-Graduação de Artes e Ciências de Harvard), descreveu esse objetivo como desenvolver “comunicadores estatísticos eficazes”:

Nos anos recentes, adotamos visão mais ampla da formação em estatística, no nível de graduação, deslocando o foco, até então concentrado em preparar poucos, que prosseguiriam nos estudos quantitativos de pós-graduação, como PhDs, para ajudar a muitos, que desenvolverão acuidade básica na abordagem e na argumentação estatística, como parte do desenvolvimento do raciocínio crítico em artes liberais. Curiosamente, a jornada, conduzida pela filosofia de que é possível tornar-se *connoisseur* de vinhos, sem ter a menor ideia de como fazer vinhos, aparentemente nos levou a preparar muito mais vinicultores do que quando nos

empenhávamos apenas na produção de safras excelentes.<sup>1</sup>

Com base nessa filosofia, Meng e colegas desenvolveram um curso de graduação denominado Estatística para a Vida Real: as chances de ser feliz (ou infeliz), que abrange módulos sobre a abordagem estatística de questões como “Romance”, “Vinho e Chocolate”, “Finanças”, “Medicina” (inclusive testes clínicos do Viagra) e “Mercado de ações”. Meng está tentando tornar a estatística, “mais que palatável, deliciosa!”<sup>2</sup>

As informações deste capítulo decerto lhe serão úteis, não importa que você seja especialista ou usuário de analítica (em outros termos, vinicultor analista ou consumidor enófilo). Os analistas, evidentemente, podem tornar os resultados de suas pesquisas mais interessantes e atraentes, a fim de inspirar mais ação. Os consumidores de analítica, como, por exemplo, gestores que encomendaram um projeto analítico, devem insistir em receber resultados em formatos interessantes e abrangentes. Se os destinatários da análise quantitativa se mostrarem entediados e confusos, provavelmente não será por culpa deles. Os consumidores de analítica podem trabalhar com os analistas quantitativos na tentativa de tornar os resultados mais compreensíveis e aproveitáveis. E, evidentemente, em geral são os consumidores de analítica que tomam decisões e partem para a ação, com base nos resultados.

A essência desse estágio é descrever o problema e contar a história por trás dele, o modelo, os dados usados e as relações entre as variáveis da análise. Depois de identificar essas relações, o significado delas deve ser interpretado, descrito e apresentado de maneira relevante para o problema. Quanto mais clara for a apresentação dos resultados, maior será a probabilidade de que a análise quantitativa conduza a decisões e a ações – que, em geral, são o objetivo derradeiro da análise em si.

A apresentação dos resultados deve abranger o delineamento do processo de pesquisa, o resumo dos resultados e as recomendações para a solução do problema – embora, provavelmente, não nessa ordem. Geralmente, é melhor começar com o resumo e as recomendações. A melhor maneira de apresentar os resultados é convocar uma reunião com os stakeholders e expor as conclusões, como introito à seção de perguntas e respostas, ou elaborar um relatório formal. Se o problema e os resultados tiverem importância acadêmica, talvez seja o caso de escrever um artigo e tentar publicá-lo no âmbito da respectiva disciplina.



Conforme observamos no [Capítulo 2](#), a apresentação dos dados na forma de tabelas numéricas em preto e branco é maneira muito eficaz de garantir que se ignorem os resultados – mesmo que se trate de mera história analítica do tipo “eis o que aconteceu”. Essa modalidade de narrativa quase sempre pode ser exposta em formato gráfico simples, como de barras ou de setores (pizza), ou algo visualmente mais ambicioso, como telas interativas. Algumas pessoas preferem linhas e colunas de números a exposições mais estimulantes, mas são poucas. Se for possível usar cores e movimentos para animar a apresentação, de maneira a tornar os resultados mais claros e convincentes, tanto melhor.

## Contando uma história com dados

Os analistas mais bem-sucedidos são aqueles que “contam uma história com os dados”. No [Capítulo 2](#), analisamos diferentes tipos de histórias a serem narradas pela analítica. Qualquer que sejam, porém, o tipo de história e o meio de transmiti-la, os elementos de todas as histórias analíticas são semelhantes. Elas se caracterizam por narrativa forte, quase sempre envolvendo problemas ou objetivos de negócios. A apresentação de uma história analítica sobre lealdade dos clientes poderia começar mais ou menos assim: “Como todos vocês sabem, há muito tempo queremos identificar nossos clientes mais leais e descobrir maneiras de torná-los ainda mais leais – e agora podemos fazê-lo.”

As boas histórias apresentam as descobertas em termos compreensíveis para o público. Se este for altamente quantitativo e técnico, expressões estatísticas e matemáticas – até eventuais equações – podem ser usadas. Com mais frequência, porém, o público não será de especialistas, razão por que as descobertas devem ser apresentadas na forma de conceitos que o público tenha condições de compreender e com que seja capaz de identificar-se. Em negócios, isso geralmente toma a forma de ganho ou perda de dinheiro e de retorno sobre o investimento.

As boas histórias terminam com as ações a praticar e com suas consequências previsíveis. Evidentemente, isso significa que os analistas precisam consultar de antemão os stakeholders para discutir vários cenários alternativos. Ninguém quer que um analista quantitativo lhe diga o que fazer.

David Schmitt, chefe de um grupo de analítica da IHG (Intercontinental Hotels Group), cadeia hoteleira global, acredita com convicção na importância

de contar histórias analíticas, e postou em blog alguns dos atributos de uma boa história:<sup>3</sup>

Então, o que faz uma boa história? Sempre que possível, recorro aos especialistas, e hoje os melhores contadores de histórias estão na Pixar, casa de grandes histórias, como *Procurando Nemo*, *Os Incríveis* e, evidentemente, *Toy Story*. Emma Coats, artista de storyboard na Pixar, lançou no Twitter uma lista de 22 regras sobre contação de histórias. Embora todas elas sejam aplicáveis diretamente à analítica, considere as três seguintes especialmente relevantes:

- “Imagine o final, antes de conceber o meio. Falando sério. Os finais são difíceis, pense no final desde o começo.” A conclusão da análise é a única razão pela qual você a está contando. Ao terminá-la, o que você quer que o público compreenda? Mais importante, o que você espera que faça? Siga esse critério como teste de todas as outras partes da história, e só mantenha as que contribuam para o final.
- “Colocar a história no papel o ajuda a fixá-la. Por mais perfeita que seja, se a ideia ficar só na sua cabeça, você não a compartilhará com ninguém.” A história criativa pode nascer na sua cabeça; ao passá-la para o papel, entretanto, ela o levará para lugares que você talvez nunca tenha imaginado. Às vezes não saio do lugar, pensando e repensando uma história, mas depois que me esforço para escrevê-la (papel, Word ou PowerPoint, não importa!) a seiva criativa realmente começa a fluir.
- “Qual é a essência da história? Qual é a maneira mais econômica de narrá-la? A resposta poderá ser seu ponto de partida.” Você deve ser capaz de contar o básico da história em não mais do que três a cinco períodos. Depois de definir esse núcleo, envolvê-lo com detalhes fica muito mais fácil.

Talvez também seja útil desenvolver uma estrutura para a comunicação com os stakeholders, capaz de esclarecer as tarefas dos analistas, do decisor e do patrocinador. Por exemplo, na Intuit, George Roumeliotis chefia um grupo de

ciência dos dados, que analisa e sugere atributos para os produtos, com base em vasta quantidade de dados on-line coletados pela empresa. Em todos os projetos de que o grupo participa com um cliente interno, ele recomenda uma metodologia para executar e para comunicar a análise. Todos os passos se caracterizam por forte orientação para os negócios:

1. Minha compreensão do problema de negócios.
2. Mensuração do impacto para os negócios.
3. Disponibilidade de dados.
4. Hipótese da solução inicial.
5. Solução.
6. Efeitos da solução para os negócios.

Os cientistas de dados que usam essa metodologia são encorajados a criar wiki (software colaborativo), para postar os resultados de cada passo. Os clientes podem rever e comentar o conteúdo do wiki. Roumeliotis diz que, mesmo quando se dispõe de recursos on-line para revisar os resultados, esse recurso encoraja a comunicação direta entre os cientistas de dados e o cliente.

## O que não comunicar

Uma vez que os analistas quantitativos se sentem confortáveis com termos técnicos – descrever os métodos estatísticos usados, especificar os coeficientes de regressão, indicar o nível do  $R^2$  (a porcentagem de variância nos dados que é explicada pelo modelo de regressão usado; ver [Capítulo 3](#)) e assim por diante – eles em geral presumem que também o público se sentirá à vontade. Esse, porém, é um erro trágico. Grande parte do público não compreende apresentação ou relatório altamente técnico. É como disse um analista da IHG: “Ninguém quer saber do seu  $R^2$ .”

Não raro, os analistas são tentados a descrever os resultados analíticos em termos da sequência de atividades que adotaram para alcançá-los: “Primeiro removemos os *outliers* (valores atípicos) dos dados; depois fizemos uma

transformação logarítmica. Daí resultou elevada autocorrelação; o que nos levou a criar 'lag variable' de um ano..." Isso é só uma amostra. Mais uma vez, o público dos resultados analíticos realmente não quer saber do processo adotado – só se importa com os resultados e consequências. É até possível que seja útil oferecer essas informações em um apêndice do relatório ou apresentação, mas não deixe que esses detalhes interfiram na narrativa de uma boa história com os dados – e comece com o que o público realmente precisa saber.

## **Exemplos Históricos de Comunicação de Resultados, Bons e Maus**

A apresentação de resultados quantitativos é técnica usada há muito tempo; e em todas as épocas, como agora, é capaz de convencer o público-alvo ou comprometer de todo a percepção da importância dos resultados. Vejamos um exemplo de cada situação.

### **Florence Nightingale: Bom exemplo de comunicação de resultados**

Florence Nightingale é conhecida como fundadora da profissão de enfermagem e reformadora dos métodos de higienização de hospitais, mas também foi das primeiras usuárias de métodos quantitativos. Quando Nightingale e 38 enfermeiras voluntárias foram enviadas, em outubro de 1854, para um hospital militar britânico na Turquia, durante a Guerra da Crimeia, lá encontraram condições terríveis nas instalações improvisadas. Quase todas as mortes no hospital eram atribuíveis a doenças epidêmicas, endêmicas e contagiosas, não aos ferimentos sofridos em batalha. Em fevereiro de 1855, a mortalidade dos casos tratados no hospital era de 43%. Nightingale identificou com clareza a necessidade de melhorar as condições básicas de higiene no hospital e acreditava que a estatística poderia ajudar a resolver o problema. Passou, então, a coletar dados e a fazer registros, mantendo anotações diárias minuciosas de todas as ocorrências, como entradas, ferimentos, doenças, tratamentos e mortes.

A maior inovação de Nightingale, contudo, foi na apresentação dos resultados. Desde cedo, Nightingale se interessou por números e pela tabulação

de informações. Além de reconhecer a importância de provas baseadas em números, também compreendia que as tabelas numéricas não despertavam grande interesse (mesmo quando eram muito menos comuns do que são hoje!) e que o leitor médio não as leria e, portanto, não entenderia as evidências. Como queria que os leitores recebessem sua mensagem estatística, ela desenvolveu diagramas para dramatizar as mortes desnecessárias provocadas pela falta de higiene e a urgência de providências corretivas (ver [Figura 4.1](#)). Embora rotineiro hoje, tratava-se, na época, de método novo de apresentação de dados.

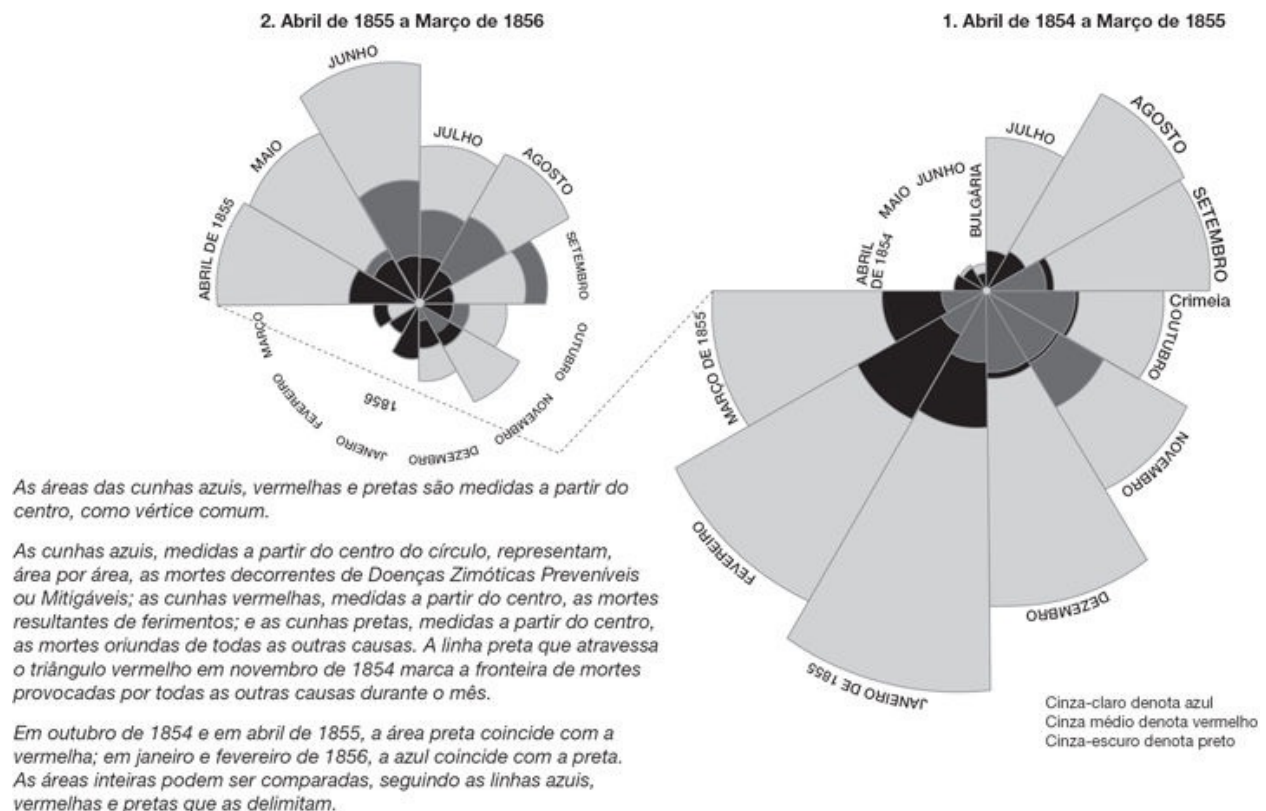
Os diagramas inovadores dela eram uma espécie de gráfico de setores, com os dados representados como que por fatias de pizza. Nightingale os produzia em várias cores, para mostrar com nitidez como a mortalidade resultante de diferentes causas mudava mês a mês. As evidências oferecidas pelos números e diagramas eram claras e inquestionáveis.

Nightingale sempre enviava relatórios à Inglaterra sobre as condições que encontrava, insistindo na necessidade de reformas urgentes. Para tanto, recorria exaustivamente aos diagramas criativos para demonstrar a natureza e a magnitude das condições da assistência médica na Crimeia aos membros do Parlamento, que dificilmente leriam ou compreenderiam apenas as tabelas numéricas. Os destinatários ficavam chocados ao constatar que os soldados feridos morriam nos hospitais, em vez de convalescer. Finalmente, as taxas de mortalidade apresentaram forte queda, como mostram os dados coletados sistematicamente por Nightingale, e se mantiveram em queda. Ao voltar à Inglaterra, em junho de 1856, depois do fim da Guerra da Crimeia, Nightingale foi recebida como celebridade e enaltecida como heroína.

#### **FIGURA 4.1**

---

### **Diagrama de Florence Nightingale sobre as causas da mortalidade no “Exército do Leste”**



Em 1859, tornou-se Fellow of the Royal Statistical Society – primeira mulher aceita na organização –, e, em 1874, ingressou como membro honorário na American Statistical Association. Karl Pearson, estatístico famoso e fundador do primeiro departamento de estatística de uma universidade no mundo, exaltou Nightingale como “profetiza” no desenvolvimento da estatística aplicada.<sup>4</sup>

## Gregor Mendel: Mau exemplo de comunicação dos resultados

Como exemplo menos impressionante de comunicação dos resultados – e lembrete de como o tópico é importante – considere o trabalho de Gregor Mendel.<sup>5</sup> Mendel, pai do conceito de herança genética, afirmou, poucos meses antes de morrer, em 1884: “Meus estudos científicos proporcionaram-me enorme satisfação; e estou convencido de que não demorará muito para que todo o mundo reconheça os resultados de meu trabalho.” O mundo acabou descobrindo o trabalho de Mendel, mas só depois de várias décadas. Se ele tivesse sido mais eficaz na comunicação dos resultados, é possível que a adoção de suas ideias tivesse acontecido com muito mais rapidez – talvez ainda em vida.

Mendel, um monge, foi precursor da nova ciência da genética. A principal teoria em biologia na época era que os traços herdados se misturavam de geração para geração; os estudos de Mendel sobre a herança de certas características mostraram que, em vez disso, a transmissão da informação genética entre as gerações seguia determinadas leis (mais tarde batizadas com o nome dele). O significado do trabalho de Mendel não foi reconhecido até a virada para o século 20; a redescoberta independente dessas leis constituiu os fundamentos da moderna ciência da genética.

Mendel desenvolveu os princípios matemáticos da herança genética por meio de experimentos escrupulosos nas terras do mosteiro. Escolheu para estudar com detalhes a pera comum, cujo cultivo era barato, exigia pouco espaço e proliferava com rapidez (duas gerações por ano). Mendel conseguiu cruzar seletivamente cepas puras com traços específicos e observar os resultados ao longo de muitas gerações.

Entre 1856 e 1863, Mendel dedicou-se com paciência a cruzamentos cuidadosos, isolando cada planta para evitar polinização acidental por insetos. Depois de oito anos de trabalho, Mendel cultivara 12.980 espécimes de peras, oriundos de 225 matrizes originais, dos quais extraiu dados complexos, a fim de detectar padrão ou regularidade que governasse a aquisição de traços herdados. Assim identificou características recessivas e dominantes, mais tarde denominadas genes.

Pena que na comunicação dos resultados Mendel não tivesse sido tão eficaz quanto nos experimentos. Ele publicou suas conclusões em obscuro periódico científico da Morávia. O artigo foi distribuído a mais de 130 instituições científicas, na Europa e além-mar, mas produziu pouco impacto na época, sendo citado apenas três vezes nos 35 anos seguintes. O trabalho complexo e detalhado de Mendel não foi compreendido mesmo por especialistas influentes no mesmo campo. Fosse ele cientista profissional, em vez de monge, talvez tivesse conseguido divulgar seu trabalho com muito mais eficácia, publicando-o, possivelmente, até no exterior. Mendel de fato tentou contatar cientistas de outros países, enviando a Darwin e a outros cópias de seu trabalho; hoje, porém, conhece-se o paradeiro de apenas alguns deles.

Embora, pouco antes da morte, Mendel tenha dito “Minha hora chegará”, não há como saber se ele realmente acreditava nessas palavras; morreu sem jamais saber como suas descobertas mudariam a história.

Não obstante o trabalho de Mendel fosse ao mesmo tempo brilhante e inédito, só depois de 30 anos a comunidade científica descobriu a realidade – apenas no começo do século 20 percebeu-se a importância das ideias dele, e os cientistas da área reconheceram o pioneirismo de Mendel. Qual é a lição aqui? Quem não quiser que os resultados notáveis de seu trabalho analítico seja ignorado durante tanto tempo, ou até mais, deve dedicar muita atenção à comunicação das descobertas.

## Métodos modernos de comunicação dos resultados

Hoje, dispõe-se de várias maneiras de comunicar os resultados, que não existiam na época de Nightingale e de Mendel. Caso se esteja apenas apresentando números e gráficos simples, no papel ou na tela, é provável que não se estejam usando todas as ferramentas disponíveis. Evidentemente, a adequação das ferramentas depende da situação e do público, e ninguém quer recorrer a imagens sensuais como mera apelação.

Felizmente, contudo, a analítica visual (também conhecida como *visualização de dados*) progrediu muito nos últimos anos. Quem recorre apenas a gráficos de barras ou de setores só está arranhando a superfície de todo o potencial das apresentações visuais. O quadro “Propósitos e tipos da analítica visual” talvez não inclua todas as opções possíveis, mas é uma das exposições mais abrangentes que já vimos.<sup>6</sup> Talvez pareça difícil decidir que tipos de gráficos usar, para que propósitos; alguns softwares de analítica visual, porém, até fazem essa escolha para o usuário, com base nas espécies de variáveis a que se referem os dados. O SAS Visual Analytics, por exemplo, é ferramenta que já presta esse serviço, por meio do atributo denominado “Autochart”. Se os dados incluírem, por exemplo, “Uma categoria de data / tempo e qualquer número de outras categorias ou medidas”, o programa gerará automaticamente o gráfico apropriado.<sup>7</sup>

Os tipos de analítica visual listados nas planilhas eletrônicas são estáticos, mas a disciplina se torna cada vez mais dinâmica e interativa. Hans Rosling, professor sueco, popularizou essa abordagem com sua TED Talk, muito vista, em que usou a analítica visual para mostrar a evolução comparativa da saúde pública em países desenvolvidos e em desenvolvimento.<sup>8</sup> Rosling criou um site na internet, denominado Gapminder ([www.gapminder.org](http://www.gapminder.org)), que exhibe muitos exemplos de analítica visual interativa. É provável que, com o passar do tempo,



se desenvolvam muito mais ferramentas de analítica comparativa, capazes de mostrar a movimentação dos dados ao longo do tempo, mas elas não são adequadas ou necessárias para todos os tipos de dados e análises.

Às vezes, é até possível apresentar os resultados de maneira mais tangível que com gráficos. Por exemplo, Vince Barabba, pesquisador e estrategista de marketing, que presta serviços a grandes empresas, como General Motors, Xerox e Kodak (ele não tem culpa dos problemas desses clientes), se revelou pensador criativo ao apresentar a essas empresas os resultados de suas pesquisas de mercado. Na GM, por exemplo, ele sabia que os executivos estavam acostumados a avaliar o potencial de modelos de carros tridimensionais. Assim, a certa altura, quando precisou expor alguns resultados de pesquisas de mercado extremamente importantes, desenvolveu um modelo 3-D dos gráficos, que os executivos podiam ver e sentir sob perspectiva de 360°. Observar e apalpar, de vários ângulos, um “espigão” na demanda do mercado despertava impacto muito mais intenso.

## Propósitos e tipos da analítica visual

### SE VOCÊ QUISER:

Ver as relações entre pontos de dados:

- **Gráfico de dispersão (*scatterplot*):** mostra a relação entre duas variáveis em grade bidimensional.
- **Matriz de dispersão (*matrix plot*):** Mostra relações e frequências de variáveis hierárquicas.
- **Mapa de calor (*heat map*):** Os valores individuais contidos na matriz são representados em cores.
- **Diagrama de rede (*network diagram*):** Mostra as relações entre entidades e a intensidade dos caminhos entre elas.

Comparar um conjunto de frequências ou valores, tipicamente para uma variável:

- **Gráfico de barras (*bar chart*):** O comprimento da barra representa valores.
- **Histogram (*histogram*):** Tipo de gráfico de barras que mostra a frequência dos dados a intervalos especificados.
- **Gráfico de bolhas (*bubble chart*):** Exibe um conjunto de valores numéricos, representados por círculos, com o tamanho dos círculos correspondendo a valores.

Mostrar as altas e baixas de uma variável em relação a outra (tipicamente tempo):

- **Gráfico de linha (*line graph*):** Gráfico bidimensional, quase sempre apresentando uma variável ou variáveis múltiplas, com valores de dados padronizados.
- **Gráfico de pilha (*stack graph*):** Gráfico de linha com áreas preenchidas na parte inferior, tipicamente mostrando mudanças em variáveis múltiplas; também pode expor alterações em várias categorias, por meio de cores diferentes.

Ver as partes de um todo e como elas se relacionam entre si:

- **Gráfico de setores ou de pizza (*pie chart*):** Exibe a distribuição de valores de uma variável no formato de pizza; as porcentagens dos valores correspondem ao tamanho das fatias.
- **Mapa de árvore (*tree map*):** Mostra tamanho dos valores de variáveis hierárquicas, como mundo / continentes / países / população em cada país.

Compreender dados ao longo de área geográfica:

- Sobrepor dados resumidos em mapas geográficos, com cores, bolhas ou espigões, representando diferentes valores.
- Analisar ocorrências em textos:
- **Nuvem de tags (*tag cloud*):** Visualização da frequência de palavras; as palavras usadas com mais frequência aparecem com tipos maiores.
- **Rede de frases (*phrase net*):** Mostra a constância de combinações de palavras usadas juntas; as palavras mais frequentes aparecem com tipos maiores.

O grupo de analítica de David Schmitt, na IHG, é denominado Performance Strategy and Planning, da área de finanças. A equipe deve contar histórias do tipo “eis o que aconteceu”, sobre o desempenho da empresa, na forma de relatórios. Portanto, o foco deles é “apresentar narrativas com dados”, usando todas as ferramentas possíveis, com o objetivo de atrair a atenção e estimular a ação, com base nos resultados. Para tanto, exploram numerosas técnicas, dependendo do público. Uma das abordagens é criar “vídeos de música” – clipes autônomos, de cinco minutos, que transmitem os conceitos amplos por trás dos resultados, usando imagem e som. Prosseguem com uma apresentação verbal que acrescenta informações de apoio, a fim de transmitir os significados subjacentes aos conceitos.

Por exemplo, o grupo de Schmitt criou recentemente um vídeo que descreve as previsões de demanda no verão. Denominado “Summer Road Trip”, ele mostra um carro numa estrada, passando por sinais onde se lê “Demanda alta à

frente”.<sup>9</sup> O objetivo da apresentação era levar o público a refletir sobre quais seriam os principais indutores do desempenho no período vindouro, e como atuariam diferentes em partes do país. Conforme Schmitt observou, “os dados não são o ponto, os números não são o ponto – o que importa são as ideias”. Depois de comunicar a ideia básica, Schmitt faz apresentações mais convencionais para mergulhar nos dados, esperando, porém, que as mentes dos participantes já tenham sido preparadas e condicionadas pelo vídeo que acabaram de ver.

Jogos são outra abordagem para comunicar os resultados e os modelos analíticos. Podem ser usados para transmitir a maneira como as variáveis interagem em relações complexas. Por exemplo, o “Jogo da Cerveja”, simulação baseada nos processos de distribuição de uma empresa de cerveja, foi desenvolvido no MIT, na década de 1960, e já foi usado por milhares de empresas e de escolas para ensinar modelos e princípios de gestão de cadeias de fornecimento, como o “Efeito Chicote” (*bullwhip effect*) – as oscilações nos volumes de pedidos resultantes de más informações entre os participantes da cadeia de fornecimento. Outras empresas estão começando a criar os próprios jogos para comunicar objetivos específicos. A empresa de caminhões Schneider National desenvolveu jogo de simulação para comunicar a importância do raciocínio analítico na mobilização de caminhões e reboques. O objetivo do jogo é minimizar os custos variáveis para determinado nível de receita, assim como reduzir o tempo de ociosidade dos motoristas. Os participantes decidem aceitar cargas ou movimentar caminhões vazios, com a ajuda de ferramentas de apoio. A Schneider usa o jogo no intuito de ajudar seu pessoal a compreender a importância dos métodos analíticos para comunicar a dinâmica do negócio e para mudar a mentalidade dos empregados, que deixam de ser meros “tomadores de pedidos” e passam a atuar como “geradores de lucros”. Alguns clientes da empresa também já jogaram o jogo.

As empresas ainda podem recorrer à tecnologia contemporânea a fim de possibilitar que os decididores interajam diretamente com os dados. Por exemplo, a Deloitte Consulting criou um sistema de questionários e relatórios sobre operações de aeroportos como aplicativo do iPad. O sistema usa o Google Maps para mostrar nos mapas os aeroportos em que opera determinada empresa de aviação – nesse caso, a Delta. Diferentes cores de aviões (vermelho, para ruim; verde, para bom) indicam desempenho positivo ou negativo em

determinado aeroporto. O toque no símbolo dos aeroportos no mapa gera dados financeiros e operacionais referentes àquelas instalações específicas. Diferentes botões oferecem indicadores referentes a pessoal, níveis de serviços, finanças, operações e áreas problemáticas. Esse aplicativo é só um exemplo do que pode ser feito com as atuais tecnologias interativas e amigáveis.

## Além do relatório

As apresentações ou relatórios não são os únicos produtos possíveis dos projetos analíticos. Até seria melhor se os analistas se empenhassem na produção de resultados mais próximos da criação de valor. Por exemplo, muitas empresas, cada vez mais, estão incluindo a analítica em contextos de tomada de decisões automáticas.<sup>10</sup> Em seguradoras, em bancos e em situações de formação de preços para consumidores (como em empresas de aviação e em redes de hotéis), sistemas analíticos de decisões automáticas são muito comuns – com efeito, às vezes é até difícil conseguir que um ser humano analise seu pedido de empréstimo ou de seguro. Nessas condições, sabemos que o uso da analítica é inevitável pela falta de método alternativo (ou, ao menos, as exceções são muito raras; é possível que, em casos excepcionais, seres humanos interfiram). Se você for analista quantitativo ou dono de importante processo decisório, incumbido de desenvolver e implementar um desses sistemas, sua tarefa será muito mais complexa que produzir um relatório.

Na indústria de informações on-line, o *Big Data* das empresas envolve muitos petabytes de dados. As novas informações chegam em tais volumes e velocidades que os seres humanos teriam dificuldade em processá-las e compreendê-las. Nesse contexto, os cientistas de dados dessas organizações (basicamente, analistas quantitativos com qualificações acima do normal em TI) geralmente se concentram em áreas de desenvolvimento de produtos. O objetivo deles é desenvolver protótipos de produtos e novos atributos de produtos, em vez de relatórios ou apresentações.

Por exemplo, o grupo Data Science, no site de networking de negócios LinkedIn, é parte da organização de produtos, e, como tal, desenvolveu vários novos atributos e funções de produtos, com base nas relações entre redes sociais e empregos. Aí se incluem “Pessoas que você talvez conheça”, “Empregos em que você talvez esteja interessado”, “Grupos que podem ser de seu interesse”, e

outros. Alguns desses atributos (em especial “Pessoas que você talvez conheça”) produziram efeito dramático em termos de crescimento e persistência da base de clientes da LinkedIn.

Se você ou os analistas quantitativos de sua organização estiverem trabalhando basicamente em problemas e processos internos, o produto a ser entregue não precisa ser necessariamente apenas um relatório ou apresentação. Como o objetivo é, efetivamente, melhorar o processo ou decisão, você ou seus analistas precisarão esforçar-se ao máximo para promover essas melhorias. Quando Tom fez algumas pesquisas sobre como as empresas melhoraram 57 decisões diferentes, poucos anos atrás, concluiu que a analítica foi o fator mencionado com mais frequência como meio para executar o aprimoramento.<sup>11</sup> Mas “mudanças de cultura e liderança” foi o segundo; “melhores dados”, o terceiro; e “alterações nos processos de negócios”, o quarto. Em média, os entrevistados pelas empresas mencionaram mais de cinco fatores que contribuíram para a melhoria das decisões. Isso significa que os analistas devem ser mais que analistas; também precisam ser consultores em mudanças nas empresas.

## Quando os resultados não impõem ações

Espera-se que os resultados de análises quantitativas recomendem providências; às vezes, porém, resultados perfeitamente bons não exigem ações específicas, embora ainda sejam úteis. Por exemplo, o coautor Jinho colaborou, certa vez, em um estudo sobre a importância de mover-se primeiro no jogo Go. Um dos mais antigos jogos estratégicos de tabuleiro do mundo, Go reúne elementos simples (linhas e peças pretas e brancas, em forma de lentes, ou “pedras”) e aplica regras simples. Não obstante a aparente simplicidade, contudo, o jogo gera situações de sutileza, que encantam os jogadores há milênios.<sup>12</sup> Na Ásia, especialmente na Coreia, no Japão e na China, dezenas de milhares de pessoas jogam Go, e quase mil profissionais ganham a vida competindo em torneios que oferecem milhões de dólares em prêmios.

No começo do jogo, o tabuleiro está vazio. Um jogador pega as pedras pretas; o outro, as brancas. As pedras brancas e pretas alternam as jogadas, com as pretas movimentando-se primeiro. Isso gera certa vantagem, mas nunca ninguém determinou o tamanho da vantagem. Desde 1974, nos jogos

profissionais, as pedras brancas começam com 5,5 pontos de vantagem, para compensar o benefício da movimentação inicial das pretas. É o chamado *Komi*.

Jinho analisou dados de 577 jogos profissionais, na Coreia, para verificar se 5,5 pontos eram o valor certo para o komi. Os resultados sugeriram que esse handicap inicial era muito pequeno. Surgiu, contudo, uma complicação na análise. Os dados indicavam que movimentar-se primeiro era vantagem apenas se o jogador fosse bastante hábil para aproveitá-la. Providência teoricamente plausível a ser adotada com base nesses resultados seria a diferenciação do komi, que consistiria em (1) manter o atual nível de komi (5,5 pontos) no caso de primeiros moventes cujo desempenho indicasse que o handicap vigente, embora baixo, seria suficiente para compensar o benefício da movimentação inicial, e (2) aumentar o komi para primeiros moventes com melhor performance. Dificilmente, contudo, essa diferenciação do komi seria aceita em torneios internacionais. Portanto, uma boa análise (que resultou em publicação do trabalho de Jinho e colega em periódico acadêmico) não redundou em nenhuma ação específica.<sup>13</sup>

## Resumo

Acabamos de descrever cada um dos três estágios e os seis passos do modelo de raciocínio analítico. A execução sequencial de cada um dos passos descritos nos [capítulos 2, 3 e 4](#) possibilita a análise e solução de quase qualquer problema quantitativo. Mesmo quem não tiver sólida formação quantitativa será capaz de executar a maioria desses passos. A verdadeira análise estatística ou matemática ocorre em apenas dois deles. E os computadores podem executar grande parte do trabalho árduo na maioria dos casos.

A chave do sucesso é planejar e executar bem o passo inicial e o passo final do processo. Se o problema for bem reconhecido e formulado, os passos intermediários serão relativamente simples. A comunicação eficaz dos resultados também é decisiva, pois, do contrário, não se tomará nenhuma providência – por isso é que o passo final é realmente importante.

## Exemplo de raciocínio analítico: Teoria prática para prever divórcios

Casal recém-casado procurou um matemático brilhante, famoso por prever se novos casamentos durariam toda uma vida de felicidade ou se escoariam ralo abaixo, até o divórcio inevitável. Pediu-se que os dois se sentassem um diante do outro, numa sala, sozinhos, e conversassem entre si sobre questão controversa durante 15 minutos. Analisando a conversa gravada do casal, o matemático sentenciou: “Vocês bem que poderiam divorciar-se imediatamente, para evitar maiores aborrecimentos.” Sabia-se que as previsões dele eram de exatidão impressionante. O matemático que desenvolveu modelos para determinar se casais apaixonados passariam a vida juntos ou romperiam o relacionamento em divórcio é James Murray, professor da Universidade de Oxford, que trabalhou com John Gottman, professor e psicólogo da Universidade de Washington, na pesquisa. Gottman forneceu as hipóteses e os dados (observações codificadas, em vídeo e áudio, de muitos casais), além de longa experiência no que contribui para o sucesso dos casamentos. Murray entrou com a expertise em modelos não lineares. Depois da realização da pesquisa, Gottman foi extremamente eficaz na comunicação dos resultados a casais de todo o mundo. Examinemos a abordagem dele, sob a perspectiva do referencial de seis passos – com ênfase na comunicação dos resultados.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** O aumento das taxas de divórcio nos países desenvolvidos, inclusive nos Estados Unidos, é fenômeno generalizado, importante e mal compreendido. No entanto, sem o conhecimento teórico dos processos relacionados com a estabilidade e a dissolução da sociedade conjugal, é difícil projetar e avaliar novas formas de intervenções para ajudar casais com problemas de relacionamento. O desenvolvimento de um arcabouço matemático para casamentos duradouros fornecerá os fundamentos de uma teoria científica sobre as relações conjugais. Murray e Gottman empenharam-se em conceber uma fórmula para as interações de marido e mulher.

**REVISÃO DAS DESCOBERTAS ANTERIORES:** Calcular as chances de um casamento feliz não é nada novo, mas as previsões de divórcio não têm sido nem rigorosas nem exatas. Murray e Gottman foram mais longe que a maioria: juntaram forças para criar um modelo matemático capaz de prever com exatidão espantosa se os casais se divorciarão. Gottman vinha pesquisando casamentos e comportamentos associados a divórcio havia anos; portanto, conhecia o

território.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Murray e Gottman observaram que as conversas refletiam problemas subjacentes dos casais, e que os debates, as risadas, as provocações, e as demonstrações de afeição durante a conversa criam laços emocionais. Especificamente, eles se concentravam nas seguintes reações: humor, concordância, alegria, afeição, interesse, raiva, dominação, tristeza, lamentação, beligerância, defensividade, desgosto, autoproteção, desprezo.

**COLETA DE DADOS (MENSURAÇÃO).** Os pesquisadores convidaram casais para participar de um estudo de laboratório. Setecentos recém-casados aceitaram o convite. Foi pedido para sentar-se um defronte ao outro, numa sala, sozinhos, e para debater questões contenciosas, como dinheiro, sexo, ou relações com afins. Recomendou-se a cada casal que conversassem sobre o tópico escolhido – algo controverso entre os dois havia algum tempo – durante 15 minutos, enquanto eram filmados. Observadores, então, assistiam ao filme e atribuíam pontos positivos ou negativos ao marido e à mulher, dependendo do que dissessem. Os pares que demonstravam afeição, humor ou felicidade durante o debate recebiam a pontuação máxima, enquanto os que exibiam desprezo ou beligerância recebiam o mínimo. Desprezo era considerado mais destrutivo que desgosto, tristeza ou raiva. A pontuação variava de + 4 a – 4, conforme a seguir:

Humor: +4	Tristeza: –1
Concordância: +4	Lamentação: –1
Alegria: +4	Beligerância: –2
Afeição: +4	Atitude defensiva: –2
Interesse: +2	Autoproteção: –2
Raiva: –1-1	Desgosto: –3
Dominação: –1	Desprezo: –4

**ANÁLISE DOS DADOS.** A tarefa de Murray e Gottman era identificar padrões preditivos de divórcio ou de casamento feliz. Primeiro, os escores eram lançados em gráfico; o ponto em que as duas linhas se cruzavam ilustrava as chances de sucesso ou fracasso do casamento. Se o marido ou a mulher era consistentemente negativo, presumia-se que a união estava fadada ao divórcio. O importante era quantificar a razão entre interações positivas e negativas durante a conversa. A razão mágica é 5:1, e o casamento pode enfrentar



problemas quando cair abaixo disso. Os resultados foram inseridos em modelo matemático, com base em equações de diferenças que destacam traços de caráter subjacentes, capazes de influenciar o sucesso do casamento.

Usando os escores de resultados, eles dividiram os casais em cinco grupos, conforme a seguir:

*Validadores:* Calmos, íntimos, apoiam-se um no outro e cultivam relações de companheirismo. Preferem vivências comuns a experiências individuais.

*Evasores.* Fazem o possível para evitar confrontos e conflitos. Só respondem positivamente um ao outro.

*Voláteis:* Embora românticos e apaixonados, têm discussões acaloradas. Misturam instabilidade e estabilidade; no entanto, em geral, tendem mais à infelicidade.

*Hostis:* Um parceiro não fala sobre certos assuntos e o outro se submete; portanto, não há comunicação.

*Antagônicos segregados:* Um é irascível e litigante. O outro não está interessado em discussões.

O modelo matemático caracterizou as diferenças entre dois tipos de casais estáveis (validadores e evasores), cujos casamentos tendem a durar, e dois tipos de casais instáveis (hostis e antagônicos segregados). Previu-se que o grupo dos voláteis continuaria casado – apesar da instabilidade do casamento.

Os 700 casais que participaram do estudo foram procurados a cada dois anos, durante 12 anos. A fórmula de Murray e Gottman indicou corretamente a incidência de divórcio, com 94% de acertos. A exatidão do teste não passou dos 94% porque alguns casais voláteis, que, na opinião dos pesquisadores, continuariam casados e infelizes, efetivamente acabaram divorciando-se.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** O modelo foi publicado em um livro de Gottman, Murray e colegas, intitulado *The Mathematics of Marriage: Dynamic Nonlinear Models*. O livro, a princípio, destinava-se a outros acadêmicos. Diferentemente, porém, de muitos colegas de

academia (e de Murray), Gottman estava muito interessado em influenciar os hábitos dos casais pesquisados. Publicou, então, vários livros e artigos sobre a pesquisa e fundou (com a esposa Julie) o Gottman Relationship Institute ([www.gottman.com](http://www.gottman.com)), que oferece treinamento, vídeos e várias outras ferramentas de comunicação para a melhoria do relacionamento. O trabalho também é promissor para terapeutas, proporcionando-lhe meios para ajudar casais a superar padrões de comunicação destrutivos, que podem lançá-los ladeira abaixo rumo ao divórcio. O Instituto Gottman também desenvolveu um conjunto de material de apoio e de oficinas de orientação para terapeutas. Finalmente, o modelo ainda cria condições para que pesquisadores simulem as reações dos casais em várias situações. Daí resultaram experimentos mentais do tipo “e se” (*what if*), que podem ser úteis no desenvolvimento de estratégias de intervenção com base científica para casamentos problemáticos.

Gottman ajudou a elaborar o maior experimento clínico aleatório (história de cientista maluco), ao pesquisar mais de 10 mil casais. Eis o que diz Gottman sobre como a pesquisa ajuda casais na vida real: “Nos últimos oito anos, envolvi-me intensamente no problema, trabalhando com minha esposa, extremamente talentosa, na tentativa de juntar as ideias e aproveitar nossa teoria para ajudar casais e bebês. E agora sabemos que essas intervenções de fato fazem grande diferença. Conseguimos salvar 75% de casamentos em dificuldade, por meio de workshops de dois dias e de nove sessões de terapia de casais.”<sup>14</sup>

## Exemplo de Raciocínio Analítico: Escore Fico

O escore FICO, número de três dígitos entre 300 e 850, é um instantâneo da situação financeira de uma pessoa em determinado ponto do tempo.<sup>15</sup> Quando alguém entra com um pedido de crédito – no cartão de crédito, para compra de automóvel ou de financiamento hipotecário – os emprestadores querem avaliar o risco da transação. Escores FICO são avaliações de crédito, praticadas pela maioria dos emprestadores para estimar o risco de crédito, que sugerem o valor e as condições do empréstimo (juros, prazos, garantias e outras). É exemplo impressionante da conversão da analítica em ação, uma vez que quase todos os emprestadores nos Estados Unidos – e, cada vez mais, no resto do mundo – o adotam como critério. Vejamos seu desenvolvimento no referencial de seis passos.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Os escores de crédito oferecem aos emprestadores medida rápida e objetiva do risco de crédito do cliente. Antes do uso de escores, o processo de concessão de crédito não raro era lento, inconsistente e distorcido. O engenheiro Bill Fair e o matemático Earl Isaac tiveram a ideia de que as decisões de negócios seriam mais eficazes caso se calculasse, por metodologia estatística, o risco de calote, considerando vários fatores da história pessoal e financeira dos candidatos. Em 1956, constituíram uma empresa, desenvolveram o modelo e começaram a vender o primeiro sistema de avaliação de crédito. Em 1989, lançaram o primeiro escore FICO, de propósitos gerais.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Os relatórios de crédito surgiram mais de 100 anos atrás, quando pequenos comerciantes varejistas se reuniram para trocar informações sobre os clientes. As associações comerciais se converteram em pequenos birôs de crédito, que, em seguida, com o advento da computação, fundiram-se em organizações maiores. Algo parecido com um escore de crédito de propósitos gerais, contudo, ainda estava longe do estágio inicial.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Entre os diferentes dados de um relatório de crédito, Fair e Isaac selecionaram variáveis, que podem ser agrupadas em cinco categorias (fatores como idade, gênero, raça, nacionalidade e estado civil não foram considerados), conforme mostra a [Tabela 4.1](#).

**TABELA 4.1**

---

### **Variáveis de análise de crédito de Fair e Isaac**

Histórico de pagamentos	Informações sobre pagamentos de tipos específicos de dívidas; existência de registros públicos negativos, itens em cobrança e/ou de inadimplência; gravidade da inadimplência; quantias devidas em créditos inadimplentes ou em itens em cobrança; tempo decorrido desde o vencimento das dívidas, desde o lançamento do registro público negativo e do atraso dos itens em cobrança; quantidade de contas atrasadas em arquivo; quantidade de contas pagas mediante acordo.
Quantias devidas	Quantias em contas a pagar; saldos devedores em tipos específicos de contas; falta de tipos específicos de saldos, em alguns casos; número de contas com saldos; proporção de uso das linhas de crédito; proporção do saldo devedor dos empréstimos em aberto.
	Tempo decorrido desde a abertura da conta; tempo decorrido

Escopo do histórico de crédito	desde a abertura de tipos específicos de contas; número de análises de crédito recentes; antiguidade das contas ativas.
Novo crédito	Número de contas abertas recentemente e proporção destas por tipo; número de análises de crédito recentes; tempo decorrido desde a abertura das contas recentes por tipo; tempo decorrido desde a última análise de crédito; restabelecimento de histórico de crédito positivo, depois de problemas de pagamento passados.
Uso de tipos de crédito	Número dos vários tipos de contas.

---

**COLETA DE DADOS (MENSURAÇÃO).** Os hábitos financeiros dos americanos são monitorados por uma ou mais de três agências nacionais de relatórios de crédito (credit-reporting agencies – CRAs): Equifax, Experian e Trans Union. Todos os meses, instituições financeiras e credores diversos enviam relatórios de crédito às CRAs, abrangendo numeração das contas, tipos de crédito (cartões de crédito, financiamento de veículos, empréstimos hipotecários), saldos devedores, ações de cobrança e histórico do pagamento de contas. Os dados sobre os consumidores individuais podem variar entre as agências. Uma vez que o score FICO se baseia em informações que as CRAs mantêm sobre cada um, é possível ter três scores FICO: um de cada agência.

**ANÁLISE DOS DADOS.** Em geral, o score de crédito leva em conta ampla variedade de informações do relatório de crédito do consumidor, mas nem todas são tratadas da mesma maneira. Alguns aspectos do histórico de crédito são mais importantes que outros e terão peso maior na avaliação final. O score FICO é calculado com base em fórmula proprietária, de uso exclusivo de Fair e Isaac. Embora os componentes exatos da fórmula e os métodos de cálculo não estejam disponíveis para publicação (e cada CRA use algoritmo específico), o FICO revelou os seguintes componentes e respectivos pesos:

Histórico de pagamentos: 35%

Utilização de crédito: 30%

Escopo do histórico do crédito: 15%

Tipos de crédito: 10%

Análises recentes: 10%

APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS. A FICO não mantém um banco de dados de escores FICO. Em vez disso, quando um empréstador pede uma avaliação de crédito, o escore é calculado pela CRA à qual o empréstador pediu o relatório. A FICO fornece às CRAs software com o algoritmo – fórmula matemática derivada de amostras aleatórias de informações de crédito dos consumidores – que é usada para calcular o escore. Como, porém, cada CRA modifica um pouco o algoritmo, os escores das CRAs podem ser diferentes para o mesmo devedor, na mesma data.

O escore FICO varia entre 300 e 850. As melhores avaliações de crédito são atribuídas a pessoas com escore acima de 770, mas um escore de 700 já é considerado bom. O escore médio é de mais ou menos 725. Quando o escore cai para menos de 650, os consumidores em geral se qualificam para empréstimos “subprime”, e a taxa de juros começa a subir em forte aclave.

Os escores FICO, as mais usadas avaliações de crédito do mundo, aumentaram drasticamente a eficiência dos mercados de crédito dos Estados Unidos, ao melhorarem a avaliação de risco. Os empréstadores passam a dispor de melhor previsão sobre se o empréstimo será pago, com base exclusivamente no histórico de crédito do consumidor. Número crescente de empresas que não têm nada a ver com oferta de crédito (como seguradoras, operadoras de telefonia móvel, administradoras de imóveis e subsidiárias de empresas de serviços financeiros) estão analisando relatórios de crédito e usando as informações neles contidas para decidir sobre fazer negócios com determinados clientes e para classificar consumidores em diferentes níveis. Alguns empregadores também verificam avaliações de crédito ao contratar pessoal, para melhor avaliar os candidatos. Muitas seguradoras de automóveis consideram as informações dos relatórios de crédito, para estimar o nível geral de responsabilidade financeira dos segurados, que apresentaria forte correlação com a responsabilidade na condução de veículos. Em síntese, este foi um dos maiores sucessos da história na aplicação do raciocínio analítico.

**Exemplo de raciocínio analítico: Nascimento de um novo conceito em *value stores***

Em maio de 1999, a Homeplus entrou no setor de hipermercados, na Coreia do Sul, ao formar joint venture com a Tesco. Na época, o setor de hipermercados no país era um “oceano vermelho” – cheio de sangue, onde batalhavam 11 concorrentes em competição intensa. Não só varejistas locais, como E-mart, Lotte Magnet, Kim’s Club, Hanaro Mart, LG Mart, Mega Mart, Aram Mart e Top Mart, mas também varejistas ocidentais, dotados de capital, aderiram à competição ilimitada. A Homeplus lançou-se nesse setor difícil como o décimo segundo concorrente, mas alcançou sucesso “10-10” que virou lenda, ao alcançar vendas de 10 trilhões de won (moeda nacional) em apenas 10 anos. Nesse período, a Homeplus gerou taxas anuais de crescimento das vendas de 47% e de crescimento do lucro de 175%, ultrapassando concorrentes mais antigos. Os segredos por trás da grande vitória da Homeplus são a capacidade de liderança do CEO, SH Lee, e sua estratégia de “nada menos que a vitória”, além de nova arma desenvolvida com base em pesquisa de mercado: o conceito de “value store”, que reverteu completamente a noção convencional de hipermercado. Embora, basicamente, a pesquisa de mercado não fosse estatística, ainda é possível analisá-la à luz do modelo de seis passos.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** A Homeplus tinha plena consciência de que não seria bem-sucedida fazendo lojas iguais às dos concorrentes. Portanto, precisava criar conceito de loja totalmente diferente, que se destacasse com nitidez dos hipermercados existentes. Que conceito seria esse? Em meio a uma enxurrada de ideias e teorias de marketing, a Homeplus encontrou a resposta no princípio de “os clientes em primeiro lugar”. Os hipermercados eram “armazéns” que copiavam o modelo europeu ou americano. As lojas eram amontoados monótonos de gôndolas e prateleiras apinhadas de caixas. Pesquisas e análises exaustivas das necessidades dos clientes eram a única maneira de compreender o que os clientes realmente queriam dos hipermercados.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** As pesquisas de mercado existentes sobre clientes de hipermercados não eram acessíveis, por serem dados confidenciais de outros varejistas. O mote do CEO Lee pregava que a maneira de tornar-se o mais eficaz varejista do mundo era “Parta do melhor e torne-o ainda melhor”. Isso significava, claramente, compreender o que o número 1 do mundo estava fazendo, e superá-lo. Com o objetivo de compreender as mais

recentes tendências do mercado varejista global, a Homeplus fez o benchmarking de vários canais varejistas, como hipermercados, supermercados, category killers, lojas de departamentos, shopping centers multicomplex, em 25 países, abrangendo Estados Unidos, Japão, Europa e Sudeste Asiático. O objetivo era explorar o aprendizado e aprimorar a Homeplus.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** As variáveis da pesquisa de varejistas e clientes eram:

- Lugar de compras e motivo para escolher o lugar de compras.
- Propósito da compra, nível de satisfação, queixas.

**COLETA DE DADOS (MENSURAÇÃO).** Os dados para a pesquisa foram reunidos por meio de dois métodos básicos:

- Visitas a e análises de vários varejistas em todo o mundo.
- Pesquisa de mercado abrangente e estruturada, e estudo de clientes da Coreia do Sul por empresa independente.

**ANÁLISE DOS DADOS.** O resultado da ampla análise de dados foi impressionante. Preço não era a única coisa que os clientes queriam. Era fator importante, mas o que os clientes realmente almejavam era alto valor – preço razoável que levasse em conta o tempo e uso dos produtos. Também esperavam que seu valor como cliente fosse reconhecido. Portanto, procuravam a aparência e a sensação premium de uma loja de departamentos e qualidade de serviços compatível com o ambiente. Ainda mais importante foi a constatação de que os clientes queriam não só comprar vários produtos a preços baixos, mas também desfrutar dos “one-stop-living services” (serviços de qualidade em parada única), em ambiente agradável. Portanto, o novo conceito de hipermercado Homeplus tinha de buscar a “value store”, que não só venderia ampla variedade de produtos a preços baixos, mas também ofereceria os “living services” que os clientes realmente buscam – serviço gentil, ambiente agradável, instalações convenientes, vários programas educacionais (forte valor cultural na Coreia do Sul) e ofertas empolgantes de cultura e entretenimento.

APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS. O projeto da primeira loja Homeplus, em Ansan, incluía ambiente tão agradável quanto o de uma loja de departamentos e instalações convenientes que ofereciam “one-stop-living services” no andar térreo, inclusive praça de alimentação com 400 assentos, farmácia, clínica, óptica, lavanderia, banco, playground, creche e até Centro de Serviços Públicos do governo, para emissão de carteiras de identidade. Com a inclusão de um Centro Cultural, a loja deixava de ser apenas um lugar de compras, para transformar-se no maior centro comunitário da região. Todos os especialistas em varejo que observavam as instalações físicas da Homeplus, antes da inauguração, balançavam a cabeça, afirmando que os fundamentos do projeto de hipermercado tinham sido ignorados. Não fazia sentido incluir instalações de conveniência relativamente não lucrativas no espaço dispendioso do andar térreo. O negócio de hipermercado já foi descrito como “jogo de espaço”, e espaço é igual a dinheiro; portanto, a regra é maximizar as vendas no espaço disponível. Não admira que a colocação de um Centro Cultural no primeiro andar fosse considerada “absurdo”. Os visitantes da Tesco, parceira Homeplus na joint venture, também questionavam o arranjo: “Não seria melhor receber os clientes com boa comida, ao entrarem no andar térreo? Não é esse o layout mais adequado à natureza do varejo?” E SH Lee respondia: “O CEO decidiu. Os clientes são o CEO”, e prosseguia na execução do plano.

A loja da Homeplus, em Ansan, a primeira value store na Coreia, quebrou recordes de vendas no dia da inauguração. As vendas correspondiam a uma vez e meia o total de duas lojas concorrentes nas imediações, e os 100 mil clientes que compareceram à loja no dia da inauguração quase equivaliam ao número de residentes no raio de cinco quilômetros. A Homeplus, porém, voltava a bater recordes sempre que abria nova loja, e continuava crescendo em ritmo acelerado. A loja de Yongdeongpo, em Seul, que abriu em 2001, registrava vendas que superavam até as de uma loja de departamentos. O crescimento explosivo melhorou ainda mais a reputação da loja, em âmbito não só local, mas também internacional. Alguns analistas de investimentos observaram:

“A Homeplus harmonizou artisticamente a sensação das feiras livres com a sofisticação das técnicas de varejo.”

“O conceito da loja era notável, em comparação com o dos concorrentes.”



“As vendas da Homeplus eram inacreditáveis.”

Outros concorrentes começaram a converter suas lojas armazéns em value stores, mas a Homeplus continuava a oferecer serviços diferenciados, inclusive escolas para a educação de adultos. Em 2011, havia Homeplus Schools em 110 das 124 lojas espalhadas por todo o país, com mais de um milhão de alunos por ano e seis mil instrutores. Trata-se da maior iniciativa de educação de adultos do mundo. O mais importante benefício da Homeplus School é o impacto na lealdade dos clientes. Muitos clientes citam a Homeplus School como a principal razão para gostarem da Homeplus. Em média, os alunos da Homeplus School, em comparação com os clientes normais, comparecem às lojas com o dobro da assiduidade e, em cada visita, permanecem nas lojas o dobro do tempo. A propaganda boca a boca difundida pelos alunos da Homeplus School é poderoso motor para aumentar o brand value da Homeplus.

## Análise quantitativa e criatividade

Por volta de 275 a.C, um jovem foi a um banho público. Ao entrar numa banheira, a água transbordou. De repente, ele saltou da banheira e saiu correndo pela rua, nu, gritando: “Eureca!, eureka!” Quem o viu naquele estado imaginou que ele estivesse louco. O jovem era Arquimedes, e *eureka* em grego significa “descobri”. Descreveremos em poucas palavras o que ele descobriu e por que ficou tão empolgado. Basta dizer, porém, que ele usou criatividade e análise quantitativa na solução do problema.

Em geral se supõe que criatividade é o oposto de análise quantitativa. Considera-se criatividade algo exploratório e visionário, envolvendo imaginação e inspiração. Já a análise quantitativa seria monótona, rotineira e numérica. Temos a forte impressão, contudo, que criatividade e analítica não são, de modo algum, conceitos antagônicos, mas, sim, não raro, ideias correlatas e complementares. As aplicações mais bem-sucedidas da analítica são altamente criativas (como esperamos tenha ficado claro nos exemplos já apresentados), e criatividade é componente importante das abordagens analíticas à solução de problemas. Também diríamos que a criatividade sozinha – sem dados ou análises – geralmente não oferece base suficiente para boas decisões. Sempre achamos que as pessoas e as organizações mais bem-sucedidas conjugam criatividade e analítica.

Por exemplo, a Apple é considerada não raro uma das organizações mais criativas do planeta. Com efeito, os seus produtos sempre envolveram muita criatividade. A mesma empresa, contudo, adota análises e processos rigorosos nas operações de cadeia de fornecimento, para dispor dos produtos certos, no momento certo. Nas lojas de varejo, por exemplo, a Apple coleta e analisa grandes volumes de dados. “Desde o lançamento dos produtos, a empresa passa a

monitorar a demanda por loja e por hora, ajustando diariamente a programação da produção”, diz um artigo.<sup>1</sup> A descrição de cargo vago em gestão de cadeia de fornecimento na Apple exige, entre outros atributos, a combinação de “conhecimento e experiência das melhores práticas de cadeia de fornecimento, forte capacidade analítica e tarimba em negócios”.<sup>2</sup> Se até empresas altamente criativas como a Apple exigem inclinação analítica em muitos de seus empregados, provavelmente veremos muito mais combinações de criatividade e analítica no futuro.

Evidentemente, criatividade e analítica pode ser mistura perigosa. Você provavelmente conhece o livro famoso de Darrell Huff, de 1954, *How to Lie with Statistics* (Como mentir com estatística). O título sugere corretamente que é possível usar métodos quantitativos pelo menos para flexibilizar a verdade.<sup>3</sup> Ouvimos dizer, com frequência, em geral de brincadeira, que “torturamos os estatísticos até confessarem”. A linha entre adequação e inadequação no uso da criatividade em analítica é tênue. O principal fator diferenciador é a sinceridade na busca da verdade. Quando se explora a criatividade em analítica com o intuito de demonstrar a correção das próprias ideias ou das do chefe – e se os números tornam a tentativa extremamente difícil – é melhor deixar a criatividade de lado e partir para outra hipótese.

## Rápida revisão dos seis passos

Vale a pena rever como a criatividade se encaixa no processo de raciocínio analítico de seis passos, que descrevemos nos [Capítulos 2, 3 e 4](#). Em seguida, apresentaremos um processo criativo capaz de iluminar todo o conjunto.

No passo de *reconhecimento e formulação do problema*, a criatividade é extremamente útil e importante. Metade da batalha da solução do incômodo e da tomada de decisão consiste em equacionar o problema e/ou a decisão com criatividade, de modo a adotar abordagem eficaz. Esse é o ponto do processo de raciocínio analítico em que o analista cria uma hipótese para o que está vendo nos dados – ato, em geral, muito intuitivo e criativo. Em vista do contexto organizacional ou empresarial e do conjunto de restrições, a formulação criativa pode mudar o contexto, abordá-lo de maneiras diferentes e reduzir ou eliminar as restrições. Por exemplo, na equação do valor do vinho, descrita no [Capítulo 3](#), Orley Ashenfelter formulou o problema do valor do vinho como algo

previsível, com base apenas em variáveis climáticas e na antiguidade do vinho. Do mesmo modo, no estudo sobre o casamento, descrito no [Capítulo 4](#), James Murray e John Gottman tiveram o insight criativo de que o sucesso contínuo no relacionamento conjugal era previsível com base somente no conjunto de comportamentos entre os cônjuges. Em geral, a criatividade aplicada às decisões analíticas é o reconhecimento de que resultados aparentemente complexos podem ser previstos ou explicados por fatores relativamente simples e mensuráveis.

Em tese, a *revisão das descobertas anteriores* não tende a ser o passo mais criativo, mas a identificação das descobertas anteriores e a escolha das técnicas analíticas que talvez sejam relevantes para o problema em questão envolvem criatividade. Por exemplo, usava-se, tradicionalmente, a técnica de “análise da sobrevivência” para compreender as taxas de mortalidade e as circunstâncias da morte de seres humanos e de outros organismos vivos. Um pesquisador, contudo, Junxiang Lu, aplicou-a com sucesso à previsão do valor vitalício dos clientes na indústria de telecomunicações.<sup>4</sup> Outros pesquisadores também a usaram para a solução de outros problemas de marketing, como o de determinar quando um cliente está no mercado em busca de um produto.<sup>5</sup>

O passo da *modelagem (seleção de variáveis)*, apesar de o nome soar como termo de analítica, também pode ser muito criativo – em especial para os primeiros a adotar determinado modelo. A escolha de variáveis em um modelo às vezes é óbvia, dependendo da intuição ou das variáveis anteriores, mas, às vezes, pode ser ato criativo. Lembre-se, por exemplo, do uso da extensão das palavras como indicador de autoria, nas cartas de Mark Twain, conforme descrito no [Capítulo 3](#), escolha não muito criativa de Claude Brinegar, que lera sobre precedentes na literatura especializada, mas abordagem extremamente inteligente de Thomas Mendenhall, que, aparentemente, foi o primeiro a adotá-la em estudos sobre as obras de Shakespeare. Evidentemente, quando se usam as mesmas abordagens e variáveis de modelagem para resolver os mesmos tipos de problemas, é provável que se chegue aos mesmos resultados. Assim sendo, por que se dar ao trabalho de fazê-lo?

A *coleta de dados* em si talvez seja processo monótono, mas decidir que dados coletar pode ser muito criativo. Não importa que se queira estudar o comportamento de humanos, de ratos ou de átomos, é provável que exista alguma maneira ainda inédita de medir e observar os fenômenos relevantes, que

ninguém já tenha imaginado.

Por exemplo, os psicólogos sociais Mihaly Csikszentmihalyi e Reed Larson queriam pesquisar as atitudes e sentimentos de adolescentes. A fim de coletar dados sobre como os pessoas nessa faixa etária se sentiam durante todo o dia, o projeto recorreu a um método de pesquisa singular. Os pesquisadores deram pagers a 75 alunos de ensino médio e incumbiram alunos de ensino superior de enviar-lhes mensagens aleatórias ao longo do dia para descobrir os sentimentos deles, conforme os próprios relatos. O método de coleta de dados, denominado amostragem de experiências, hoje é muito usado.<sup>6</sup>

Embora não tenham ficado surpresos ao constatarem que os adolescentes se sentiram infelizes quase todo o tempo, os pesquisadores se surpreenderam ao descobrir que as emoções deles se tornavam positivas ao se dedicarem a tarefas desafiadoras. O livro daí resultante, *Being Adolescent: Conflict and Growth in the Teenage Years*, foi o primeiro relato público do estado mental criativo e engajado que Csikszentmihalyi denominou *fluxo*, foco de boa parte de seus trabalhos posteriores.<sup>7</sup> Na verdade, ele estava sendo criativo na coleta de dados sobre criatividade!

O passo da *análise de dados* não é aquele no qual, geralmente, mais se exige criatividade, a não ser que o analista seja realmente bom em matemática e estatística, e, ainda por cima, seja conservador. Esse é o primeiro ponto no processo de raciocínio analítico em que a criatividade pode causar problema. Todo teste estatístico ou análise matemática adota premissas e aceita restrições, e não se deve atenuá-las caso não se tenha certeza do que se está fazendo.

O passo de *apresentação de resultados* e de adoção de providências, por outro lado, é etapa em que a criatividade se reveste de importância vital, embora nem sempre receba a atenção devida. Como o público de mentalidade não analítica nem sempre compreende resultados analíticos apresentados em termos técnicos ou matemáticos, o bom analista precisa pensar com criatividade em como apresentar os resultados de maneira compreensível e até divertida. Por exemplo, não fale em valores de coeficientes nem em porcentagens de variâncias. Em vez disso, prefira expressar-se de maneira mais coloquial, como: “Se aumentarmos as despesas com propaganda em US\$1, obteremos, em média, US\$1,29 em receita adicional.” Essa maneira de falar talvez não pareça assim tão criativa, mas é muito mais conducente à compreensão e à aceitação que o abuso de termos técnicos, e esse processo de simplificação exige alguma criatividade.

## Os quatro estágios do raciocínio analítico criativo

Não queremos empanziná-los com muitos estágios e passos, mas talvez seja útil discutir como o pensamento analítico criativo se insere nos seis passos que analisamos. Em geral, o processo criativo segue esses quatro passos sequenciais:

*Preparação*: Construção dos fundamentos para a solução do problema.

*Imersão*: Engajamento intenso com a solução do problema e com os dados disponíveis; inicia-se longa luta em busca de respostas.

*Incubação*: internalização do problema na mente subconsciente, na expectativa de que se desenvolvam prováveis conexões inusitadas abaixo do nível da consciência (em geral, quase no momento em que já se sente frustrado e pronto para desistir!).

*Insight*: O grande avanço na compreensão de como o problema pode ser resolvido por meio da análise quantitativa.

O mais comum é que todo o estágio de *análise das descobertas anteriores* e parte do de *seleção de variáveis* se enquadrem na etapa de *preparação* do processo de raciocínio analítico criativo. A etapa de *imersão* corresponde a uma parte do estágio de modelagem, ao todo do de coleta de dados e a parte do estágio de análise de dados. A etapa de *incubação* ocorre quando a análise de dados chega a um impasse. Finalmente, quando de repente espoca o insight, todas as peças do quebra-cabeça se encaixam. A [Figura 5.1](#) é uma representação gráfica de tudo isso.

### Exemplo de raciocínio analítico: Arquimedes e a coroa

Aplicaremos o processo criativo de quatro passos ao raciocínio analítico de seis passos no contexto da história de Arquimedes – um dos primeiros exemplos conhecidos de raciocínio analítico criativo.<sup>8</sup>

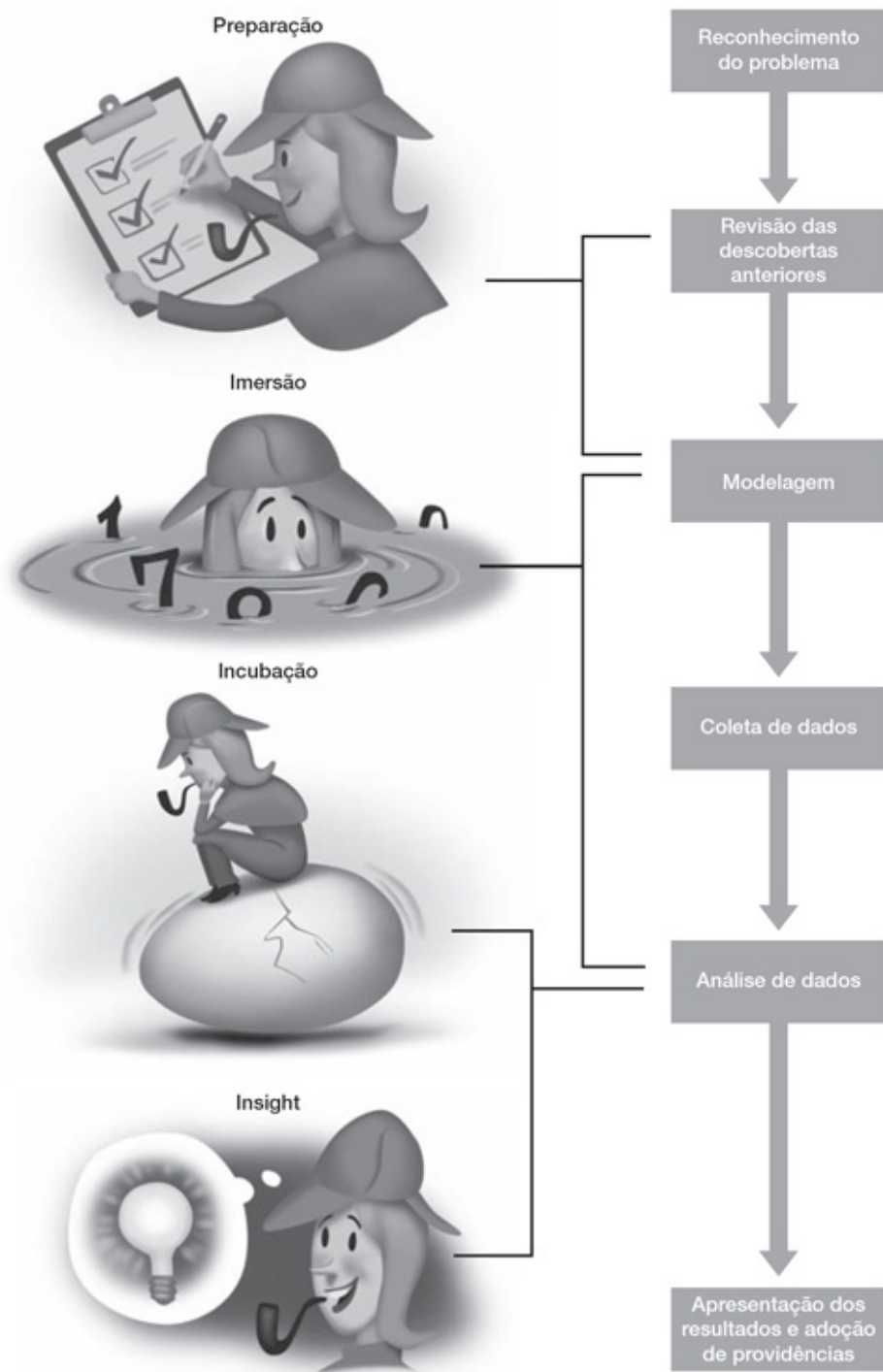
RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA. Hierão, Rei de Siracusa (na

Sicília), fez o voto de colocar uma coroa de louros de ouro no templo dos deuses imortais. Para tanto, pesou a quantidade exata de ouro e o confiou a um ourives. No dia combinado, o ourives entregou, para a satisfação do rei, uma coroa elegantemente confeccionada, cujo peso era igual ao do ouro recebido. Quando, porém, se preparava para a cerimônia de oferecer a coroa às divindades, o rei ouviu rumores de que a coroa não era de ouro puro; de que o ourives substituíra parte do ouro por peso igual de prata (na época, como hoje, era comum que os ourives diluíssem ouro com metais mais baratos). Hierão receou que tivesse sido enganado. Não sabendo como comprovar a fraude, expôs a questão a seus conselheiros, que se reconheceram incapazes de resolver o mistério. Finalmente, o rei pediu a Arquimedes para elucidar a questão, *sem danificar a coroa*, antes da data estabelecida para a cerimônia. Arquimedes, com 22 anos e já famoso por seus trabalhos em matemática e física, aceitou o desafio.

#### **FIGURA 5.1**

---

### **Análise quantitativa e criatividade**



REVISÃO DAS DESCOBERTAS ANTERIORES. Não havia método específico para medir o volume de nenhum objeto com forma irregular; o próprio Arquimedes tinha de descobrir como fazê-lo. Foi aqui que ele mergulhou no estágio criativo da *preparação*. Arquimedes, talvez um dos maiores matemáticos e inventores de todos os tempos, era capaz de medir o volume da maioria dos objetos regulares.



Ele raciocinou que, como o ouro pesa mais que a prata, uma coroa de ouro misturada com prata teria de ser mais volumosa (ou seja, ter mais volume) para que seu peso fosse igual ao de outra feita só de ouro. O problema era como medir com exatidão o volume de objetos com forma irregular, como a coroa, sem danificá-lo.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Ainda no modo de *preparação*, Arquimedes sabia que o volume da coroa seria a principal variável para prever a pureza do metal. Na verdade, contudo, a mensuração bem-sucedida do volume exigia *imersão* profunda no problema e nos dados. A maneira mais simples de determinar o volume da coroa era derretê-la, moldá-la em um cubo e medir suas dimensões. Ele, porém, não podia fazer isso, pois recebera ordens explícitas de não danificar a coroa. Diante do desafio, ficou obcecado pela solução do problema e refletiu sobre como resolvê-lo durante muito tempo. Todavia, mesmo depois de longo esforço, o problema continuava insolúvel.

**COLETA DE DADOS.** Imerso no problema, Arquimedes começou a coletar dados. Apurou as densidades do ouro e da prata e formulou a hipótese de que uma mistura dos dois poderia conter até 30% de prata. Ainda assim, contudo, era preciso determinar o volume da coroa de louros, e ele não sabia como fazê-lo.

**ANÁLISE DOS DADOS.** O dia da cerimônia se aproximava. Frustrado, Arquimedes estava a ponto de desistir, mesmo consciente de que sua reputação como gênio em matemática e física seria muito prejudicada. Entretanto, tanta era a intensidade de seu envolvimento com o problema que a questão se internalizara em seu subconsciente: ele estava no modo de *incubação*. Um dia, para reconfortar o espírito e descansar o corpo fatigado, ele entrou na banheira de um banho público (outra forma de imersão!) e observou que a água derramava-se pelas bordas. De repente, se deu conta de que o volume de água deslocado deveria ser igual ao volume da parte de seu corpo que estava submersa. Isso significava que o volume de objetos irregulares poderia ser medido com precisão, bastando mergulhá-los em água e calcular o volume de água deslocado. Portanto, ao mergulhar em água a coroa de louros e outro objeto de volume mensurável, feito de ouro puro, se a coroa de louros deslocasse maior volume de água, chegaria à conclusão de que o volume dela era superior ao do outro objeto. Empolgado com a descoberta, pulou da banheira e correu nu pelas ruas de

Siracusa, gritando de alegria “eureka, eureka!” Sem dúvida, Arquimedes finalmente tivera o *insight*. Fez, então, o teste, mergulhou em água a coroa suspeita e o cubo de ouro puro, e os volumes de água deslocados foram diferentes. Concluiu, então, que a coroa não era de ouro puro e que o ourives de fato misturara outro metal com o ouro para enganar o rei.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Arquimedes expôs os resultados ao Rei Hierão, que se admirou com o brilhantismo do insight de Arquimedes. Em consequência, a reputação de Arquimedes como gênio em matemática e em física aumentou ainda mais. Os resultados de seu trabalho, todavia, não foram igualmente bons para todos; o ourives desonesto foi executado.

Criatividade pode ser definida como *capacidade de desenvolver ideias originais e úteis*. Os exemplos de Arquimedes, porém, sugerem que, sob a perspectiva da análise quantitativa, criatividade é apenas *facilidade de captar nova relação entre as variáveis*, depois de ter identificado, selecionado e mensurado variáveis relevantes. De acordo com essa definição, é possível aprender, ensinar, exercitar e praticar a criatividade. Também se melhora a criatividade por meio do desenvolvimento de qualificações analíticas. As empresas podem desenvolver essas capacidades, oferecendo treinamento em analítica aos empregados e construindo cultura organizacional que estimule o raciocínio analítico no trabalho.

## **Imersão e trabalho árduo como fonte de criatividade e insight**

É famosa a afirmação de Thomas Edison de que “gênio é 99% transpiração e 1% inspiração”. Criatividade é insight que de súbito ilumina a solução do problema. Nessas condições, qual é a fonte da criatividade? Acreditamos que a criatividade é produto do esforço intenso. O raciocínio analítico criativo e o insight dele decorrente não são características inatas e aleatórias; mas, sim, a recompensa pelos 99% de transpiração – o labor empreendido no processo de análise quantitativa: a pesquisa de descobertas anteriores, a seleção e mensuração de variáveis relevantes, a busca ingente de padrões subjacentes às variáveis. Quando se imerge no esforço estrênuo, com persistência, um dia, quando menos se espera, vislumbra-se o insight eureka. Foi exatamente o que aconteceu com

Arquimedes, com Newton e com muitos outros gênios.

Barbara McClintock, cientista americana, ganhou o Prêmio Nobel de 1983 pela descoberta da transposição genética. Foi a única mulher a receber sozinha o Prêmio Nobel de Fisiologia ou Medicina. Durante longas pesquisas de campo, ela não raro experimentava momentos de eureka, a ponto de afirmar, peremptoriamente, que sua fonte de inspiração era a imersão absoluta no trabalho, a identificação completa com a pesquisa, esquecendo-se até de si mesma:

McClintock passou a conhecer tão intimamente cada um de seus pés de milho que, ao estudar seus cromossomos, ela de fato se identificava com eles: “Descobri que, quanto mais trabalhava com eles, mais cresciam; e quando eu realmente trabalhava com eles eu não era forasteira, eu de fato estava presente. Era parte do sistema. Eu até conseguia ver as partes internas dos cromossomos – tudo, efetivamente, estava lá. Fiquei surpresa, porque realmente me sentia como se estivesse lá e como se aqueles fossem meus amigos... Ao observar certas coisas, elas se tornam parte de você. E você se esquece de si mesma. O principal aqui é esquecer-se.”<sup>9</sup>

Evidentemente, criatividade e trabalho árduo não se associam apenas no reino quantitativo. Essa conjugação se aplica à arte e à literatura, por exemplo. Cho Jung Rae, famoso autor coreano, também assevera que a inspiração não surge de repente, mas, sim, jorra de longa acumulação de trabalho exaustivo e de imersão persistente:

Você se mantém focado em uma única coisa; acumula incessantemente várias ideias, até que, a certa altura, como num flash, a inspiração o acomete e você compreende num átimo o que vinha procurando com tanta ansiedade. Diz-se que a inspiração irrompe de repente. Sim, é verdade, quando se considera apenas o momento exato da inspiração. Entretanto, no processo, é preciso combinar trabalho exaustivo com imersão persistente, durante longo período. É possível, portanto, afirmar o seguinte: a profundidade do trabalho exaustivo e a intensidade da imersão persistente determina a fecundidade da inspiração.<sup>10</sup>

Em geral, a intuição é vista como entendida como pensamentos que vêm à

mente com rapidez e sem aparente reflexão. Daí se supõe que a intuição leva à percepção direta da verdade, sem processo de raciocínio nem a indispensável justificativa. Acreditamos, porém, que a intuição é habilidade que podemos desenvolver basicamente pela repetição contínua da análise quantitativa e por outros meios. A neurobiologia sugere que os neurônios podem ser “treinados” por meio da análise contínua. Portanto, acreditamos que a intuição se desenvolve naturalmente, quando se é capaz de imaginar, com base na experiência acumulada ao longo do tempo, um padrão subjacente entre as variáveis, sem coletar e analisar dados. Georg Hegel, filósofo alemão, achava que apenas quem raciocina de maneira altamente analítica é capaz de ter intuições puras e verdadeiras.<sup>11</sup>

## O papel da descoberta do padrão na criatividade analítica

A essência da análise de dados criativa é descobrir padrões entre as variáveis nos dados. Esses padrões denotam relações regulares ocultas entre as variáveis. Nesse estágio, o raciocínio matemático é muito útil, porque a matemática em si é a ciência dos padrões: a geometria euclidiana, a primeira grande ciência na Grécia Antiga, resultou da extração de padrões geométricos na natureza; o teorema de Pitágoras é o padrão entre os três lados de um triângulo retângulo.

A descoberta de padrões nas organizações contemporâneas tende a resultar mais da análise estatística que da matemática pura. Certas ferramentas estatísticas são excelentes para revelar a incidência de padrões nos dados, a taxas superiores às decorrentes do mero acaso. Determinada análise pode sugerir, por exemplo, que, ao demonstrarem certo padrão de compra (ou a falta de), os clientes tendem a parar de comprar. Ou que, ao comprarem determinado livro, os clientes também se mostrem propensos a comprar alguma outra coisa – por exemplo, as tão conhecidas recomendações da Amazon. (Um amigo de Tom chegou a receber uma delas, segundo a qual quem comprava um dos livros de Tom tendia a se interessar por imitações de excremento de cachorro, para oferecer de brincadeira a alguém.) Ver [Tabela 5.1](#), que descreve diferentes tipos de padrões e indica softwares comerciais e de códigos abertos para a identificação de padrões.

**TABELA 5.1**

---

## Software de mineração de dados para a descoberta de padrões em dados

		Código aberto		
Tarefas	Descrição	Propósito único	Multipropósito	Comerciais (todos multipropósitos)
Regressão	Encontrar uma função que ajuste os dados		RapidMiner R KNIME Weka Orange	Microsoft Analysis Services SAS: Enterprise Miner Oracle Data Mining IBM-SPSS Modeler STATISTICA KNIME LIONsolver Datameer Analytics Solution
Classificação	Classificar os dados em uma de várias classes predeterminadas	CART ML-Flex		
Clustering (análise de clusters)	Encontrar grupos de objetos (clusters)	Carrot2 ELKI		
Análise de associação sequencial	Busca de relações entre variáveis	JHepWork ARtool		
Deteção de anomalias	Identificar registros de dados inusitados	ELKI JHepWork		
Mineração de textos	Extrair padrões e tendências de textos	GATE NLTK UIMA RapidMiner		

### Exemplo de Raciocínio Analítico: Cerveja e Fraldas

Como exemplo mais detalhado da identificação de padrões, considere a descoberta de mineração de dados mencionada com mais frequência – segundo a qual os homens que entram em mercearias nos fins de semana tendem a comprar fraldas e cerveja. A constatação ocorreu não na Grécia Antiga, mas em Chicago, em 1992. É, definitivamente, um exemplo de identificação de padrões, mas não é bom exemplo de raciocínio analítico, de um modo geral (embora achemos que os exemplos negativos também podem ser instrutivos).<sup>12</sup>

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Thom Blischok, então gerente de um grupo de consultoria de varejo na empresa de armazenamento de dados Teradata, e equipe estavam analisando dados de pontos de venda (*point-of-sale* –

POS) para um cliente, Osco Drug. Como disse um dos membros da equipe, John Earle, o objetivo era “buscar afinidades entre itens comprados na mesma ida à loja”. Prosseguindo, explicou: “Sugerimos, então, testes de rearrumação de mercadorias nas lojas, para ver como os novos arranjos afetavam as afinidades.”<sup>13</sup> Por fim, Earle insinua que o principal objetivo da análise era demonstrar o valor da tecnologia, em vez de resolver algum problema específico ou abordar determinada decisão na Osco.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Não havia muito a percorrer em descobertas anteriores, à exceção da certeza de que os itens para bebês eram muito lucrativos, daí a importância de descobrir artigos correlatos que, de alguma maneira, levassem à compra ou aumentassem os volumes de compras desses produtos.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Frequência da compra de vários itens, conforme os dados sobre ponto de venda na mercearia.

**COLETA DE DADOS.** A Osco tinha os dados de seus sistemas de pontos de venda e os disponibilizou para a equipe de analistas da Teradata. O banco de dados sob análise continha informações de compras referentes a 1,2 milhão de cestas de mercado (todos os itens comprados em determinada transação de compra) extraídos de 25 lojas da Osco Drug.

**ANÁLISE DOS DADOS.** Dispõe-se hoje de métodos de mineração de dados muito mais sofisticados, mas, em 1992, esses novos recursos ainda não estavam disponíveis ou eram pouco usados. A equipe da Teradata elaborou queries (consultas) para identificar os itens comprados por associação com outros, acima dos níveis normais. A análise de certo caso mostrou que os clientes (eles, na verdade, não eram identificados como homens) que compareciam às lojas entre 17 e 19 horas das quintas e sábados compravam cervejas e fraldas com muita frequência. Não se realizaram testes estatísticos para garantir que os níveis de afinidade observados não ocorriam por acaso.

**APRESENTAÇÃO DE RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Esse é o ponto em que a análise deixa a desejar. Não raro se afirma em relatos apócrifos da história que cervejas e fraldas eram colocadas em locais próximos nas lojas ou que, ao

contrário, eram dispostas estrategicamente em locais distantes para obrigar os clientes a atravessar toda a loja. De fato nada disso aconteceu. As descobertas foram consideradas divertidas, e ninguém tentou, sejam os analistas da Teradata, sejam os gestores da Osco, fazer alguma coisa com base nos dados, nem mesmo analisar a possibilidade de tomar alguma providência a partir desses resultados.

Não temos informações suficientes para compreender por que esse esforço criativo de descoberta de padrões não foi levado adiante, mas o exemplo ilustra o fato de que todos os aspectos do processo de raciocínio analítico devem ser eficazes para que daí decorra alguma mudança. Os computadores até podem ser capazes de identificar padrões nos dados, mas são seres humanos que os interpretam – e fazem alguma coisa com base neles (ver “Computadores e padrões: O caso do pi” e “Padrões no dígito inicial – Maneira de detectar fraude”).

## Computadores e padrões: O caso do pi

Pi ( $\pi$ ) é o quociente da divisão da circunferência pelo diâmetro. O valor de pi é de, aproximadamente, 3,141592, na notação decimal usual. Muitas fórmulas em matemática, ciência e engenharia incluem o pi, o que o torna o conceito matemático mais importante, além do teorema de Pitágoras.<sup>a</sup>

O pi é número irracional, o que significa que seu valor não pode ser expressado exatamente como fração, e, por conseguinte, sua representação decimal nunca deixa de repetir-se. Isso não significa, contudo, que os estudiosos – desde os babilônios antigos até os matemáticos do presente – tenham desistido de encontrar padrão nas muitas casas decimais com que o pi 0, tem sido calculado.

Evidentemente, o advento dos computadores digitais no século 20 resultaram em novos cálculos do pi. O recorde de casas decimais nos cálculos hoje é de cinco trilhões de dígitos. O aumento da quantidade de dígitos é consequência mais de avanços na capacidade de computação que de progressos na capacidade matemática dos seres humanos. Mesmo assim, entretanto, não se encontraram padrões iterativos na sucessão de dígitos.

Embora nem os humanos nem os computadores tenham identificado padrões na sequência das casas decimais do pi, os matemáticos descobriram muitos padrões nas aplicações do pi, que se destacam como nova função da criatividade humana no raciocínio analítico. Por exemplo, eis uma história narrada pelo matemático David Acheson:

*Imagine a surpresa, então, em meados do século XVII, quando os matemáticos descobriram o  $\pi$  irrompendo em todos os tipos de situações que, aparentemente, nada tinham a ver com círculos. Um desses resultados mais famosos é a extraordinária conexão entre pi e números ímpares:*

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$$

...são conexões espantosas como essa que deixam os matemáticos realmente surpresos.<sup>b</sup>

Descobriu-se outra conexão entre pi e números pares. Além disso, pi aparece nas “cinco famosas” equações desenvolvidas por Leonhard Euler, matemático suíço do século 18, consideradas “a mais bonita formulação matemática de todos os tempos”, em pesquisa de 1988, entre matemáticos.

A lição dos estudos sobre pi é que, mesmo quando os computadores assumem importante função matemática, a criatividade humana ainda encontra maneira de manifestar-se. Nas aplicações da análise quantitativa às organizações e às empresas, os computadores passaram a executar a grande maioria dos cálculos estatísticos. É no uso desses cálculos, em apoio ao processo decisório, onde ainda reside a criatividade.

a. Ver “Pi”, Wikipedia, <http://en.wikipedia.org/wiki/Pi>; “Ask Dr. Math FAQ: About Pi”, Math Forum, <http://mathforum.org/dr.math/faq.pi.html>; “Facts About Pi”, Math Forum, <http://mathforum.org/library/drmath/view/57543.html>; Finding pi: <http://mathforum.org/library/drmath/view/55815.html>; Pre-computer History of pi: <http://personal.bgsu.edu/~carother/pi/Pi2.html>.

b. David Acheson, *1089 and All That: A Journey into Mathematics* (Oxford: Oxford University Press, 2002, 14.

## Últimas palavras sobre analítica e criatividade

Esperamos ter demonstrado que raciocínio analítico e criatividade não só podem coexistir, mas também são processos altamente correlatos. Não se pode ser nem bom analista quantitativo nem bom consumidor de raciocínio quantitativo quando não se exploram todas as faculdades criativas. Lembre-se, contudo, que quão se é criativo na análise dos dados e na explicação dos resultados está sujeito a limites. A criatividade é importante, mas não tanto quanto a verdade.

## Padrões no dígito inicial – Maneira de detectar fraude

Ted Hill, professor de Matemática no Georgia Institute of Technology, passa para os alunos um dever de casa no começo de cada curso: (a) lançar uma moeda 200 vezes e registrar os resultados reais de cara ou coroa ou (b) meramente não lançar a moeda e forjar os 200 resultados. Na aula seguinte, ele passa os olhos pelos dados apresentados pelos alunos e identifica, para espanto geral, quase todos os casos de resultados simulados. Como ele consegue esse prodígio? Ele sabe, por meio de cálculos muito complexos, que, a certa altura de uma série de 200 lançamentos de moeda, cara ou



coroa aparecerá seis ou mais vezes em seguida. A maioria dos simuladores não sabe disso e evita apresentar longas repetições de cara ou de coroa, que, equivocadamente, consideram improvável. De relance, Hill localiza a sucessão de resultados idênticos nos 200 lançamentos dos alunos. A falta dessas repetições sugere que os resultados são falsos. Não se trata, porém, de simples truque em sala de aula: caso não se encontre nos dados alguma espécie de padrão esperado, é provável que tenha ocorrido simulação ou fraude.

Como todos sabemos muito bem que nosso sistema numérico usa dígitos de 1 a 9, presumimos intuitivamente que a chance de ocorrer aleatoriamente qualquer um dos dígitos é de 1/9. Contraintuitivamente, porém, a suposição não é verdadeira. A *lei do primeiro dígito*, também denominada Lei de Benford, afirma que nas listas de números extraídas de muitas fontes de dados da vida real, o dígito inicial se distribui de maneira específica, não uniforme: o primeiro dígito é 1 em 30% das vezes, e a frequência de ser o primeiro diminui à medida que aumenta o valor do dígito.<sup>a</sup> De acordo com a Lei de Benford, a probabilidade de cada dígito ser o inicial é a seguinte:

Primeiro dígito	1	2	3	4	5	6	7	8	9
Chance de ocorrer como primeiro dígito (%)	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6

Esse fato um tanto surpreendente foi descoberto em 1881, pelo astrônomo americano Simon Newcomb, ao observar que, nas tábuas logarítmicas, as primeiras páginas estavam muito mais gastas que as demais. Em 1938, o físico Frank Benford fez a mesma descoberta em quantidades de dados muito maiores que a de Newcomb, ao analisar 20.229 diferentes conjuntos de dados, envolvendo áreas de rios, estatísticas de beisebol, números de artigos em revistas e os endereços das primeiras 342 pessoas listadas na *American Men of Science*. Todos esses conjuntos de números aparentemente díspares apresentavam a mesma proporção de frequência de primeiros dígitos já observada nas tábuas logarítmicas. Esse padrão do dígito inicial acabou sendo denominado Lei de Benford, em honra dele. Hoje, reconhece-se que a Lei de Benford se aplica, inequivocamente, a muitas situações do mundo real.

Muitos estatísticos e contadores acreditam firmemente que a Lei de Benford é ferramenta relativamente simples, mas poderosa, para identificar possíveis defraudações, estelionatos, sonegações e falsificações contábeis. A ideia por trás do conceito é simples: alguém que manipule um conjunto de dados, dificilmente seguirá a distribuição prevista pela Lei de Benford. Basta comparar a frequência do dígito inicial de cada dado com a frequência provável para detectar a simulação. Em geral, os dados falsificados ou simulados têm muito menos números começando com 1 e muito mais começando com 6 que os dados reais.

Em 1972, Hal Varian, economista de Berkeley, mostrou que a lei poderia ser usada para detectar possíveis fraudes em listas de dados socioeconômicos apresentados em apoio a decisões sobre políticas públicas. Mark Nigrini, perito judicial, tornou-se famoso ao aplicar um sistema que concebera, com base na Lei de Benford, a alguns casos de fraude em Brooklyn. Hoje, muitos órgãos de arrecadação e fiscalização do imposto de renda usam softwares de identificação de fraudes, com base na Lei de Benford, assim como numerosas empresas de grande porte, inclusive consultorias de contabilidade. Nos Estados Unidos, provas judiciais baseadas na Lei de Benford são aceitas nos âmbitos federal, estadual e local.

a. Benford's Law. *Wikipedia*, [http://en.wikipedia.org/wiki/Benford%27s\\_law](http://en.wikipedia.org/wiki/Benford%27s_law); Malcolm W. Browne, "Following Benford's Law, or Looking Out for N. 1", *New York Times*, 4 de agosto de 1998, <http://www.nytimes.com/1998/08/04/science/following-benford-s-law-or-looking-out-for-no-1.html?pagewanted=all&src=pm> "Benford's Law", *Wolfram MathWorld*, <http://mathworld.wolfram.com/BenfordsLaw.html>; T.P. Hill, "The First-Digit Phenomenon", *American Scientist*, julho-agosto de 1998; Mark J. Nigrini, "Benford's Law",

## Exemplo de raciocínio analítico: Capacidade linguística e doença de Alzheimer

Doença de Alzheimer, ou Alzheimer, é disfunção do cérebro que provoca problemas de memória, raciocínio e comportamento. Os sintomas em geral se desenvolvem lentamente, pioram com o tempo e se tornam graves demais a ponto de interferir nas tarefas diárias, e, finalmente, culminam com a morte. A doença de Alzheimer responde por algo entre 60% a 80% dos casos de demência. Aproximadamente 5,3 milhões de americanos e uma em cada oito pessoas com mais de 65 anos (13%) têm Alzheimer, que é a sexta principal causa de morte nos Estados Unidos. Além das aflições do paciente, os familiares e cuidadores do doente padecem de dores emocionais e físicas, uma vez que as exigências do cuidado no dia a dia, as novas atribuições impostas aos familiares e as decisões difíceis sobre internação em organizações especializadas podem ser muito dolorosas.

Não se sabe muito sobre as causas e o avanço do Alzheimer. Muitos pesquisadores tentaram compreender a associação de certas características, ou marcadores, das pessoas com maior risco de Alzheimer. Por exemplo, indivíduos com menos educação formal talvez sejam mais propensos ao desenvolvimento de demência e outras doenças, por causa das diferenças de estilo de vida associadas a instrução, nutrição, consumo de álcool e insalubridade no trabalho. David Snowdon, professor do Sandersbrown Center, em Aging, da Universidade de Kentucky, e colegas, acham que as habilidades linguísticas na juventude pode ser melhor marcador que estilo de vida.<sup>14</sup> Eles especularam que alto nível de capacidade linguística na infância pode agir como amortecedor do declínio cognitivo, ao facilitar processos mnemônicos de codificação, organização e recuperação de informações. Para chegar a essas conclusões, conduziram estudo pioneiro e altamente criativo, associando capacidade cognitiva na juventude com o risco de Alzheimer. A escolha de uma amostra da população – uma ordem de freiras – e o uso de autobiografias como fonte de dados foram abordagens muito criativas ao pensamento analítico. Examinaremos agora o trabalho deles, sob o referencial de seis passos.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Determinar se as habilidades linguísticas na juventude se associam ao funcionamento cognitivo e ao baixo risco de Alzheimer no futuro.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Muitos dos procedimentos adotados pela equipe de Snowdon basearam-se em trabalho anterior do Dr. David Wekstein e do Dr. William Markesbery. Em 1989, ambos iniciaram um estudo de mudanças associadas à idade, na cognição e na funcionalidade de um grupo de idosos que haviam concordado em doar o cérebro após a morte. O foco do estudo era compreender como mudanças no cérebro poderiam correlacionar-se com Alzheimer e outros transtornos neurológicos na idade avançada.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Os participantes do estudo de Snowdon foram freiras da congregação religiosa School Sisters of Notre Dame, em Milwaukee, Wisconsin. De 1991 a 1993, as mulheres do convento que tinham nascido antes de 1917 foram convidadas a participar do Nun Study (Estudo das Monjas), pesquisa longitudinal do envelhecimento e do Alzheimer. Das 1.027 elegíveis, 678 (66%) concordaram em participar e deram autorização por escrito. A taxa de participação no Estudo das Monjas foi relativamente alta, considerando que todos os participantes concordaram em doar o cérebro depois da morte, assim como submeter-se a avaliações anuais das funções cognitiva e física. Snowdon e equipe investigaram ainda um subconjunto de 93 participantes no Estudo das Monjas, que haviam preparado autobiografias manuscritas da vida pregressa, arquivadas no convento. As variáveis selecionadas são as seguintes:

- Habilidades linguísticas (densidade das ideias e complexidade gramatical)
- Funções cognitivas (sete dimensões diferentes) e doença de Alzheimer na vida póster

**COLETA DE DADOS (MENSURAÇÃO).** As autobiografias das irmãs foram usadas para caracterizar o nível de habilidades linguísticas na juventude. Depois de passar, em média, quatro anos em treinamento no convento, cada irmã escreveu uma autobiografia poucas semanas antes de fazer os votos religiosos. As informações constantes dos arquivos do convento indicavam que se pediu a cada irmã para “... escrever uma breve descrição da própria vida. Esse relato deve conter não

mais do que 200 a 300 palavras e caber em uma única folha de papel... incluindo local de nascimento, genitores, fatos interessantes e edificantes da infância, escolas frequentadas, influências que as atraíram para a vida religiosa, e outros eventos notáveis”.

Extraíram-se de cada biografia dois indicadores de capacidade linguística: *densidade das ideias* e *complexidade gramatical*. Definiu-se *densidade das ideias* como número médio de ideias expressas a cada 10 palavras. Consideraram-se *ideias* proposições elementares, compostas tipicamente de verbo, adjetivo, advérbio ou frases preposicionadas. Também se contaram proposições complexas, que afirmavam ou inferiam relações causais, temporais ou outras entre ideias. Avaliou-se a *complexidade gramatical* com base no critério *Nível de Desenvolvimento*, que classifica os períodos sintáticos em oito níveis, de 0 (períodos simples com uma oração) a 7 (períodos complexos, com várias orações coordenadas ou subordinadas).

Mediu-se a função cognitiva por meio de uma bateria de sete testes neuropsicológicos, abrangendo memória, concentração, linguagem, capacidade visual-espacial e orientação temporal e espacial. As 93 freiras do estudo escreveram as autobiografias com a idade média de 22 anos, e foram avaliadas quanto à função cognitiva 58 anos depois, com idade entre 75 e 87 anos.

**ANÁLISE DOS DADOS.** Baixa densidade de ideias e baixa complexidade gramatical nas autobiografias escritas na juventude se associaram a baixos escores nos testes cognitivos na velhice. Baixa densidade de ideias na juventude se correlacionou de maneira mais intensa e consistente que baixa complexidade gramatical com deficiências na função cognitiva. Entre as 14 irmãs que morreram, confirmou-se por meio de exame neuropatológico a presença de Alzheimer em todas aquelas com baixa intensidade de ideias na juventude e em nenhuma com alta intensidade de ideias.

**APRESENTAÇÃO DE RESULTADOS E TOMADA DE PROVIDÊNCIAS.** Snowdon e colegas concluíram que o desempenho linguístico escrito, medida da capacidade cognitiva na juventude adotada pelo estudo, “... é poderoso marcador de problemas cognitivos, doença de Alzheimer e lesões no cérebro na velhice”. Isso indica que a baixa capacidade linguística na juventude pode ser sintoma sutil de mudanças muito precoces no cérebro, que acabariam levando ao Alzheimer. Esses tipos de estudos nos permitirão prever com mais eficácia o risco de sofrer

várias doenças relacionadas com a idade. Com o título “Linguistic Ability in Early Life and Cognitive Function and Alzheimer’s Disease in Late Life”, o trabalho foi publicado no importante *Journal of the American Medical Association*.<sup>15</sup> Snowdon também escreveu um livro popular sobre as freiras, intitulado *Aging with Grace: What the Nun Study Teaches Us About Leading Longer Healthier, and More Meaningful Lives*. Eis o que se lê em resenha do livro no *Library Journal*:

Snowdon escreve com empatia e afeição sobre essas irmãs, que, com generosidade, também concordaram em doar os cérebros para estudos patológicos post-mortem. Com base na pesquisa, explica Snowdon, ficou claro que mudanças patológicas nem sempre se correlacionam com mudanças observáveis, que a capacidade linguística parece proteger contra a doença de Alzheimer, que a prevenção de ataques e doenças cardíacas pode ajudar a evitar a demência, e que hereditariedade, dieta e exercícios físicos também são importantes. Misturando histórias pessoais com fatos científicos, esse olhar inspirador e fascinante sobre envelhecer é altamente recomendado.<sup>16</sup>

O estudo de Snowdon com as freiras também foi tema de reportagem de capa da revista *Time*, em 2009.<sup>17</sup> O trabalho mostra que o raciocínio analítico criativo pode alcançar público muito amplo.

### **Exemplo de raciocínio analítico: Simon Hannes e informações privilegiadas**

Simon Hanes já trabalhara como analista de investimentos no Macquarie Bank. Sob o pseudônimo “Mark Booth”, adquiriu opções de compra sobre ações da TNT, empresa transportadora, por cerca de US\$90 mil. O Macquarie prestava consultoria à TNT. Pouco antes de deixar a instituição financeira, Hannes teve acesso indireto a informações de que a TNT talvez viesse a ser adquirida. A empresa de fato foi adquirida dois dias depois de Hannes comprar as opções, e lucrou mais de US\$2 milhões na transação. Hannes disfarçou com inteligência suas operações, exigindo trabalho analítico criativo por parte dos investigadores da Australian Securities and Investments Commission (ASIC) para descobrir a

fraude.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Irregularidades nas operações com opções sobre ações da TNT foram detectadas três dias antes do anúncio da aquisição da empresa. A divulgação da informação resultou em aumento de 200 vezes no retorno sobre o investimento em menos de 72 horas. Embora os órgãos fiscalizadores suspeitassem de que numerosas organizações e indivíduos haviam operado no mercado naqueles três dias, não se conseguia identificar e localizar um dos primeiros operadores. Até que levaram a questão à ASIC, agência reguladora do mercado. Apesar da mobilização, durante três meses, de muitos recursos investigativos, com base em técnicas de investigação tradicionais, a ASIC não conseguia identificar a(s) pessoa(s) por trás dessas primeiras transações. Admitiu-se o uso de identidade falsa como a causa mais provável do fracasso.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Não havia descobertas anteriores específicas sobre o caso, por se tratar de evento isolado. Pesquisas pretéritas e a experiência da ASIC sugeriam que, tipicamente, transações fraudulentas em geral envolvem redes de pessoas, o que seria fator-chave para a investigação. A ASIC já desenvolvera meios para identificar redes a partir de pessoas, empresas ou endereços existentes em bancos de dados do próprio órgão ou de acesso público.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** As duas principais variáveis do modelo eram acesso a informações sobre a possível aquisição da TNT e disponibilidade de recursos (conta bancária, dinheiro etc.) para efetuar a operação, envolvendo, inclusive, eventuais saques, mais ou menos coincidentes com as transações.

**COLETA DE DADOS (MENSURAÇÃO).** As técnicas investigativas tradicionais tinham condições de identificar pessoas capazes de, ou mais tendentes a, acessar informações privilegiadas (“pessoas com conhecimento”). Do mesmo modo, os investigadores dispunham de meios para descobrir saques em instituições financeiras, nas áreas geográficas mais prováveis, por indivíduos e organizações (“pessoas com dinheiro”). A partir daí, podiam extrair e coletar de maneira sistemática elementos (indivíduos, empresas, endereços e ativos) afastados cinco elos das “pessoas com conhecimento” e das “pessoas com dinheiro”, mediante

processos de extração de rede, obtendo, assim, conjuntos de dados analíticos com mais de 160 mil indivíduos, empresas, endereços, ativos e saques, interligados através de um milhão de conexões.

**ANÁLISE DOS DADOS.** Os conjuntos de dados continham várias redundâncias. Em consequência, antes de dar o próximo passo da análise, os investigadores resolveram fundir entidades semelhantes no conjunto de dados (ou seja, indivíduos, empresas, endereços e ativos que pareciam idênticos foram unificados para fins de análise). A ASIC usou mais de uma centena de algoritmos proprietários para realizar a tarefa. Anthony Viel, então investigador-chefe da ASIC no caso, e hoje sócio de analítica e criminalística da Deloitte Australia, observou o seguinte sobre o restante do processo: “Depois da unificação de todas as possíveis redundâncias, usamos o algoritmo de caminho mínimo para identificar associações entre pessoas com conhecimento, de um lado, e saques de recursos, de outro. Separamos os tipos de conexões em associações ‘hard’ e ‘soft’, para refinarmos os resultados. No primeiro passe, identificamos 65 pessoas potenciais, com conexões hard e soft afastadas três elos. No segundo passe, identificamos duas pessoas potenciais, afastadas quatro elos, considerando apenas conexões hard. Uma dessas pessoas era falso positivo, por ter decorrido de unificação incorreta. Só restou aquela que de fato nos interessava.”<sup>18</sup>

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Simon Hannes, a pessoa de interesse, identificada por meio da análise de redes, foi abordada com mandado de busca. Vários elementos incriminadores foram encontrados em sua casa, que possibilitaram o indiciamento dele por vários crimes correlatos. Ele foi julgado e condenado pelo júri, sentença que foi mantida em recurso. Ficou preso durante dois anos e meio, foi multado em US\$100 mil, e não pode manter o produto das transações ilícitas, pelo uso de informações privilegiadas. Viel ainda usa técnicas semelhantes para analisar fraudes financeiras e de outros tipos.

## 6

# Desenvolvimento de recursos de análise quantitativa

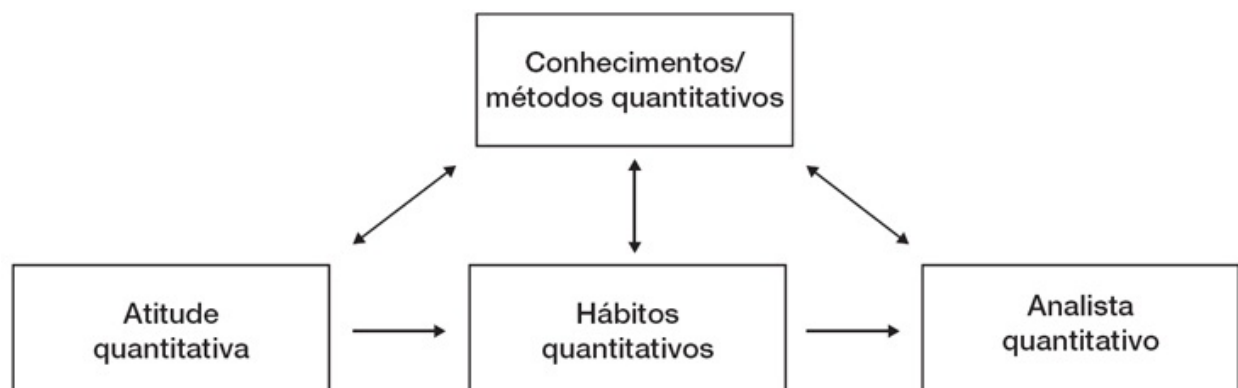
Como se determina o destino de alguém ou um estilo de vida? Conforme Aristóteles afirmou muito tempo atrás, nossos hábitos, ou o que fazemos repetidamente, define o que somos e configura nosso destino. Essa asserção pode ser representada pelo seguinte fluxograma:

Pensamentos → ações → hábitos → caráter → destino

Os pensamentos determinam as ações. As ações geram os hábitos. O que se faz reiteradamente forma a imagem que se projeta para os outros. A maneira como se é visto pelos outros traça o próprio destino. O processo que o leva a tornar-se analista quantitativo competente segue padrão semelhante, como mostra a [Figura 6.1](#).

**FIGURA 6.1**

### Processo de tornar-se um analista quantitativo competente





---

As atitudes quantitativas usuais formam os hábitos quantitativos. O trabalho reiterado com números desenvolve o analista quantitativo. Os conhecimentos e métodos quantitativos se aprimoram gradualmente pelas interações com atitudes e hábitos quantitativos. Vejamos que atitudes e hábitos quantitativos é preciso cultivar para tornar-se analista quantitativo proficiente, e que conhecimentos e métodos quantitativos é necessário desenvolver em cada estágio.

## **Atitude quantitativa**

Embora o conhecimento de métodos quantitativos decerto seja útil quando se quer tomar decisões analíticas, é igualmente importante manter-se aberto ao aprendizado de números e persistir em altos padrões de comprovação. Quando não se foge espavorido de números e equações, é possível dominar de maneira surpreendente ferramentas importantes de análise quantitativa.

## **Não tenha medo de números**

Como já sugerimos, o inumerismo, ou incapacidade de executar e entender operações aritméticas, aflige muita gente esclarecida e bem informada. Muitas pessoas “normais” têm aversão patológica à matemática, o que não é apenas infortúnio, mas também empecilho. A matemática, contudo, não é fundamental para o bom raciocínio quantitativo. O que distingue os bons pensadores quantitativos não são suas qualificações em matemática pura, mas, sim, a abordagem ao tratamento da informação quantitativa. Os pensadores quantitativos eficazes cultivam certas atitudes, habilidades e hábitos a que recorrem sempre que precisam tomar decisões baseadas em números. Alguns especialistas até afirmam que a matemática que se aprende no ensino fundamental é suficiente para tornar-se bom pensador quantitativo: “Conhecimento de matemática e raciocínio quantitativo são coisas muito diferentes... Uma das razões para salientar a distinção entre conhecimento de matemática e raciocínio quantitativo é que ser bom pensador quantitativo exige muito pouca matemática além do nível de 5ª série do ensino fundamental.”<sup>1</sup>

Para os incrédulos, cite-se o exemplo de Emily Rosa. Ela ainda nem alcançara

a 6ª série, quando, aos 9 anos, realizou uma análise quantitativa do tipo cientista maluco a respeito do conceito de toque terapêutico. O trabalho dela foi considerado bastante sofisticado e útil para ser publicado no *Journal of the American Medical Association*.<sup>2</sup> (Publicado ininterruptamente desde 1883, o JAMA está entre os mais prestigiosos periódicos médicos do mundo, destacando-se como o de maior circulação.)

Como será que um projeto de ciências da 4ª série do ensino fundamental chegou ao JAMA? Emely viu a mãe, Linda, assistindo a um videotape sobre uma atividade em expansão, denominada *toque terapêutico* (TT), método controverso de tratar doenças, mediante a manipulação de “campos de energia” do paciente. Os proponentes afirmam que mais de 100 mil pessoas em todo o mundo já foram treinadas nas técnicas de TT, lecionada em mais de 100 faculdades e universidades em 75 países. É considerada a terapia mais reconhecida dentre as praticadas por profissionais de enfermagem holista. Durante a sessão de TT, o terapeuta movimentava as mãos a apenas alguns centímetros do corpo do paciente, para detectar e remover “energias estagnadas”, que, assim se supõe, provocam várias doenças. Emily sugeriu à mãe que ela talvez fosse capaz de realizar um experimento para verificar a veracidade de tais alegações. Linda, que era enfermeira, deu a Emily alguns conselhos sobre o método. Depois de algumas pesquisas, Linda e Emily descobriram que nenhum estudo anterior tentara confirmar se os praticantes de TT realmente eram capazes de detectar campos de energia humana.

Emily concentrou-se na seguinte questão básica: se os terapeutas realmente são capazes de fazer o que alegam, eles, então, pelo menos sentirão o campo de energia. Do contrário, o valor médico da TT será duvidoso. Portanto, se a porcentagem da percepção correta dos campos de energia não for consideravelmente superior à correspondente a acertos aleatórios, as pretensões dos adeptos do toque terapêutico são infundadas.

Por meio de anúncios e de outras pistas, Emily localizou 25 praticantes de TT no nordeste do Colorado, dos quais 21 concordaram em submeter-se ao teste. Os terapeutas foram informados de que o estudo seria publicado como projeto dela na feira de ciências da 4ª série. Durante cada teste, os terapeutas deixavam as mãos, com as palmas para cima, sobre uma superfície plana, afastadas 25 a 30cm uma da outra. Para evitar que as mãos do experimentador fossem vistas, instalava-se uma placa alta e opaca, com aberturas na base, sobre os braços do

paciente, prendendo-se uma toalha à tela, que lhe cobria os braços.

Emily lançava uma moeda para determinar qual das mãos do paciente seria o alvo (algo conhecido na área de testes quantitativos como *randomized assignment to groups* [atribuição aleatória]). Ela, então, pousava a mão direita, com a palma para baixo, de 8 a 10cm acima do alvo e dizia “o.k”. O terapeuta, então, afirmava qual das duas mãos dele ou dela estava mais perto da mão do experimentador. Os 21 terapeutas fizeram 280 tentativas.

Embora todos os participantes tivessem afirmado que seriam capazes de passar no teste, os terapeutas localizaram corretamente a mão de Emily em apenas 122 (44%) das 280 tentativas, resultado não superior ao número de acertos prováveis, por mera adivinhação (ver <http://keepingupwiththequants.weebly.com>). Algo em torno de 50% seria o esperado por mero acaso. Emily concluiu que a incapacidade dos terapeutas de comprovar o princípio mais fundamental do toque terapêutico era evidência inquestionável de que as pretensões do método careciam de fundamento e de que seu uso profissional era injustificável. Em abril de 1998, quando Emily já tinha 11 anos e estava na 6ª série, o experimento foi publicado na JAMA. George Lundberg, editor do periódico, afirmou que os estatísticos do JAMA ficaram surpresos com a simplicidade e com a clareza dos resultados.<sup>3</sup> Emily conquistou seu lugar no *Guinness Book of World Records* como a pessoa mais jovem a ter uma pesquisa publicada em importante periódico científico.

Emily não é gênio nem criança prodígio. Só usou bom senso e raciocínio criativo claro. A capacidade mais importante para ser analista quantitativo competente é extrair informações dos dados, por meio não de conhecimentos matemáticos, mas, sim, de raciocínio quantitativo. O primeiro passo nessa direção é superar o medo dos números e sentir-se à vontade com eles, ou seja, a certeza de já possuir ou de ser capaz de desenvolver os conhecimentos matemáticos necessários. O requisito básico para ser analista quantitativo competente é ser amigo dos dados, em vez de fugir deles.

Mesmo que suas experiências com matemática na escola não tenham sido boas, é sempre possível recuperar-se; ninguém está condenado a adotar perspectiva não quantitativa durante toda a vida. O fato de ter descarrilado em matemática no começo não significa que seja impossível aprender o necessário mais adiante na carreira. Por exemplo, DJ Patil, que cunhou o termo *cientista de dados* para descrever sua função, ao liderar a área de analítica na LinkedIn,

empresa de networking profissional, experimentou um conjunto desastroso de reveses matemáticos na escola, que só veio a superar mais tarde: “A primeira vez que fracassei em matemática foi na 8ª série. Continuei a enfrentar dificuldades durante todo o curso de ensino médio, mal conseguindo formar-me; a ponto de precisar de autorização especial a fim de passar para o nível seguinte, no qual também fui reprovado. Como não podia entrar no curso superior, ingressei numa ‘junior college’... matriculei-me em cálculo e pedi transferência para a Universidade da Califórnia, onde constatei que não tinha ideia de nada daquilo. Eu continuava a ter problemas com matemática”, confessou. “Por fim, concluí que tudo aquilo era ridículo e que eu precisava entrar na universidade.”<sup>4</sup>

Patil pegou alguns livros na biblioteca e, em um único fim de semana, aprendeu sozinho toda a matemática do ensino médio. Saiu-se bem em cálculo e foi admitido na Universidade da Califórnia, San Diego, onde concluiu o currículo de matemática em três anos. Começou, então, a estudar para o PhD em Matemática Aplicada, na Universidade de Maryland.

Mais uma vez, teve alguma dificuldade no começo: “Eu era um dos americanos, competindo com russos, israelenses e coreanos. Meus resultados no primeiro exame de qualificação foram péssimos; acho que fiquei em penúltimo lugar, e o último foi alguém que não apareceu.”

Na segunda vez em que tentou o exame, conseguiu as notas mais altas. Patil acabou graduando-se e tornou-se membro do corpo docente de Maryland, onde trabalhou em modelagem climática. Em seguida, prestou serviços ao governo federal, em questões de inteligência. Como os financiamentos de pesquisas eram escassos na época, foi trabalhar na Skype, que, na época, era controlada pela eBay. Em seguida, tornou-se cientista-chefe de dados da LinkedIn, onde os analistas quantitativos exerciam enorme influência no desenvolvimento de produtos.

Agora, Patil é cientista de dados residente (primeiro profissional com esse título) na empresa de capital de risco Greylock Partners, ajudando as startups do portfólio a raciocinar sobre dados e analítica. Ele talvez seja o melhor exemplo de talento matemático latente.

**Recorra à internet para aprender conceitos numéricos**

Na era da globalização, muitos termos novos referentes a econômicos, financeiros, científicos e estatísticos aparecem todos os dias nos noticiários e nas conversas. Muita gente que não está familiarizada com a maioria desses neologismos simplesmente os deixa passar, sem compreendê-los. Quem pretende tornar-se analista quantitativo, contudo, deve memorizá-los ou anotá-los, para procurá-los no Google. Para tanto, não faltam verbetes na Wikipédia, cursos online e livros-textos aos montes. Arquive os resultados da pesquisa e organize-os para consultas no futuro. Trata-se de maneira eficaz de aprender e de superar aos poucos a resistência aos números. Você provavelmente não compreenderá tudo o que ler, mas decerto estará assimilando alguma coisa. Se persistir no esforço durante pelo menos seis meses, acabará surpreendendo-se ao constatar que os amigos e colegas passarão a considerá-lo pessoa bem informada, principalmente quando se tratar de conceitos quantitativos.

### Amplie sua curiosidade

Quem cultiva a curiosidade sobre números amplia os conhecimentos sobre conceitos quantitativos e passa a compreendê-los com mais profundidade. Por exemplo, ao descobrir que o homem mais veloz do mundo é Usain Bolt, da Jamaica, você provavelmente terá curiosidade em conhecer os melhores tempos dele. E descobrirá que o recorde dele em 100 metros é 9,58 segundos. Pergunte-se, então, a que isso equivale, em quilômetros por hora ou milhas por hora:

$$9,58 \text{ segundos} \approx 10 \text{ segundos} = \frac{1}{6} \text{ minuto} = \frac{1}{360} \text{ hora}$$

$$100 \text{ m} = 0,1 \text{ km}$$

$$\text{Em que, } \frac{0,1 \text{ km}}{\frac{1}{360} \text{ hora}} = 36 \text{ kph} = 22,4 \text{ mph}$$

Agora você talvez esteja curioso sobre o recorde da maratona, que é 2:3:59. Por meio de cálculos simples, você conclui que essa marca corresponde à velocidade média de 20,4 km/hora (12,7 milhas/hora). Comparando essas duas

velocidades médias, você descobre as velocidades máximas do ser humano na prova mais curta e na prova mais longa das corridas olímpicas.

Curiosidade sobre todos os aspectos dos números é o atributo básico do bom analista quantitativo.

## Atitudes quantitativas ↔ Conhecimento quantitativo

A essa altura, para desenvolver o conhecimento quantitativo, é preciso ler livros de introdução à estatística para o grande público. Um dos mais populares é *How to Lie with Statistics*, de Darrel Huff, best-seller da segunda metade do século 20.<sup>5</sup> O livro é compêndio ilustrado leve e fluente, que expõe os erros comuns, intencionais e não intencionais, referentes à interpretação de conceitos estatísticos e mostra como esses erros podem levar a conclusões inexatas. É texto excelente que esclarece como os especialistas mentem com estatística e como os usuários devem compreender os números. *Flaws and Fallacies in Statistical Thinking*, de Stephen K. Campbell, publicado em 1974, é ainda muito útil, considerando a data de lançamento, para quem lê jornais e revistas e depara com números estatísticos.<sup>6</sup> Livro moderno de um consultor em estatística, sobre o mesmo tópico, é *Common Errors in Statistics and How to Avoid Them*, de Phillip Good e James Hardin.<sup>7</sup> Há muitos outros trabalhos desse tipo por aí. Procure na internet títulos populares, leia algumas das resenhas e selecione os que lhe parecerem mais adequados às suas necessidades.

## Desenvolva o raciocínio probabilístico

Aspecto importante do raciocínio quantitativo é a compreensão das leis da probabilidade e da aleatoriedade. A falta dessa compreensão é uma das mais graves falhas intelectuais da maioria dos adultos. Como disse Nassim Nicholas Taleb, em livro cujo título é o próprio epítome, o ser humano, em geral, é “iludido pelo acaso” – imputa significado cósmico a acontecimentos que facilmente poderiam resultar do acaso.<sup>8</sup> Exemplo significativo é o mal-entendido da “festa de aniversário comum”. Numa festa de aniversário com apenas 23 pessoas presentes, dois convidados descobrem que fazem aniversário no mesmo dia. “Quais são as chances?”, talvez comentem entre si. Ocorre que as chances são muito altas – na verdade, pouco mais de 50% (ver

<http://keepingupwiththequants.weebly.com>).

O conhecimento de probabilidade é extremamente útil, para compreender não só as festas de aniversário, mas também numerosos empreendimentos humanos. Quem não conhece probabilidade não compreende que o mercado é um *passeio aleatório* ou *random walk* (ou seja, as variações nos preços das ações não seguem padrão ou tendência discernível) e que alguns *stock pickers* (investidores ativos que compram ações com base em determinados critérios) alcançam resultados superiores à média durante alguns anos seguidos, mas, inevitavelmente, acabam sofrendo perdas vultosas). Tampouco compreenderá o fenômeno da *regressão à média*; se sua renda for bem superior à média, por exemplo, a de seu filho provavelmente será inferior à sua. Você tenderá a perder muito dinheiro se frequentar os cassinos de Las Vegas sem conhecer probabilidade. Também terá dificuldade em conseguir emprego no setor de seguros ou em definir se e quando fazer certos investimentos em renda fixa. E evidentemente, probabilidade é a base da estatística inferencial. Em resumo, é preciso compreender probabilidade não só para ser pensador quantitativo, mas também para alcançar sucesso na vida.

Muitas são, evidentemente, as maneiras de aprender probabilidade. A maioria dos livros-textos aborda o tópico no contexto de estatística básica, pois estatística requer raciocínio probabilístico. Os livros que mencionei neste capítulo têm algumas seções sobre probabilidade. Quem quiser concentrar-se em probabilidade, não em estatística, em geral, deve interessar-se por um livro-texto introdutório sobre probabilidade, *The Pleasures of Probability*, de Richard Isaac.<sup>9</sup> Para quem detesta compêndios, uma alternativa bem escrita e divertida para o grande público é *O Andar do bêbado: Como o acaso determina nossas vidas*, de Leonard Mlodinow.<sup>10</sup>

## De volta às aulas

Não é necessário ser autodidata em analítica, probabilidade e raciocínio quantitativo a contragosto. Vários são os recursos de aprendizado, on-line ou face a face. As universidades oferecem material didático gratuito on-line, inclusive diversos cursos sobre estatística. Por exemplo, quem quiser aprender em uma das universidades com mais orientação estatística no mundo pode clicar no curso de probabilidade e estatística do MIT.<sup>11</sup> Em termos de educação on-line



remunerada, é possível comprar da Harvard Business Publishing, por pouco mais de US\$100, um curso sobre métodos quantitativos, de um colega de Tom na Harvard Business School, Jan Hammond. Quem não se importar em gastar muito mais, pode obter o grau de master of science em analítica preditiva, por meio de cursos totalmente on-line, da Northwestern University.

Para quem prefere relacionamentos face a face com os professores e colegas, aumenta cada vez mais a oferta de programas de pós-graduação em Analítica nas universidades americanas, a maioria outorgando título de mestrado. O vovô de todos eles, por ter sido fundado cinco anos atrás, e um dos mais bem concebidos, é o de Master of Science in Analytics (MSA), da North Carolina State University.<sup>12</sup> Por que NC State? Bem, Jim Goodnight, CEO e fundador da SAS, empresa de software de analítica, foi aluno e professor da instituição, e, generosamente, doou os recursos para o desenvolvimento do programa.

O MSA é programa profissional que se empenha em aprofundar a compreensão das ferramentas, métodos, aplicações e práticas de analítica avançada. Em nossa opinião, os princípios do projeto são bem adequados ao tema e às necessidades de pessoas com orientação quantitativa em empresas e organizações:

- *Curta duração*: dez meses de estudo intensivo – três semestres (verão, outono e primavera); matrícula em julho e graduação em maio do ano seguinte; aprendizado em tempo integral.
- *Carga de estudos árdua*: em tempo integral (Se-Se/9-5) no *campus*; currículo integrado, compartilhado com os outros alunos; trabalho em equipe durante todo o programa; e, em geral, atuação em projetos, quando não em sala de aula.
- *Foco em conteúdo amplo e prático*: currículo integrado e multidisciplinar (envolvendo várias escolas e departamentos da NC State), com o propósito de promover o desenvolvimento de habilidades práticas, a serem aplicadas na solução de problemas do mundo real, incluindo áreas como estatística, matemática aplicada, ciências da computação, pesquisa operacional, finanças, economia e marketing.
- *Aprendizado em ação*: avaliação eminentemente prática, em lugar da tese



tradicional. (Os alunos trabalham em equipes de cinco, resolvem problemas do mundo real, recorrem a dados fornecidos por empresas; trabalho substantivo altamente estruturado, com a duração de sete meses, que culmina com um relatório final para o patrocinador.)

O NCSU MSA tem novo currículo, composto de turmas desenvolvidas exclusivamente para o programa. Os tópicos incluem mineração de dados, mineração de textos, previsão, otimização, banco de dados, visualização de dados, privacidade e segurança de dados, analítica financeira e analítica de clientes.

Os participantes do programa apresentam ampla diversidade de antecedentes, embora seja desejável certo grau de orientação quantitativa. A idade média dos alunos é de 27 anos, e cerca de 26% dos matriculados já é pós-graduado. Mais ou menos metade dos alunos trabalhava até então em tempo integral. Apesar das incertezas da economia e não obstante a lenta recuperação dos níveis de emprego, a turma de 2011 do MSA alcançou o recorde de 469 entrevistas de emprego (média de 20 para cada um dos 40 alunos – que aumentou para 80 em 2012-2013). O programa registrou 100% de colocação no mercado de trabalho durante cinco anos consecutivos. Considerando que a demanda do mercado por pós-graduados em MSA vem aumentando, é natural que outras universidades desenvolvam programas de pós-graduação semelhantes. A SAS patrocinou novas iniciativas nas universidades Louisiana State e Texas A&M. A Universidade de San Francisco também criou programa congênere na Costa Oeste, e a Universidade de Nova York iniciou sua própria versão em Nova York. Pesquisa recente revelou que 59 universidades ofereciam graduação ou especialização em Analítica de Negócios ou em Business Intelligence – 37 em nível de mestrado e 22 em nível de graduação.<sup>13</sup> As escolas também estão começando a oferecer cursos em ciência de dados, a que em breve se seguirão programas de pós-graduação.

## Hábitos quantitativos

As atitudes são importantes, mas também os hábitos. Já se disse várias vezes que é mais fácil agir para desenvolver nova maneira de pensar que pensar para desenvolver nova maneira de agir. Praticando com regularidade e persistência os

seguintes hábitos quantitativos, você acabará cultivando atitudes quantitativas.

## **Exija números**

Os bons pensadores quantitativos (e as organizações que pretendem fomentá-los), devem sempre pedir números quando alguém expõe ideias, palpites e observações casuais. Sempre pergunte “Você tem dados para sustentar essa hipótese?”. Se você for realmente corajoso, não hesite em afirmar “O plural de caso não é dados”. Exigir dados é importante porque responde às seguintes perguntas: Como essa pessoa está pensando? Por meio de que processos e com a ajuda de que ferramentas alcançou-se esse curso de ação? Além disso, para agir conforme o discurso, combata o próprio impulso de se precipitar em conclusões sem dispor de dados suficientes.

As mesmas recomendações se aplicam quando você apresenta suas ideias aos outros. Na falta de dados, faça algum experimento de cientista maluco para reunir alguns. É importante sentir a necessidade de buscar dados objetivos antes de desenvolver teorias. Ao fazê-lo, você reforçará suas ideias com análises quantitativas cuidadosas e aumentará suas chances de convencer os outros com suas ideias. O hábito de exigir números e de validar suas ideias com números é indispensável para tornar-se analista quantitativo profissional.

## **Nunca confie nos números**

Acabamos de afirmar que você sempre deve reunir números e outros dados para sustentar suas teorias e visões de mundo. Agora, diremos que você sempre deve suspeitar dos números e dados. Da mesma maneira como você age com alguém que acabou de conhecer, não confie nos dados até conhecê-los melhor. Nunca aceite de imediato, sem questionamentos, quaisquer números que lhe sejam apresentados. Várias são as razões pelas quais os números talvez não sejam confiáveis. As pessoas (principalmente os políticos) às vezes mentem e trapaceiam com os números, ou, ao menos, os distorcem para promover intenções ocultas. Como observou Andrew Lang, poeta e crítico escocês, “ele usa estatística como bêbados usam postes, mais para apoio que para iluminação”.<sup>14</sup> Os números não raro estão ultrapassados e, portanto, já não

refletem a realidade. Também é possível que não sejam boa amostra (isto é, não representam a população que se tenta descrever). Adotar perspectiva crítica em relação aos números é procedimento adequado, em especial quando os números são surpreendentes e atípicos. A melhor maneira de adquirir confiança nos números é ser cético de início e verificar suas bases. Mais especificamente, as dúvidas em relação aos números devem concentrar-se nas três categorias seguintes:

**RELEVÂNCIA.** Os números devem ser relevantes para as questões a que se referem e representativos do grupo ou entidade que se pretende analisar. Se os números não oferecerem respostas, eles não fazem sentido.

**EXATIDÃO.** Se os números forem relevantes, mas inexatos, é necessário descartá-los. É possível avaliar a exatidão dos números questionando quem os produziu e como foram produzidos. Os números que não passam no teste de credibilidade são inúteis.

**INTERPRETAÇÃO CORRETA.** Mesmo quando exatos, os números podem ser enganosos, se forem mal interpretados. Mormente as pessoas com segundas intenções tendem a manipular os números deliberadamente. Vejamos um exemplo em que o mesmo número foi interpretado de maneira a justificar preconceitos:

Um crítico da revista *Newsweek*, que fez a resenha do livro *The Better Half*, sobre as primeiras sufragistas, concluiu a crítica com uma observação instigante, ao indagar, retoricamente, o que Susan B. Anthony e outras sufragistas teriam dito sobre a descoberta de um sociólogo da Universidade de Columbia, quase 50 anos depois do reconhecimento do direito de voto das mulheres nos Estados Unidos, de que apenas uma esposa em cada 22 afirmava nem sempre votar nos mesmos candidatos preferidos do marido. Um leitor enviou a seguinte observação: “Acho que ficariam muito felizes. O movimento feminista decerto percorreu longo caminho, se, menos de 50 anos depois, apenas um marido em cada 22 tem a coragem de votar contra a mulher.”<sup>15</sup>

Em suma, sempre se deve questionar se os números apresentados foram

interpretados de maneira adequada, em relação ao problema em foco.

## Desconfie principalmente dos argumentos de causalção

Uma das coisas mais importantes sobre as quais ser cético no raciocínio analítico envolve a dificuldade de identificar relações de causa e efeito. Como já mencionamos ao descrever experimentos de cientista maluco, quando se criam grupos de teste e de controle e se distribuem os participantes entre eles ao acaso, tende-se a atribuir eventuais diferenças nos resultados entre os dois grupos aos fatores que estão sendo testados. A simples relação estatística entre dois fatores, porém, dificilmente configura causa e efeito. Você já deve ter ouvido a frase “correlação não é causalção”, e é importante lembrar-se dela.

Christopher Chabris e Daniel Simons, psicólogos cognitivos, no livro *The Invisible Gorilla and Other Ways Our Intuitions Deceive Us*, sugerem método útil para verificar a causalidade: “Ao ouvir ou ler sobre associação entre dois fatores, reflita sobre a possibilidade de a distribuição entre os dois grupos representativos ter sido realmente aleatória. Caso tenha sido impossível, dispendioso demais ou eticamente duvidoso distribuir aleatoriamente os participantes entre os dois grupos, o estudo não preencheu os requisitos de experimento confiável e a inferência causal é injustificável.”<sup>16</sup>

Se, por exemplo, você ler que, “segundo estudo de 10 anos, a ingestão intensa de bebida alcoólica provoca câncer”, indague se os sujeitos do teste realmente foram distribuídos aleatoriamente entre os grupos e se, de fato, beberam muito ou beberam pouco durante todo o período. Isso parece improvável. O mais provável é que o pesquisador tenha identificado correlação entre ingestão intensa de álcool (provavelmente autorrelatada) e câncer numa população que foi monitorada durante 10 anos. O pesquisador deveria ter advertido que a correlação encontrada poderia ser explicada por outras variáveis (por exemplo, os bebedores contumazes talvez também fossem fumantes inveterados), mas omitiu-se a esse respeito.

Se você reacear que alguém em sua organização incorreu na falácia *cum hoc ergo propter hoc* (A se correlaciona com B, portanto A causa B), a maioria das soluções envolve analisar com detalhes o projeto do experimento ou conhecer em profundidade estatística e econometria. Portanto, talvez seja esse o momento de consultar um especialista.

## Faça perguntas

O objetivo de fazer perguntas é compreender com mais clareza o problema e o processo. Do mesmo modo, quando se têm dúvidas sobre os números apresentados, convém perguntar sem hesitação. Muita gente reluta em fazer perguntas sobre números, pois receiam parecer tolas; medo quase sempre indevido ou exagerado. Para algumas ideias sobre as perguntas mais apropriadas, ver “Boas perguntas sobre análise quantitativa”.

Imagine que alguém pergunte sobre os números apresentados – essa pessoa provavelmente parecerá corajosa e respeitável, em vez de burra. Portanto, ao deparar com números incompreensíveis, sinta-se à vontade para questionar. Além disso, no caso de alguns números, convém fazer perguntas complementares. Por exemplo, ao lhe apresentarem médias, pergunte sobre a dispersão da distribuição, ou desvio-padrão. Também cabe indagar sobre outliers (valores atípicos) nos dados, em consequência de erros, discrepâncias ou faltas, que poderiam ser significativos. Também cabe pedir também a mediana, caso forneçam apenas a média. Considerar apenas a média, sem conhecer a dispersão, pode ser muito ilusório, uma vez que as observações individuais podem ser muito diferentes entre si. Ademais, quando alguém se refere a números de certa pesquisa, é preciso perguntar quem fez a pesquisa, como foi planejada e como se formularam as perguntas. Resultados de pesquisas sem esses esclarecimentos prévios podem ser enganosos. Em suma, é bom desenvolver o hábito de questionar e sondar, fatores importantes para aprimorar habilidades quantitativas.

## Práticas de análises quantitativas

Como o jovem que perguntou a alguém como chegar ao Carnegie Hall, em Nova York, quem quiser desenvolver a arte, a capacidade e a disciplina da análise quantitativa deve “praticar, praticar e praticar”. De início, talvez seja difícil avançar na solução do problema. No entanto, essa é uma fase natural do aprendizado, e apenas enfrentando-a é possível melhorar a capacidade analítica – aos poucos, mas com constância. Para tanto, pode-se recorrer ao mesmo

processo de seis passos já descrito para a solução de qualquer problema específico:

- *Reconhecimento do problema:* Você provavelmente tem muitos itens na agenda. Priorize-os de acordo com a urgência e a aplicabilidade da análise quantitativa. O mais importante no estágio de reconhecimento do problema é compreender totalmente o problema e avaliar sua importância. A definição dessas duas questões não só esclarece o que se pretende com a solução do problema, mas também facilita a execução dos estágios subsequentes.

### Boas perguntas sobre análise quantitativa

A lista a seguir certamente não inclui todas as perguntas possíveis a serem feitas, mas é pelo menos o ponto de partida. As sugestões aqui apresentadas são eficazes em quase todas as análises quantitativas, ou nas análises não quantitativas que deveriam ser qualitativas.

1. Quais são os dados que sustentam a hipótese?
2. Quais são as fontes dos dados da análise?
3. A amostra que forneceu os dados é representativa da população?
4. Há valores atípicos na distribuição? Como eles afetam os resultados?
5. Quais são as premissas da análise?
6. Que condições, se houver, invalidariam as premissas e o modelo?
7. Por que você adotou esse método de análise específico?
8. Como você ajustou os dados ao modelo?
9. Você considerou algum outro método para analisar os dados e, em caso positivo, por que o rejeitou?
10. Qual é a probabilidade de que as variáveis independentes realmente estejam afetando a variável dependente? Haveria outras análises a serem feitas para dar melhor ideia da causalidade?

- *Revisão das descobertas anteriores:* Depois que se reconhece o problema, é preciso investigar todas as descobertas anteriores correlatas. Embora seja necessário gastar muito tempo nesse estágio, explorar ao máximo motores

de busca como o Google pode ser muito útil. A pesquisa de informações pertinentes é muito importante para chegar ao âmago do problema, assim como a identificação das variáveis relevantes e a captura de associação genérica entre as variáveis identificadas. Depois de identificar com abrangência as descobertas anteriores, descortinar-se-á panorama mais nítido de como resolver o problema. Podemos resumir a situação nos seguintes termos: “A questão está bem definida, todas as descobertas anteriores foram revisadas; mais da metade já foi feito.”

- *Modelagem (seleção de variáveis)*: Se você identificou todas as variáveis relevantes ao revisar as descobertas anteriores, comece a descartar, uma a uma, as que não são diretamente relevantes para a solução do problema. Quais dispensar e quais preservar dependerá principalmente do propósito do modelo. Se for fazer um trem de brinquedo, as proporções do trem de verdade são importantes. Em contraste, se a intenção for calcular a viabilidade econômica do trem, variáveis como capacidade de tração, velocidade e consumo de combustível irão se tornar muito relevantes. Ao concentrar-se em certos aspectos do problema, como numa caricatura, você se aproxima da solução e melhora suas perspectivas.
- *Coleta de dados (mensuração)*: Para medir as variáveis selecionadas, é preciso primeiro verificar se os dados necessários já foram coletados por outrem na organização. Em muitos casos, as variáveis já foram medidas e acumuladas em outro departamento ou até por você mesmo. É até possível que os dados já sejam de domínio público. Mesmo que seja necessário comprá-los, eles decerto serão muito mais baratos do que se coletados e reunidos a partir do zero. Quando você não dispuser de dados diretamente relevantes para o problema, escolha cuidadosamente o método a ser usado para coletá-los. Caso resolva fazer uma pesquisa, a composição do formulário e até a redação do questionário devem ser estudadas com todo o cuidado. Se você optar por um experimento para reunir dados, consulte um especialista em projeto de experimentos. Como a obtenção de dados exatos e oportunos é muito importante para a solução do problema, é preciso investir tempo e esforço nesse estágio.
- *Análise dos dados*: A análise dos dados consiste em identificar padrões

consistentes, ou relações entre variáveis. O método estatístico a ser usado é escolhido no passo de reconhecimento do problema. Quando se compreende em profundidade a natureza do problema, o método a ser adotado no estágio seguinte de análise dos dados torna-se notório. Se o problema for de comparação entre grupos, é preciso usar técnica específica. Se for de dependência entre variáveis, recomenda-se regressão ou método semelhante. Por se tratar de técnicas populares, usadas em muitas situações, é preciso dispor de tempo suficiente para familiarizar-se com os respectivos conceitos e aplicações. E sem dúvida haverá especialistas em sua organização capazes de orientá-lo.

- *Apresentação de resultados e adoção de providências:* Não negligencie a execução desse componente extremamente importante do raciocínio quantitativo. Grupos de analítica bem-sucedidos dedicam tanto tempo a esse passo quanto aos cinco anteriores. Converse com outras pessoas na organização sobre como elas comunicam os resultados das análises e discuta com elas algumas de suas ideias a esse respeito. E aprenda a linguagem das apresentações gráficas de estatísticas – recorra aos livros de Edward Tufte como se fossem amigos, por exemplo. E se Tufte com um de seus seminários de um dia “Presenting Data and Information” (Apresentação de dados e de informações), reserve sua vaga com antecedência.

## Hábitos quantitativos ↔ Métodos/Conhecimentos quantitativos

Nessa fase, é preciso estudar os fundamentos para compreender e executar análises quantitativas. Os cursos a fazer primeiro são Estatística Elementar e Métodos de Pesquisa. O preferível é encontrar esses cursos on-line. Também é possível aprender em sucessivos estágios, pela internet ou livros-textos, ao deparar com conceitos ou enfrentar dificuldades. Muita gente não consegue entender alguns conceitos de estatística elementar porque muitos livros-textos não explicam como eles se aplicam a problemas do mundo real. Portanto, é preciso selecionar com cuidado os textos que o introduzem a aplicações práticas de estatística, em vez de se limitar à teoria. Sob esse enfoque, recomendamos, em consequência, o livro *Statistics for Business and Economics*, de Heinz Kohler,



como texto principal.<sup>17</sup> Eis o que um resenhista comentou na [Amazon.com](https://www.amazon.com):

O melhor livro de estatística de todos os tempos.

Eu estava com muito medo daquele curso de estatística... Tinha a impressão de que seria muito difícil aprender conceitos e inferências em estatística... É muito fácil seguir os layouts e as estruturas dos capítulos... Diferentemente do esperado, estatística nunca foi problema para mim! Foi uma grata surpresa ver como Kohler tornou a estatística tão interessante e tão fácil.<sup>18</sup>

O mesmo se pode dizer de métodos de pesquisa; também há muitos livros-textos sobre o assunto. Vá on-line e leia algumas das resenhas antes de escolher o texto. O mais importante é fazer esses cursos básicos e tentar resolver todos os problemas e responder a todos os questionários dos livros. Como a maioria desses exercícios se destina a ligar teoria e prática, você apreende com mais profundidade os conceitos em exame. Se você ler integralmente os textos selecionados pelo menos três vezes, e se dedicar em seguida a toda a parte prática, você decerto dominará todos os conhecimentos básicos necessários para ser analista quantitativo proficiente. Agora, o que resta fazer é praticar, praticar, praticar.

## O analista quantitativo

Em negócios, as atividades que envolvem análise quantitativa ocorrem como parte de processos e de contextos sociais. Tipicamente, incluem não só fazer a análise, mas também escrever um relatório (ou fazer uma apresentação, trabalhando com outros analistas, aprendendo uns com os outros em atividades como seminários, e trabalhando com especialistas). Essas atividades são fundamentais para desenvolver habilidades analíticas.

## Escreva um relatório

O primeiro passo para adquirir prática é escrever um relatório, tarefa muito importante para familiarizar-se com a prática de “contar histórias com dados”.

Você provavelmente aprenderá muito com a leitura de textos, mas sem dúvida aprenderá muito mais quando se empenhar, escrupulosamente, em resolver um problema da vida real e, depois, em escrever um relatório expositivo e propositivo. Seria bom escolher um problema que fosse instigante para a maioria das pessoas na organização, mas não se preocupe se não for. Atente tanto quanto possível em selecionar o problema certo, pois, como disse Voltaire, as pessoas o julgam mais pelas perguntas que pelas respostas. Depois, zelosamente, siga o referencial de seis passos da análise quantitativa e use o relatório formal para apresentar suas descobertas e sugestões. Três são os propósitos de escrever o relatório:

- Aprender mais e melhorar a capacidade analítica, solucionando um problema real por conta própria.
- Contribuir com seriedade para resolver um problema no trabalho.
- Ajudar a criar uma atmosfera analítica no trabalho e motivar os outros a pensar e a agir analiticamente.

Não espere muito de sua primeira tentativa, mas empenhe-se para que o relatório demonstre que você fez a revisão crítica de todas as descobertas anteriores, esforçou-se mais que qualquer outra pessoa para resolver o problema e que produziu um trabalho analítico rigoroso. Lembre-se de que você também será julgado não só pelas conclusões, mas também pela maneira como resolver a questão.

## **Forme uma comunidade de análise quantitativa**

Arregimente amigos e colegas para formar um comitê de análise quantitativa. Muitas são as boas razões para fazê-lo: estudar juntos com eficácia, trabalhar em projetos com eficiência e desempenhar papel-chave na promoção de abordagens analíticas no trabalho. O grupo deve definir prioridades para enfrentar os desafios no trabalho, discutir os temas, dividir as tarefas de acordo com o referencial de seis passos, explicar os próprios métodos de trabalho e revezar em apresentações uns aos outros. Esse tipo de trabalho em equipe é a maneira mais

garantida e rápida de aprender. Como disse Henry Ford, “juntar-se é o começo; manter-se juntos é progresso; trabalhar juntos é sucesso”. Além disso, esse pequeno grupo com orientação analítica pode mudar a empresa, sobretudo promovendo a cultura analítica e motivando os outros a pensar e a agir de maneira analítica.

## **Realize seminários com regularidade**

Os objetivos dos seminários são óbvios: ter a oportunidade de apresentar os resultados dos projetos dos membros do comitê e discutir com as pessoas certas, para acelerar o desenvolvimento e a execução dos projetos e para reforçar o clima organizacional propenso ao raciocínio analítico e a decisões fáticas. Muitas empresas também recorrem a palestrantes externos para debater questões analíticas. Seminários regulares e específicos proporcionam experiências esclarecedoras e inspiradoras, motivando os participantes a se superarem continuamente, inclusive para melhor se apresentarem em eventos futuros. Possibilite o acesso de pessoal não técnico a alguns deles.

## **Analista quantitativo ↔ Métodos / Conhecimentos quantitativos**

Nesse estágio, o principal foco do estudo em grupo é aprender, examinando como se executam análises quantitativas para solucionar problemas da vida real. Você pode começar com os casos apresentados neste livro. Preparem os papéis ou materiais originais; revezem-se fazendo apresentações uns para os outros; discutam os detalhes sob o referencial dos seis passos. Ao depararem com métodos quantitativos avançados, aproveitem a oportunidade para aprendê-los juntos, aplicando-os em situações reais. Além disso, recursos muito úteis são convidar especialistas para explicar o método em detalhes e para discutir as vantagens e as desvantagens da técnica. Talvez seja possível encontrar esse especialista na própria empresa ou ao menos na universidade local.

Seguindo todos ou muitos dos passos descritos neste capítulo, você fará grandes avanços no esforço para tornar-se pelo menos analista quantitativo semiprofissional. Sem dúvida, também melhorará suas perspectivas de carreira e ampliará sua capacidade mental. E ainda será participante ativo de importante

transformação que, neste exato momento, está sacudindo o mundo empresarial e organizacional, em geral.

## **Exemplo de raciocínio analítico: Facilitar a obtenção de bolsas de estudos**

Na academia, a competição pela estabilidade como docente pressiona cada vez mais os acadêmicos a produzir com frequência novos trabalhos para publicação. A frase “publicar ou perecer” descreve muito bem, sobretudo em universidades prestigiosas, voltadas para pesquisas, a premência de sempre publicar trabalhos para impulsionar ou sustentar uma carreira no magistério superior. Uma vez que preparar trabalhos para publicação, porém, sobretudo nos periódicos mais importantes, não é tarefa fácil e exige muito tempo, a coautoria com outros acadêmicos se tornou a maneira mais produtiva e eficaz de sobreviver e avançar. A colaboração entre professores e alunos de pós-graduação também é comum. A chave, portanto, para receber bolsas de estudos em programas de mestrado, doutorado e outros é desenvolver capacidade de estudo e pesquisa para ajudar os professores em suas atividades acadêmicas.

Jinho vive na Coreia do Sul e tem duas filhas estudando nos Estados Unidos. Sozinho, ele não teria recursos para isso, mas fez análises sistemáticas sobre como conseguir bolsas de estudos para pós-graduação. Agora, as duas filhas são bolsistas de programas de doutorado nas universidades de Stanford e Michigan, respectivamente. De início, Jinho hesitava em discutir em público a história da família dele, mas resolveu ir adiante e mostrar como estudantes com formação quantitativa poderiam conseguir bolsas de estudos e como o raciocínio intencional e sistemático sobre as próprias habilidades quantitativas produz resultados.

Quando estava terminando a faculdade, a filha mais velha de Jinho, Nuri, queria concluir os estudos nos Estados Unidos, mas sabia que, para tanto, precisaria de bolsa de estudos. Embora isso fosse quase impossível para estrangeiro com especialização em comunicações, Jinho encorajou-a, insistindo em que a filha talvez conseguisse realizar seu objetivo caso se preparasse de maneira sistemática para candidatar-se a bolsas de estudos.

Nessas condições, que tipo de preparação seria necessária para convencer os membros dos comitês de admissão de programas de pós-graduação que Nuri

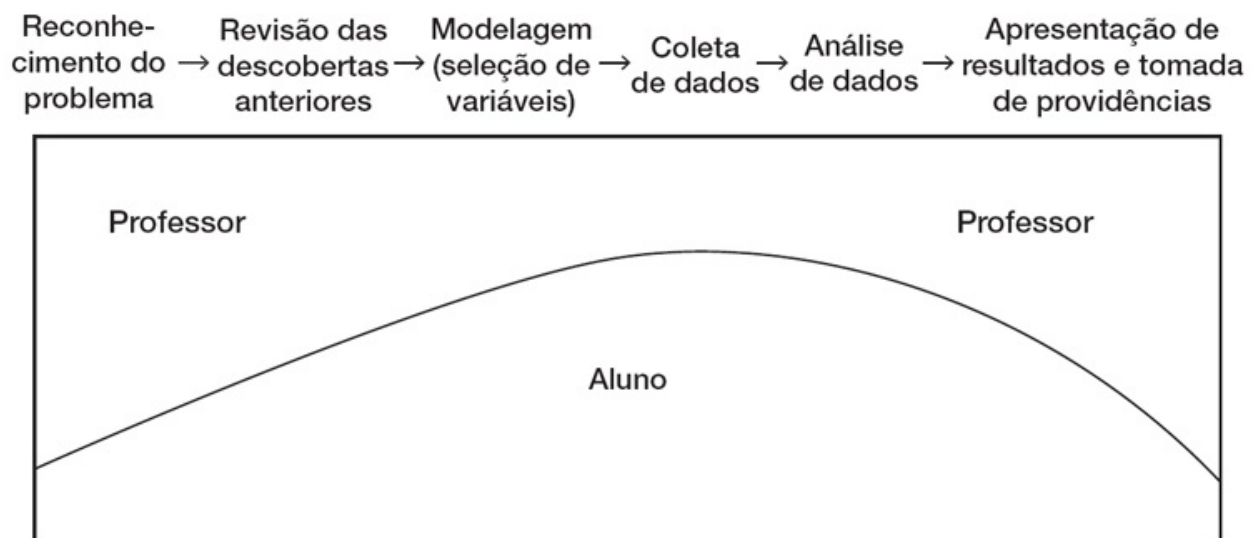
tinha condições de ajudá-los em suas pesquisas e que, portanto, valeria a bolsa de estudos? Apliquemos os seis passos da análise quantitativa e examinemos os papéis dos professores, como protagonistas, e dos alunos de pós-graduação, como coadjuvantes, nas atividades acadêmicas. (ver [Figura 6.2](#)).

Atividade típica no estágio de reconhecimento do problema é identificar um tema de pesquisa que tenha grandes chances de ser aceito em um periódico acadêmico. Os professores em geral são protagonistas nesse estágio. Já o propósito do estágio de revisão das descobertas anteriores é demonstrar como o novo tema de pesquisa difere dos estudos já concluídos ou em curso e como pode contribuir de maneira singular para os conhecimentos existentes. Também nesse estágio os professores desempenham o papel principal.

**FIGURA 6.2**

---

**Teria Nuri os requisitos para conseguir bolsa de estudos? Os seis passos da análise quantitativa.**



---

Normalmente, depois que os professores definem os dados a serem coletados, os alunos de pós-graduação, em geral, ajudam os professores não só na coleta de dados, mas também na análise dos dados. Essas tarefas quase sempre são delegadas a alunos de pós-graduação, que, quase sempre, têm tempo e disposição para o trabalho. A maneira mais segura de conseguir bolsa de estudos, portanto, é, primeiro, ter capacidade para compreender as razões da escolha de

determinado tema e para escolher os métodos de pesquisa adequados, e, segundo, ter qualificações para coletar e analisar dados sob a supervisão de um professor. Finalmente, os professores normalmente escrevem o artigo a ser submetido aos periódicos, depois que os alunos de pós-graduação coletam e manipulam os dados.

A fim de preparar-se para o processo de seleção de bolsistas, Nuri fez cursos de estatística introdutória e de métodos de pesquisa. Para aprofundar ainda mais o conhecimento dos conceitos, Jinho e Nuri se reuniam depois de cada capítulo dos livros-textos, para sessões de discussão e de perguntas e respostas. Ela também participou de programas intensivos sobre o uso de SPSS (pacote de estatística) para familiarizar-se com a análise de dados. Finalmente, Jinho e Nuri leram juntos 20 artigos recentes, publicados em alguns dos periódicos especializados em telecomunicações, e discutiram as ideias dos trabalhos, métodos de pesquisa, coleta de dados, análise de dados e interpretação de resultados. Esse processo foi necessário para Nuri compreender com clareza como os conceitos que aprendeu nos três cursos preparatórios se aplicavam em contextos de pesquisa reais. Todos os preparativos foram concluídos durante o último ano de Nuri na faculdade. No fim do ano, ela já acumulara tal capacidade de pesquisa que estava qualificada e preparada para apoiar os trabalhos de pesquisa de qualquer professor.

Nuri detalhou seus preparativos e capacidades na seção de “Descrição de Propósitos” dos pedidos de inscrição em programas de mestrado e, por fim, ganhou uma bolsa de estudos da Universidade de Wisconsin. Lá, graças ao seu empenho, Nuri trabalhou com muitos professores para a publicação de vários trabalhos. Depois, recebeu financiamento (*fellowship*) da Universidade de Wisconsin, e hoje está ocupada na elaboração de sua tese de doutorado. Youngri, filha mais moça de Jinho, depois de seguir os mesmos passos da irmã, recebeu financiamento da Universidade de Michigan. Está trabalhando no laboratório de engenharia química.

### **Exemplo de raciocínio analítico: Daryl Morey e Shane Battier, do Houston Rockets**

Este exemplo é sobre duas pessoas com orientação quantitativa, na área de basquete profissional. Daryl Morey é gerente geral do Houston Rockets, time de

basquete da National Basketball Association (NBA). Para quem conhece o livro e o filme *Moneyball: O homem que mudou o jogo*, Morey tem boas razões para reivindicar o título de “Billy Beane do basquete profissional”.<sup>19</sup> Especializou-se em Ciências da Computação e em Estatística na faculdade, em Northwestern, e tem MBA pela Sloan School of Management, do MIT. Sempre quis aplicar as ideias de Bill James sobre estatísticas de basquete a outros esportes profissionais, e tornou-se vice-presidente de Operações e Informação do Boston Celtics. Aos 35 anos, tornou-se gerente geral do Rockets e passou a adotar vários métodos estatísticos e quantitativos para melhorar o desempenho do time. Também preside, anualmente, a Sports Analytics Conference, do MIT, que hoje atrai mais de dois mil participantes.

Shane Battier é jogador da NBA – ala (*forward*) – que hoje joga no Miami Heat. Jogou no Houston Rockets de 2006 a 2011. É relativamente analítico, como acontece com os jogadores de basquete profissionais, e foi considerado o sétimo jogador mais inteligente nos esportes profissionais pela revista *Sporting News*.<sup>20</sup> Daryl Morey observa (em artigo de Michael Lewis, autor de *Moneyball*) que Battier

...recebeu seu pacote especial de informações. “Ele foi único jogador que o recebeu”, diz Morey. “Demos-lhe uma mangueira de bombeiros jorrando dados e deixamos que ele bebesse nela. A maioria dos jogadores é como golfistas: não devem pensar para jogar.” Os dados basicamente dividiram a quadra em muitas áreas distintas e previram as chances de Bryant fazer arremessos de diferentes pontos, sob diversos graus de pressão ofensiva, em diferentes relações com outros jogadores, e o desempenho provável dele nessas várias condições. Battier aprende muito com o estudo dos dados sobre os superastros que ele, em geral, é incumbido de marcar.<sup>21</sup>

Battier, contudo, foi jogador controverso dos Rockets, conforme revela o referencial de seis passos.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Deveriam o Rockets adquirir Battier como jogador, embora as estatísticas sobre seu desempenho individual não fossem assim tão impressionantes?

**REVISÃO DAS DESCOBERTAS ANTERIORES.** A analítica dos esportes é cada vez

mais comum, e muitos são os livros (e, ainda mais, os sites) sobre o tema hoje. Essa análise, contudo, é muito mais fácil em esportes nos quais as estatísticas sobre o desempenho individual dos jogadores são altamente relevantes para o sucesso do time. Em basquete, a dinâmica do desempenho de toda a equipe e a interação dos jogadores são mais difíceis de medir. Os números referentes ao desempenho individual de Battier eram relativamente fracos (em cinco anos jogando para o Memphis Grizzlies, ele alcançou a média de 10 pontos por jogo, e pouco menos de cinco rebotes por jogo). Mas o time e os colegas pareciam jogar melhor – às vezes muito melhor – quando ele estava no jogo.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** As variáveis da decisão de adquirir Battier dos Grizzlies seria o custo de aquisição em si (na compra direta ou na troca por outros jogadores), a quantia que ele receberia, várias medidas de desempenho individual e, idealmente, algumas medidas do desempenho do time, com e sem a participação de Battier.

**COLETA DE DADOS (MENSURAÇÃO).** Os indicadores de desempenho e os números financeiros eram fáceis de conseguir. E havia como avaliar o impacto do jogador sobre o desempenho da equipe. A estatística “plus/minus”, adaptada por Roland Beech de [82games.com](http://82games.com), partindo de estatística semelhante usada em hóquei, compara o desempenho do time em duas situações: com o jogador na equipe durante o jogo e com o jogador no banco de reservas.

**ANÁLISE DOS DADOS.** Morey e os estatísticos recorreram à análise plus/minus para avaliar Battier. Várias são as versões da estatística plus/minus que podem ser ajustadas (usando análise de regressão) para diversos fatores, inclusive a qualidade dos colegas e dos adversários do jogador. Não sabemos exatamente como Morey calculou as estatísticas para adquirir Battier. Mas, no campeonato de 2006-2007, o primeiro dele com o Rockets, Morey disse que o plus/minus ajustado de Battier era de 8, significando que, se ele fosse substituído por um jogador médio da NBA, o diferencial de pontos do Rockets em relação a outros times diminuiria em oito pontos.<sup>22</sup> Com base nesses números, Battier situava-se em 17º lugar na NBA; o salário dele, no entanto, coincidia com o nível médio da liga, e o jogador pelo qual foi trocado classificava-se em 45º em plus/minus.

**APRESENTAÇÃO DE RESULTADOS E AÇÃO.** Morey, evidentemente, consumou a



aquisição de Battier, que, na opinião dele, jogou bem no Rockets. No campeonato antes da chegada de Battier, o Rockets terminou 34-48. No campeonato de 2007-2008, já com Battier, o time terminou 52-30, e, em 2008-2009, atingiu a marca de 55-27 – sucessão de 22 vitórias consecutivas, incluindo muitos jogos em que os jogadores mais famosos do Rockets estavam contundidos. Battier jogou durante cinco anos para o Rockets, até machucar-se e ser trocado, voltando para o Grizzlies. Depois, juntou-se ao Heat. Battier ainda atuou bem no Miami, levando o time a conquistar o campeonato da NBA, em 2012.

## Trabalhando com os quants

Como este livro é para não quants, achamos que seria útil descrever a melhor maneira de relacionar-se com profissionais de analítica e com cientistas de dados. Embora já se aprenda muito lendo este livro e executando algumas das atividades sugeridas no capítulo anterior, isso não será suficiente para realizar sozinho as análises sofisticadas típicas dos trabalhos de analítica. Vez por outra, será necessário trabalhar com especialistas em áreas quantitativas. Os analistas quantitativos e os cientistas de dados geralmente têm PhD ou mestrado em disciplinas como Estatística, Matemática e mesmo Física. Só isso já sugere o nível de habilidades quantitativas necessárias para a execução de trabalhos sérios em analítica.

Basicamente, o que descrevemos neste capítulo é um conjunto de relacionamentos entre três conjuntos de atores:

- Decididores de empresas e outras organizações
- Profissionais de negócios ou membros de equipes organizacionais
- Analistas quantitativos ou cientistas de dados

Nossa premissa neste livro é que você se inclui numa das duas primeiras categorias e precisa trabalhar em estreito entrosamento com profissionais da terceira categoria. Se você for membro do terceiro grupo, é possível que ainda considere útil este capítulo, pois ele lhe oferecerá algumas diretrizes sobre como trabalhar de maneira eficaz com os não quants.

Há boas razões para a interação desses três grupos, em vez de um deles predominar sobre os demais. Caso você ainda não tenha percebido, defendemos

com veemência que o processo decisório se fundamente em dados e em analítica. Intuição e experiência, entretanto, para muitos executivos, ainda são recursos importantes na tomada de decisões. É até possível que esses dois atributos acarretem vieses nas decisões, mas, inegavelmente, ambos são importantes para definir critérios financeiros adequados, para propor cenários alternativos cruciais do tipo “what if”, e para estabelecer as condições em que os modelos analíticos são relevantes.

O objetivo, portanto, é tomar decisões analíticas e, ao mesmo tempo, preservar o papel do instinto dos executivos e de outros profissionais. Poucos executivos são aptos ao mesmo tempo em analítica e em intuição. Isso significa que precisam trabalhar em estreito entrosamento com os analistas quantitativos, para decidir com eficácia. Com efeito, diríamos que a qualidade do relacionamento entre executivos e assessores quantitativos é fator crítico de decisões eficazes.

Como diz Karl Kempf, *fellow* da Intel (*Intel Fellow*, um dos títulos com que se distinguem cientistas detentores de grande autonomia na organização), que chefia um grupo de *decision engineering* (referencial que integra melhores práticas de processo decisório em organizações), “não se trata de matemática; trata-se de relacionamento”.<sup>1</sup> É uma afirmação notável de Kempf, conhecido informalmente na Intel como “UberQuant” e “Chief Mathematician”. Se alguém conhecido como matemático-chefe diz que não se trata de matemática, devemos prestar atenção.

Kempf observou que os algoritmos matemáticos e estatísticos aplicados pelos analistas quantitativos na execução de suas atividades podem ser de simplicidade surpreendente ou de complexidade espantosa. Em todos os casos, porém, os algoritmos resultaram do trabalho de pessoas muito inteligentes, ao longo de décadas (ou séculos, como já salientamos em alguns exemplos), que foram verificados, dissecados e comprovados sucessivas vezes, por outras pessoas muito inteligentes.

A matemática funciona, mas o lado humano dessas decisões é, em geral, muito menos refinado. As organizações que mantêm grupos de análise quantitativa – o que todas deveriam fazer nessa era da analítica – precisam examinar com cuidado os tipos de analistas que contratam e as funções que eles desempenham na interação com executivos. Os analistas com a combinação certa de habilidades tendem a ser escassos. Além disso, os executivos precisam

mudar as expectativas e as interações no relacionamento com os analistas. Finalmente, é preciso ampliar e refinar os métodos e ferramentas dos grupos quantitativos para que se integrem plenamente na maneira como as pessoas e as organizações tomam decisões.

## **Integração negócios-analítica na engenharia de decisões da Intel**

Karl Kempf e a equipe de analítica da Intel observaram em primeira mão muitas das recomendações que apresentamos nesta seção. O grupo se concentra intensamente nas relações entre analistas e decidores. Esforça-se para descobrir maneiras de promover o respeito mútuo – de convencer os decidores a demonstrar certo interesse e algum respeito pelas atividades dos analistas quantitativos e de persuadir os matemáticos a demonstrar grande interesse e enorme respeito pelos insights dos decidores. A assimetria no relacionamento é intencional. Embora seja bom que os decidores desenvolvam sensibilidade pela matemática, Kempf e equipe consideram absolutamente vital para os matemáticos compreender em profundidade a intuição dos executivos e dos profissionais de negócios. Isso não significa que os decidores devam tornar-se especialistas em matemática, mas talvez sugira que os matemáticos precisam pensar como especialistas em negócios.

Embora seja possível que os matemáticos jamais compreendam em plenitude a natureza da intuição em negócios, é preciso, por definição, que reconheçam o papel da intuição e falem a linguagem dos executivos. O método da Intel consiste em enviar seus “matemáticos” para as áreas de negócios – ao menos para ouvir, para aprender e, depois de algum tempo, para perguntar. Na melhor das hipóteses, o analista pode ser treinado, como alguém recém-admitido, para participar dos processos de negócios. Em ambos os casos, a missão é compreender a organização formal e informal, a maneira como o grupo é incentivado e recompensado, e assim por diante.

Para Kempf, já se alcança algum sucesso quando o pessoal de matemática compreende os problemas de negócios; mas só se conquista grande sucesso quando o pessoal de negócios acredita que o pessoal de matemática compreende os problemas da empresa. Daí resulta, geralmente, certo respeito do pessoal de negócios pelo pessoal de matemática (“alguém que de fato se aproxima e se

esforça para compreender nossos problemas mostra que realmente está interessado em ajudar”) e, com muita frequência, certo respeito do pessoal de matemática pelo pessoal de negócios (“não é tão fácil quanto pensei – esse cara é realmente muito inteligente”).

Admitindo que o maior defensor do uso da analítica criou condições para que o pessoal de matemática, durante algum tempo, observasse o grupo de negócios ou até se aliasse a ele, efeito colateral útil aqui seria identificar e engajar os principais negativistas do grupo. Na pior das hipóteses, talvez estejam certos, e a pretensão é impossível; na melhor das hipóteses, passamos a conhecer os indivíduos mais propensos a criticar os resultados ou as soluções. Anne Robinson, que dirigiu o grupo de analítica da Cisco Systems e agora exerce a mesma função na Verizon Wireless, também enfatiza a importância de incluir os céticos na lista de stakeholders. “Eles preservam a honestidade e mantêm o alto desempenho da equipe”, diz. “Se você conseguir convencê-los, não será difícil persuadir qualquer outra pessoa.” (No fim deste capítulo, apresentamos um exemplo do trabalho de Anne, na Cisco.)

Na Intel, o próximo passo no relacionamento será a colaboração entre decididores e quants na construção do modelo básico. O principal quant promove brainstorms para extrair inputs (elementos de dados, fontes de dados, ideias para detectar e eliminar ou consertar dados ruins), outputs (que soluções são mais desejáveis, que métodos de apresentação intuitivamente parecem mais satisfatórios para os usuários de negócios almejados), quais são as variáveis-chave e quais são as principais relações entre as variáveis.

Mais uma vez, nesses exercícios, os profissionais de negócios não precisam compreender conceitos matemáticos complexos; mas, no mínimo, precisam apurar a sensibilidade quanto a aspectos básicos como os seguintes:

- Sendo A e X correlatos, se A sobe, em que direção X se movimenta?
- Quais são os valores mais altos e mais baixos a serem assumidos pela variável B?
- Se houver intervalo ou defasagem temporal entre a causa Y e o efeito Q, qual será sua duração provável?

Como com qualquer outro tipo de modelo, alguns exemplos concretos (reais

ou fictícios) são extremamente úteis para avançar na estruturação do modelo básico. Nesse exercício, o quant deve ouvir com atenção, fazer perguntas esclarecedoras e absorver o máximo possível dos conhecimentos do decisor sobre o negócio. Isso tem a ver tanto com construção de relacionamentos quanto com desenvolvimento do modelo.

A essa altura, a equipe de quants deve estar preparada para entrar em ação em alta velocidade. Para isso, precisa selecionar os métodos matemáticos certos, formalizar o modelo para representá-lo no computador, coletar dados e introduzi-los no computador. Em seguida, os analistas podem testar o modelo, executando análises de sensibilidade das variáveis e das relações, e experimentar alternativas. Quando o decisor tiver condições de fornecer alguns problemas para testes, os analistas podem começar a detectar vieses nas percepções dos decisores e deles próprios, ajustando o modelo em conformidade. O aspecto mais importante desse estágio é rodar o protótipo tão cedo quanto possível e apresentá-lo aos usuários para receber feedback. Em geral, é boa ideia repetir o processo com diferentes participantes da área de negócios, para testar até que ponto o modelo está completo e para promover o consenso na equipe de negócios.

Refinam-se, então, o modelo e o sistema, com base no feedback, para nova demonstração. Em outras palavras, é importante cometer muitos erros no começo. Em cada rodada, haverá aspectos que os quants esqueceram, não compreenderam ou simplesmente expressaram de maneira errada; e que os decisores não mencionaram; ou mencionaram, mas voltaram atrás, e assim por diante. Não importa que o projeto seja uma análise única ou um processo contínuo, o sucesso na solução do problema geralmente promove o sucesso no relacionamento. O decisor provavelmente precisará de alguma fé no começo; se as coisas derem certo, porém, a fé inicial logo é substituída pela credibilidade baseada na experiência. O desenvolvimento da confiança, do respeito e da compreensão entre as partes exige tempo e esforço, mormente pelo analista, pois o pessoal de negócios não tem muito desses recursos para investir. Não raro, porém, ocorre que um primeiro relacionamento bem-sucedido desencadeia uma sequência de análises poderosas e de ferramentas úteis, aprofundando a confiança e a compreensão.

## O retorno da combinação de arte e ciência na Intel

A abordagem que acabamos de descrever foi desenvolvida e aprimorada continuamente na Intel ao longo dos últimos vinte e tantos anos, abrangendo todo o espectro da analítica na Intel, onde rende benefícios consideráveis. O trabalho inicial se concentrou na fabricação, envolvendo projeto, construção, ampliação e operação de fábricas. Os problemas analisados e as decisões implementadas incluíram ampla variedade de questões, desde a especificação de máquinas e layouts até a gestão de operações e a manutenção de equipamentos.

Um segundo conjunto importante de atividades analíticas se concentrou na integração de produção, estoques e logística entre a rede de fábricas da Intel. Com instalações de produção nos Estados Unidos, Costa Rica, Irlanda, Israel, China, Malásia e, mais recentemente, Vietnã – todas operando 24/365 – a analítica nas fábricas se estende ao longo não só de fusos horários, mas também de culturas e línguas.

Projetos subsequentes se expandiram corrente acima e corrente abaixo na cadeia de fornecimento. Alguns sobre estruturas de contratos referentes a equipamentos e materiais se empenharam em relacionamentos win-win (ganha-ganha) com os fornecedores e na otimização da agilidade da Intel. Outros sobre previsão de demanda e distribuição de produtos se concentraram nos serviços aos clientes e na contenção dos custos da cadeia de fornecimento. Esses sistemas continuam em uso ou foram aprimorados por ferramentas de análise e por processos decisórios de segunda e terceira geração. Os trabalhos se estenderam mais recentemente para as amplas organizações de desenvolvimento de produtos da Intel. Os analistas de engenharia de decisões estão desenvolvendo relacionamentos com o pessoal de negócios sênior, com intuição e experiência sobre criação de novos produtos. Os projetos de análise mais recentes variam da seleção de produtos à elaboração de cenários para a alocação de recursos de engenharia, de importância crítica para a gestão do portfólio de projetos.

Esses projetos e os respectivos impactos renderam à Intel o prêmio de 2009 do Institute for Operations Research and the Management Sciences, pela “aplicação reiterada dos princípios de pesquisa operacional e das ciências administrativas, de maneira desbravadora, diversificada, inovadora e duradoura”. Ao receber o prêmio, o então presidente do conselho de administração da empresa, Craig Barrett, atribuiu o aumento da competitividade da empresa nas duas décadas anteriores à aplicação constante de analítica avançada, adicionando bilhões (com “b”) aos resultados financeiros.

## Suas atribuições analíticas

Para que a solução de problemas por meio da analítica seja bem-sucedida, os analistas quantitativos e o pessoal de negócios (você) devem assumir algumas atribuições. Passamos boa parte deste livro descrevendo as tarefas dos quants e mostrando como o pessoal de negócios pode compreender melhor os dados e relatórios deles. Agora é hora de tratar das funções analíticas a serem exercidas pelo pessoal de negócios, tenham eles ou não mentalidade matemática (ver “O que os analistas quantitativos devem esperar do pessoal de negócios”).

### O que os analistas quantitativos devem esperar do pessoal de negócios

Como decididor ou como profissional de negócios, você deve:

- Dedicar aos analistas tempo e atenção suficientes para que compreendam o problema sob sua perspectiva.
- Disponibilizar tempo e atenção dos membros de sua equipe para que ajudem os analistas a compreender os detalhes da situação.
- Compreender com clareza as necessidades de tempo e dinheiro para o desenvolvimento da solução, promovendo a concordância mútua quanto à proposta.
- Aprender o suficiente de matemática e de estatística básica para ter uma ideia geral de como o modelo funciona e de quando se torna inválido.
- Pressionar, com habilidade, se não compreender algo, pedindo explicação diferente e mais clara.
- Participar de todos os eventos importantes, como apresentações, demonstrações e lançamentos.
- Deixar claro para os membros da equipe que o uso eficaz do modelo é importante para o seu sucesso e para o deles.

### Aprender alguma coisa de matemática e de estatística

No [Capítulo 6](#), sugerimos várias maneiras de o pessoal de negócios aprender alguma coisa de estatística. Acreditamos que essa atribuição se aplica a todos os níveis da organização, inclusive aos altos executivos. Por quê? Em nossa



sociedade repleta de dados e imbuída de cultura empresarial, é impossível compreender de que maneira os dados e a analítica se aplicam ao processo decisório sem um mínimo de refinamento matemático.

Quem carecer dessa base provavelmente enfrentará problemas, como mostra o exemplo de Joe Cassano, da AIG Financial Products, no [Capítulo 1](#). Muitas empresas, usam, cada vez mais, modelos estatísticos e matemáticos em suas operações. Portanto, princípio de extrema importância é que os gestores não adotem em suas empresas modelos analíticos que não compreendem. Como disse Robert Schiller, economista de Yale, ao explicar algumas das razões da crise financeira de 2008-2009, que previu, “é preciso ter mentalidade quantitativa para gerenciar uma empresa; os detalhes quantitativos realmente importam”.<sup>2</sup>

Algumas organizações enfatizam a importância da familiaridade com matemática e com modelos quantitativos. Ed Clark, CEO do TD Bank Group e PhD em Economia, por Harvard, evitou os problemas que muitos banqueiros americanos enfrentaram durante a crise financeira. E assim se manifestou sobre as dificuldades dos outros bancos no *Toronto Star*: “O que achei assustador foi que, ao conversar com meus colegas da área [de produtos estruturados], percebi aos poucos que eles não compreendiam o real significado daqueles produtos. Eles nunca se deram ao trabalho de analisar a matemática de cada um dos produtos. E isso se explica, em parte, por terem delegado para níveis muito baixos a compreensão dos produtos.”<sup>3</sup>

À medida que todos os setores se tornam mais voltados para dados e para análises, cabe aos altos executivos compreender até certo ponto a complexidade da analítica, pois, do contrário, não serão capazes de questionar quando operadores de mercado sugerirem que assumam riscos desproporcionais e desconhecidos ou quando profissionais de marketing recomendarem modelo preditivo que empregue excesso de dados sobre clientes, para que, do contrário, não exponham a grandes perigos suas instituições e clientes.

Alguns dos conceitos que todos os executivos devem compreender são:

- Medidas de tendência central (média, mediana, moda).
- Probabilidade e distribuições.
- Amostragem.

- Fundamentos de correlação e de regressão.
- Rudimentos de projeto de experimentos (*experimental design* ou *design of experiments*).
- Interpretação de analítica visual.

Os métodos para a aquisição desses conhecimentos podem ser os mesmos usados por pessoal menos graduado, à exceção de que os altos executivos podem ter recursos para a contratação de professores ou de consultores, que darão treinamento a grupos de executivos ou até a executivos isolados.

## Compreender e questionar as premissas

Já recorreremos à famosa citação de George Box de que “todos os modelos são falhos, mas alguns são úteis”. Também dissemos na ocasião que é importante saber quando os modelos já não são úteis. Em geral, isso ocorre quando as premissas do modelo – e todos os modelos têm premissas – deixam de ser corretas ou válidas. O mundo está sempre mudando, e o trabalho do executivo cético é determinar se as mudanças comprometem o modelo. Eis alguns exemplos de premissas em modelos quantitativos que, efetivamente, já foram usadas por organizações:

- A disposição dos clientes para comprar um produto a certo preço (conhecida como modelo de elasticidade preço) não mudou, mesmo que a economia tenha deteriorado.
- As amostras de clientes com que testamos várias versões de páginas da internet, muitos anos atrás, ainda representam as preferências dos clientes de hoje.
- O modelo preditivo sobre a probabilidade de inadimplência de devedores hipotecários que desenvolvemos quando os preços estavam em alta ainda é válido agora que os preços estão em queda (obviamente, esta é um tanto problemática).
- A probabilidade de um furacão atingir o Sul da Flórida não aumentou,

embora pareça que estamos enfrentando algum grau de mudança climática global.

- Números de telefones fixos ainda oferecem amostra representativa para pesquisas de opinião políticas, embora muitas pessoas já não tenham telefones fixos (como já sugerimos, também essa é problemática).

Nem todas essas premissas eram inválidas. Na verdade, como quase todos os modelos se baseiam em dados do passado (lembre-se de que é difícil conseguir dados sobre o futuro), eles partem da premissa de que o futuro será como o passado na maioria dos aspectos. E esses modelos, com muita frequência, permanecem válidos por longos períodos. Conforme observou Charles Duhigg, em livro recente, *The Power of Habit*, o comportamento humano, uma vez estabelecido, pode persistir durante muito tempo.<sup>4</sup> Isso nos permite, como que por mágica, prever o futuro com base no passado.

Algumas organizações recorrem a talentos de alto preço apenas para fazer perguntas instigantes sobre as premissas. Veja o caso de Larry Summers, por exemplo. O ex-assessor econômico dos governos Clinton e Obama e ex-presidente da Universidade de Harvard trabalhou como consultor de D. E. Shaw, fundo de hedge quantitativo. Tom se encontrou com Summers em evento social e perguntou-lhe o que ele fazia na empresa. Ele respondeu: “Vou lá uma vez por semana e percorro as mesas dos quants que desenvolvem os modelos matemáticos de operações de mercado. Pergunto-lhes quais são as premissas que fundamentam seus modelos e quais são as circunstâncias que as tornariam inválidas. É surpreendente a frequência com que os quants não conseguem responder com objetividade.” Diz-se que Summers recebeu US\$5 milhões de dólares por esse trabalho, o que, em si, já é motivo suficiente para que sua contribuição seja considerada muito valiosa.

Também você pode agir como Larry Summers. Se alguém lhe apresentar um modelo, você sempre demonstrará inteligência se perguntar quais são as premissas em que se baseiam e quais são as circunstâncias que as invalidariam. Se a resposta for demasiado técnica, insista na pergunta de como o mundo teria de mudar para que o modelo não mais fosse útil.

**Pressionar quando não compreender**

O último ponto da seção anterior pode ser generalizado; é importante pressionar quando não compreender. A forma mais importante de pressionar é pedir dados e análises, em vez de casos ou opiniões. É como disse Gary Loveman, CEO da Caesars Entertainment: “Meu trabalho não é ter todas as respostas, mas, sim, fazer muitas perguntas penetrantes, perturbadoras e, às vezes, agressivas, como parte do processo analítico que conduz a insights e a refinamentos.”<sup>5</sup>

Os tipos específicos de perguntas que estimulam uso mais intenso da analítica poderiam ser as seguintes:

- “Você esqueceu seus dados?”
- “Como você acha que essa hipótese poderia ser testada com dados?”
- “Você pensou em uma análise empírica dessa ideia?”
- “Temos cerca de xx clientes. Você testou essa ideia com algum deles?”
- “Talvez pudéssemos imaginar algum conceito pequeno, mas rigoroso, sobre esse conceito.”

Você pegou a ideia. Se muitas pessoas em sua organização sempre fizerem perguntas desse tipo, a mudança cultural será rápida e intensa. Isso não significa que você precisa ouvir ou falar nos mesmos termos. Como bom exemplo, o filme *O dia antes do fim* dramatiza alguns dos eventos que levaram à crise financeira de 2008-2009. O enredo se baseia em um banco de investimentos que parece o Lehman Brothers. O personagem quant na trama, PhD em Engenharia de Propulsão, desenvolve novo algoritmo para o cálculo da exposição do banco ao risco. Quando ele apresenta o algoritmo ao chefe de operações de mercado, representado por Kevin Spacey, o czar explosivo diz: “Você sabe que eu não consigo ler essas coisas. Fale comigo em inglês.”<sup>6</sup> Todos os gestores deveriam ser igualmente exigentes.<sup>7</sup>

Liam Fahey, professor de Marketing e Estratégia, descreveu em *Strategy and Leadership* o papel dos executivos no desenvolvimento de trabalhos analíticos, por meio de uma série de perguntas a serem feitas.<sup>8</sup> Elas são um bom resumo de o que esperar dos executivos. Eis a lista completa das perguntas que ele recomenda:

- Qual é a questão ou a necessidade de negócios que o trabalho analítico pretende esclarecer?
- Quais são os insights centrais relevantes para a compreensão do tema e do contexto?
- Como posso alavancar esses insights no trabalho que faço?
- Como esses insights afetam as decisões com que nos defrontamos?
- Como esses insights ajudam a tomar decisões emergentes e futuras?

Quando as primeiras descobertas começarem a surgir, os executivos devem perguntar:

- O que há de surpreendente nessa descoberta?
- Poderia você fazer novas análises para confirmar ou refutar a descoberta?
- Deveríamos envolver outras pessoas no questionamento dessa nova descoberta?
- Haveria algum insight significativo emergindo aqui?
- Caso a descoberta se confirme, como ela afetaria meu raciocínio sobre este e outros tópicos e questões?

Para cada novo insight, os executivos devem perguntar:

- O que há de novo em cada insight?
- Qual era o velho entendimento?
- Quão significativa é a diferença?
- Qual é o raciocínio ou “argumento” que conecta o conjunto de dados ao insight?

Depois da comunicação dos insights, os executivos devem perguntar:

- Quem está ou estava envolvido na formulação do novo entendimento?
- Como poderiam eles ter influenciado os resultados?
- Quais poderiam ser as principais diferenças entre indivíduos ou unidades?

Se você, como executivo, fizer todas essas perguntas, seu envolvimento no trabalho analítico será muito maior e os analistas o verão como pessoa interessada e esclarecida. E se os analistas forem capazes de responder a todas elas com clareza, também eles estarão fazendo bom trabalho!

## O que esperar dos profissionais de analítica?

Depois de passar algum tempo descrevendo as atribuições dos decididores na solução de problemas quantitativos, faz sentido explicar também o que os profissionais de analítica devem fazer para superar as expectativas dos decididores. Resumimos essas atividades no box “O que os decididores devem esperar dos analistas quantitativos”.

### O que os decididores devem esperar dos analistas quantitativos

Se você for um executivo de empresa que estiver trabalhando com analistas quantitativos, eis o que legitimamente esperar deles:

- Conhecer bem os negócios da empresa como um todo e, principalmente, os processos de negócios específicos a serem melhorados com o apoio deles.
- Compreender seu estilo de pensamento e os tipos de análises e produtos que influenciarão seu raciocínio.
- Desenvolver relações de trabalho eficazes com as pessoas-chave da organização.
- Usar linguagem de negócios para explicar os benefícios e melhorias a serem fornecidos pela analítica.
- Fornecer estimativa exata do prazo e do custo para desenvolver o modelo e os recursos correlatos.
- Ser paciente e persistente, adotando linguagem diferente, se necessário, caso você não compreenda o que estão propondo e continue cético quanto aos benefícios previstos.
- Ter processo estruturado para conseguir as informações e as regras necessárias à

construção do modelo.

- Ajudá-lo a pensar sobre aspectos amplos do problema, como formulação das decisões envolvidas, identificação dos stakeholders e definição das capacidades necessárias para implementar a nova solução.
- Desenvolver novos modelos e conjuntos de ferramentas, de rápida prototipagem, de modo a oferecer-lhe algo substancial com rapidez, sobre o qual você possa oferecer feedback imediato, desde que não haja forte razão em contrário.
- Reconstruir e melhorar o modelo, até que atenda a todas as suas especificações de desempenho.
- Concordar com um cronograma, durante o qual você revisará o modelo, e explicar-lhe o que procurar como indícios de que o modelo está funcionando bem e que não precisa ser alterado.

## Aprender o negócio e interessar-se pelos seus problemas

Alguns analistas quantitativos estão interessados basicamente nos métodos e nas análises em vez de nos problemas de negócios a serem resolvidos. Essa mentalidade decorre, em parte, de nosso sistema educacional, que tende a ensinar matemática e estatística de maneira não contextualizada. Se, porém, não se concentrarem nos problemas de negócios, os quants não serão capazes de resolvê-los com eficácia nem de fornecer muito valor aos decididores.

O momento mais oportuno para verificar se o analista está interessado na solução de problemas de negócios é na fase de recrutamento e seleção. Depois da contratação é difícil promover mudanças de mentalidade. Anne Robinson, chefe do grupo de analítica da Verizon Wireless, por exemplo, pede a todos os candidatos para descrever um problema de negócios específico de que ele ou ela tenha tratado no passado e para indicar o que lhe pareceu mais interessante na questão. Karl Kempf, da Intel, faz perguntas semelhantes. Se o candidato empaca na resposta – o que, infelizmente, segundo Robinson e Kempf, é muito comum – a pessoa é rejeitada.

## Conversar em linguagem de negócios

Abordamos esse tema mais de uma vez ao longo do livro, e, sem dúvida, expressar-se em termos de negócios nem sempre é tarefa fácil para não

especialistas. Os analistas quantitativos, no entanto, precisam aprender a traduzir seus métodos e descobertas para o jargão de negócios. Nesse intuito, em muitos casos, será necessário usar a terminologia com que os profissionais de negócios estão familiarizados – LIFO, FIFO, ROI, comportamento dos clientes, dinheiro economizado e gerado. Falar em dinheiro o tempo todo talvez pareça um pouco mercenário para algumas pessoas, mas essa é a linguagem dos negócios. Em órgãos públicos e em entidades sem fins lucrativos, geralmente se fala em termos de cidadãos, constituintes e orçamentos.

Patrick Moore, chefe de um grupo de analítica comercial na Merck (no fim do capítulo, apresentamos um exemplo do trabalho desse grupo), diz que tenta seguir três regras práticas ao explicar os resultados analíticos, de modo a ajudar os clientes internos a tomar melhores decisões empresariais:

- Evitar a ideia de que a análise é uma “caixa-preta”. Esse preconceito levará o cliente a evitá-la. Portanto, ele tenta ser muito transparente.
- Reforçar a ideia de que a análise se baseou em dados e métodos apropriados; em outras palavras, ele e os analistas dele tentam ser e parecer confiantes quanto a ter trabalhado analiticamente da maneira mais eficiente e eficaz possível.
- Oferecer aos clientes internos “degustações” ou “miniaturas” dos resultados, a serem usados para convencer os céticos e para persuadir os próprios líderes.

O grupo de Moore faz amplo uso de gráficos para transmitir, por exemplo, a importância relativa de diferentes variáveis do modelo. Mesmo que não compreendam inteiramente os critérios ou estatísticas, os clientes internos captam a importância relativa dos fatores em um gráfico de barras.

## Explicar qualquer termo técnico

Em certas ocasiões, os quants talvez tenham de usar alguns termos técnicos para explicar o que fizeram. Nessas circunstâncias, devem estar preparados para explicar os respectivos conceitos em vernáculo, e isso significa não ser pego



desprevenido. Se houver alguma ferramenta ou método analítico que seja usado com frequência na organização, os quants talvez queiram reunir-se com colegas para juntos determinarem como explicá-los em linguagem objetiva e clara. Evidentemente, para análises relativamente simples – as que envolvem não mais que duas variáveis – a analítica visual dispõe de meios poderosos para explicar as relações entre os dados.

## **Dispor-se a desenvolver relacionamentos**

Como já mencionamos neste capítulo, as melhores decisões não envolvem matemática, mas relacionamentos. Se os quants não quiserem relacionar-se com o pessoal de negócios, talvez devam mudar de profissão e trabalhar como astrofísico, guarda florestal ou em alguma outra atividade solitária, em vez de como analistas quantitativos de empresas.

Isso é fácil de dizer, mas é verdade que muitos quants demonstram preferência histórica por números a pessoas. No entanto, caso se recrutem candidatos em programas universitários de analítica para empresas (como os da North Carolina State University) e se selecionem os que demonstrarem mais orientação para pessoas, é possível enfrentar o problema.

## **Não fazer os executivos se sentirem estúpidos**

Conhecemos numerosas organizações em que o pessoal quantitativo parece gostar de fazer os profissionais de negócios “normais” se sentirem estúpidos. Eles dizem coisas como “Sem dúvida, você sabe o que é análise de regressão, não?” ou “Lamento, mas o teste do chi-quadrado é elementar demais para eu ter de explicar”. Alguns “heavy-quants” (como uma organização apelidava seus analistas quantitativos mais sofisticados) chegavam ao ponto de tripudiar sobre os “light-quants” da mesma empresa.

Evidentemente, esse comportamento é inaceitável e altamente prejudicial para a eficácia da solução de problemas. Como ocorre com muitos outros maus comportamentos, no entanto, achamos que essa atitude negativa resulta de as pessoas não se sentirem respeitadas. Nas organizações em que os quants se envolvem estreitamente nos processos de negócios e se consideram acatados

pelos decididores, eles tendem a ser excelentes colegas de trabalho. Diferentemente, esses maus comportamentos não raro se difundir-se nas organizações que, por algum motivo, contrataram analistas quantitativos, mas os ignoram na hora das decisões importantes. Os quants, como todas as outras pessoas, respeitam quando são respeitados.

## Exemplo de raciocínio analítico: Previsão de demanda na Cisco

Prever a demanda dos clientes é problema para muitas empresas, em especial para as de manufatura.<sup>9</sup> Trata-se de questão extremamente importante para a Cisco Systems, líder do mercado em fornecimento de equipamentos de telecomunicações. Como a empresa em si não fabrica a maioria dos produtos que vende, sua cadeia de fornecimento global é muito complexa. É como diz Kevin Harrington, vice-presidente de operações de negócios globais, na área de gestão de cadeia de valor para os clientes: “Evidentemente, a previsão da demanda dos clientes é parte central da gestão da cadeia de fornecimento e fator importante da fabricação enxuta (*lean manufacturing*). A disciplina torna-se ainda mais desafiadora em épocas como a atual, caracterizada por rápidas mudanças na macroeconomia e por alta volatilidade na oferta e na demanda. Com efeito, a necessidade da Cisco de baixar parcela significativa do estoque [no valor de US\$2,25 bilhões], depois do estouro da bolha pontocom em 2001, foi uma das causas do ímpeto para a transformação mais ampla de nossa cadeia de valor.”<sup>10</sup>

O projeto daí resultante é bom exemplo não só do raciocínio analítico, mas do bom relacionamento entre analistas quantitativos e profissionais de negócios.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** O problema para a Cisco era possibilitar melhor projeção da demanda, envolvendo mais de 10 mil produtos diferentes. Os gestores de várias partes da empresa, abrangendo vendas, marketing e finanças, já tinham desenvolvido algo denominado “previsão por consenso” (*consensus forecast*), com base na combinação de intuição e extrapolação de tendências. Karl Braitberg, vice-presidente de gestão e planejamento da demanda da Cisco, porém, achava que previsões estatísticas, baseadas nos padrões de pedidos em carteira e de demanda histórica, gerariam uma “segunda opinião”, a ser comparada com a “previsão por consenso”, que podia ser afetada pelo excesso de entusiasmo de marketing. Incumbiu, então,

Anne Robinson, gerente sênior de previsão e modelagem analítica, e sua equipe de seis pessoas, de desenvolver o sistema de previsão estatística. Robinson percebeu que, para ser bem-sucedida, seria necessário não só desenvolver modelo de previsão de alta qualidade, mas também convencer a administração da Cisco de aceitar e usar as previsões estatísticas. Para tanto, identificou os principais stakeholders e adotou processo de desenvolvimento “ágil”, em que, gradualmente, produtos eficazes seriam entregues com frequência regular, ao longo dos 18 meses do projeto. Em cada entrega de produto, ela apresentava os resultados aos stakeholders, instruía-os quanto à operação do modelo, e, assim esperava, convencia-os a usá-lo.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Várias são as abordagens à previsão estatística. As descobertas anteriores sugerem que se alcançam os melhores resultados por meio de um conjunto de métodos de previsão – abordagem denominada *ensemble forecasting*. Com base em pesquisas e investigações, Robinson conhecia o potencial do método, e fez questão de que todas as ferramentas de previsão desenvolvidas pela equipe explorassem esse recurso.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** As principais variáveis do modelo provavelmente seriam o nível de pedidos atual e a demanda histórica. Essas variáveis são de uso comum em processos de previsão nos vários setores.

**COLETA DE DADOS (MENSURAÇÃO).** Embora as variáveis a serem usadas fossem claras desde o começo, muitas eram as fontes de pedidos atuais a serem avaliadas, a fim de determinar os dados mais importantes para o modelo. Por exemplo, a Cisco monitora os pedidos de clientes por segmentos, como setores de atividade, tamanho dos clientes, área geográfica, expedição, e outros. Os dados nem sempre se complementam com perfeição. Felizmente, todas as possíveis fontes de dados constavam de banco de dados preexistente, abrangendo toda a empresa. A equipe de Robinson, contudo, também precisava criar novas métricas de atendimento da demanda, centradas no cliente, que se tornassem vetores do que significava a centralização nos clientes em toda a cadeia de fornecimento da Cisco.

**ANÁLISE DE DADOS.** As previsões estatísticas resultam em faixas de demanda estimadas, com um intervalo de confiança para cada estimativa. Ela pode sugerir,

por exemplo, que a demanda mensal por determinado roteador seria de 3.000 a 3.500 unidades, com 95% de chance de que a demanda real se situaria dentro desse espectro. A abordagem “ágil” ao desenvolvimento do modelo envolvia uma série de passos, cada um com duração de dois a três meses, para mostrar a possibilidade de um modelo bem-sucedido, suscetível de ampliação, para lidar com a quantidade e variabilidade dos produtos da Cisco. Alguns dos passos consistiam no seguinte:

- Escolher ferramenta compatível com as necessidades (a Cisco selecionou o SAS Forecast Service, que oferece apoio para modelos ensemble).
- Determinar se os modelos estatísticos gerariam previsões mais exatas que as resultantes do consenso, o que foi comprovado.
- Sintonizar os modelos para aumentar a exatidão.
- Verificar se o método de previsão poderia ser ampliado para milhares de produtos, classificados em três centenas de famílias (poderia).
- Automatizar os modelos (a operação manual seria muito trabalhosa, mas os gestores e especialistas da Cisco ainda podem assumir o controle, se necessário).

Em cada passo, previa-se a revisão pelos stakeholders, o que propiciava o comprometimento com o novo método, durante todo o processo.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** O método de previsão estatística hoje produz faixas de previsões semanais para mais de 18 mil produtos, envolvendo horizonte temporal de 24 meses. A combinação de previsões estatísticas e previsões por consenso resultou em aumento médio de 12% na exatidão. De acordo com Kevin Harrington, o projeto foi um sucesso:

Os resultados foram previsões mais exatas, giro dos estoques mais rápido, e melhoria geral do equilíbrio oferta demanda, gerando benefícios tanto para a Cisco quanto para os clientes, na forma de redução do excesso de estoques e aumento da rapidez e confiabilidade dos produtos. No pior momento da crise econômica recente, a Cisco conseguiu reduzir os

estoques na cadeia de fornecimento, sem baixas contábeis nem comprometimento do nível dos serviços. Hoje, nossos especialistas em previsões estatísticas estão empenhados em melhorar ainda mais todo o processo e em gerenciar o aumento da demanda resultante da recuperação da economia global.<sup>11</sup>

Além dos resultados descritos por Harrington, Robinson observa que agora os gestores da Cisco se sentem à vontade com o uso de faixas e probabilidades para descrever a demanda. Na verdade, preferem previsões de faixas a previsões de pontos (números isolados), e se referem a faixas em todas as conversas importantes. Em suma, a cultura de previsões na Cisco mudou drasticamente, tomando rumo mais analítico.

Ao longo de todo o projeto, Robinson tentou envolver toda a comunidade de previsões no processo analítico. Nesse intuito, realizou processo estruturado de brainstorming com os stakeholders, para identificar novas métricas centradas em clientes. Também promoveu discussões em grupo com públicos mais amplos, road shows (apresentações itinerantes) para vários públicos e a exposição “Forecasting 101” (Introdução à Previsão), que fez numerosas vezes. E ainda desenvolveu recursos visuais para a apresentação dos resultados do modelo e estimulou a equipe a “contar a história com dados” extraídos dos resultados. Robinson formou parceria estreita com a organização de TI da Cisco, inclusive observando que, às vezes, era difícil distinguir as tarefas da equipe de analítica e as tarefas do pessoal de TI.

## **Exemplo de raciocínio analítico: Otimização da força de vendas na Merck**

A definição do tamanho ideal da força de vendas em grande empresa farmacêutica como a Merck é tarefa analítica difícil. Novos produtos são lançados com regularidade, aumentando a demanda e a necessidade de pessoal de vendas; patentes caducam, reduzindo a demanda e a necessidade de pessoal de vendas. A falta de histórico de demanda para novos produtos impossibilita a previsão exata das necessidades de pessoal de vendas.

Muitas empresas farmacêuticas recorrem a consultores externos para dimensionar a força de vendas. Quando, porém, Paul Kallukaran, PhD em

Análise Quantitativa, com experiência em análise de dados sobre vendas de empresas farmacêuticas, se juntou à equipe de analítica comercial da Merck, os executivos resolveram fazer a análise com recursos próprios.

**RECONHECIMENTO E FORMULAÇÃO DO PROBLEMA.** Com certos medicamentos perdendo a patente e outros entrando no ciclo de vendas, qual deveria ser o tamanho ótimo da força de vendas? Como a força de vendas não é monolítica, mas, sim, segmentada por marca e área geográfica, o problema consistia em dimensionar os segmentos, não a força de vendas total.

**REVISÃO DAS DESCOBERTAS ANTERIORES.** Considerando que as empresa de consultoria oferecem serviços de dimensionamento da força de vendas, dispõe-se de alguma literatura sobre o assunto. Embora seu trabalho anterior não se enquadrasse diretamente nessa área, Kallukaran já conhecia as abordagens de outras empresas. Estas, contudo, tendiam a ser “caixas-pretas”, orientadas pela perspectiva dos executivos de marketing e vendas. Kallukaran e Patrick Moore, chefe de analítica comercial, não gostavam desse aspecto. No passado, diferentes grupos da Merck tinham recorrido a vários consultores e a diversos métodos para dimensionar a força de vendas; essa era a primeira vez em que adotavam uma abordagem central.

**MODELAGEM (SELEÇÃO DE VARIÁVEIS).** Kallukaran decidiu usar vários métodos para determinar o tamanho ótimo da força de vendas. Além de fazê-lo da maneira tradicional, com modelos estatísticos, ele e a equipe adotaram abordagem mais detalhada, tentando compreender o que era necessário para servir a cada cliente. Para tanto, conversaram com o pessoal de vendas sobre suas atividades com clientes médicos e estimaram a carga de trabalho provável com cada um. Também calcularam as previsões de vários produtos e desenvolveram modelos não lineares de respostas a promoções de vendas e de mudanças consequentes prováveis nas prescrições dos médicos. E ainda analisaram o impacto da força de vendas em comparação com outros fatores que afetam os comportamentos dos médicos ao prescreverem medicamentos, como hábito, marca, equidade e pressão dos pacientes. Os analistas também consideraram dados referentes aos pacientes, para compreender padrões de aderência ao longo do tempo. Muitos pacientes apresentaram quedas significativas no uso, o que afeta as vendas do medicamento no longo prazo.

Finalmente, empregaram modelo de programação linear inteira para otimizar os recursos por médico, por produto e para cada um das centenas de territórios.

**COLETA DE DADOS.** A indústria farmacêutica em geral obtém dados sobre prescrições médicas de outras associações, e a Merck tinha esses dados. O modelo de atividade da força de vendas do projeto, porém, exigia que se pesquisasse a força de vendas com base nos diferentes comportamentos e no tempo dedicado aos clientes. Era importante preservar a confiança da força de vendas – ou seja, pesquisá-la de maneira a não parecer exercício de redução –, de modo que o pessoal de vendas não se sentisse sob a ameaça de perder o emprego e tivesse confiança suficiente para dar respostas exatas.

**ANÁLISE DOS DADOS.** Conforme observamos, o exercício complexo envolvia várias abordagens analíticas, inclusive modelos de otimização inteira não paramétricos (que não assumiam determinado tipo de distribuição), capazes de computar curvas de respostas para cada segmento de produto, com base no histórico de respostas às promoções. Como o projeto exigia que se rodassem modelos para cada marca e território, todo o processamento demorava de início 16 horas. Como, porém, o grupo de Kallukaran queria respostas mais rápidas, a computação foi distribuída entre centenas de computadores laptop, cada um se limitando a determinado território. Com todo esse poder de computação, o modelo inteiro era rodado em 20 minutos.

**APRESENTAÇÃO DOS RESULTADOS E ADOÇÃO DE PROVIDÊNCIAS.** Embora se tratasse de nova abordagem ao problema de dimensionamento da força de vendas na Merck, a questão não era totalmente desconhecida para os clientes internos, o que facilitou a aceitação do modelo e a adoção de providências com base nos resultados. No lado de vendas, o VP de planejamento estratégico tinha mentalidade analítica e já vivenciara no passado a atuação dos consultores na área. No lado de marketing, os analistas já haviam desenvolvido modelos de avaliação das respostas às promoções, mas os usuários não se acostumaram aos novos métodos. Alguém sempre se queixava de que faltava algo no modelo; o ótimo se tornara inimigo do bom. No projeto de dimensionamento da força de vendas, Kallukaran encorajou os clientes de marketing a “usar o que temos”. De início, ele trabalhou primeiro com uma das menores equipes de marca e mostrou por que a nova abordagem era melhor que a alternativa mais intuitiva que

usavam antes. Para tanto, comparou o método intuitivo com os resultados dos dados e das análises, mas não tentou impor-lhes o modelo. “Usem-no como apenas mais um input para suas decisões”, dizia-lhes. O esforço para convencer diferentes grupos a usar o modelo foi atenuado pela abordagem consistente em toda a empresa. O presidente da Merck na época gostou da possibilidade de comparar as necessidades de diferentes equipes de marcas para avaliar com mais facilidade os pedidos de recursos. Com o passar do tempo, quase todos os grupos da Merck adotaram o modelo. A equipe de analítica de Kallukaran também recebia pedidos de recálculo do modelo sempre que surgiam vagas em determinado território, e os gerentes de território agora desfrutavam de mais autonomia e eram responsáveis pelos resultados financeiros. Assim, a equipe, que incluía desenvolvedores de sistemas, criou um “aplicativo analítico” (*analytical app*) que sugere ao gerente de vendas preencher ou não preencher a vaga. O modelo é muito usado e facilita as decisões locais, sem exigir muito do grupo central de Kallukaran.

## Considerações finais sobre o raciocínio analítico

A essa altura – efetivamente o fim do livro, a não ser para quem gosta de esmiuçar notas de rodapé – esperamos ter convencido os leitores de muitas coisas. Primeiro, que o raciocínio analítico e que as decisões baseadas em dados e em analítica desempenharão papel cada vez mais importante nos negócios e na sociedade. Precisaremos de muitos gestores e profissionais que se sintam à vontade com o raciocínio analítico, e queremos que você seja um deles. Segundo, esperamos, agora, que você se considere capaz de atuar no jogo, mesmo que não seja especialista em estatística ou em matemática. Compreendendo os estágios e os passos do raciocínio analítico e os atributos de um bom processo decisório analítico, é possível integrar-se com os melhores quants e ajudá-los a aprimorar os resultados. Ao longo do caminho, você também passará a raciocinar e a decidir melhor.

Terceiro, embora a maioria das pessoas ache que o núcleo do raciocínio analítico seja a “solução do problema”, este é apenas um dos passos que compõem a decisão analítica bem-sucedida. Caso se formule o problema de maneira incorreta ou subótima, a solução não será muito útil. E se os resultados não forem comunicados com eficácia, é improvável que sirvam para tomar



decisões ou para adotar providências. Se você estiver trabalhando em um problema analítico, pensando em como distribuir seu tempo, comece alocando-o igualmente entre os três estágios.

Quarto, e finalmente, muita gente acha que o mundo do raciocínio e das decisões analíticas envolve apenas números, estatísticas rigorosas e pensamento com o lado esquerdo do cérebro. O lado direito do cérebro, porém, também precisa envolver-se seriamente. Mostramos – sobretudo no [Capítulo 5](#) – que a criatividade é importante para o raciocínio analítico e argumentamos neste capítulo (com base em muitos experimentos e observações) que os relacionamentos são tão ou mais fundamentais para o progresso na analítica que a simples capacidade de processar números.

Quem leu todo o livro e refletiu um pouco sobre as ideias e exemplos já está preparado para cerrar fileiras com os pensadores analíticos. Parabéns! Esta é uma época excitante para participar deste grupo. A quantidade e a importância dos dados nas organizações tende a disparar ao longo do tempo, e você também terá condições de avançar sob esse forte empuxo. Esperamos que seu novo foco analítico seja benéfico tanto para a sua carreira quanto para o sucesso de sua organização.

# Notas

## Capítulo 1

1. Xiao-Li Meng, resumo, “Statistical Education and Educating Statisticians: Producing Wine Connoisseurs and Master Winemakers”, seminário apresentado na Universidade de Minnesota, 28 de outubro de 2011, <http://catalystsumn.blogspot.com/2011/11/statistics-education-seminar-presented.html>.
2. “Big Data: The Next Frontier for Innovation, Competition, and Productivity,” McKinsey Global Institute, maio de 2011, [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation).
3. R. M. Dawes, D. Faust e P. E. Meehl, “Clinical Versus Actuarial Judgment,” *Science* 243, n. 4899 (March 1989): 1668–1674.

## Capítulo 2

1. S. Babou, “What Is Stakeholder Analysis?” The Project Management Hut, <http://www.pmhut.com/what-is-stakeholder-analysis>.
2. “Expedia Travels to New Heights,” *SASCom Magazine*, Third Quarter 2011, 14.
3. Todos esses casos de testes são exemplos de clientes da Applied Predictive Technologies, empresa de software, embora tenhamos tratado de cada empresa isoladamente. Para mais informações sobre os métodos, ver Thomas H. Davenport, “How to Design Smart Business Experiments”, *Harvard Business Review*, novembro de 2009.
4. Mireya Navarro, “For Many Latinos, Race Is More Culture Than Color”, *New York Times*, 13 de janeiro de 2012.
5. W. A. Wallis e H.V. Roberts, *Statistics: A New Approach* (New York: Free Press, 1960).
6. Lisa Carley-Baxter et al., “Comparison of Cell Phone and Landline Surveys: A Design Perspective”, *Field Methods* 22, n. 1 (fevereiro de 2010): 3-15.
7. Saul Hansell, “Google Answer to Filling Jobs Is an Algorithm”, *New York Times*, 3 de janeiro de 2007, [www.nytimes.com/2007/01/03/technology/03google.html](http://www.nytimes.com/2007/01/03/technology/03google.html).
8. “Joseph Jagger”, Wikipedia, [http://en.wikipedia.org/wiki/Joseph\\_Jagger](http://en.wikipedia.org/wiki/Joseph_Jagger); “Joseph Jagger: The Man Who Broke the Bank”, [www.wildjackcasino.com/joseph-jagger.html](http://www.wildjackcasino.com/joseph-jagger.html); “Joseph Jagger”, [www.realmoneycasinos.net/joseph-jagger.html](http://www.realmoneycasinos.net/joseph-jagger.html); “Roulette The Men Who Broke the Bank at Monte Carlo – Joseph Jagger”, [www.wiseguyroulette.com/roulette-history/joseph-jagger/](http://www.wiseguyroulette.com/roulette-history/joseph-jagger/).
9. Rama Ramakrishnan, “Three Ways to Analytic Impact”, blog *The Analytic Age*, 26 de julho de 2011, <http://blog.ramakrishnan.com/>.
10. *People v. Collins*, 68 Cal. 2d 319 (1968); [http://scholar.google.com/scholar\\_case?case=2393563144534950884](http://scholar.google.com/scholar_case?case=2393563144534950884); “*People v. Collins*”, [http://en.wikipedia.org/wiki/People\\_v.\\_Collins](http://en.wikipedia.org/wiki/People_v._Collins).

## Capítulo 3

1. A. M. Starfield, Karl A. Smith e A. L. Bleloch, *How to Model It: Problem Solving for the Computer Age* (New York: McGraw-Hill, 1994), 19.
2. George Box and Norman R. Draper, *Empirical Model-Building and Response Surfaces* (New York: Wiley, 1987), 424.
3. Garth Sundem, *Geek Logik: 50 Foolproof Equations for Everyday Life*, (New York: Workman, 2006).
4. Minnie Brashears, *Mark Twain, Son of Missouri* (Whitefish, MT: Kessinger Publishing, 2007).
5. Ernest E. Leisy, ed., *The Letters of Quintus Curtius Snodgrass* (Irving, TX: University Press of Dallas, 1946).
6. Claude S. Brinegar, “Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship,” *Journal of the American Statistical Association* 58 (1963).
7. R. C. Rosen et. al., “The International Index of Erectile Function (IIEF): A Multidimensional Scale for Assessment of Erectile Function,” *Urology* 49, n., 6 (1997): 822–830. R. C. Rosen et. al., “Development and Evaluation of an Abridged, 5-item Version of the International Index of Erectile Function (IIEF-5) as a Diagnostic Tool for Erectile Dysfunction,” *International Journal of Impotence Research* 11 (1999): 319–326.
8. Rama Ramakrishnan, “Three Ways to Analytic Impact,” *The Analytic Age* (blog), 26 de julho de 2011, <http://blog.ramakrishnan.com/>.
9. Anand Rajaraman, “More Data Usually Beats Better Algorithms,” *Datawocky* (blog), 24 de março de 2008, <http://anand.typepad.com/datawocky/2008/03/more-datausual.html>.
10. Daryl Morey, “Success Comes from Better Data, Not Better Analysis”, postagem em blog, *Harvard Business Review*, 8 de agosto de 2011, [http://blogs.hbr.org/cs/2011/08/success\\_comes\\_from\\_better\\_data.html](http://blogs.hbr.org/cs/2011/08/success_comes_from_better_data.html).
11. “Tycho Brahe”, *Wikipedia*, [http://en.wikipedia.org/wiki/Tycho\\_Brahe](http://en.wikipedia.org/wiki/Tycho_Brahe); Michael Fowler, “Tycho Brahe”, <http://galileoandstein.physics.virginia.edu/1995/lectures/tycho.html>; Arthur Koestler, *The Watershed: A Biography of Johannes Kepler* (Doubleday, 1960); “Johannes Kepler”, *Wikipedia*, [http://en.wikipedia.org/wiki/Johannes\\_kepler](http://en.wikipedia.org/wiki/Johannes_kepler); “Johannes Kepler: The Laws of Planetary Motion”, <http://csep10.phys.utk.edu/astr161/lect/history/kepler.html>; Michael Fowler, “Tycho Brahe and Johannes Kepler”, <http://galileoandstein.physics.virginia.edu/lectures/tycho.htm>; Michael Fowler, “Johannes Kepler”, <http://galileoandstein.physics.virginia.edu/1995/lectures/kepler.html>; “Johannes Kepler”, *Encyclopædia Britannica Online Academic Edition*, <http://www.britannica.com/EBchecked/topic/315225/Johannes-Kepler>; Ann Lamont, “Johannes Kepler: Outstanding Scientist and Committed Christian”, 1:1, <http://www.answersingenesis.org/creation/v15/i1/kepler.asp>, 1 de dezembro de 1992.
12. Bill Franks, “Why Nobody Is Actually Analyzing Unstructured Data”, International Institute for Analytics, postagem em blog, 9 de março de 2012, <http://iianalytics.com/2012/03/why-nobody-is-actually-analyzing-unstructured-data/>.
13. Peter Passell, “Wine Equation Puts Some Noses Out of Joint”, *New York Times*, 4 de março de 1990.
14. “Alternative Rich List”, *FT.com*, 22 de setembro de 2006.
15. Fischer Black e Myron Scholes, “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy* 81, no. 3 (1973): 637–654; “Black–Scholes”, *Wikipedia*, <http://en.wikipedia.org/wiki/Black–Scholes>; “The Prize in Economics 1997”, press release, [http://nobelprize.org/nobel\\_prizes/economics/laureates/1997/press.html](http://nobelprize.org/nobel_prizes/economics/laureates/1997/press.html).
16. Fischer Black e Myron Scholes, “The Pricing of Options and Corporate Liabilities”, *Journal of Political Economy* 81, no. 3 (maio de 1973).
17. R. J. Larsen e M. L. Marx, *An Introduction to Mathematics Statistics and Its Applications*

(Englewood Cliffs, NJ: Prentice-Hall, 1981), 159. A coluna citada foi republicada em muitos compêndios e cursos.

## Capítulo 4

1. Xiao-Li Meng, resumo, “Statistical Education and Educating Statisticians: Producing Wine Connoisseurs and Master Winemakers”, seminário realizado na University of Minnesota, 28 de outubro de 2011, <http://catalystsumn.blogspot.com/2011/11/statistics-education-seminar-presented.html>.

2. Xiao-Li Meng, “Statistics: Your Chance for Happiness (or Misery),” descrição de curso, [http://www.stat.harvard.edu/Academics/invitation\\_chair\\_txt.html](http://www.stat.harvard.edu/Academics/invitation_chair_txt.html).

3. David Schmitt, “Tell a Story”, 27 de junho de 2012, [http://www.allanalytics.com/author.asp?id=2092&doc\\_id=246428](http://www.allanalytics.com/author.asp?id=2092&doc_id=246428).

4. I. Bernard Cohen, *The Triumph of Numbers: How Counting Shaped Modern Life* (New York: W.W. Norton, 2006), capítulo 9; “Florence Nightingale”, *Wikipedia*, [http://en.wikipedia.org/wiki/Florence\\_Nightingale](http://en.wikipedia.org/wiki/Florence_Nightingale); P. Nuttall, “The Passionate Statistician,” *Nursing Times* 28 (1983): 25–27.

5. Gregor Mendel, “Experiments in Plant Hybridization”, <http://www.mendelweb.org/>; “Gregor Mendel”, *Wikipedia*, [http://en.wikipedia.org/wiki/Gregor\\_Mendel](http://en.wikipedia.org/wiki/Gregor_Mendel); Seung Yon Rhee, Gregor Mendel, *Access Excellence*, [http://www.accessexcellence.org/RC/AB/BC/Gregor\\_Mendel.php](http://www.accessexcellence.org/RC/AB/BC/Gregor_Mendel.php); “Mendel’s Genetics”, [anthro.palomar.edu/mendel/mendel\\_1.htm](http://anthro.palomar.edu/mendel/mendel_1.htm); David Paterson, “Gregor Mendel”, [www.zephyrus.co.uk/gregormendel.html](http://www.zephyrus.co.uk/gregormendel.html); “Rocky Road: Gregor Mendel”, *Strange Science*, [www.strangescience.net/mendel.htm](http://www.strangescience.net/mendel.htm); Wolf-Ekkehard Lönnig, “Johann Gregor Mendel: Why His Discoveries Were Ignored for 35 Years”, [www.weloennig.de/mendel02.htm](http://www.weloennig.de/mendel02.htm); “Gregor Mendel and the Scientific Milieu of His Discovery”, [www.2iceshs.cyfronet.pl/2ICESHS\\_Proceedings/Chapter\\_10/R-2\\_Sekerak.pdf](http://www.2iceshs.cyfronet.pl/2ICESHS_Proceedings/Chapter_10/R-2_Sekerak.pdf); “Mendelian Inheritance”, *Wikipedia*, [http://en.wikipedia.org/wiki/mendelian\\_inheritance](http://en.wikipedia.org/wiki/mendelian_inheritance).

6. Esta lista foi adaptada e modificada de outra no site ManyEyes, da IBM; ver [http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization\\_Options.html](http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization_Options.html).

7. Este exemplo foi extraído de SAS Visual Analytics 5.1 User’s Guide, “Working with Automatic Charts”, <http://support.sas.com/documentation/cdl/en/vaug/65384/HTML/default/viewer.htm#n1xa25dv4fiyz6n1etsfk>.

8. Hans Rosling, “Stats That Reshape Your Worldview”, TED talk, fevereiro de 2006, [http://www.ted.com/talks/hans\\_rosling\\_shows\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen.html](http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html)

9. Embora o grupo de Schmitt às vezes crie esses vídeos na empresa, este foi feito por uma empresa de produção.

10. James Taylor, “Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics”, IBM Press, 2011.

11. Thomas H. Davenport, “How Companies Make Better Decisions”, International Institute of Analytics, 2010, [www.sas.com/decisions](http://www.sas.com/decisions).

12. Ch’ananda, S. (1990), “The Origin of Go”, *British Go Journal*, 79, Spring, 18-19.

13. Jinho Kim e Hyung Jun Kim, “The Advantage of Playing First in Go”, *CHANCE* 8, n. 3 (1995): 22–26.

14. Entrevista com Gottman in [http://www.edge.org/3rd\\_culture/gottman05/gottman05\\_index.html](http://www.edge.org/3rd_culture/gottman05/gottman05_index.html).

15. As informações sobre o Escore FICO foram extraídas de: “FICO Credit Score Chart”, [www.myfico.com/crediteducation/whatsinyourscore.aspx](http://www.myfico.com/crediteducation/whatsinyourscore.aspx); “Credit Score in the United States”, *Wikipedia*,

[http://en.wikipedia.org/wiki/Credit\\_score\\_in\\_the\\_United\\_States](http://en.wikipedia.org/wiki/Credit_score_in_the_United_States); Malgorzata Wozniacka and Snigdha Sen, “Credit Scores: What You Should Know About Your Own,” *PBS Frontline*, November 2004, <http://www.pbs.org/wgbh/pages/frontline/shows/credit/more/scores.html>.

## Capítulo 5

1. Adam Satariano e Peter Burrows, “Apple’s Supply Chain Secret? Hoard Lasers”, *Bloomberg BusinessWeek*, 3 de novembro de 2011, <http://www.businessweek.com/magazine/apples-supplychain-secret-hoard-lasers-11032011.html>.
2. Descrição de cargo on-line em: <http://jobs.apple.com/index.ajs?BID=1&method=mExternal.showJob&RID=81197>, acessado em 5 de janeiro de 2012.
3. Darrell Huff, *How to Lie with Statistics* (New York: Norton, 1954).
4. Junxiang Lu, “Predicting Customer Churn in the Telecommunications Industry: An Application of Survival Analysis Modeling Using SAS”, *SAS User Group International Proceedings*, 2002.
5. Tim Drye, Graham Wetherill e Alison Pinnock, “When Are Customers in the Market? Applying Survival Analysis to Marketing Challenges”, *Journal of Targeting, Measurement and Analysis for Marketing* 10 (2001): 179–188.
6. Joel M. Hektner, Jennifer A. Schmidt, Mihaly Csikszentmihalyi, *Experience Sampling Method: Measuring the Quality of Everyday Life* (Thousand Oaks, CA: Sage, 2007).
7. Mihaly Csikszentmihalyi and Reed Larson, *Being Adolescent: Conflict and Growth in the Teenage Years* (New York: Basic Books, 1984).
8. “Archimedes”, Wikipedia, <http://en.wikipedia.org/wiki/Archimedes>; “Eureka”, Wikipedia, <http://en.wikipedia.org/wiki/Eureka>; Rohini Chowdhury, “Eureka!': The Story of Archimedes and the Golden Crown,” Long, Long Time Ago, [http://www.longlongtimeago.com/llta\\_greatdiscoveries\\_archimedes\\_eureka.html](http://www.longlongtimeago.com/llta_greatdiscoveries_archimedes_eureka.html); John Monahan, “Archimedes Coins ‘Eureka!’ in the Nude—and Other Crazy Ah-Ha Moments of Science,” *Scientific American*, December 7, 2010, <http://www.scientificamerican.com/article.cfm?id=archimedes-coins-eureka-eureka>.
9. Robert e Michèle Root-Bernstein, *Sparks of Genius: The Thirteen Thinking Tools of the World’s Most Creative People* (Boston: Houghton Mifflin, 1999), 4.
10. Cho Jung Rae, *The Enchanted Prison of Writing* (Seoul: Sisain Books, 2009), 394.
11. Evidentemente, as ideias de Hegel são complexas e difíceis de resumir. Análise profunda desses conceitos pode ser encontrada em Tom Rockmore, *Hegel, Idealism, and Analytic Philosophy* (New Haven: Yale University Press, 2005).
12. Muitas são as versões dessa história; o relato mais definitivo, porém, é de Dan Power’s DSS News, 10 de novembro de 2002, <http://dssresources.com/newsletters/66.php>.
13. John Earle, citado no blog de Riggs Eckleberry “Of Interest”, 21 de dezembro de 1998, [http://www.riggs.com/archives/1998\\_12\\_01\\_OIarchive.html](http://www.riggs.com/archives/1998_12_01_OIarchive.html).
14. David Snowdon et. al., “Linguistic Ability in Early Life and Cognitive Function and Alzheimer’s Disease in Late Life”, *JAMA*, 275 (1996): 528–532. “Landmark Study Links Cognitive Ability of Youth with Alzheimer’s”, <http://archive.hhs.gov/news/press/1996pres/960220b.html>.
15. Snowdon, et. al., “Linguistic Ability in Early Life.”
16. Jodith Janes, resenha de *Aging with Grace*, de David Snowdon, *Library Journal* 126, n. 11 (junho de 2001): 96.
17. Tiffany Sharples, “Can Language Skills Ward Off Alzheimer’s? A Nuns’ Study,” *Time*, 9 de julho de

2009, <http://www.time.com/time/health/article/0,8599,1909420,00.html#ixzz1p7bYUKSu>.

18. Correspondência por e-mail com Anthony Viel, 9 de novembro de 2011.

## Capítulo 6

1. Derrick Niederman e David Boyum, *What the Numbers Say: A Field Guide to Mastering Our Numerical World* (New York: Broadway Books, 2003), 233.

2. Linda Rosa, Emily Rosa e Larry Sarner, “A Close Look at Therapeutic Touch”, *Journal of American Medical Association* 279, n. 13 (1998): 1005–1010; Douglas W. Hubbard, *How To Measure Anything* (Hoboken, NJ: Wiley, 2010), 13–17.

3. Kevin Courcey, “Therapeutic Touch: Further Notes”, [www.quackwatch.org/01QuackeryRelatedTopics/tt2.html](http://www.quackwatch.org/01QuackeryRelatedTopics/tt2.html).

4. From “The Phenom List—The Top Big Data Scientists: DJ Patil”, [http://thephenomlist.com/Big\\_Data\\_Scientists/DJ\\_Patil](http://thephenomlist.com/Big_Data_Scientists/DJ_Patil), acessado em 23 de fevereiro de 2012. Outras informações foram extraídas de entrevistas com Patil.

5. Darrell Huff, *How to Lie with Statistics* (New York: Norton, 1954).

6. Stephen K. Campbell, *Flaws and Fallacies in Statistical Thinking* (Englewood Cliffs, NJ: Prentice-Hall, 1974).

7. Phillip I. Good e James W. Hardin, *Common Errors in Statistics* (New York: Wiley, 2003).

8. Nassim Nicholas Taleb, *Fooled by Randomness: The Hidden Role of Chance in Life and the Markets* (New York: Random House, 2005).

9. Richard Isaac, *The Pleasures of Probability* (New York: Springer, 1995).

10. Leonard Mlodinow, *The Drunkard's Walk: How Randomness Rules Our Lives* (New York: Pantheon, 2008).

11. Ver <http://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2005/>.

12. Informações sobre NC State MSA foram extraídas de várias reuniões com administradores do programa com membros do corpo docente e de “Institute for Advanced Analytics, North Carolina State University”, [analytics.ncsu.edu](http://analytics.ncsu.edu).

13. Esses números preliminares do relatório “State of Business Intelligence in Academia 2011” foram apresentados por Barbara Wixom no Third BI Congress, Orlando FL. 16 de dezembro de 2012.

14. Citado em Thomas Dietz e Linda Kalof, *Introduction to Social Statistics: The Logic of Statistical Reasoning* (New York: Wiley, 2009), xxviii.

15. Campbell, *Flaws and Fallacies in Statistical Thinking*, 108.

16. Christopher Chabris e Daniel Simons, *The Invisible Gorilla and Other Ways Our Intuitions Deceive Us* (New York: Crown, 2010), 163.

17. Heinz Kohler, *Statistics for Business and Economics* (London: Longman, 1988).

18. Resenha da Amazon de “A ‘Umea University’ student (Sweden) give ratings”, 24 de agosto de 1999, [http://www.amazon.com/review/R2LQ3TGC1PC51D/ref=cm\\_cr\\_dp\\_title?ie=UTF8&ASIN=0673184447&channel=detail-glance&nodeID=283155&store=books](http://www.amazon.com/review/R2LQ3TGC1PC51D/ref=cm_cr_dp_title?ie=UTF8&ASIN=0673184447&channel=detail-glance&nodeID=283155&store=books), recuperado em 30 de dezembro de 2012.

19. Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York: Norton, 2003).

20. “SN Names the 20 Smartest Athletes in Sports”, *The Sporting News*, 23 de setembro de 2010, <http://aol.sportingnews.com/mlb/feed/2010-09/smart-athletes/story/sporting-news-names-the-20-smartest->



athletes-in-sports.

21. Michael Lewis, “The No-Stats All Star,” *New York Times*, February 13, 2009, [www.nytimes.com/2009/02/15/magazine/15Battier-t.html](http://www.nytimes.com/2009/02/15/magazine/15Battier-t.html).

22. Frances X. Frei e Mathew Perlberg, “Discovering Hidden Gems: The Story of Daryl Morey, Shane Battie e the Houston Rockets (B)”, estudo de caso da Harvard Business School (Boston: Harvard Business Publishing, setembro de 2010), 1.

## Capítulo 7

1. Comunicação pessoal com o autor.

2. “Surveying the Economic Horizon: A Conversation with Robert Shiller”, *McKinsey Quarterly*, April 2009,

[http://www.mckinseyquarterly.com/Surveying\\_the\\_economic\\_horizon\\_A\\_conversation\\_with\\_Robert\\_Shille](http://www.mckinseyquarterly.com/Surveying_the_economic_horizon_A_conversation_with_Robert_Shille)

3. David Olive, “Getting Wise Before That ‘One Big Mistake’”, *Toronto Star*, 17 de dezembro de 2007.

4. Charles Duhigg, *The Power of Habit: Why We Do What We Do in Life and Business* (New York: Random House, 2012).

5. Gary Loveman, “Foreword,” in Thomas H. Davenport e Jeanne G. Harris, *Competing on Analytics: The New Science of Winning* (Boston: Harvard Business School Press, 2007), x.

6. Mais contexto sobre o filme e os personagens pode ser encontrado em <http://business-ethics.com/2011/11/23/0953-margin-call-a-small-movie-unveilsbig-truths-about-wall-street/>.

7. *Margin Call*, filme escrito e dirigido por J. C. Chandor, 2011.

8. Liam Fahey, “Exploring ‘Analytics’ to Make Better Decisions: The Questions Executives Need to Ask”, *Strategy and Leadership* 37, no. 5 (2009): 12–18.

9. As informações para este exemplo foram extraídas de várias entrevistas com Anne Robinson; e Blake Johnson, “Leveraging Enterprise Data and Advanced Analytics in Core Operational Processes: Demand Forecasting at Cisco”, estudo de caso, Stanford University Management Science and Engineering Department.

10. Kevin Harrington, “Seeing the Future in Value Chain Management”, *Analytics Magazine*, março/abril de 2010.

11. Ibid.